

Risk *(of)* Communication

Ethics, Futures, and Strategic Complexity

Guest editor

Alessandro Lovari



Editor

Giovanni Boccia Artieri
Università degli Studi di Urbino Carlo Bo

Comitato editoriale

Sara Bentivegna (*"Sapienza" Università di Roma*)
Federico Boni (*Università degli Studi di Milano*)
Vanni Codeluppi (*Università di Modena e Reggio Emilia*)
Stefano Cristante (*Università del Salento*)
Giovanni Fiorentino (*Università degli Studi della Tuscia*)
Gino Frezza (*Università di Salerno*)
Laura Gemini (*Università degli studi di Urbino Carlo Bo*)
Guido Gili (*Università degli Studi del Molise*)
Gianluca Manzoli (*Università di Bologna*)
Alberto Marinelli (*"Sapienza" Università di Roma*)
Andrea Minuz (*"Sapienza" Università di Roma*)
Francesca Pasquali (*Università degli Studi di Bergamo*)
Gianfranco Pecchinenda (*Università "Federico II", Napoli*)
Giovanni Ragone (*Sapienza Università di Roma*)
Mario Ricciardi (*Politecnico di Torino*)
Massimo Scaglioni (*Università Cattolica del Sacro Cuore, Milano*)
Luca Toschi (*Università di Firenze*)

Comitato scientifico

Maurice Aymard (*Fondation Maison des Sciences de l'Homme, Parigi*)
Sebastiano Bagnara (*Università di Sassari*)
Roberta Bartoletti (*Università di Urbino Carlo Bo*)
Nancy Baym (*University of Kansas*)
Sergio Brancato (*Università di Napoli "Federico II"*)
Jean Burgess (*Queensland University of Technology*)
Enrique Bustamante (*Universidad Complutense, Madrid*)
Gustavo Cardoso (*ISCTE, Lisbona*)
Gianni Corino (*University of Plymouth*)
Patrice Flichy (*Université Paris Est Marne-la-Vallée*)
Jeremy Hight (*California Institute of Arts, Los Angeles*)
Paula Levine (*San Francisco State University*)
Jorge Lozano (*Universidad Complutense, Madrid*)
Roberto Maragliano (*Università di Roma 3*)
Sara Monaci (*Università degli Studi di Torino*)

Xavier Perez (*Universitat Pompeu Fabra, Barcellona*)
Mike Phillips (*University of Plymouth*)
Beppe Richeri (*Università della Svizzera Italiana, Lugano*)
Massimo Riva (*Brown University, Providence*)
Gino Roncaglia (*Università della Tuscia, Viterbo*)
Roberto Saracco (*EIT ICT Labs, Trento*)
Donald Sassoon (*Queen Mary College, Londra*)
Chris Speed (*Edinburgh College of Art*)
Yvon Thiec (*Eurocinema, Bruxelles*) Ugo Volli (*Università di Torino*)

Journal Manager

Manolo Farci
Università degli studi di Urbino Carlo Bo

Editorial Staff

Stefano Brilli
Università degli studi di Urbino Carlo Bo
Elena Ceccarelli
Università degli studi di Urbino Carlo Bo
Dom Holdaway
Università degli studi di Urbino Carlo Bo
Stefania Parisi
Sapienza Università di Roma
Rossella Rega
Università di Siena
Anna Sfardini
Università Cattolica del Sacro Cuore
Chiara Spaggiari
Università degli studi di Urbino Carlo Bo
Elisabetta Zurovac
Università degli studi di Urbino Carlo Bo

Direttore responsabile ai termini di legge

Fabio Tarzi
"Sapienza" Università di Roma

Hanno diretto la rivista

Alberto Abruzzese (2013-2015)

Casa Editrice Sapienza

ISSN: 2282-2542

Rivista telematica scientifica della Sapienza Università di Roma registrata presso il Tribunale Civile di Roma (N. 256 del 30/10/2013)



Questa rivista è pubblicata sotto una licenza Creative Commons Attribution 3.0

Risk communication and risks of communication: Ethics and strategic approaches to communicative complexity

Alessandro Lovari*
University of Cagliari

In recent years, growing scholarly interest has surrounded the notion of risk (Beck, 1992; Giddens, 1999), which has gradually expanded across multiple fields, including communication, journalism and media studies (Covello *et al.*, 1986; Heath, 2025; Heath & O'Hair, 2010; Sellnow & Sellnow, 2024). This evolution not only underscores the increasing centrality of the concept within scholarly debate, but also reflects a broader transformation in how societies identify, interpret, and categorize risks, involving an expanding range of actors within the public sphere, from organizations to citizens and media systems (Schwarz, Seeger & Kim, 2025). Most contemporary risks are now deeply shaped by media dynamics, as digital platforms and online environments amplify their visibility and impact. Within this complex communicative landscape, institutions play a pivotal role in both prevention and management of crisis, working to inform and guide citizens amid an overload of messages often blurred by disinformation (Lovari & Bowen, 2020; OECD, 2021).

In this context, the special issue "Risk (of) Communication: Ethics, Futures, and Strategic Complexity" investigates how communication of risk can itself become a source of risk, ethical tension, and strategic uncertainty for institutions, media, and citizens. The title "Risk (of) Communication" reflects the dual analytical perspective that has guided the development of this special issue. On the one hand, it critically addresses the significant expansion of risk communication in contemporary scenarios marked by climate and environmental changes, socio-political turbulence, and the pervasive impact of digital platforms and generative artificial intelligence (La Rocca & Lovari, 2024). Risk communication today involves both public institutions and private organizations and increasingly intersects with crisis, disaster and emergency communication, revealing the fluidity and hybridization of these strategic communication flows. For a long time, such domains have remained marginal in Italy and often confined within rigid disciplinary

* alessandro.lovari@unica.it

boundaries, not only in Italian academia but also at the international level (Diers-Lawson et al., 2024).

On the other hand, this special issue focuses on the risk of communication itself, that is, the risks arising when communication is not strategically planned and managed by organizations in a best-practice approach of ethically managing issues (Bowen *et al.*, 2024). In contemporary communicative ecologies profoundly shaped by social media and digital platforms, such communication practices may have negative impacts on stakeholders, be hijacked by digital publics (Hautala *et al.*, 2025) undermining organizational reputation and triggering crisis scenarios. These challenges call for new competencies, a renewed sense of responsibility and ethical awareness among communication professionals (Bowen & Erzikova, 2025).

By foregrounding this dual perspective, the special issue invites reflections on communication not only as a response to risk or a strategic leverage for communicators, but also as a potential source of risk that must be carefully managed in increasingly complex and hybrid media environments. In particular, the special issue addresses the dual nature of risk (of) communication, ethical implications in crisis contexts, speculative and future-oriented perspectives on natural disasters, and the profound impact of disinformation and artificial intelligence on strategic and organizational communication.

This special issue emerged within the activities of the PRIN 2022 PNRR INSULANDER¹, a national research project - involving the University of Cagliari and University of Enna “Kore” - that investigates the strategic role of risk and crisis communication in insular and structurally vulnerable contexts, like Sardinia and Sicily (Italy). The idea was initiated by Alessandro Lovari (University of Cagliari) and Gevisa La Rocca (University of Enna “Kore”), with the aim of fostering a critical body of research around the transformations of risk and crisis communication models and practices at the international level. Building on this initial insight, the editorial project was further developed by Alessandro Lovari through an enriching collaboration with the Global Strategic Communication Consortium² and it was linked to the 2025 edition of the CONCLAVE, a conference which brought together leading experts and international scholars in the field of strategic communication in Cagliari and Carloforte (Sardinia, Italy). The CONCLAVE, entitled “Risk (of) Communication”, provided a valuable forum for examining the multifaceted dimensions of risk communication within the broader field of strategic communication. It contributed to shaping key research questions related to cascading environmental crises, evolving media ecologies, and the governance of risk communication across different organizational contexts, with particular attention to the emergence of new forms of vulnerability created by insularity, environmental risks, digital and data use hazards, or risks even instigated by the strategy of organizations themselves.

Starting from this dialogue between empirical research and theoretical reflections - and connecting with the main objectives of the INSULANDER project - the articles of this special issue aim at advancing international debate on the challenges and responsibilities of strategic communication in contemporary risk societies³. The 13 manuscripts, authored by 27 scholars working in 16 universities across Europe, Asia, and the United States, collectively rethink risk and crisis communication across environmental disasters, AI-driven infrastructures, geopolitical conflicts, organizational practices, and normative ethics.

The first set of papers proposes new theoretical lenses and models for understanding risk communication under conditions of cascading crises and mega-crises. Brett Robertson introduces the concept of “speculative risk communication” arguing for an interdisciplinary, future-oriented agenda, that treats communication as a core infrastructure capable of anticipating breakdowns across interdependent systems. This contribution highlights not only the anticipatory power of communication but also its ethical and political stakes, calling for the dismantling of disciplinary silos to build coherent and responsible communication futures.

Marta Lukacovic and Deborah Sellnow-Richmond propose the concept of “strategic communication literacy” (SCL), designed to empower diverse and marginalized stakeholders to participate in mitigating risks that escalate into AI-exacerbated mega-crises. Positioned at the intersection of media literacy, AI literacy, and culture-centred approaches, SCL is proposed as both an educational and structural agenda aimed at strengthening voice infrastructures and countering global challenges, while emphasizing the necessity of broad, interdisciplinary coalitions to achieve systemic change.

Bianca Persici Toniolo and Gisela Gonçalves contribute with a systemic and integrative Loop Model for risk and crisis communication in wildfire disasters, developed through an extensive literature review and applied to the Portuguese context. Their framework, complemented by an Informational Quality Indicator (IQI), offers both theoretical refinement and practical tools for assessing whether public organizations prioritize public protection over political reputation. The model is also adaptable to a broader range of climate and weather-related emergencies.

Within the field of environmental risk, another article focuses on the relational dynamics among different actors involved in communicating natural disasters, particularly in insular contexts. Alessandro Lovari and Daniela Pisu analyse the complex relationship between journalists and public communicators in the coverage of risks and natural disasters in Sardinia and Sicily, revealing a hybrid model in which formal procedures intersect with informal practices. Their qualitative study shows that while collaboration is perceived as necessary by both professional groups, it is constrained by bureaucratic regulations, journalistic routines, algorithmic pressures, logics of visibility and sensationalism. Thus, journalists and public communicators emerge as “imperfect allies,” jointly responsible for providing timely and reliable information, yet simultaneously exposed to ethical risks and disinformation that may further endanger vulnerable territories.

Taken together, this first set of papers shows that environmental risks and disasters are not only material events, but also communicative constructions shaped by infrastructures, professional relationships, and normative expectations. These contributions foreground both innovative conceptual tools and concrete models aimed at making risk communication more anticipatory, inclusive, and accountable in the face of escalating environmental and climate crises.

The second cluster of contributions explicitly addresses artificial intelligence, datafied infrastructures, and digital platforms as both mediators and generators of risk for organizations, publics, and democratic systems. Kulsawasd Jitkajornwanich and Kerk Kee develop a framework for “smart alerts” explaining how AI and social media analytics are

reshaping crisis communication by accelerating detection and personalization, while simultaneously generating secondary technological risks such as bias, false positives, privacy violations, and deepfakes. Their speculative design approach and technical case study of Twitter(X)-based flood detection in Thailand identify distinct models of human-AI collaboration and propose governance mechanisms including graduated confidence communication, multi-source verification, and adaptive oversight architectures. These mechanisms seek to rebalance the speed-accuracy dilemma, reframe temporality in predictive messaging, and sustain public trust in AI-enabled warning systems, thereby articulating the dual risk posed by both hazards and their mediated communication.

Qi Chen and Yicheng Zhu analyse how AI-based fact-checking interfaces influence the processing of fabricated foreign-risk misinformation, focusing on the interplay between verification source (human vs AI), blind nationalism, conspiratorial orientation, and third-person perception. Their experimental study shows that blind nationalism, rather than conspiratorial orientation, predicts perceived social norms surrounding foreign-risk misinformation, which in turn shape third-person perception and information-seeking intentions, with interface design moderating these psychological pathways.

Anli Xiao and Yang Cheng examine how international companies should respond to misinformation para-crises on social media amid geopolitical conflict, considering contextual priming (positive vs negative portrayals of the country of origin), crisis narrative type, and communication strategy. Their findings suggest that narrative responses and accommodative strategies may be more effective than defensive, informative responses, particularly when geopolitical tensions and media representations influence public judgments. The study thus underscores the strategic risks inherent in communicative choices during politically charged crises.

Several contributions explicitly connect the risk (of) communication to broader social problems such as gender inequality, institutional abuse, and cyber harms. Elina Erzikova, Shannon Bowen, Lana Ivanitskaya, Kerk Kee, and Mary Beth West reconceptualize gaslighting as an organizational and strategic communication risk, moving beyond interpersonal abuse to examine how institutions obscure truth, discredit critics, and normalize doubt to preserve power. Their bibliometric analysis of 235 publications maps a fragmented yet expanding field of gaslighting research and identifies four thematic clusters encompassing structural inequities, identity and power, emotional harm, and control and help-seeking. Building on this mapping, the authors propose the Organizational Gaslighting as a Strategic Communication Risk (OGSCR) framework, highlighting how organizations generate risks not only by communicating about issues, but through opaque and manipulative communication practices themselves.

Katerina Tsetsura, HM Murtuza, Mark Raymond, and Typhaine Joffe focus on deepfake-related initiatives in G7 countries, the United Nations, and the European Union, applying the Social Amplification of Risk Framework to assess how regulatory efforts address AI-generated misleading content in crises such as natural disasters. Their comparative analysis shows that while existing and proposed regulations emphasize responsible innovation and transparency, they largely fail to provide concrete strategies for managing the harms of

deepfakes in crisis scenarios or for mitigating their potential to escalate conflicts and distort public perception.

Gea Ducci, Lucia D'Ambrosi, Paola De Rosa, Camilla Folena, and Marica Spalletta examine the risks associated with gender-sensitive communication strategies in the European public sector, situating their analysis within a broader understanding of public communication as a strategic tool for democracy, trust, and inclusion. Analysing four EU campaigns conducted between 2020 and 2024, they identify structural, socio-cultural, and political risks that constrain the transformative potential of these initiatives, highlighting tensions between ambitious gender equality policies and polarized media environments in which gender issues are heavily contested.

The final set of papers moves into explicitly political and ethical terrain, where risk communication intersects with war, peace, and the moral language of transparency. Juyan Zhang analyses the Biden administration's rhetoric on China policy through the Mixed-Motive Model of Public Relations, identifying a shift from a purely advocacy-oriented "decoupling" stance toward a more nuanced combination of competition and cooperation. Using Burkean rhetorical criticism, the study demonstrates how cooperative rhetoric functions as a form of risk communication aimed at preventing US-China competition from escalating into armed conflict, articulating "God" and "Devil" terms that anchor a complex strategic narrative. The paper illustrates how foreign policy discourse manages risk not only through policy decisions, but also through carefully constructed communicative frameworks that align liberal, realist, and constructivist worldviews.

Gregg Simons offers a critical analysis of Western war-justifying discourse, starting from the George Orwell's reflections on doublespeak and deceptive political language. Focusing on the Global War on Terrorism, the Arab Spring and the Ukraine conflict, the article shows how emotional appeals, selective values, and vague notions such as "humanitarian war" operate as ritualistic and symbolic mechanisms for engineering consent to military actions that may ultimately undermine public security.

Shannon Bowen's contribution closes the issue critically by interrogating the concept of transparency, a topic widely invoked in media ethics and strategic communication but often weakly theorized and, in some cases, ethically problematic. Drawing on an interdisciplinary literature review and moral philosophy, the study highlights persistent scholarly concerns from moral philosophy regarding the uncritical use of transparency as an ethical principle. Rather than constituting an autonomous moral construct, transparency emerges as an amoral practice whose ethical value depends on the intentions that guide its use. As such, relying on transparency alone may represent a risk of communication, particularly in organizational and institutional contexts, where ethical responsibility is at stake. The article argues that transparency can equally serve ethical or unethical ends and should therefore be conceptually subordinated to stronger moral foundations, such as good intention, duty, dignity, and respect.

Across these thirteen contributions, the special issue highlights that the risk (of) communication is not a marginal concern but a defining feature of contemporary societies, in which environmental disasters, AI-driven tools, geopolitical conflicts, and organizational practices are mediated through digital platforms. The papers collectively show how

communication can anticipate, create, resolve, or amplify crises, empower or marginalize publics, legitimize or contest power, and clarify or obscure ethical obligations thus situating risk communication at the core of strategic and societal futures. Together, these contributions invite scholars and practitioners to rethink risk communication as a deeply ethical, future-oriented, and strategically complex practice that requires interdisciplinary collaboration and renewed attention to power, justice, duties, and responsibility.

The editor extends his sincere thankfulness to all the authors and reviewers involved, as well as the two research teams of the project INSULANDER (University of Cagliari and University of Enna “Kore”) for their support for the special issue. Moreover, he wants also to express his special gratitude to Giovanni Boccia Artieri for accepting the idea of this special issue, as well as to Paolo Brescia and Stefano Brilli for their collaboration, dedication and editorial support.

Biographical note

Alessandro Lovari is an Associate Professor of Sociology of Communication at the Department of Political and Social Sciences, University of Cagliari (Italy) where he teaches Public Communication, and Institutional Communication, Digital Technologies and Media Relations (graduate program). He is the coordinator of the PhD program in Research and Social Innovation and PI of several national research projects on risk and emergency communication in environmental disasters (INSULANDER) as well as on AI impact on communication and citizens (IATI). Lovari's research focuses on public sector communication, public relations, and health communication, investigating digital technologies' impact on organizations' practices and citizens' behaviors in ordinary and crisis situations. He was a visiting scholar at Purdue University, University of Cincinnati, University of South Carolina, Virginia Commonwealth University (USA), and Jyväskylä University (Finland). He is author of more than hundred publications in monographs, book chapters and journal articles.

References

- Beck, U. (1992). *Risk society: Towards a new modernity*. Sage.
- Bowen, S. A., & Erzikova, E. V. (Eds.). (2025). *Handbook of innovations in strategic communication: AI, futurism and directions*. Edward Elgar Publishing.
- Bowen, S. A., Lovari, A., & Timmermans, A. (2024). Issues management: A new model for a key function in public affairs. In A. Timmermans (Ed.), *Research handbook on public affairs* (pp. 282–297). Edward Elgar Publishing. <https://doi.org/10.4337/9781803920283.00031>
- Covello, V. T., von Winterfeldt, D., & Slovic, P. (1986). Risk communication: A review of the literature. *Risk Abstracts*, 3(4), 171–182.
- Diers-Lawson, A., Schwarz, A., Meissner, F., & Ravazzani, S. (Eds.). (2024). *Risk and crisis communication in Europe: Towards integrating theory and practice in unstable and turbulent times*. Routledge.

- Giddens, A. (1999). Risk and responsibility. *The Modern Law Review*, 62(1), 1–10. <https://doi.org/10.1111/1468-2230.00188>
- Hautala, M., Luoma-aho, V., & Brown, J. C. (2025). Communication hijacking: Strategic communication gone dark. *Journal of Communication Management*. Advance online publication. <https://doi.org/10.1108/JCOM-09-2024-0171>
- Heath, R. L. (2025). Strategic communication beyond the amphitheater: Future strategic issues management. In S. A. Bowen & E. V. Erzikova (Eds.), *Handbook of innovations in strategic communication: AI, futurism and directions* (pp. 13–26). Edward Elgar Publishing.
- Heath, R. L., & O'Hair, H. D. (Eds.). (2010). *Handbook of risk and crisis communication*. Routledge.
- La Rocca, G., & Lovari, A. (Eds.). (2024). *Comunicazione del rischio insulare: Prospettive in comunicazione, politiche pubbliche e analisi dei contesti*. FrancoAngeli.
- Lovari, A., & Bowen, S. A. (2020). Social media in disaster communication: A case study of strategies, barriers, and ethical implications. *Journal of Public Affairs*, 20(1), e1967. <https://doi.org/10.1002/pa.1967>
- Organisation for Economic Co-operation and Development. (2021). *OECD report on public communication: The global context and the way forward*. OECD Publishing. <https://doi.org/10.1787/22f8031c-en>
- Sellnow, T. L., & Sellnow, D. D. (2024). *Before crisis: The practice of effective risk communication*. Cognella Academic Publishing.
- Schwarz, A., Seeger, M. W., & Kim, S. (Eds.). (2025). *The handbook of international crisis and risk communication research* (2nd ed.). John Wiley & Sons.

Notes

¹ The research project PRIN 2022 PNRR INSULANDER (INvestigating the Strategic role of commUnication for resiLient islAnds copiNg DisastErs Risk management), is funded by the European Union – Next Generation EU, PNRR, Mission 4, Component 2, Project Code: P2022A4YAL.

² The mission of the Global Strategic Communication Consortium is to move strategic communication forward by focusing research and collaborations on future challenges, technological innovations, crises, and ethics in the field. Members are leading scholars from five continents who attend the annual Conclave as a venue for intense discussion about research, as well as collaborative thought leadership to drive the field forward. More information can be found at https://sc.edu/study/colleges_schools/cic/journalism_and_mass_communications/conference_symposium/gsc/

³ Several participants of the 2025 CONCLAVE, and authors of this special issue, contributed in Carloforte to a specific Delphi-like session related to the Working Package 7 of the PRIN PNRR INSULANDER (Discussion of communities' guidelines for disaster risk communication with international experts in order to extend their possible impact and application to other Mediterranean islands).

Anticipating Breakdown: A Call for Interdisciplinary Futures in Cascading Crises and Speculative Risk Communication *

Brett W. Robertson^{**}
University of South Carolina

This essay calls for an urgent reorientation of risk communication research and practice toward interdisciplinary futures in an era of cascading crises. Hazards today unfold across interdependent infrastructures, multiplying disruptions that no single discipline can address in isolation. The concept of speculative risk communication captures this challenge by emphasizing anticipation: messages that project possible breakdowns into the present, guiding publics to prepare before crises fully materialize. However, anticipatory communication also carries ethical, political, and strategic stakes that demand an integrated perspective. The central argument of this research agenda is that communication must be treated as core infrastructure, sustaining trust and coordinating action across strained systems. Knowledge infrastructures that often reinforce disciplinary silos must instead incentivize collaboration and cooperation. This essay issues a call to action for scholars, practitioners, and institutions: dismantle silos, integrate insights, and build interdisciplinary futures capable of anticipating breakdown with coherence and ethical responsibility.

Keywords: Cascading crises, speculative risk communication, interdisciplinary collaboration, risk communication

* Article submitted on 22/09/2025. Article accepted on 18/12/2025.

** br31@mailbox.sc.edu

The twenty-first century has ushered in a new risk environment defined by cascading crises. Hazards once treated as isolated incidents now unfold in interconnected sequences that strain infrastructures, institutions, and publics (Pescaroli & Alexander, 2018). A wildfire may destroy homes and landscapes, but its consequences rarely stop there. Smoke spreads across regions, triggering respiratory illnesses and straining hospitals. Roads close, delaying emergency response. Electricity grids falter, cutting power to households that depend on cooling or medical equipment. Digital networks are filled with contradictory claims and misinformation, complicating protective action. What begins as a hazard expands into multiple disruptions across overlapping systems.

Similar dynamics unfold across hazards as varied as floods, hurricanes, heatwaves, and pandemics. Each illustrates how disruption in one system cascades into failures across others, multiplying consequences through interdependencies that are difficult to predict and even harder to manage. Scholars of hazards, geography, and systems science use terms such as cascading crises, compound events, or concatenated disasters to capture this interdependence. These frameworks emphasize how hazards trigger chains of consequences that amplify one another (Cutter, 2021). Related scholarship has described large-scale, system-wide emergencies as “mega-crises,” emphasizing their scope, duration, and capacity to overwhelm governance and response systems (Helsloot et al., 2012). Cascading crises overlap with this concept but differ in analytic emphasis. While mega-crisis highlights magnitude and institutional overload (Sellnow-Richmond & Lukacovic, 2026), cascading crises foreground the relational dynamics through which disruption propagates across interdependent infrastructures, organizations, and information systems. This focus on interdependency and sequence is central to understanding how breakdown unfolds and how communication must anticipate strain across systems rather than respond to a singular, bounded event. Focusing on relational propagation also clarifies why the metaphor of dominoes oversimplifies the phenomenon, since collapse rarely follows a straight line. Failures emerge through dense networks of dependency where one strain produces multiple fractures. Energy systems rely on water resources, hospitals require electricity, and information systems depend on both. A single disruption can trigger secondary breakdowns that exceed the original hazard in severity (Brunner et al., 2024).

This complexity presents a challenge for risk communication. Messages about hazards do not circulate in isolation. They are interpreted under stress, filtered through cultural and organizational trust, and refracted across digital platforms (Tierney, 2007). Confusion or contradiction in messaging can accelerate cascading dynamics by slowing protective action or undermining organizational credibility (Stephens et al., 2013). Warnings now spread instantly across multiple media channels, and missteps circulate far beyond intended audiences (Bean et al., 2016). Communication plays a central role in shaping how hazards are understood, how risks are prioritized, and how the public decides to act. Clarity, timeliness, and accuracy remain essential, but they are insufficient when infrastructures themselves are unstable. The public must be prepared not only for hazards but for the possibility that protective systems will fail (Stephens & Faust, 2026). Advisories about hospital capacity limits, warnings of potential power shortages, and guidance on evacuation

delays illustrate this shift. These are messages about strain, uncertainty, and possible failure.

To capture this communicative transformation, I use the term speculative risk communication. The literature generally defines speculative communication as the warnings, advisories, and calculated silences that circulate when infrastructures approach their limits (Boin & 't Hart, 2003; Pang et al., 2021; Palen et al., 2007). However, when considering risk perception, speculative communication enables us to project potential futures of breakdown into the present (Fischhoff, 1995). This form of communication asks the public to prepare for possibilities that may not yet materialize but demand attention (Watanabe & Fujimi, 2025). Speculative risk communication centers on anticipating the impact of hazards alongside the limits, breakdowns, and cascading failures of the infrastructures designed to manage them. It foregrounds what might occur when protective systems reach their breaking point (Ross et al., 2024).

Risk communication has long been understood as forward-looking, insofar as it addresses hazards that may or may not materialize into crises (Krimsky, 2007; Seeger, 2025). From this perspective, all risk communication entails some degree of uncertainty and projection. Speculative risk communication does not dispute this foundation. Instead, it identifies a distinct mode of risk communication that becomes salient when institutions must communicate not only about hazards, but about the potential failure of the protective systems themselves. This distinction matters because communicating the possibility of infrastructural breakdown introduces qualitatively different ethical, political, and strategic challenges than communicating probabilistic risk alone. Unlike crisis communication, which responds to realized breakdown, speculative risk communication operates in the liminal space whereby institutions must acknowledge strain and uncertainty without confirmation that failure has occurred.

Prior scholarship has also advanced the concept of multidimensional risk communication to describe how public discourse integrates multiple domains of risk during emerging crises, particularly in complex socio-technical environments (Gui et al., 2018). This work usefully foregrounds complexity and the interaction of diverse risk dimensions. Speculative risk communication differs in analytic focus. It does not center on the number or variety of risks communicated, but on how institutions invite the public to anticipate cascading breakdown across interdependent systems before crisis conditions fully unfold. The emphasis is not on additive complexity, but on anticipatory projection under infrastructural strain, where communication prepares publics for the possible failure of protective systems rather than discrete hazards alone. At its core, speculative risk communication refers to anticipatory messaging that prepares publics for the possible failure of protective systems under cascading strain, as well as communicating probabilistic hazards.

The stakes of speculative risk communication are sizable. Ethical questions arise when deciding which risks to emphasize and which to minimize. A message that stresses energy demand may signal to some residents that rolling blackouts are imminent, while others assume they are unaffected. A warning about strained hospital capacity suggests that some patients may not receive timely care. Each message reflects institutional judgments about whose safety matters, what uncertainties are tolerable, and how responsibility is distributed.

These judgments are ethical choices, even when delivered in technical language (Barbour & Manly, 2016; Kasperson, 2014).

Speculative risk communication also constructs futures by guiding the public to imagine breakdown and act in anticipation of strain. Institutions must present these futures credibly and persuasively, striking a balance between uncertainty and actionable guidance (Sellnow & Seeger, 2013). Too much emphasis on failure risks fatalism, while too little leaves people unprepared (Mileti & Sorensen, 1990; Paten, 2008). The challenge lies not only in tone but in anticipating how messages circulate across networks. Extant research shows that institutional messages rarely move directly to individuals, instead diffusing through journalists, social media platforms, civil society groups, and interpersonal networks, where meaning is reshaped through processes of amplification, reframing, and selective retransmission (Fellenor et al., 2018; Houston, 2012; Sutton et al., 2020). These patterns are especially pronounced among low-trust publics, who may rely on alternative intermediaries such as social media influencers rather than official sources (Schmuck & Harff, 2023). A message crafted for one audience may be amplified, reframed, or distorted by another (Stephens et al., 2020). Cascading crises magnify this complexity, as hazards, infrastructures, and information systems intersect in unpredictable ways. These dynamics underscore the themes of ethics, futures, and strategic complexity central to this special issue. Speculative risk communication highlights the ethical stakes of anticipatory messaging, the projection of possible futures into the present, and the challenge of guiding publics across multiple systems under stress (Bowen, 2016). Addressing this challenge requires confronting a deeper problem: the persistence of disciplinary silos.

Despite the interdependence of hazards and infrastructures, academic research remains segmented. Engineers and planners model vulnerabilities in physical systems, but rarely examine how the public interprets warnings. Social scientists study trust and behavior, yet often fail to incorporate technical constraints in energy grids or health systems. Health experts design protective measures but may overlook how misinformation distorts uptake. Communication scholars analyze meaning-making under stress but frequently neglect spatial or infrastructural data. Each field generates valuable insights, yet each remains partial. The fragmentation of disciplines mirrors the fragmentation of infrastructures. Silos of expertise replicate the very conditions that cascading crises exploit.

The concept of cascading crises also builds on earlier scholarship on risk society, which emphasized the growing complexity, interdependence, and reflexivity of modern risks (Beck, 1992). Beck argued that contemporary hazards increasingly emerge from the very systems designed to manage them, producing uncertainty that exceeds traditional institutional control. Cascading crises make this condition empirically visible by tracing how failures proliferate across infrastructures, organizations, and communication systems. Thus, cascading crises operationalize risk society by specifying how abstract systemic risk materializes through interconnected breakdowns that demand anticipatory communication rather than reactive response (Beck, 1999; Beck et al., 1994).

This fragmentation cannot continue. The risks of our time demand approaches that integrate across disciplinary boundaries. Cascading crises ignore academic divisions of labor, and knowledge must be created across fields in the same way that infrastructures

operate across systems. The future of risk communication depends on intentional, sustained, and central interdisciplinary collaboration. This requires a transformation in how research is conceived and practiced. The urgency of this shift is evident: communication functions as infrastructural upkeep. Messages sustain the connection between institutions and the public. They maintain trust, signal readiness, and distribute responsibility. When communication fails, infrastructures unravel more quickly. Treating communication as core infrastructure underscores the importance of interdisciplinarity (Stephens & Faust, 2026). Engineers cannot maintain power grids without understanding how the public responds to conservation advisories. Health experts cannot manage pandemics without insight into how trust and misinformation shape compliance. Communication scholars cannot theorize warnings without accounting for technical and institutional constraints. Each discipline depends on the others.

This essay advances the concept of speculative risk communication to argue that anticipating breakdown requires interdisciplinary futures. By framing communication as infrastructural upkeep and by recognizing that cascading crises expose the limits of siloed expertise, the essay calls for a new agenda in risk communication research.

The guiding research question behind this agenda is: *How can speculative risk communication be conceptualized as an interdisciplinary practice for anticipating breakdown in cascading crises?* This question underscores the importance of collaboration across disciplines. It asks scholars to confront the limits of disciplinary approaches and to imagine futures in which communication is integrated with technical, social, and institutional systems.

This manuscript presents a focused research agenda with a conceptual and integrative analytic approach. It advances theory through critical synthesis across risk communication, crisis communication, hazards research, infrastructure studies, and ethics. The analysis uses illustrative cases from recent disasters and public health crises to reveal recurring patterns of cascading breakdown and communicative strain. These cases serve as analytic reference points that sharpen theoretical insight, selected not for representativeness but for their capacity to reveal how anticipatory communication operates under cascading infrastructural strain across different domains. The agenda is guided by explicit analytic commitments to conceptual coherence, interdisciplinary integration, and normative relevance, which structure the synthesis and clarify the basis for assessing the theoretical contribution. As a research agenda, this work contributes by reframing problems, identifying gaps across disciplines, and articulating priorities for interdisciplinary futures in risk communication.

The remainder of the agenda develops this question in three parts. First, it outlines the nature of cascading crises and the consequences of disciplinary fragmentation. Second, it elaborates the features of speculative risk communication, emphasizing its ethical, political, and strategic stakes as a hinge concept for anticipating breakdown. Third, it issues a call to action for scholars, practitioners, and institutions to build interdisciplinary futures by treating communication as core infrastructure.

Cascading Crises and the Limits of Siloed Knowledge

Building on this framing, cascading crises expose the vulnerabilities of infrastructures, institutions, and information systems that appear stable when hazards remain isolated. Once disruption begins, it rarely stops within the boundaries of a single domain. Cascading crises encompass both the cascading consequences of an initial hazard and multi-hazard or compound events, where multiple hazards interact in sequence or simultaneously, as in the 2011 earthquake–tsunami–nuclear disaster in Fukushima. This understanding aligns with models of cascading effects that conceptualize disasters as sequences of interdependent events triggered by an initial hazard, producing amplified impacts across physical, social, and infrastructural systems due to their interconnections (Pescaroli & Alexander, 2015; Zuccaro et al., 2018). A wildfire may destroy homes and landscapes, and its impact radiates outward. Smoke spreads across states, causing respiratory illnesses and overwhelming hospitals already working at capacity. Roads close, cutting off mobility for both emergency responders and ordinary residents. Electricity grids weaken under stress, leaving households without cooling or medical devices. Digital networks are filled with contradictory claims, misinterpretations, and misinformation, eroding the clarity of official warnings. In this sequence, the hazard serves as the catalyst, and the crisis emerges when multiple systems fail simultaneously (Cutter, 2018).

Similar dynamics occur across a wide range of hazards. A flood may disable a water treatment plant, leaving residents reliant on bottled supplies at the exact moment transportation systems are disrupted, and emergency deliveries are delayed. A heat wave strains electricity networks, forcing power companies to issue conservation advisories as cooling systems fail and hospitals admit a rising number of patients with heat-related illnesses. Hurricanes destroy critical infrastructure and disrupt health services, while also degrading the communication channels used to coordinate protective actions. Pandemics overwhelm health systems, and their impact extends into economies, governance, and public life. In each case, the hazard initiates the sequence, and the crisis unfolds when consequences accumulate across interdependent systems (Brunner et al., 2024).

These examples illustrate a defining characteristic of contemporary risks: failures rarely remain confined to one system. They move across infrastructures, institutions, and information systems in ways that are difficult to anticipate and even harder to manage. Scholars use terms such as cascading crises, compound events, and concatenated disasters to describe this interdependence (Alexander & Pescaroli, 2019; Cutter, 2018). Each label highlights how hazards interact with and amplify one another. The metaphor of falling dominoes oversimplifies the complexity of collapse. Crises rarely unfold in a straight line. They ripple through dense networks of dependency where one strain generates multiple fractures. Energy systems depend on water availability. Hospitals require electricity for life-saving equipment. Information systems rely on both. A single disruption can trigger secondary failures that surpass the severity of the original hazard.

Historical Lessons in Fragmentation

The 2021 blackout in Texas highlighted how the failure of one system can cascade across others (Busby et al., 2021). As the power grid faltered during extreme cold, millions of residents lost heat and electricity. Hospitals struggled to maintain care as backup generators failed. Water treatment facilities froze, forcing boil-water advisories. Conflicting information about outage durations and the reliability of backup systems circulated online, leaving residents unsure how to respond. Engineers identified the technical causes of the grid failure, health experts warned of hypothermia and carbon monoxide poisoning, and communication professionals scrambled to issue guidance. Each account contributed to the story, yet the lived reality was a convergence of all these failures (Flores et al., 2023). No single field could capture the crisis in its entirety.

The 2021 Texas power blackout offers substantial analytical value because it combines large-scale failure with detailed empirical evidence on communication during cascading infrastructure collapse. Survey research by Day et al. (2024) showed how power outages disrupted access to official messaging and complicated compliance with boil water notices, with perceived efficacy, income, and race shaping protective action. Infrastructure-focused studies further demonstrated that managed outages and system failures disproportionately harmed low-income and minority communities, clarifying how speculative advisories about energy conservation and water safety interacted with existing social vulnerabilities (Lee et al., 2021). Crisis communication data extended this picture beyond the immediate event, as analysis of crisis text line usage reveals sustained increases in mental health distress in high-impact regions for months after the blackout (Sugg et al., 2023). These studies show how speculative risk communication functions during cascading failure by projecting uncertainty about system capacity, unevenly allocating the burdens of anticipation, and carrying communicative consequences across time and social structure.

The COVID-19 pandemic revealed another dimension of fragmentation. Health experts generated protective guidance on masking, distancing, and vaccination (Malecki et al., 2021). Social scientists studied patterns of compliance, distrust, and politicization (Dada et al., 2021). Communication scholars analyzed misinformation flows across digital platforms (Yoo et al., 2023). Each line of inquiry produced valuable insight, but often without sufficient integration. As a result, protective guidance sometimes failed to account for the ways misinformation reshaped public decision-making (Zuo et al., 2025). Communication analyses highlighted the distortions of public discourse but often did not fully account for the organizational and infrastructural pressures driving health recommendations. The result was a fragmented response to a broad-based crisis.

The COVID-19 pandemic provides a powerful case for this research agenda, given its extensive empirical record of anticipatory communication under prolonged systemic strain. Studies using surveys, content analysis, platform data, and mixed-methods approaches show how the public interpreted and acted on forward-looking guidance on masking, vaccination, hospital capacity, and uncertainty (Dekker et al., 2025; Zhang et al., 2025; Unlu et al., 2023). Anticipatory messaging circulated across multiple channels simultaneously,

shaping perceptions and behaviors in distinct ways. Authoritative media reinforced institutional trust while reducing perceived risk. Social media heightened risk perception while fragmenting credibility. Platform-specific dynamics drove uneven patterns of mistrust and compliance. This body of research shows how speculative risk communication during COVID-19 brought imagined futures of breakdown into the present and unevenly distributed the burdens of anticipation across publics. As an illustrative case, COVID-19 reveals how speculative risk communication operates where infrastructure strain, uncertainty, trust, and circulation converge, underscoring the need for interdisciplinary approaches to understand how anticipatory messages shape behavior before crises fully unfold.

European heat waves underscore the costs of disciplinary silos. Meteorologists and climate scientists forecast temperature extremes with increasing precision (Korhonen et al., 2025). Urban planners model the vulnerabilities of infrastructure under sustained heat. Health experts warn of increased mortality among at-risk groups. Each of these insights often circulates separately. Communication about heat risks tends to emphasize immediate protective measures, such as staying hydrated or avoiding outdoor activities, while overlooking infrastructural limitations, including inadequate housing or overwhelmed healthcare systems (Robertson et al., 2024). The absence of integrated knowledge results in gaps in public guidance, leaving at-risk groups vulnerable.

Empirical research on European heat waves sharpens these dynamics. Survey and experimental studies show high awareness of heat warnings alongside uneven knowledge of current alerts and uneven confidence in how to respond, shaped by demographic, psychological, and contextual factors (Heidenreich et al., 2024). Comparative analyses of heat health warning systems across Europe reveal wide variation in alert thresholds, communication channels, and language accessibility, which weakens anticipatory guidance for mobile and at-risk populations (Casanueva et al., 2019). Governance-focused studies of national heat health action plans show that fragmented roles and unclear cross-sector responsibilities constrain the practice of anticipatory heat risk communication (Vanderplanken et al., 2021). Together, this research demonstrates how speculative risk communication during heat waves distributes future strain unevenly across publics and exposes the costs of disciplinary and institutional fragmentation when anticipating breakdown before systems enter crisis.

Historical disasters reinforce the point. Hurricane Katrina in 2005 revealed how infrastructural breakdown, organizational failure, and communication missteps converged to magnify harm (Leavitt & Kiefer, 2006). Engineers debated the failure of levees. Emergency managers tracked logistical bottlenecks. Communication professionals criticized the conflicting and delayed messages that left residents confused about evacuation orders. Each field produced analyses of what went wrong, but the disaster itself unfolded as a single, converging chain of events. Similarly, Hurricane Maria in 2017 devastated Puerto Rico through physical destruction and prolonged failures of energy, health, and information infrastructures (Andrade et al., 2020). Months after the storm, inconsistent communication about death tolls, service restoration, and federal response eroded trust. The disciplinary fragments of analysis explained parts of the story, but the crisis itself exceeded those boundaries.

Studies of Hurricane Katrina show why these disasters remain analytically instructive. Failures in information and communication technologies disrupted interorganizational coordination and reduced the public sector networks' ability to adapt as conditions deteriorated (Comfort et al., 2007; Davis et al., 2015). Research on Hurricane Maria shows a similar pattern. The collapse in power and telecommunications infrastructure weakened crisis communication, fueled rumor circulation, and damaged institutional credibility because contingency planning for cascading failures was absent (Andrade et al., 2020). Katrina and Maria stand as critical empirical cases that expose how anticipatory communication falters as systems near failure and demonstrate the need for speculative risk communication that addresses infrastructural strain, organizational uncertainty, and the limits of institutional control.

Limits of Single-Discipline Approaches

Across these empirical cases, siloed research proves insufficient for anticipating breakdown when communication operates under infrastructural strain and uncertainty. Engineers and planners excel at modeling vulnerabilities in physical systems. They can forecast when energy grids may reach their limits or when storm surges may overwhelm levees. These models are sophisticated and indispensable for identifying points of strain. However, they rarely address how the public interprets technical information or how organizational credibility influences compliance with advisories.

Social scientists study trust, behavior, and perception. They analyze how organizational credibility shapes protective action and how social networks facilitate the dissemination of information. Their research explains why some publics respond quickly to official warnings while others resist or reinterpret them. However, this work often lacks grounding in the physical constraints that shape the possible responses. A household may want to comply with evacuation guidance but find that routes are blocked or fuel is unavailable. A community may trust a warning but lack the resources to act on it. Without integrating infrastructural realities, behavioral insights remain incomplete.

Health experts design protective measures that translate technical knowledge into actionable guidance. They develop protocols to reduce exposure, manage disease spread, or protect at-risk groups. These measures often assume that the public can follow instructions as intended. They may not account for how misinformation distorts uptake or how infrastructural collapse renders protective guidance difficult to implement. Advisories to boil water, for example, fail when electricity is unavailable to power stoves.

Communication scholars examine how meaning is constructed in the face of uncertainty. They study framing, amplification, and distortion, revealing how publics interpret and reframe messages under stress. This work sheds light on why some warnings are trusted and others are dismissed. However, communication research often neglects the infrastructural and organizational conditions that shape why warnings are issued in the first

place, why inevitable silences are chosen, and why some messages spread more rapidly than others.

Each of these disciplinary perspectives provides a partial view of cascading crises. The challenge is that they remain siloed. Engineers, social scientists, health experts, and communication scholars generate knowledge that is internally rigorous but externally incomplete. This fragmentation of scholarship mirrors the fragmentation of infrastructures. Silos of expertise replicate the very weaknesses that cascading crises exploit.

Why Interdisciplinarity Is Urgent

The consequences of disciplinary fragmentation are measured in organizational failure and public harm. A blackout becomes a health crisis when hospitals lose power, and patients are left without care. A communication crisis arises when the public receives mixed messages about what to expect and how to prepare. A flood becomes an infrastructure crisis when roads collapse, a health crisis when sanitation systems fail, and an information crisis when digital networks are disrupted. A pandemic becomes a governance crisis when public authorities and related governance organizations struggle to maintain legitimacy and coordinate a response across interdependent public, private, and nonprofit systems, a dynamic documented in comparative analyses of COVID-19 governance across Germany, Sweden, and the United Kingdom (Hanson et al., 2021). In each case, the hazard is only part of the story. Crisis conditions intensify as weaknesses across multiple interdependent systems accumulate and converge, allowing localized disruptions to propagate across infrastructure networks and scale into system-wide strain (Pescaroli & Alexander, 2016).

This reality makes interdisciplinarity a requirement. Cascading crises do not respect disciplinary boundaries, and scholarship cannot afford to maintain them. Knowledge must be generated across fields in the same way that infrastructures operate across systems. A power grid depends on a water supply. A hospital depends on electricity. Risk communication research relies on integrating knowledge across various disciplines, including engineering, health, and other social science fields. Without this integration, research produces insights as brittle as the infrastructures it seeks to explain.

Interdisciplinary collaboration must be intentional, continuous, and central to risk scholarship. It cannot be limited to occasional joint projects or symbolic gestures toward integration. It requires frameworks that actively combine methods, theories, and insights into shared practices. Engineers must understand how the public interprets conservation advisories. Social scientists must recognize the infrastructural thresholds that constrain protective action. Health experts must anticipate how misinformation will shape the uptake of guidance. Communication scholars must situate message design within the technical and organizational systems that issue warnings. Each discipline must rely on the others, not as an afterthought but as a condition of effective research.

Communication makes the urgency of this shift especially visible. It serves as infrastructure maintenance rather than merely as a means of information transfer. Messages

sustain the link between organizations and publics: they maintain trust, signal readiness, and distribute responsibility. When communication falters, infrastructures unravel more quickly. Engineers cannot stabilize a grid if households ignore or misinterpret advisories. Health experts cannot manage a pandemic if misinformation reshapes protective behaviors. Communication scholars cannot analyze warning systems without accounting for the infrastructural thresholds that dictate when alerts are issued. Each discipline holds a piece of the puzzle, but only their integration produces a coherent response. Cascading crises also remind us that risks are system-wide. They reveal that failures in scholarship parallel failures in infrastructure. The silos of academia perpetuate weaknesses in the very systems they seek to secure. The urgency of interdisciplinarity is therefore both intellectual and practical. It is about producing knowledge that reflects the realities of contemporary risks. It is about ensuring that scholarship does not replicate the very conditions of fragmentation that hazards exploit.

Speculative Risk Communication and Interdisciplinary Convergence

Recognizing the limits of siloed knowledge sets the foundation for the next step. If cascading crises demand integrated perspectives, then risk communication must be reconceptualized as a site where interdisciplinary insights converge. This is where the concept of speculative risk communication becomes critical. It highlights how warnings, advisories, and silences about infrastructures under strain serve as anticipatory practices that both depend on and reveal interdisciplinary knowledge. Speculative risk communication is where engineering models, social and behavioral insights, health protocols, and communication strategies intersect. This form of communication demonstrates how breakdown is anticipated, managed, and made meaningful before it occurs entirely. The following section develops this concept as the hinge between cascading crises and interdisciplinary futures. By examining speculative risk communication, it becomes possible to see why collaboration across fields is indispensable for anticipating breakdown in an era defined by cascading risks.

Speculative risk communication differs from traditional notions of hazard messaging because it directs attention toward anticipation rather than the aftermath. The conventional model assumes that risk communication begins once a hazard has been identified and quantified, then moves outward to the public through warnings, advisories, or instructions (Heath & Gay, 1997). Speculative risk communication unsettles this timeline. It introduces a vocabulary of strain, of systems approaching their thresholds, of possible futures pressing against the present. A conservation advisory issued during a heat wave signals not only a current hazard but also the potential collapse of the grid. A notice about hospital bed availability conveys more than just the current capacity. It signals the likelihood that protective measures may fail under demand. These forms of communication do not simply report conditions. They should ideally construct futures, ask the public to prepare for them,

and redistribute responsibility across institutions and residents (Muiderman et al., 2023; Sandstig & Eriksson, 2024).

Understanding speculative risk communication in this way reveals its power as a hinge. It makes visible the interdependence of disciplines that often work in parallel but rarely in concert. Engineers provide the models that predict strain on power networks. Social scientists trace how publics interpret advisories and warnings. Health experts translate exposure thresholds into guidelines for protective behavior. Communication scholars analyze how these guidelines are framed and reframed as they move across networks. The communicative act sits at the intersection of all these practices. A conservation advisory is only meaningful if the public understands and complies with it. Compliance is shaped by trust, cultural context, and material capacity, a consistent finding across risk and crisis communication research (Mileti & Sorensen, 1990; Paton, 2008; Sellnow & Seeger, 2013). Trust, in turn, is shaped by past performance and credibility. Credibility rests on whether the advisory reflects real infrastructural conditions. The advisory itself is a communicative artifact that emerges from engineering analysis, health risk thresholds, social patterns of interpretation, and strategic communication design. No single discipline can account for all these dimensions, and yet the communicative moment depends on their convergence.

The Ethical, Political, and Strategic Stakes of Anticipation

This dependence highlights the ethical stakes of anticipation. Institutions decide which futures to emphasize and which to withhold (Fathollahzadeh et al., 2024). A message about strained hospital capacity might underscore the urgency of vaccination campaigns, but it also risks panic among vulnerable populations. A warning about limited evacuation routes may encourage early departures, but also overwhelm traffic systems before they are necessary. Every anticipatory message involves judgments about uncertainty, responsibility, and consequence. These judgments are rarely framed as ethical choices, but they carry ethical weight. They allocate visibility to some risks while leaving others in silence. They distribute burdens unevenly across populations. They shape whether the public prepares with agency or waits in fatalism. Ethical analysis is therefore inseparable from the practice of speculative risk communication.

One ethical consequence of speculative risk communication lies in how the public processes anticipatory messages under uncertainty. Research on risk perception using the Extended Parallel Process Model (EPPM) shows that when messages emphasize threat without a corresponding sense of efficacy, individuals are more likely to engage in fear control rather than danger control, dismissing or avoiding the message rather than acting on it (Witte, 1992; Witte & Allen, 2000). Speculative risk communication is especially vulnerable to this dynamic because it asks the public to prepare for futures that have not yet materialized and for failures that institutions themselves may not fully control. If anticipatory messages foreground breakdown without articulating feasible pathways for action, they risk overwhelming audiences, eroding perceived self-efficacy, and producing disengagement

rather than preparedness or compliance. The ethical weight of speculative risk communication, therefore, lies not only in what futures are projected, but in whether publics are positioned as capable agents within those futures. Communicating about infrastructural strain carries an obligation to balance uncertainty with actionable capacity, ensuring that anticipation mobilizes adaptive behavior rather than resignation.

Speculative risk communication also highlights the political dimensions of futures (Rabe et al., 2024; Li et al., 2025). Messages do not only report conditions. They shape imagination. A conservation advisory envisions a future where energy scarcity is a possibility. A pandemic guideline envisions a future in which contagion can spread unchecked unless collective action intervenes. These constructions are political because they privilege certain risks, specific interpretations, and particular publics. They emphasize one possible trajectory of crisis while sidelining others (Bhuller & Trevithick-Sutton, 2024). The act of communication does not simply describe reality; it also shapes it. It performs reality by making certain futures actionable. This privileged power to choose one future over another reinforces the need for interdisciplinary collaboration. Without it, speculative risk communication may only highlight the blind spots of a single discipline. A grid advisory framed without behavioral insight is likely to be disregarded or resisted. A health guideline crafted without infrastructural awareness risks impossibility in practice. A message about evacuation that neglects social inequality risks exposing vulnerable populations to heightened danger. The political stakes of anticipation demand the breadth of disciplinary insight that only collaboration can provide.

Speculative risk communication also reveals the strategic complexity of circulation. Research in risk and crisis communication shows that messages rarely travel directly from institutions to individuals, instead moving through journalists, social media platforms, community organizations, and interpersonal ties, where meaning is reframed, amplified, or distorted (Fellenor et al., 2018; Houston, 2012; Sutton & Fischer, 2021; Sutton et al., 2020; Slater et al., 2024). Each stage alters meaning and introduces new interpretations. Anticipating circulation is therefore as important as anticipating breakdown. Institutions must ask not only what to communicate, but also how the communication will be received. Without interdisciplinary perspectives, these questions remain unanswered. Engineers may know the limits of the grid but not how conservation messages resonate with different cultural groups. Social scientists may understand the dynamics of trust but not how technical thresholds drive decisions to issue warnings. Health experts may understand the biological consequences of heat exposure, but they may not be aware of the infrastructural barriers to implementing protective measures. Communication scholars may know how frames circulate but not why institutional silences occur. Anticipating circulation requires integrating all of these insights into a communicative strategy.

From Hinge to Imperative: A Call to Action for Interdisciplinary Futures

The hinge function of speculative risk communication directly leads to a call to action. Interdisciplinary collaboration is essential. It is the condition of effective communication in an era defined by cascading crises. The fragmentation of scholarship mirrors the fragmentation of infrastructure. When each discipline operates in isolation, the result is brittle. The blackout becomes a metaphor for academic silos. The pandemic reflects the costs of fragmented expertise. The hurricane illustrates how disciplinary limits compound public harm. These crises deliver the same lesson: anticipation without integration lacks traction.

At the core of this challenge lie knowledge infrastructures that act as gatekeepers of what scholarship is valued and legitimized. Universities are entrenched sites of disciplinary division, where tenure lines, budgets, and evaluation metrics reward insular excellence while discouraging cross-field collaboration. Funding agencies shape which projects flourish, and when they privilege single-discipline proposals, they reinforce silos. Professional associations and journals also serve as boundary-setters. Editorial criteria, publication practices, and conference structures often replicate disciplinary divisions by marginalizing interdisciplinary work. Together, these three arenas (universities, agencies, and associations) form the institutional framework that either enables integration or perpetuates fragmentation. Recognizing them as interconnected knowledge infrastructures highlights the need for structural change across the entire ecosystem. Without this reorientation, the systems responsible for guiding research and practice amplify the vulnerabilities they seek to study.

Beyond the academy, institutions that manage risks must treat communication as infrastructure. Energy companies must recognize that conservation advisories are lifelines. Hospitals must see bed availability notices as signals of possible collapse. Governments must understand that evacuation advisories are acts of collective imagination. Speculative risk communication functions as the connective tissue across systems: maintaining trust, coordinating action, and distributing responsibility. Institutions must integrate communicative strategy into every stage of risk management. They must staff interdisciplinary teams that combine technical experts, behavioral scientists, health professionals, and communication specialists. They must treat message design as infrastructural planning. They must anticipate circulation as much as they anticipate strain. Without this integration, institutional communication fails under stress.

At the heart of this call to action is a recognition that speculative risk communication is a product of the convergence of disciplines. This concept emerges at the intersection of technical knowledge, behavioral insight, health expertise, and communicative design. Its success will depend on the integrity of this intersection, and its failure could reflect the absence of collaboration. Speculative risk communication is both a concept and a demand. Scholarship must move beyond silos. Institutions must invest in integration. The public must be treated as active participants in shaping their futures. Risk communication must be understood as essential infrastructure maintenance, as vital as the systems it sustains.

The urgency of this demand is escalating. Cascading crises are accelerating. Climate change intensifies the frequency and severity of hazards. Globalization accelerates the spread of failures across systems. Digital platforms amplify the circulation of misinformation on an unprecedented scale. The twenty-first century is already defined by systemic risks that exceed disciplinary boundaries. The future will be determined by whether scholarship and practice effectively address those risks through equal systemic integration. Delay leaves the public vulnerable. Fragmentation replicates failure. Integration anticipates breakdown with coherence and ethical responsibility.

I hope that this call to action presents a challenge to scholars, practitioners, and institutions alike. Studying communication about risks must be paired with practicing it as risk management. Analyzing infrastructures must go hand in hand with treating communication as infrastructure. Disciplinary silos must be critiqued and dismantled in practice. Interdisciplinary futures must be built to withstand cascading crises. This is the purpose of speculative risk communication, and it is the imperative of our time.

Biographical note

Brett W. Robertson, Ph.D., is an Associate Professor in the College of Information and Communications (CIC) and Associate Director of the Hazards Vulnerability and Resilience Institute (HVRI) at the University of South Carolina. His research focuses on disaster preparedness, risk communication, and improving access to critical information during emergencies. He has published extensively on topics such as disaster resilience, emergency communication, and collaborative efforts to strengthen public safety. His work has earned numerous accolades, including the W. Charles Redding Outstanding Dissertation Award from the International Communication Association. In 2025, he received the Michael J. Mungo Undergraduate Teaching Award, the University of South Carolina's highest honor for undergraduate teaching.

References

- Alexander, D., & Pescaroli, G. (2019). What are cascading disasters? *UCL Open Environment*, 1. <https://doi.org/10.14324/111.444/ucloe.000003>
- Andrade, E. L., Barrett, N., Edberg, M. C., Rivera, M. I., Seeger, M. W., & Mercado, R. (2020). Mortality reporting and rumor generation: An assessment of crisis and emergency risk communication following Hurricane María in Puerto Rico. *Journal of International Crisis and Risk Communication Research*, 3(1), 15-48. <https://doi.org/10.3316/INFORMIT.101283164195893>
- Barbour, J. B., & Manly, J. N. (2016). Redefining disaster preparedness: Institutional contradictions and praxis in volunteer responder organizing. *Management Communication Quarterly*, 30(3), 333-361. <https://doi.org/10.1177/0893318916629101>
- Bean, H., Liu, B. F., Madden, S., Sutton, J., Wood, M. M., & Mileti, D. S. (2016). Disaster warnings in your pocket: How audiences interpret mobile alerts for an unfamiliar

- hazard. *Journal of Contingencies and Crisis Management*, 24(3), 136-147. <https://doi.org/10.1111/1468-5973.12108>
- Beck, U. (1992). *Risk Society: Towards a New Modernity*. Sage.
- Beck, U. (1999). *World Risk Society*. Polity Press.
- Beck, U., Giddens, A., & Lash, S. (1994). *Reflexive Modernization: Politics, Tradition and Aesthetics in the Modern Social Order*. Stanford University Press.
- Bhuller, Y., & Trevithick-Sutton, C. C. (2024). Risk communication: lessons from an ethnographic, pragmatic, and Canadian regulatory perspective. *Frontiers in Communication*, 9. <https://doi.org/10.3389/fcomm.2024.1235055>
- Boin, A., & Hart, P. T. (2003). Public leadership in times of crisis: Mission impossible? *Public Administration Review*, 63(5), 544-553. <https://doi.org/10.1111/1540-6210.00318>
- Bowen, S. A. (2016). Clarifying ethics terms in public relations from A to V, authenticity to virtue: BledCom special issue of PR review sleeping (with the) media: Media relations. *Public Relations Review*, 42(4), 564-572. <https://doi.org/10.1016/j.pubrev.2016.03.012>
- Brunner, L. G., Peer, R. A. M., Zorn, C., Paulik, R., & Logan, T. M. (2024). Understanding cascading risks through real-world interdependent urban infrastructure. *Reliability Engineering & System Safety*, 241, 109653. <https://doi.org/10.1016/j.ress.2023.109653>
- Busby, J. W., Baker, K., Bazilian, M. D., Gilbert, A. Q., Grubert, E., Rai, V., Rhodes, J. D., Shidore, S., Smith, C. A., & Webber, M. E. (2021). Cascading risks: Understanding the 2021 winter blackout in Texas. *Energy Research & Social Science*, 77, 102106. <https://doi.org/10.1016/j.erss.2021.102106>
- Casanueva, A., Burgstall, A., Kotlarski, S., Messeri, A., Morabito, M., Flouris, A., Nybo, L., Spirig, C., & Schwierz, C. (2019). Overview of existing heat-health warning systems in Europe. *International Journal of Environmental Research and Public Health*, 16. <https://doi.org/10.3390/ijerph16152657>
- Comfort, L., & Haase, T. (2006). Communication, coherence, and collective action. *Public Works Management & Policy*, 10, 328-343. <https://doi.org/10.1177/1087724x06289052>
- Cutter, S. L. (2018). Compound, cascading, or complex disasters: What's in a name? *Environment: Science and Policy for Sustainable Development*, 60(6), 16-25.
- Cutter, S. L. (2021). The changing nature of hazard and disaster risk in the Anthropocene. *Annals of the American Association of Geographers*, 111(3), 819-827. <https://doi.org/10.1080/24694452.2020.1744423>
- Dada, S., Ashworth, H. C., Bewa, M. J., & Dhatt, R. (2021). Words matter: Political and gender analysis of speeches made by heads of government during the COVID-19 pandemic. *BMJ Global Health*, 6(1). <https://doi.org/10.1136/bmjgh-2020-003910>
- Davis, G., & Robbin, A. (2015). Network disaster response effectiveness: The case of ICTs and Hurricane Katrina. *Journal of Homeland Security and Emergency Management*, 12, 437-467. <https://doi.org/10.1515/jhsem-2014-0087>
- Day, A., O'Shay, S., Islam, K., Seeger, M., Sperone, F., & McElmurry, S. (2024). Boil water notices as health-risk communication: Risk perceptions, efficacy, and compliance

- during winter storm Uri. *Scientific Reports*, 14. <https://doi.org/10.1038/s41598-023-50286-y>
- Dekker, R., Engbersen, G., Snel, E., & De Boom, J. (2025). The effect of institutional trust on the relationship between social media as an information resource and policy non-compliance: Dutch survey evidence from the COVID-19 pandemic. *International Review of Administrative Sciences*, 91, 259-274. <https://doi.org/10.1177/00208523241306410>
- Fathollahzadeh, A., Babaie, J., Salmani, I., Morowatisharifabad, M. A., & Khajehaminian, M. R. (2024). Challenges of disaster risk communication from the perspectives of experts and affected people: A conventional content analysis. *International Journal of Disaster Risk Reduction*, 108, 104553. <https://doi.org/10.1016/j.ijdr.2024.104553>
- Fellenor, J., Barnett, J., Potter, C., Urquhart, J., Mumford, J., & Quine, C. (2018). The social amplification of risk on Twitter: the case of ash dieback disease in the United Kingdom. *Journal of Risk Research*, 21, 1163-1183. <https://doi.org/10.1080/13669877.2017.1281339>
- Fischhoff, B. (1995). Risk perception and communication unplugged: twenty years of process. *Risk Analysis*, 15(2), 137-145. <https://doi.org/10.1111/j.1539-6924.1995.tb00308.x>
- Flores, N. M., McBrien, H., Do, V., Kiang, M. V., Schlegelmilch, J., & Casey, J. A. (2023). The 2021 Texas power crisis: Distribution, duration, and disparities. *Journal of Exposure Science & Environmental Epidemiology*, 33(1), 21–31. <https://doi.org/10.1038/s41370-022-00462-5>
- Hanson, C., Luedtke, S., Spicer, N., Sörensen, J., Mayhew, S., & Mounier-Jack, S. (2021). National health governance, science and the media: drivers of COVID-19 responses in Germany, Sweden and the UK in 2020. *BMJ Global Health*, 6. <https://doi.org/10.1136/bmjgh-2021-006691>
- Heath, R. L., & Gay, C. D. (1997). Risk communication: Involvement, uncertainty, and control's effect on information scanning and monitoring by expert stakeholders. *Management Communication Quarterly*, 10(3), 342-372. <https://doi.org/10.1177/0893318997010003004>
- Heidenreich, A., & Thieken, A. (2024). Individual heat adaptation: Analyzing risk communication, warnings, heat risk perception, and protective behavior in three German cities. *Risk Analysis*, 44, 1788-1808. <https://doi.org/10.1111/risa.14278>
- Houston, J. B. (2012). Public disaster mental/behavioral health communication: Intervention across disaster phases. *Journal of Emergency Management*, 10(4), 283-292. <https://doi.org/10.5055/jem.2012.0106>
- Kasperson, R. (2014). Four questions for risk communication. *Journal of Risk Research*, 17(10), 1233-1239. <https://doi.org/10.1080/13669877.2014.900207>
- Korhonen, N., Balen, D., Prill, F., Salonen, M., Venäläinen, A., & Bär, M. (2024). The probabilistic skill of extended-range heat wave forecasts in Europe. *Natural Hazards and Earth System Sciences*, 25, 1865–1884. <https://doi.org/10.5194/nhess-25-1865-2025>

- Krimsky, S. (2007). Risk communication in the internet age: The rise of disorganized skepticism. *Environmental Hazards*, 7, 157-164. <https://doi.org/10.1016/j.envhaz.2007.05.006>
- Leavitt, W. M., & Kiefer, J. J. (2006). Infrastructure interdependency and the creation of a “normal disaster”: The case of Hurricane Katrina and the City of New Orleans. *Public Works Management & Policy*, 10(4), 306–314. <https://doi.org/10.1177/1087724X06289055>
- Lee, C., Maron, M., & Mostafavi, A. (2021). Community-scale big data reveals disparate impacts of the Texas winter storm of 2021 and its managed power outage. *Humanities & Social Sciences Communications*, 9. <https://doi.org/10.1057/s41599-022-01353-8>
- Li, N., Zeng, W., Yin, S., & Zhao, L. (2025). How risk communication affects public trust in government: The moderating role of policy expectations. *Frontiers in Public Health*, 13, 1557786. <https://doi.org/10.3389/fpubh.2025.1557786>
- Malecki, K. M., Keating, J. A., & Safdar, N. (2021). Crisis communication and public perception of COVID-19 risk in the era of social media. *Clinical Infectious Diseases*, 72(4), 697-702. <https://doi.org/10.1093/cid/ciaa758>
- Mileti, D. S., & Sorensen, J. H. (1990). Communication of emergency public warnings: A social science perspective and state-of-the-art assessment. *Oak Ridge National Laboratory*. <https://doi.org/10.2172/6137387>
- Muiderman, K., Vervoort, J., Gupta, A., Norbert-Munns, R. P., Veeger, M., Muzammil, M., & Driessen, P. (2023). Is anticipatory governance opening up or closing down future possibilities? Findings from diverse contexts in the Global South. *Global Environmental Change*, 81, 102694. <https://doi.org/10.1016/j.gloenvcha.2023.102694>
- Palen, L., Hiltz, S. R., & Liu, S. B. (2007). Online forums supporting grassroots participation in emergency preparedness and response. *Communications of the ACM*, 50(3), 54-58. <https://doi.org/10.1145/1226736.1226766>
- Pang, A., Jin, Y., Seo, Y., Choi, S. I., Teo, H. X., Le, P. D., & Reber, B. (2022). Breaking the sound of silence: Explication in the use of strategic silence in crisis communication. *International Journal of Business Communication*, 59(2), 219-241. <https://doi.org/10.1177/23294884211046357>
- Paton, D. (2008). Risk communication and natural hazard mitigation: How trust influences its effectiveness. *International Journal of Global Environmental Issues*, 8(1), 2-16. <https://doi.org/10.1504/IJGENVI.2008.017256>
- Pescaroli, G., & Alexander, D. (2018). Understanding compound, interconnected, interacting, and cascading risks: A holistic framework. *Risk Analysis*, 38(11), 2245-2257. <https://doi.org/10.1111/risa.13128>
- Pescaroli, G., & Alexander, D. (2016). Critical infrastructure, panarchies and the vulnerability paths of cascading disasters. *Natural Hazards*, 82, 175-192. <https://doi.org/10.1007/s11069-016-2186-3>
- Rabe, L., Sataøen, H. L., Lidskog, R., & Eriksson, M. (2024). Making risk communication in practice: dimensions of professional logics in risk and vulnerability

- assessments. *Journal of Risk Research*, 27(3), 389-403. <https://doi.org/10.1080/13669877.2024.2328199>
- Robertson, B. W., Dow, K., Salinas, J., & Cutter, S. L. (2024). Heat risk perceptions and coping strategies of the unhoused. *International Journal of Environmental Research and Public Health*, 21(6), 737. <https://doi.org/10.3390/ijerph21060737>
- Ross, A. D., Siebeneck, L., Wu, H. C., Kopczynski, S., Nepal, S., & Saucedo, M. (2024). Seven challenges for risk communication in today's digital era: The emergency manager's perspective. *Sustainability*, 16(24), 11306. <https://doi.org/10.3390/su162411306>
- Sandstig, G., & Eriksson, M. (2024). Shared responsibility or not? A responsibility messages experiment during a cyber crisis in designing for risk communication. *Journal of Risk Research*, 27(10), 1163-1179. <https://doi.org/10.1080/13669877.2024.2387351>
- Seeger, M. W. (2025). History of crisis and risk communication: The development of a global community of multi-disciplinary research, theory and practice. In A. Schwarz, M. W. Seeger, S. Kim (Eds.), *The Handbook of International Crisis and Risk Communication Research* (pp. 15-27). Wiley. <https://doi.org/10.1002/9781394180844.ch2>
- Sellnow, T. L., & Seeger, M. W. (2013). *Theorizing crisis communication*. Wiley.
- Sellnow-Richmond, D., & Lukacovic, M. (2026). *Communication amid Mega-Crises*. Routledge.
- Schmuck, D., & Harff, D. (2023). Popular among distrustful youth? Social media influencers' communication about COVID-19 and young people's risk perceptions and vaccination intentions. *Health Communication*, 39, 2730-2743. <https://doi.org/10.1080/10410236.2023.2286408>
- Slater, M. D., Coleman, S., & Freiburger, N. (2024). Using the "media as mediator" approach to understand the influence of communication channel trust on COVID-19 protective behaviors: England, January 2022. *Journal of Health Communication*, 29(10), 635-643. <https://doi.org/10.1080/10810730.2024.2404912>
- Stephens, K. K., Barrett, A. K., & Mahometa, M. J. (2013). Organizational communication in emergencies: Using multiple channels and sources to combat noise and capture attention. *Human Communication Research*, 39(2), 230-251. <https://doi.org/10.1111/hcre.12002>
- Stephens, K. K., & Faust, K. (2026). *The handbook of infrastructure communication*. Wiley.
- Stephens, K. K., Robertson, B. W., & Murthy, D. (2020). Throw me a lifeline: Articulating mobile social network dispersion and the social construction of risk in rescue communication. *Mobile Media & Communication*, 8(2), 149-169. <https://doi.org/10.1177/2050157919846522>
- Sutton, J., & Fischer, L. M. (2021). Understanding visual risk communication messages: An analysis of visual attention allocation and think-aloud responses to tornado graphics. *Weather, Climate, and Society*, 13(1), 1-12. <https://doi.org/10.1175/WCAS-D-20-0042.1>
- Sutton, J., Renshaw, S., & Butts, C. (2020). COVID-19: Retransmission of official communications in an emerging pandemic. *PLoS ONE*, 15. <https://doi.org/10.1371/journal.pone.0238491>

- Tierney, K. J. (2007). From the margins to the mainstream? Disaster research at the crossroads. *Annual Review of Sociology*, 33(1), 503-525. <https://doi.org/10.1146/annurev.soc.33.040406.131743>
- Unlu, A., Truong, S., Tammi, T., & Lohiniva, A. (2023). Exploring political mistrust in pandemic risk communication: Mixed-method study using social media data analysis. *Journal of Medical Internet Research*, 25. <https://doi.org/10.2196/50199>
- Vanderplanken, K., Van Den Hazel, P., Marx, M., Shams, A., Guha-Sapir, D., & Van Loenhout, J. (2021). Governing heatwaves in Europe: comparing health policy and practices to better understand roles, responsibilities and collaboration. *Health Research Policy and Systems*, 19. <https://doi.org/10.1186/s12961-020-00645-2>
- Watanabe, M., & Fujimi, T. (2025). Causal effects of time-sensitive risk communication on disaster preparedness. *International Journal of Disaster Risk Reduction*, 105598. <https://doi.org/10.1016/j.ijdrr.2025.105598>
- Witte, K. (1992). Putting the fear back into fear appeals: The extended parallel process model. *Communication Monographs*, 59, 329-349. <https://doi.org/10.1080/03637759209376276>
- Witte, K., & Allen, M. (2000). A meta-analysis of fear appeals: Implications for effective public health campaigns. *Health Education & Behavior*, 27, 591-615. <https://doi.org/10.1177/109019810002700506>
- Yoo, W., Oh, S. H., & Choi, D. H. (2023). Exposure to COVID-19 misinformation across instant messaging apps: Moderating roles of news media and interpersonal communication. *International Journal of Communication*, 17, 712-734. <https://ijoc.org/index.php/ijoc/article/view/17594/4021>
- Zhang, H., Yang, C., Deng, X., & Luo, C. (2025). How authoritative media and personal social media influence policy compliance through trust in government and risk perception: Quantitative cross-sectional survey study. *Journal of Medical Internet Research*, 27. <https://doi.org/10.2196/64940>
- Zuccaro, G., Gregorio, D., & Leone, M. (2018). Theoretical model for cascading effects analyses. *International Journal of Disaster Risk Reduction*, 30, 199-215. <https://doi.org/10.1016/j.ijdrr.2018.04.019>
- Zuo, T., He, L., Zheng, Y., Lin, Z., Xu, Z., & Li, N. (2025). Communication analysis of the COVID-19 infodemic by medical practitioners in China: A mixed methods study. *Social Science & Medicine*, 374, 118051. <https://doi.org/10.1016/j.socscimed.2025.118051>

Toward ‘*Strategic Communication Literacy*’: Situating the Concept within Crisis Communication Scholarship amid AI-Exacerbated Mega-Crises*

Marta N. Lukacovic**
Angelo State University

Deborah D. Sellnow-Richmond***
Southern Illinois University Edwardsville

This contribution situates the proposed concept of “strategic communication literacy” (SCL) within the field of strategic crisis communication, particularly as it applies amid mega-crises. We postulate that it is vital to define and advance a type of literacy initiative that encompasses communication knowledge and skills among a range of stakeholders, including those who are marginalized. SCL aims to empower stakeholders to participate in efforts to mitigate risks that spiral toward mega-crises, as well as to contribute to structural changes necessary to counter mega-crises, which are the grand challenges faced globally that undermine human well-being, security, and the natural environment. In the contemporary era, AI represents a technological field that exacerbates existing mega-crises but also potentially introduces new security risks. Accordingly, we use AI-exacerbated mega-crises as the primary cases to illustrate our central arguments about SCL. We outline how SCL overlaps with, yet remains distinct from, media literacy and AI literacy. SCL can be further enriched by insights from communication frameworks such as the culture-centered approach. While elevating SCL represents one pathway to strengthening culture-centered voice infrastructures in terms of marginalized stakeholders’ agency, the need for organizational and political change at a structural level is paramount and will require a broad coalition capable of enacting change at the large scale where mega-crises arise. Hence, within the ongoing quest to define and refine the key components of SCL, an inclusive and interdisciplinary approach is absolutely fundamental. Furthermore, we propose a research agenda to refine the concept of SCL and assess its relevance in relation to broader categories of mega-crises.

Keywords: AI risks, critical literacy, marginalized stakeholders, mega-crises, strategic communication

*Article submitted on 25/03/2025. Article accepted on 15/06/2025.

** marta.lukacovic@angelo.edu

*** dsellno@siue.edu

Proliferation of AI and Rising Risks of Mega-Crises

Mega-crises are characterized by far-reaching, even global, consequences, and are composed of complex, entangled, and self-exacerbating sets of issues (Alpaslan & Mitroff, 2011; Sellnow-Richmond & Lukacovic, 2026). As such, mega-crises represent wicked problems of international dimensions and inherent difficulties that impede attempts to counter their negative impacts. Mega-crises thrive on a plethora of systemic failures at different levels, ranging from local to global levels. Moreover, structural changes necessary to mitigate mega-crises are difficult to implement at the scale required. A variety of factors (such as differing legislative jurisdictions, cultural tendencies, political constraints, economic interests, and more) complicate the situation and hinder the kind of complex, coordinated approach vital to meaningfully mitigate or resolve mega-crises.

The already existing crises and looming risks associated with advances in and applications of new artificial intelligence (AI) technologies epitomize a challenge of mega-crisis proportions. For example, the massive amounts of energy and water required to operate supercomputers and data centers for certain contemporary AI models contribute to environmental degradation and climate change, both of which are mega-crises in their own right (Bender et al., 2021; Endter-Wada & Ingram, 2012; Hao, 2025). AI incarnations such as bots and algorithms also contribute to phenomena such as deepfakes that intensify information disorder, including the spread of misinformation, disinformation, and malinformation via social media and content-sharing platforms (Samoilenko & Suvorova, 2023; Wardle & Derakhshan, 2017). Another example of a looming risk is the possible massive disruption to labor markets and rising unemployment if job displacement due to the AI-driven automations is implemented in an unchecked and unregulated manner (Hao, 2025; McChesney & Nichols, 2016). The list of issues and risks tied to current or future mega-crises that AI can exacerbate continues to grow (e.g., Gebru & Torres, 2024; Morozov, 2024; Narayanan & Kapoor, 2024).

Stakeholders Facing AI-Exacerbated Risks and Mega-Crises

The emphasis on the need for extra caution when proceeding with AI is reverberated across academia, punditry, business, the non-profit sector, and government. The late US statesman Henry Kissinger and his coauthors (2021) proposed global multi-stakeholder discussions to establish appropriate guardrails for preventing AI-linked threats. Similarly, proponents of international regulation, such as Moroccan legal scholar Roumate (2024), underscore the importance of multi-stakeholder cooperation in establishing norms and laws around AI to prevent large-scale risks. Thus, across the appeals voiced by international and interdisciplinary leaders, the prevention of AI risks and the mitigation of AI-related crises is dependent upon effective *communication* and *stakeholder* participation.

Kruckeberg and Vujnovic (2025) assert that PR and strategic communication fields will continue to be relevant and possibly become increasingly more pertinent amid AI technology

proliferation. Bowen (2024) argues that public relations (PR) and communication professionals should actively pursue learning about contemporary AI technologies and applications. First, to actively participate on executive decisions on the integration of AI technologies within their organizations. Second, to effectively communicate those AI adoption decisions and implications to all relevant stakeholders. This includes crisis communication, when PR and communication professionals should connect with stakeholders to explain the problem, mitigation measures, support resources, and other key information. With proliferation of integration of these technologies into operations of many organizations, there will be instances when particularly AI tools significantly contribute to exacerbation or even are the key reason behind the emergence of a crisis. Hence, in such instances, AI literacy will be essential for PR and other communication professionals' ability to craft the most accurate and effective crisis communication strategies and tactics. From Bowen's (2024) perspective, PR and communication professionals who are well-versed in AI can play a key role in management and serve as the vital link to stakeholders. Thus, Bowen presents a useful model from organizational and managerial perspectives.

Sellnow-Richmond and Lukacovic (2026) suggest that when addressing AI-exacerbated mega-crises, prioritizing stakeholder agency will be essential. Mega-crises represent such an extensive set of problems that their mitigation cannot be confined to just one or even several organizations. Therefore, organizational and managerial approaches often failed when applied to previous mega-crisis mitigation efforts (Sellnow-Richmond & Lukacovic, 2024). Ultimately, a paramount step toward mitigating mega-crises depends on wide-reaching systemic change. For such change, the participation of diverse stakeholder groups, including vulnerable and marginalized communities, at scale, is imperative.

In this article, we work with the definition of marginalized stakeholders as groups that are systematically excluded from benefits and access to securities and opportunities that other groups enjoy due to their specific social, economic, or political status. Vulnerable stakeholders are groups that are more susceptible to risks, likely to suffer greater damages, or less likely to recover once a crisis has occurred. Marginalization tends to intensify vulnerability, particularly in the context of mega-crises, which are inherently composed of layers of deep-rooted social, economic, environmental, and political issues (Sellnow-Richmond & Lukacovic, 2026).

Thus, it is vital to chart the best possible pathways to enabling stakeholder participation in the mitigation of mega-crises. In crisis communication scholarship, one of the best practices emphasizes stakeholder collaboration and dialogue (Seeger & Sellnow, 2019). However, not all stakeholders can equally participate in a genuine dialogue in many cases. This has been evident in the ongoing mega-crises exacerbated by AI, where those most affected and harmed are rarely given a voice, let alone an opportunity to participate on decision-making (Hao, 2025; Rowe, 2023; West & Kak, 2024).

Lovari and Bowen (2019) emphasize the importance of using social media platforms to engage in two-way communication, allowing public affairs specialists and citizens to partake in dialogue about crisis-related issues. Hence, ensuring physical access to such platforms is one way to overcome the barriers that many face in participating in dialogue. However, even communities with rather robust digital access are often subjected to other disparities,

which further limit their ability to fully utilize available technologies (Robinson et al., 2015). Many of those victimized by today's Big Tech in connection with AI do have physical access to devices and platforms that could facilitate connection (West & Kak, 2024). However, they still lack the necessary knowledge, skills, other key resources, and/or leverage to take part in these conversations and advocate for themselves within genuine, two-way symmetrical communication.

One of the additional disparities limiting stakeholders' ability to partake in dialogues with those in power is the lack of AI literacy. For example, many social media users are unaware that their data have been and continue to be used to train AI systems without compensation for their input (Hao, 2025; Varoufakis, 2023). Another example relates to AI in weapons and warfare. The general public's lack of awareness about the prevalence and deployment of lethal autonomous weapons systems and other war-related uses of AI creates a situation where powerful commercial actors and nation-states are not pressured by the public to regulate or potentially ban these systems (Solovyeva & Hynek, 2023). Indeed, there is a growing international and interdisciplinary call for increased *AI literacy* across populations, professions, and contexts (e.g., Veldhuis et al., 2025).

Even with high levels of AI literacy within a population, various stakeholders will still face significant disadvantages that prevent them from participating in the necessary dialogues with the decision-makers. These disadvantages include the inability to effectively articulate one's concerns in a way that resonates with the listening party, or a lack of skills in using social media, advocacy, and persuasion to ensure their demands are considered in decision-making. Furthermore, it requires specific knowledge and skills to verify whether organizations and institutions are truly enacting the promised changes. If the changes are not materializing effectively, it may be important to pivot, adjust actions, and modify communication tactics. These examples highlight the knowledge and skills related to strategic communication. Hence, we propose the conceptualization of *strategic communication literacy (SCL)* as an essential factor to empower stakeholders amid mega-crises, as evidenced by AI-related tribulations of today. Here, SCL can be defined as a set of knowledge, skills, and abilities that empowers stakeholders to collect and assess incoming messages, as well as to contemplate, develop, and deploy outgoing messages, enabling them to meaningfully and effectively participate in key dialogues around important issues.

Strategic communication theories and practices must continue to evolve to address the needs of stakeholders within information environments increasingly reshaped by algorithms, generative AI, and other emerging technologies (Lovari & De Rosa, 2025). The concept of SCL directly responds to these pressing challenges. It can productively draw on existing scholarship in public relations, strategic communication, crisis communication, as well as complementary theoretical and analytical work across multiple disciplines. In the sections that follow, we position the SCL concept within the most relevant theoretical frameworks, identify existing gaps, and recommend directions for future research and practical application to further refine the concept.

Identifying Suitable Theoretical and Practical Locus of Strategic Communication Literacy towards Mitigation of Mega-crises

First, we must identify the main theoretical dimensions of a general approach to strategic communication that can be effective in addressing mega-crises. Mega-crises are fundamentally rooted in systemic failures that occur across societies at both local and international levels (Alpaslan & Mitroff, 2011; Sellnow-Richmond & Lukacovic, 2026). Consequently, effective mega-crisis mitigation depends on meaningful structural change at these broader systemic levels. The type of strategic crisis communication required in the context of mega-crises must therefore contribute to advancing such structural transformations.

To ensure that the structural changes, which will be pursued, reflect the concerns and needs of all stakeholders, genuine dialogue is indispensable (Seeger & Sellnow, 2019). Furthermore, genuine dialogue must remain central throughout the inevitably lengthy process of mega-crisis mitigation, as it facilitates understanding, acceptance, and active participation of the diverse stakeholders in the broad structural changes required.

To recap, the overlap between (a.) *crisis communication*, (b.) *genuine dialogue practices*, and (c.) *structural change* dimensions represent the key theoretical locus for positioning *strategic communication aimed at mitigating mega-crises* (see Fig. 1). Crisis communication and genuine dialogue have been richly theorized and empirically and critically examined across the broader field of communication. However, openness to interdisciplinary input is a strength of the discipline of communication (Eadie, 2023), and we argue interdisciplinarity should be continuously embraced in this case as well due to multidimensionality of mega-crises necessitating equally complex disciplinary lenses of examination. Especially, the conceptualization of the dimension of structural change can (and should) be drawn from a range of interdisciplinary scholarship. As poorly managed mega-crises of recent past have indicated, disciplinary siloing would be tremendously counterproductive (Sellnow-Richmond & Lukacovic, 2024).

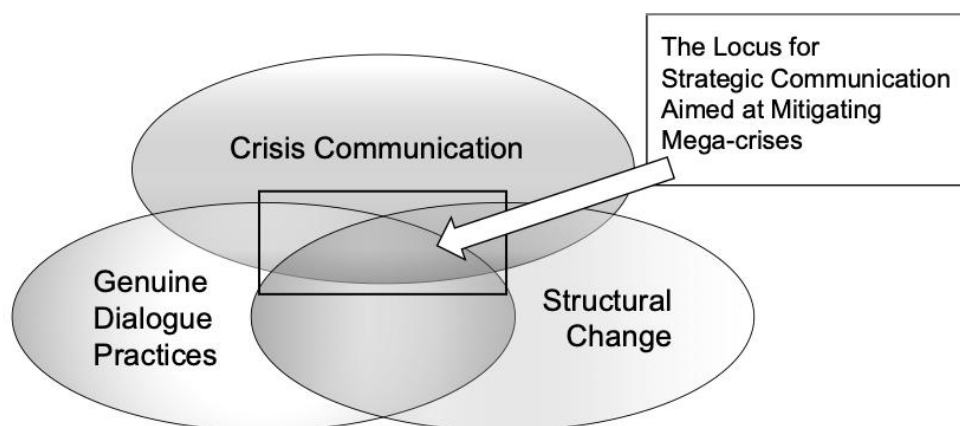


Fig. 1 - The central theoretical locus for situating strategic communication aimed at mitigating mega-crises lies at the intersection of the dimensions of crisis communication, genuine dialogue, and structural change

As applied to current practices such as mortification, structural changes can and do occur at the organizational level during crises (Benoit, 1995; Seeger & Sellnow, 2019). In these cases, the organization's culpability is openly acknowledged, corrective actions are implemented, and crisis communication is used to keep stakeholders informed throughout the process and about the changes that stem for it. This approach reflects a symmetrical model of communication between organizations and their stakeholders. In such instances, strategic communication professionals act as both the conscience of the organization and the essential communicative bridge between management and various stakeholder groups (Bowen, 2024).

In the context of emerging and ongoing risks and mega-crises linked to AI, we are encountering large-scale disruptive forces that impact a wide range of industries and sectors globally. In such circumstances, focusing solely on organizational perspectives is insufficient; broader perspectives are vital. Increasingly, we must anticipate situations where key organizations neither embrace symmetrical approaches nor permit their communication professionals to serve as enablers of stakeholder-driven, democratic change-making. In these situations, stakeholders on the margins can benefit from possessing a type of SCL, which can empower them to participate in available dialogues, especially when these stakeholders do not enjoy the access to representation by professional strategic communication practitioners.

Indeed, instances of SCL enacted by ordinary citizens or netizens can be observed in various grassroots movements that have instrumentalized strategic communication to effectively advance change. The literature highlights examples such as the savvy use of social media platforms and online socio-communicative action by loosely organized international initiatives (Oliveira et al., 2023). However, the communication scholarship reviewed thus far does not provide concrete cases of SCL employed by marginalized stakeholders specifically to counter AI-related mega-crises. Some adjacent contributions are beginning to emerge, for example, work emphasizing the role of university educators and students in using AI tools responsibly and in advocating for improved conditions for workers exploited by the Big Tech industry (Graham et al., 2025).

Practical examples of SCL employed by marginalized stakeholders resisting Big Tech's aggressive push for algorithmic development, despite its potential to exacerbate mega-crises, are well documented in the investigative reporting of tech journalist Karen Hao (2025). She traced effective activism, including the strategic use of communication by Chilean activists from the Mosacat collective, who successfully blocked the environmentally irresponsible construction of Google's data center in the Cerrillos district of Santiago de Chile. Metaphorically, this case resembled a David versus Goliath struggle; in practice, as Hao notes, it was diligent effort and effective communication that mobilized the community, triggered a chain of events, and eventually led to a legal halt of the data center project. Yet, this Chilean "David" merely won a single battle, while the broader struggle against Big Tech's extractive practices targeting marginalized communities and nations continues across Latin America and globally (Jiménez Arandia et al., 2025).

A Working List of Components of Strategic Communication Literacy

The pervasive role of media in contemporary life necessitates higher levels of media literacy across broader populations (Bulger et al., 2023). By similar logic, as AI is increasingly considered transformative across multiple sectors, a broad chorus of voices has called for enhanced AI literacy among professionals (e.g., Zerfass et al., 2020) as well as the wider public (e.g., Chiu et al., 2024). These literacies partially overlap with the concept of SCL proposed in this article. However, SCL must extend beyond media and AI literacies to become truly empowering for marginalized individuals who are often excluded from key dialogues and decision-making. In particular, confronting extremely powerful actors such as Big Tech requires challengers to possess additional sets of knowledge and skills within their strategic communicative repertoires. Hence, it is crucial to outline a working list of the critical components of SCL.

Media Literacy as Component of Strategic Communication Literacy

Media, digital, and information literacies correspond to overlapping concepts that underscore the importance of understanding the structures, processes, and effects of today's media technologies. Basic examples include recognizing the persuasive intentions of advertising versus the informative purpose of balanced reporting. However, media literacy cannot be reduced to a single finite benchmark. Instead, it can be outlined on a continuum of multiple levels (Mensonides et al., 2024; Yildiz Durak & Saritepeci, 2019). While regular consumers or citizens benefit from specific sets of media-related understandings, different professionals must cultivate specialized forms of media literacy, which may include deeper knowledge of technology or legal regulations. The complexity of today's media environments is marked by the information disorder, which is the proliferation of misinformation and disinformation (Wardle & Derakhshan, 2017). This reality necessitates an acute, ongoing, and flexible focus on enhancing media literacy among the wider public (Bulger et al., 2023). Accordingly, media literacy represents one of the critical components of SCL.

SCL must also be understood as a continuum with varying levels. More advanced forms of media literacy correspond to more sophisticated levels of SCL. For instance, a basic ability to distinguish between what is and is not an advertisement represents a lower level of media literacy, useful for members of the wider public in discerning the strategic goals behind different types of media content. By contrast, the type of SCL that enables stakeholders to engage in key dialogues about structural changes to counter mega-crises requires more advanced skills. These include knowing how to select among different media options, how to employ social media and online content-sharing platforms effectively, how to craft impactful media messages both technically and substantively, and how to adapt communication strategies in response to rapidly changing circumstances.

Artificial Intelligence (AI) Literacy as Component of Strategic Communication Literacy

Given the wide variety of technologies covered by the label of AI and the anticipated significant societal reverberations of its application, there is an active debate and multiple proposals on how to define the different types, levels, and target audiences for AI literacy relevant to various groups (e.g., Du, 2023; Veldhuis et al., 2025). For instance, in the context of journalism education, Lopezosa et al. (2023) argue that beyond emphasizing the productive use of AI tools, equal attention must be given to understanding the origins of AI and the ethical concerns surrounding it. Other solid arguments suggest that elevating levels of AI literacy among broader populations yields general societal benefits (Moravec et al., 2024). In the specific context of our focus, AI literacy represents another critical component of SCL.

There are two important ways in which AI literacy factors into SCL. The first parallels media literacy, as it encompasses boosting one's capacity to be a critical consumer on a more basic level, and developing the ability to evaluate, select, and effectively employ available AI tools in strategic communication efforts on a more sophisticated level of SCL.

The second important aspect concerns the role of the industry behind today's ubiquitous AI tools in amplifying wide-reaching risks and exacerbating mega-crises. On one hand, a specific AI tool (such as a chatbot) can be effectively utilized by stakeholders in their strategic communication efforts. On the other hand, the same tool can easily backfire, as it collects data to be reported back to its parent corporation. These data are often appropriated in ways that either directly harm the mentioned marginalized stakeholders or contribute to mega-crises with disproportionately negative impacts on others at the margins (Hao, 2025; Varoufakis, 2023). For example, a growing number of scholars, critics, and activists, such as Canadian-born American writer Cory Doctorow (QAA Podcast, 2025), warn about the emerging reality in which workers' data are appropriated to train AI systems that will eventually replace those very workers. Actual cases are already described, such as Amazon's use of driver footage to develop a fleet of driverless vehicles (Williams et al., 2022). This tendency places vulnerable workers from marginalized groups at high risk of job loss to automation and robots. For activists, this raises a dilemma in their strategic communication efforts to counter Big Tech's push toward replacing workers with robots and AI: the *catch-22* being that each possible option is comparably equally problematic. The dilemma is that some such technologies probably must be used but any use of these data-collecting technologies may inadvertently strengthen the very AI models enabling job displacement (Hao, 2025; Varoufakis, 2023). Furthermore, the Big Tech industry has a dismal record of treating and compensating vulnerable workers within its own systems, reinforcing the suspicion there is notion of a field-wide disregard for labor rights and worker well-being (Rani & Dhir, 2024; Rowe, 2023).

There are also examples of ethically and consensually collected training data used to develop AI systems that mitigate serious problems such as the extinction of Indigenous languages and knowledge systems (Hao, 2025). Such uses of AI can avoid the problematic

traps of corporate-driven systems and instead empower counter-mega-crisis efforts, for instance by applying Indigenous knowledge to promote sustainable practices to protect natural environment (Longdon, 2025). Indeed, AI technologies are being envisioned and developed globally through diverse and innovative approaches rooted in various cultural frameworks, offering pathways toward alternative practices that resist exploitative trends (Tang, 2024). Exploring these efforts can help identify effective ways of incorporating AI into the concept of SCL without inadvertently contributing to harms or abuses.

In light of ongoing events, recent exposés, and the poignant critiques budding across interdisciplinary academic circles and beyond, it is evident that the issues surrounding AI today are highly complex. Cultivating both the vision and practice of AI literacies is therefore an ongoing project that demands comprehensive interdisciplinary input, along with continuous study and refinement, given the fluid and rapidly evolving nature of the field. This suggests that while the incorporation of AI literacy into SCL is essential, its refinement must be guided by evidence from mindful applications, rigorous scholarly inquiry, and continual critical reflection.

Identifying Other Key Components of Strategic Communication Literacy

Botan (2018) outlines strategic communication practice as consisting of two vital parts: (1) “collecting *strategic information*” and (2) “developing *planned strategic communication campaigns* based on the information collected” [emphasis in original] (p. 7). SCL is proposed here with the strategic goal of involving marginalized stakeholders in dialogues and decisions concerning structural change. Accordingly, SCL must encompass the knowledge, skills, and abilities necessary for gathering strategic information and developing campaigns that can succeed in advancing this goal. Ultimately, SCL should enable successful persuasion paired with a type of political action that brings previously excluded stakeholders into deliberations and decisions. Hence, SCL enables persuasion to be deployed toward the creation of a new reality.

Ulmer et al. (2011) highlight the ethical obligation of organizations to stakeholders, including those groups who are at the margins and frequently remain forgotten by decision-makers within key organizations. Organizations that wish to genuinely serve their stakeholders and support them during crises should employ Dutta’s (2007) culture-centered approach, which focuses on lived experiences and on empowering the voices of marginalized stakeholders, while recognizing that culture is fluid and ever-evolving. Hence, active culture-centered work with stakeholders must be continuous. However, documented examples show that many current leading organizations in the tech and AI sectors are rather disinterested in the plight of marginalized stakeholders (Hao, 2025). Many corporations in the industry pay lip service to inclusion and care for the marginalized, yet such initiatives are often disingenuous and inconsequential (West & Kak, 2024). As a result, even marginalized stakeholders who may already possess high levels of SCL are unable to engage effectively in dialogue with the powers-that-be, because there is simply no real opportunity to

democratically participate in discussions and decision-making. Therefore, empowering disempowered stakeholders remains a project yet to be realized within the tech industry, and Big Tech in particular. Dutta et al. (2025) emphasize the need for “voice infrastructures” (p. 60). While elevating SCL represents one pathway to strengthening culture-centered voice infrastructures in terms of marginalized stakeholders’ agency, the need for organizational and political change at a structural level is paramount and will require a broad coalition capable of enacting change at the large scale where mega-crises arise.

The marginalized stakeholders impacted by today’s mega-crises are facing tremendous odds, and any realistic assessment of the status quo must acknowledge this imbalanced situation. Fortunately, a variety of tools is available such as social media, content-sharing platforms, and AI technologies. SCL, when based in a systematic strategic approach toward achieving its ultimate goal, can function as a blueprint for navigating the structural barriers and selecting appropriately from the often overwhelming range of available tools and methods. Examination of historical and more recent efforts, obstacles, and successes of activists from the Luddites (Merchant, 2024) to the Mosacat movement in Chile (Hao, 2025) offers an excellent starting point for listing, defining, and refining the key components of SCL. Eventually, rigorous scholarly inquiry and critical reflection in future studies are vital for developing a more complete set of components, both to elevate theoretical knowledge and to equip stakeholders on the ground with practical recommendations for addressing mega-crises and pursuing necessary structural change.

Still, one of the important lessons to be absorbed as soon as possible is that while access to technological tools, social media, or content-sharing platforms can be helpful in confronting crises and pursuing structural change, this access does not guarantee success or positive outcomes (Morozov, 2011). It is a fallacy of techno-determinism to assume that tools alone can ensure the success of activist efforts. Ultimately, it is human beings with their talents, creativity, and dedication that have driven all meaningful social change across both historical and more recent examples. Because humans and their collective units are inherently complex, no single academic field (as they are all arbitrary) can fully capture all significant trends and crucial nuances. Hence, within the ongoing quest to define and refine the key components of SCL, an inclusive and interdisciplinary approach is absolutely fundamental.

For instance, the philosophical framework of Ubuntu (amplified by the famous South African theologian Desmond Tutu, among other African thinkers) centers values such as community and consensus and has recently been applied to AI ethics as well (e.g., Mokoena, 2024). Ngondo and Klyueva (2023) state that Ubuntu holds promising potential for contemporary strategic communication. Hence, engagement with this philosophy can be an excellent interdisciplinary opportunity for the further evolution of SCL. Of course, the list of potential interdisciplinary endeavors in the evolution of SCL can be very long, and many of these directions hold tremendous potential. Our article draws on the work of various authors outside of communication, such as economist Varoufakis (2023) or technology and natural science researcher Longdon (2025), among others. However, this is only one way to engage with thinkers across fields. It is also paramount that we get involved in more active collaborative research, theorizing, and applied work across disciplinary borders and beyond.

Discussion: Implications of Strategic Communication Literacy for Research, Practice, and Education

Mega-crises are inherently difficult to mitigate, and many institutions struggle to address these wicked problems (Sellnow-Richmond & Lukacovic, 2024). Nonetheless, there is growing consensus across fields that, however overwhelming, mega-crisis mitigation must be prioritized as humanity confronts exacerbating existential risks. Public opinion research already indicates significant concern about the risks of AI among the public (Schwarz & Faj, 2024). History reminds us that although structural change is never too easy to bring about, it has been achieved. Such change has driven human advancement. This progress – this advancement has always involved overcoming past crises and expanding the experience of security in terms of safety, healthcare access, opportunities for self-actualization, human rights and liberties, and more. Thus, today's mega-crises, however daunting, should not discourage academics, activists, or communities from confronting them. While the process will be challenging, success remains possible. That path will be built on strong ideas and carefully employed tools. SCL represents precisely such a set of knowledge, skills, and abilities, enabling individuals and communities to evaluate and apply ideas and tools more effectively.

SCL is already practiced in applied settings, though it has not yet been explicitly coined as a concept. We argue that conceptualizing it as SCL is crucial, as this enables its manifestations, uses, and effects to be studied and utilized more systematically. The theoretical value is rooted in SCL as a concept useful for integration of ideas across disciplines, further evolving a highly germane endeavor with societal benefits in the context of mega-crises.

In general, SCL is a valuable concept for describing a set of knowledge, skills, and abilities that empower stakeholders at the margins to join important dialogues about the structural changes needed to avert mega-crises. Our analysis, particularly in light of the heightened risks of mega-crises linked to AI, has situated SCL within the broader theoretical landscape of crisis communication, other key strategic communication frameworks, and related interdisciplinary works across academic, journalistic, and activist spheres. In addition, we explored the interconnections between SCL and media and AI literacies, which are complementary to SCL and already well supported by pertinent scholarly literature and applied projects internationally.

The concept of SCL would greatly benefit from further study and theorization. A vital next step is to employ empirical tools to collect data on existing incarnations of SCL, particularly regarding how marginalized stakeholders have applied it to participate in key dialogues during mega-crises. Surveys and interviews with stakeholders engaged in such initiatives will be indispensable for developing a more precise roster of the knowledge, skills, and abilities that (ought to) constitute SCL. Equally important are interdisciplinary explorations focused on identifying complementary concepts that can strengthen the project of

marginalized stakeholders mobilizing to influence structures. For instance, collaborations with scholars and practitioners in psychology, sociology, political science, public health, public policy and administration, social services, and environmental and climate science hold tremendous potential for expanding both the theoretical and practical portrayals of SCL.

Another crucial direction for future research is working with and listening closely to various members of communities both locally and globally. Those at the margins often lack the time or energy to participate in traditional activist initiatives aimed at driving change, as their daily lives are consumed by navigating the ruthless impacts of mega-crises. Yet, this does not mean that people at the extreme margins should be excluded from efforts to conceptualize SCL or from broader visions of how to tackle mega-crises and realize structural change. On the contrary, available accounts demonstrate that individuals living in profound disparities often inspire awe through their ability to ingeniously mobilize against such behemoth as Big Tech and other global exploitative forces (Hao, 2025; Longdon, 2025).

Another promising direction for future work on SCL is exploring its implications for communication education. Strategic communication education is typically approached as a specialized area for professionals in the field. By contrast, SCL implies the need to outline curricula for a more general coursework. However, those who, due to marginalization, have limited access to formal education would still be excluded from the access to these opportunities. Exploring diverse avenues for teaching the knowledge, skills, and abilities that constitute SCL will therefore be essential. This may also require that professionals trained through existing curricula take on the role of educators within their broader communities. Such professionals will, besides practicing strategic communication as part of their professional routines, also share SCL with individuals and groups around them. In doing so, professionals can help empower marginalized communities. Our appeal here resonates with Stoeckle and Adi (2023), who emphasize that public relations research and education must be willing to undergo systematic transformation to remain relevant and to drive meaningful change in the face of what they refer to as *grand challenges* – and what we describe as mega-crises.

Conclusion

This article situates the proposed concept of *strategic communication literacy* (SCL) within the field of strategic crisis communication, particularly as it applies amid mega-crises. We postulate that it is vital to define and advance a type of literacy initiative that encompasses communication knowledge and skills among a range of stakeholders, including those who are marginalized. SCL aims to empower stakeholders to participate in efforts to mitigate risks that spiral toward mega-crises, as well as to contribute to structural changes necessary to counter mega-crises that undermine human well-being, security, and the natural environment. In the contemporary era, AI represents a technological field that exacerbates existing mega-crises but also potentially introduces new security risks. Accordingly, we use AI-exacerbated mega-crises as the primary cases to illustrate our central arguments about

SCL. Furthermore, we propose a research agenda to refine the concept of SCL and assess its relevance in relation to broader categories of mega-crises. Ultimately, SCL represents a productive route for the field of strategic communication, strengthening its position as a relevant field amid the uncertain landscape of emerging technologies and many complex mega-crises faced locally and globally.

Biographical note

Marta Natalia Lukacovic, Ph.D., received her doctorate from Wayne State University, Detroit, Michigan, USA. She is an Associate Professor of Communication and Mass Media and leads the Graduate Master's Degree Program in Communication at Angelo State University (Texas Tech University System). Her research centers on digital media and strategic communication surrounding the matters of security, such as global crises, pandemics, malicious AI, etc. She co-edited a two-volume book project on public relations, media, and communication in post-socialist nations of Eastern Europe and Central Asia. She co-authored the *Communication amid Mega-crises* book published with Routledge. Lukacovic published articles and chapters with leading global scholarly presses and presented at numerous international academic conferences. She is a co-founding member of the International Crisis and Risk Communication Association, a member of the Global Strategic Communication Consortium, and an expert affiliate of the Global Network on Artificial Intelligence & International Society, among other professional affiliations.

Deborah Sellnow-Richmond is an Associate Professor in the Department of Applied Communication Studies at Southern Illinois University Edwardsville, USA, where she teaches courses in public relations and organizational communication. She researches the efficacy and unforeseen effects of public relations messages in health and organizational crisis context, and the emerging role of social media in creating and resolving organizational crises, with a specific emphasis on the role of communication amid the growing threat of mega-crises worldwide. She holds a Ph.D. in communication from Wayne State University in Detroit, Michigan, a master's degree in Public Service from the University of Arkansas Clinton School of Public Service, and a bachelor's degree from the University of Minnesota.

References

- Alpaslan, C. M. & Mitroff, I. I. (2011). *Swans, swine, and swindlers: Coping with the growing threat of mega-crises and mega-messes*. Stanford University Press.
- Bender, E., McMillan-Major, A., Shmitchell, S., & Gebru, T. (2021). On the dangers of stochastic parrots: Can language models be too big? *FACCT '21: Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. <https://doi.org/10.1145/3442188.3445922>
- Benoit, W. L. (1995). *Accounts, excuses, and apologies: A theory of image restoration strategies*. SUNY Press.
- Botan, C. H. (2018). *Strategic communication: Theory and practice*. Wiley-Blackwell.

- Bowen, S. A. (2024). "If it can be done, it will be done:" AI ethical standards and a dual role for public relations. *Public Relations Review*, 50(5), 1-13. <https://doi.org/10.1016/j.pubrev.2024.102513>
- Bulger, M., Baleria, G., Hobbs, R., & Moffitt, K. R. (2023). The promise of media literacy education when "everything is at stake" and "everything is expected". *Journal of Media Literacy Education*, 15(1), 99-108. <https://doi.org/10.23860/JMLE-2023-15-1-8>
- Chiu, T. K. F., Ahmad, Z., Ismailov, M., & Sanusi, I. T. (2024). What are artificial intelligence literacy and competency? A comprehensive framework to support them. *Computers and Education Open*, 6, 1-9. <https://doi.org/10.1016/j.caeo.2024.100171>
- Du, Y. R. (2023). Personalization, echo chambers, news literacy, and algorithmic literacy: A qualitative study of AI-powered news app users. *Journal of Broadcasting & Electronic Media*, 67(3), 1–28. <https://doi.org/10.1080/08838151.2023.2182787>
- Dutta, M. J. (2007). Communicating About culture and health: Theorizing culture-centered and cultural sensitivity approaches. *Communication Theory*, 17(3), 304–328. <https://doi.org/10.1111/j.1468-2885.2007.00297.x>
- Dutta, M., Pokaia, V., Metuamata, S., Mandal, I., Baskey, P., Mandi, R., Elers, P., Rahman, M., Jayan, P. & Pattanaik, S. (2025). Center for Culture-Centered Approach to Research and Evaluation (CARE): Organizing to transform the social determinants of health. *Journal of Applied Communication Research*, 53(1), 54–61. <https://doi.org/10.1080/00909882.2025.2467858>
- Eadie, W. F. (2023). *When communication became a discipline*. Rowman & Littlefield.
- Endter-Wada, J., & Ingram, H. (2012). Global climate change as environmental mega-crisis. In I. Helsloot, A. Boin, B. Jacobs, & L. K. Comfort (Eds.), *Mega-crises; Understanding the prospects, nature, characteristics and the effects of cataclysmic events* (pp. 300-318). Charles C Thomas – Publisher, LTD.
- Geburu, T., & Torres, É. P. (2024). The TESCREAL bundle: Eugenics and the promise of utopia through artificial general intelligence. *First Monday*, 29(4), 1-42. <https://doi.org/10.5210/fm.v29i4.13636>
- Graham, M., Alyanak, O., & Valente, J. C. (2025). Reflection AI: feeding the machine - the hidden labour behind AI tools and ethical implications for higher education. *Frontiers in Communication*, 10. <https://doi.org/10.3389/fcomm.2025.1614817>
- Jiménez Arandia, P., Dib, D., & Alarcon, M. (2025, August 8). The backyard of AI; A map of the 21st century gold rush. *Pulitzer Center*. <https://pulitzercenter.org/stories/backyard-ai>
- Hao, K. (2025). *Empire of AI; Dreams and nightmares in Sam Altman's OpenAI*. Penguin Press.
- Kissinger, H., Schmidt, E., & Huttenlocher, D. P. (2021). *The age of AI: and our human future*. Little Brown & Company.
- Kruckeberg, D. & Vujnovic, M. (2025). Questions of power over (and control of) social, political, economic, and cultural dimensions of artificial intelligence in public relations and strategic communication. In A. V. Laskin & K. Freberg (eds.), *Public relations and*

- strategic communication in 2050; Trends shaping future of the profession* (pp. 9-21). Routledge.
- Longdon, J. (2025). *Natural connection; What Indigenous wisdom & marginalized people teach us about environmental action*. Penguin Random House.
- Lopezosa, C., Codina, L., Pont-Sorribes, C., & Vázquez, M. (2023). Use of generative artificial intelligence in the training of journalists: challenges, uses and training proposal. *El Profesional de La Información*, 32(4), 1–12. <https://doi.org/10.3145/epi.2023.jul.08>
- Lovari, A., & Bowen, S. A. (2019). Social media in disaster communication: A case study of strategies, barriers, and ethical implications. *Journal of Public Affairs*, 20(1). <https://doi.org/10.1002/pa.1967>
- Lovari, A., & De Rosa, F. (2025). Exploring the challenges of generative AI on public sector communication in Europe. *Media and Communication*, 13. <https://doi.org/10.17645/mac.9644>
- McChesney, R. W., & Nichols, J. (2016). *People get ready; The fight against a jobless economy and a citizenless democracy*. Nation Books.
- Mensonides, D., Smit, A., Talsma, I., Swart, J., & Broersma, M. (2024). Digital literacies as socially situated pedagogical processes: Genealogically understanding media, information, and digital literacies. *Media and Communication*, 12, 1-18. <https://doi.org/10.17645/mac.8174>
- Merchant, B. (2024). Learning from the Luddites. In E. Morozov et al., *AI futures* (pp. 72-76). Boston Review – Haymarket Books.
- Mokoena, K. K. (2024). A holistic ubuntu artificial intelligence ethics approach in South Africa. *Verbum et Ecclesia*, 45(1), 1-8. <https://doi.org/10.4102/ve.v45i1.3100>
- Moravec, V., Hynek, N., Gavurova, B., & Kubak, M. (2024). Everyday artificial intelligence unveiled: Societal awareness of technological transformation. *Oeconomia Copernicana*, 15(2), 367–406. <https://doi.org/10.24136/oc.2961>
- Morozov, E. (2011). *The net delusion: The dark side of internet freedom*. PublicAffairs.
- Morozov, E. (2024). The AI we deserve. In E. Morozov et al., *AI futures* (pp. 9-37). Boston Review – Haymarket Books.
- Narayanan, A., & Kapoor, S. (2024). *AI snake oil; What artificial intelligence can do, what it can't, and how to tell the difference*. Princeton University Press.
- Ngondo, P.S., & Klyueva, A. (2023). Inviting an Ubuntu-based approach to public relations theory building in Sub-Saharan Africa. In C. H. Botan & E. J. Sommerfeld (Eds.), *Public relations theory III* (pp. 374-392). Routledge.
- Oliveira, E., Rodriguez-Amat, J. R., Ruiz-Mora, I., & Zeler, I. (2023). The fluid and disruptive shape of activism: Strategic communication in #fridaysforfuture. *International Journal of Strategic Communication*, 17(4), 301–324. <https://doi.org/10.1080/1553118x.2023.2204299>
- QAA Podcast. (2025, August 30). *Cory Doctorow DESTROYS enshitification (E338)* [Video]. YouTube. https://www.youtube.com/watch?v=V-7CCKR__0k
- Rani, U., & Dhir, R. K. (2024, December 10). The Artificial Intelligence illusion: How invisible workers fuel the “automated” economy. *International Labor Organization*.

<https://www.ilo.org/resource/article/artificial-intelligence-illusion-how-invisible-workers-fuel-automated>

- Robinson, L., Cotten, S. R., Ono, H., Quan-Haase, A., Mesch, G., Chen, W., Schulz, J., Hale, T. M., & Stern, M. J. (2015). Digital inequalities and why they matter. *Information, Communication & Society*, 18(5), 569–582. <https://doi.org/10.1080/1369118x.2015.1012532>
- Roumate, F. (2024). *Artificial intelligence and the New World Order*. Springer Nature. <https://doi.org/10.1007/978-3-031-50312-2>
- Rowe, N. (2023, August 2). “It’s destroyed me completely”: Kenyan moderators decry toll of training of AI models. *The Guardian*. <https://www.theguardian.com/technology/2023/aug/02/ai-chatbot-training-human-toll-content-moderator-meta-openai>
- Samoilenko, S. A., & Suvorova, I. (2023). Artificial intelligence and deepfakes in strategic deception campaigns: The U.S. and Russian experiences. In E. Pashentsev (Ed.), *The Palgrave handbook of malicious use of AI and psychological security* (pp. 507–529). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-22552-9_19
- Schwarz, A., & Faj, T. (2024). Communicating and perceiving risks of artificial intelligence as an emerging technology. In T. L. Sellnow & D. D. Sellnow (Eds.), *Communicating risk and safety* (pp. 503–526). De Gruyter. <https://doi.org/10.1515/9783110752427-027>
- Seeger, M. W., & Sellnow, T. L. (2019). *Communication in times of trouble*. Wiley Blackwell.
- Sellnow-Richmond, D.D., & Lukacovic, M.N. (2024). The COVID-19 pandemic as exemplar of the chaos of mega-crises. In T. L. Sellnow & D. D. Sellnow (Eds.), *Communicating risk and safety* (pp. 97–107). De Gruyter. <https://doi.org/10.1515/9783110752427-006>
- Sellnow-Richmond, D.D., & Lukacovic, M.N. (2026). *Communication amid mega-crises*. Routledge.
- Solovyeva, A., & Hynek, N. (2023). When stigmatization does not work: Over-securitization in efforts of the Campaign to Stop Killer Robots. *AI & Society*, 38, 2547–2569. <https://doi.org/10.1007/s00146-022-01613-w>
- Stoeckle, T., & Adi, A. (2023). Relações públicas, pós-verdade e grandes desafios: uma melhor pesquisa para o valor social. *Organicom*, 20(42), 86–102. <https://doi.org/10.11606/issn.2238-2593.organicom.2023.210366>
- Tang, A. (2024). The real legacy of cybernetics. In E. Morozov et al., *AI futures* (pp. 43–47). Boston Review – Haymarket Books.
- Ulmer, R. R., Sellnow, T. L., & Seeger, M. W. (2011). *Effective crisis communication: Moving from crisis to opportunity* (2nd ed.). Sage.
- Varoufakis, Y. (2023). *Technofeudalism; What killed capitalism*. Melville House Publishing.
- Veldhuis, A., Lo, P. Y., Kenny, S., & Antle, A. N. (2025). Critical artificial intelligence literacy: A scoping review and framework synthesis. *International Journal of Child-Computer Interaction*, 43, 1–20. <https://doi.org/10.1016/j.ijcci.2024.100708>
- Wardle, C., & Derakhshan, H. (2017). *Information disorder: Toward an interdisciplinary framework for research and policy making*. Council of Europe.

<https://edoc.coe.int/en/media/7495-information-disorder-toward-an-interdisciplinary-framework-for-research-and-policy-making.html>

West, S. M., & Kak, A. (2024). AI's missing others. In E. Morozov et al., *AI futures* (pp. 58-62). Boston Review – Haymarket Books.

Williams, A., Miceli, M., & Gebru, T. (2022, October 13). The exploited labor behind artificial intelligence. *Noema*. <https://www.noemamag.com/the-exploited-labor-behind-artificial-intelligence/>

Yildiz Durak, H., & Saritepeci, M. (2019). Modeling the effect of new media literacy levels and social media usage status on problematic internet usage behaviours among high school students. *Education and Information Technologies*, 24, 2205–2223. <https://doi.org/10.1007/s10639-019-09864-9>

Zerfass, A., Hagelstein, J., & Tench, R. (2020). Artificial intelligence in communication management: A cross-national study on adoption and knowledge, impact, challenges and risks. *Journal of Communication Management*, 24(4), 377-389. <https://doi.org/10.1108/JCOM-10-2019-0137>

Safeguarding lives: An informational model for communicating wildfire disasters^{*}

Bianca Persici Toniolo^{**}

University of Beira Interior, LabCom

Gisela Gonçalves^{***}

University of Beira Interior, LabCom

This theoretical article introduces an original model for planning and evaluating risk and crisis communication by public organizations in disasters, with a particular focus on wildfires. Adopting a systemic and integrative perspective, the *Loop Model* synthesizes knowledge from specialized literature and draws upon well-established scientific theories and models to highlight the critical role of communication throughout every stage of forest and wildfire management. Developed through an extensive literature review and applied to the chronological pattern of wildfires in Portugal, the framework demonstrates both theoretical relevance and practical applicability. It may also be adapted as a framework for evaluating and planning communication strategies and messages in response to other climate- and weather-related disasters. The proposed model yields the Informational Quality Indicator (IQI), designed to assess the quality of public organizations' risk and crisis communication. A higher IQI value indicates that the communication prioritizes public protection over political reputation management. By advancing a comprehensive framework, this article contributes to the refinement of crisis and risk communication theory while offering practical guidance for public organizations facing disaster scenarios, which can be further tested in future research.

Keywords: Risk and crisis communication, *Loop Model*, Literature review, Wildfires, Climate change

^{*} Article submitted on 24/09/2025. Article accepted on 15/12/2025.

^{**} bianca.toniolo@ubi.pt;

^{***} gisela.goncalves@labcom.ubi.pt

Introduction

Climate change is significantly increasing the global risk of disasters, intensifying extreme weather events and altering environmental conditions. The resulting rise in wildfires, heat waves, and droughts in numerous countries is directly attributable to escalating global temperatures. Mediterranean countries have experienced significant challenges regarding wildfires in recent years. In Portugal, particularly, the devastating wildfires of 2017 claimed 117 lives, highlighting the critical need for effective public fire management. A quarter of the area burned in 2025 in Europe was in Portugal. The country's annual struggle with summer wildfires underscores the ongoing challenges of climate change and inadequate forest management.

The increased frequency and severity of disasters demand effective communication strategies. Poorly managed disasters often trigger both public and political crises, requiring a delicate balance between protecting the population and maintaining the public image of organizations and political leaders. This tension, as Frandsen and Johansen (2020b) highlight, reflects the interplay between informational and political perspectives in risk and crisis communication. The informational perspective adopts a rational, optimistic approach, assuming stakeholder consensus on risk assessment to facilitate informed, public-protective decision-making. Conversely, the political perspective is more cynical, recognizing the lack of consensus and the potential for heightened divisions during crises and framing communication as a rhetorical contest for power and legitimacy. This emphasizes the complexity of effective crisis communication management.

The central focus of this theoretical article is to introduce the *Loop of Risk and Crisis Communication on Wildfires* – the *Loop Model*, an original framework developed through an integrative literature review (Snyder, 2019) on risk, crisis, emergency, and disaster communication. The model is grounded in a systemic approach to communication (Plana et al., 2024), with a particular emphasis on the messages disseminated by public organizations regarding wildfires as recurrent events with identifiable chronological patterns. Systemic communication is characterized as strategic, planned, integrated, and structured (Plana et al., 2024). The framework is guided by the informational perspective of risk and crisis communication (Frandsen & Johansen, 2020b), which prioritizes the protection of the population from the consequences of wildfires.

Strategic Risk and Crisis Communication

Risk, crisis, and disaster communication is a highly relevant area within the discipline of strategic communication, ranking as the second most research topic in the *International Journal of Strategic Communication* from 2007 to 2017 (Werder et al., 2018). According to Palenchar and Heath (2007, p. 127), the strategic nature of risk communication lies more in its process than in its content, serving as "a tool for communicating values and identities as

much as being about the awareness, attitudes and behaviors related to the risk itself." Raupp (2015, p. 520) argued that, in the context of risk communication, strategy is an "emergent sense-making and framing activity" that serves different perceptions of danger in an uncertain scenario.

Crisis communication research draws heavily on two prominent theories: Benoit's Image Restoration Theory (IRT) (1997, 1995) and Coombs' Situational Crisis Communication Theory (SCCT) (2022). Rooted in strategic communication and public relations, these represent dominant paradigms in crisis communication research, emphasizing response strategies (Elgueta-Ruiz & Martínez-Ortiz, 2022; Avery et al., 2010). Since Lipset's (1953) seminal *Public Opinion Quarterly* article, crisis communication research has steadily grown (Diers-Lawson, 2020), experiencing significant expansion in the 1990s (Tachkova & Coombs, 2022). Elgueta-Ruiz and Martínez-Ortiz's (2022) meta-analysis of *Public Relations Review* articles (2000-2014) revealed "crisis management and communication" as the dominant topic of 57 total topics, demonstrating growth exceeding the overall increase in published articles.

Early crisis communication research focused on defining crises and best practices. Subsequently, case studies applying crisis response strategies emerged. Only recently has research shifted to analyzing stakeholder reactions and their organizational impacts (Diers-Lawson, 2020). This evolution, according to Diers-Lawson (2020), reflects both the field's maturation since the 1990s (with enhanced theoretical development) and the increased demand for practically applicable, theoretically grounded models. As Tachkova and Coombs (2022) note, the relationship between theory and practice is crucial: theory explains the link between responses and outcomes, while practice validates these relationships, testing the predictability of responses in specific crisis contexts.

The need for further research focused on the theoretical development of crisis communication became evident in An and Cheng's (2010) finding that only 13% of articles published in the *Public Relations Review* and the *Journal of Public Relations Research* posed research questions and/or hypotheses grounded in theory. Another meta-analysis, conducted by Diers-Lawson (2020, 2017), examined 690 crisis communication articles published in the Scimago database between 1953 and 2015. The results demonstrated that crisis communication changed substantially over this period, shifting from best practices to a data-driven approach and moving its focus from organizations to the human aspects of crisis communication.

Risk communication can be understood as an integral part of the risk management process, involving a dialogue between technical expertise and public risk perception. Essential for informed decision-making by stakeholders about potential risks, risk communication aims to promote the adoption of preventive behaviors to avoid risks, trust in risk management organizations and their guidance, and individual and collective preparedness to respond to risks (Toniolo, 2025). On the other hand, crisis communication can be seen as a continuous process of implementing strategies, tactics, and communication actions. It is present at all stages of crisis management and aims to prepare organizations to face crises, respond to crises with agility and transparency, account for

crisis management efforts, and inform about organizational changes resulting from crises. The primary goal of crisis communication is to protect the stakeholders from the negative effects of a crisis and then to preserve the reputation of the organization or public authority (Toniolo, 2025).

Loop Model: A New Approach to Risk and Crisis Communication

The *Loop Model* emerges from the need for a theoretically grounded model to help public organizations in developing effective communication strategies for safeguarding the population against wildfires. The conceptualization of the proposed model was based on an extensive integrative literature review (Snyder, 2019) that encompassed studies on risk, crisis, emergency, and disaster communication. This methodological approach sought to capture a comprehensive and multidisciplinary understanding of the phenomenon of disaster communication.

The selection of theoretical sources that informed the development of the *Loop Model* followed the most established frameworks in risk and crisis communication research (Sellnow & Seeger, 2021), including the Crisis and Emergency Risk Communication (CERC) model (CDC, 2014), the Situational Crisis Communication Theory (SCCT) (Coombs, 2017, 2007), and the Internalize–Distribute–Explain–Act (IDEA) model (Sellnow et al., 2017), among others. In addition, classical communication theories such as Framing Theory, applied to strategic communication (Hallahan, 2005, 1999), also guided the conceptualization of the model. However, these models were not adopted exclusively.

Given the multifaceted nature of risk and crisis phenomena, the selection of references was guided by theoretical approaches capable of capturing their systemic and dynamic interactions, as emphasized by Sellnow and Seeger (2021). While the CERC model offers a stage-based operational framework, and the SCCT focuses on reputation management and crisis response strategies, the IDEA model emphasizes message design, instructional clarity, and stakeholder engagement, particularly through the framing of recommended actions. In contrast, the framework integrates and extends these perspectives by introducing an informational–political *continuum* that explains how the communication of public organizations can oscillate between safeguarding citizens and preserving political reputation.

The informational perspective focuses on citizens' safety, emphasizing clarity, factual accuracy, and practical guidance, whereas the political perspective understands communication as a rhetorical process of contestation over meaning and framing (Frandsen & Johansen, 2020d). Public organizations therefore move fluidly between communication aimed at protecting the population and communication designed to sustain public trust and institutional credibility. In the proposed model, persuasion is oriented toward promoting protective behaviors. Thus, the informational and political dimensions are interdependent rather than oppositional: communication that informs citizens effectively also strengthens

legitimacy, while politically accountable communication enhances informational credibility. This conceptual balance clarifies how responsibility and transparency can coexist with persuasion in the management of risk and crisis communication, while also emphasizing the importance of maintaining *equilibrium* between the two perspectives and ensuring that collective interests prevail over individual ones.

In line with the informational perspective of risk and crisis communication (Frandsen & Johansen, 2020b), communication planned according to the *Loop Model* not only contributes to public safety but may also have the secondary effect of enhancing the reputation of the public organizations and political actors involved.

First, the model synthesizes existing literature and optimizes it specifically for climate- and weather-related disasters, with a particular focus on wildfires in Portugal. At its next stage, the model adopts a normative approach, providing practical implications for planning and decision-making in risk and crisis communication related to various disasters, a topic that remains underexplored (Kuipers & Welsh, 2017). The model identifies patterns that facilitate its adaptation to different phenomena caused and/or intensified by climate change.

The framework integrates multiple validated approaches to risk and crisis communication, combining them to address the unique context of public organizations' communication strategies concerning wildfires. A significant influence on the proposed model is the Crisis and Emergency Risk Communication (CERC) model (CDC, 2014), which is extensively utilized by U.S. organizations to communicate with citizens and the media during risk, crisis and emergency situations. In addition to CERC, the *Loop Model* draws on Coombs' (2017) Situational Crisis Communication Theory (SCCT), Pechta, Brandenburg and Seeger (2010) four-Channel Model of Emergency Communication, Austin and Jin (2017) Social-Mediated Crisis Communication (SMCC) model, Sellnow et al. (2017) Internalize-Distribute-Explain-Act (IDEA) model, and Frandsen & Johansen (2017a, 2017b) Rhetorical Arena Theory (RAT). Furthermore, the model incorporates Hallahan's (1999) framing approach, Olsson's (2014) perspective on the strategic and operational dimensions of public organizations' disaster communication, Bytze's (2008) thematic content categories and Reich, Bentman and Jackman (2011) Crisis Communication Guide for Public Organizations.

The innovation of the framework lies in its ability to integrate, within a single framework, multiple scientifically validated risk and crisis communication theories, establishing a communication approach that prioritizes population protection and is specifically designed for disasters requiring permanent management.

For the development of the 36 key messages that compose the proposed model, we relied on studies by Diers-Lawson (2020); Wukich (2019, 2016); CDC (2014); and Reich, Bentman and Jackman (2011). Thus, the *Loop Model* draws on diverse scientifically validated perspectives on risk and crisis communication to propose an innovative approach to planning and evaluating communication strategies, accounting for the complexity of both the communication process and disasters.

A key outcome of the model is the development of the Informational Quality Indicator (IQI), an aggregated metric for multivariate analysis of messages. This tool enables the evaluation of the quality of information disseminated by public organizations in risk and crisis

communication. Higher IQI values indicate communication that aligns more closely with an informational perspective, emphasizing public protection over political reputation management.

Phases of the *Loop Model*

The *Loop Model* is structured as a 12-month cycle comprising six chronologically defined phases that recur indefinitely: (1) *Prevention*; (2) *Preparation*; (3) *Alert*; (4) *Rescue*; (5) *Recovery*; and (6) *Evaluation*. *Rescue* and *Recovery* phases constitute a second turn of the loop, which is activated only in the event of wildfires. Each phase of the model has specific objectives aimed at addressing the challenges of communication during disasters (Reich, Bentman & Jackman, 2011). When effectively executed in accordance with the model's variables, these objectives lead to the creation of messages that enhance the protection of the population across physical, emotional and material dimensions in the context of wildfires and their aftermaths.

It is important to note that the chronology of the framework phases was established based on the patterns of wildfire occurrences in Portugal, as identified by Novo et al. (2024), in conjunction with the annual engagement level calendar set by Portugal's Special Rural Firefighting Device (Dispositivo Especial de Combate a Incêndios Rurais – DECIR)¹, the national operational system that organizes firefighting resources into seasonal phases, defining the level of mobilization and readiness required at each stage. Each year, the exact operational dates are confirmed in the National Operational Directive (Diretiva Operacional Nacional – DON) No. 2², which coordinates all civil protection entities and strategies during the wildfire season. Since the risk of wildfires can vary annually due to the effects of climate change, "periods of greater or lesser danger depend on meteorological conditions rather than calendar dates" (Viegas, 2024, p. XIII). It is therefore essential to build flexibility into these operational periods, which serve as benchmarks for planning risk and crisis communication by public organizations. Rather than being treated as fixed phases, they should be viewed as adaptable intervals that respond to the evolving duration and intensity of wildfire risks. Accordingly, the proposed model was conceived as a dynamic framework, capable of adjusting its chronological phases in alignment with the updated DECIR structure and other regional operational plans.

The application of the model to the chronological pattern of wildfire occurrences in Portugal is illustrated in *Figure 1*.

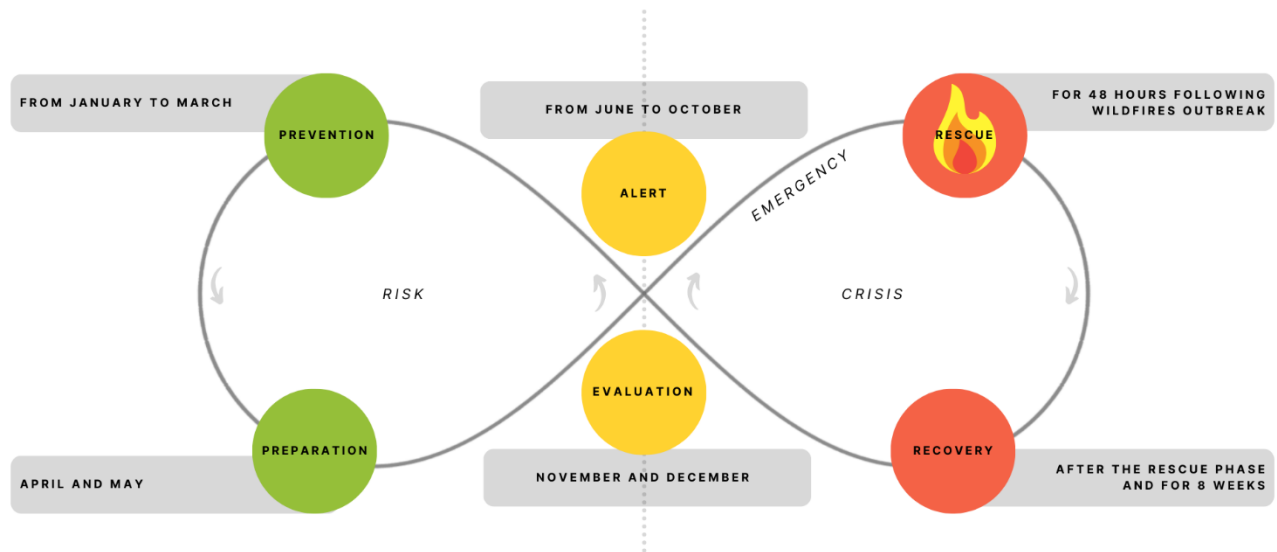


Figure 1- Application of the Loop Model to the chronological pattern of wildfires in Portugal. Source: Elaborated by the authors.

Loop Model Variables

The *Loop Model* includes 12 variables organized within the five dimensions of Lasswell's (1948) model (see *Figure 2*), which are: (1) control analysis (sender); (2) content analysis (message); (3) media analysis (channels); (4) audience analysis (receiver); and (5) effects analysis (reactions). As a message-centered model, the model places greater emphasis on content analysis.

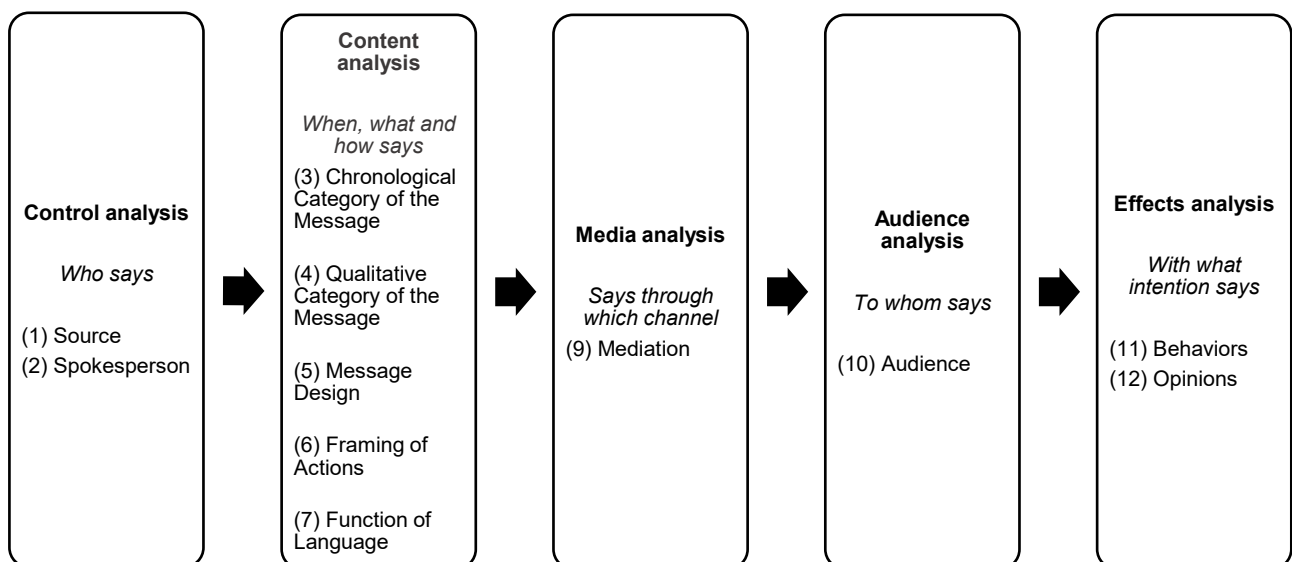


Figure 2 - Diagram of the Loop Model variables applied to the dimensions of the Lasswell Model. Source: Elaborated by the authors.

1. Source

In the communication of disasters, including wildfires, various public organizations serve as primary information sources, disseminating information through their own channels or via journalistic media. When the informational perspective dominates risk and crisis communication, these sources issue messages in accordance with their respective responsibilities for managing the disaster, reflecting the bureaucratic perspective of response coordination (Sellnow & Seeger, 2021).

2. Spokesperson

Risk and crisis communication is characterized by distinct spokesperson roles across different phases: scientists serve as the primary sources during the *Prevention* and *Alert* phases; political authorities act as spokespersons during the *Recovery* and *Evaluation* phases; and technical experts are featured during the *Preparation* and *Rescue* phases (Sellnow & Seeger, 2021; McLean & Ewart, 2015; Seeger, Sellnow & Ulmer, 2003). During the *Recovery* phase, expressing emotions, empathy and support for populations affected by a disaster can significantly enhance the reputation of political authorities among citizens, fostering greater trust in their actions and commitments (McLean & Ewart, 2015; Schnee, 2015; Olsson, 2014). Furthermore, collaborating with media personalities, including artists, athletes and other influencers, is crucial for expanding the reach and enhancing the relevance of the communicated content (UNDRR, 2022).

3. Chronological Categorization

This variable pertains to the period during which the message is disseminated. *Table 1* illustrates the six phases of the framework in relation to the chronological pattern of wildfire occurrences in Portugal, along with their corresponding objectives for risk and crisis communication.

Communication	Phase	Communication's goal	Time frame
Risk	Prevention	To minimize the risk of wildfires by involving the population in adopting precautionary measures.	January to March
	Preparation	To promote public trust in the preparedness of public organizations and the community to respond to wildfires.	April and May
	Alert	When the risk is imminent, warn the population, emphasizing the sense of urgency and the physical proximity of wildfires.	June to October

Risk -> Crisis	<i>Rescue</i>	In the event of wildfires, to protect the population, guiding them to act safely.	From the outbreak of a wildfire until 48 hours later
Crisis	<i>Recovery</i>	Once wildfires are under control, to inform affected communities about available public policies.	Up to eight weeks after the <i>Rescue</i> phase end
Crisis -> Risk	<i>Evaluation</i>	To examine risk and crisis management, involving the affected population and communities in evaluating actions taken for recovery and planned actions to prevent future occurrences.	November and December

Table 1- Application of the Loop Model to the chronological pattern of wildfires in Portugal and communication objectives for each phase. Source: Elaborated by the authors.

The first three phases of the proposed model – *Prevention*, *Preparation* and *Alert* – are primarily associated with risk communication. Prevention involves activities aimed at entirely avoiding the adverse impacts of hazardous events. Although it is not possible to eliminate all risks, prevention strategies serve to reduce vulnerability and exposure to wildfire hazards, thereby minimizing their effects (Plana et al., 2024). During the *Prevention* phase, messages aligned with an informational perspective of risk and crisis communication seek to reduce the occurrence of wildfires by encouraging the population to adopt precautionary measures.

Preparation involves developing the capacities to anticipate, respond to and recover from wildfires. Within the scope of risk management, preparation aims to create the necessary skills to manage emergencies and ensure organized transitions from rescue (or response) to recovery (Plana et al., 2024). In the *Preparation* phase of the *Loop Model*, the purpose of communication is to foster confidence in the capacity of public organizations and communities to effectively respond to wildfires.

In addition to contributing to public confidence in the ability of public organizations to respond effectively to wildfires, as well as promoting compliance with their guidelines, in the model, the *Preparation* phase is related to equipping communities with resources and training for self-protection so that they can respond autonomously to potential rural fires. This has become increasingly necessary, as noted by Almeida, Modarres and Rodrigues (2024), given the large number of annual occurrences and the saturation of civil protection agencies' capacity to respond to multiple events simultaneously.

According to Pechta, Brandenburg and Seeger (2010), to alert presupposes placing the public at the center of a multidirectional communication process, with information exchange among the public, public organizations and mass media. This flow of information, described in the Four-Channel Emergency Communication Model (Pechta, Brandenburg & Seeger, 2010), includes, among other processes, sending periodic updates to the affected public, emphasizing the urgency and imminence of the risk. Thus, when summer approaches and climatic and meteorological conditions favor the occurrence and rapid spread of wildfires, the *Alert* phase of the proposed model begins. The *Alert* phase marks the transition from risk communication to crisis communication and, together with the *Rescue* phase, form part of emergency communication. In the *Alert* phase, risk and crisis communication practiced

by public organizations, primarily aims to warn the population about the imminent danger of new rural fires.

Although some interaction between phases may occur, particularly between *Prevention* and *Alert*, they remain conceptually distinct: *Prevention* focuses on reducing vulnerability and exposure in advance, whereas *Alert* arises from an imminent and localized threat, marked by a heightened sense of urgency and the need for immediate communicative action.

To rescue or to respond consists of carrying out immediate actions before, during and after a wildfire to save lives, reduce health impacts, ensure public safety, and meet basic needs. Focused on the short term, rescue depends on proper preparation and the development of response capabilities at various levels (Plana et al., 2024). The *Rescue* phase marks the realization of risk, or in other words, the crisis itself, and, like the *Alert* phase, is part of emergency communication. In the *Loop Model*, the *Rescue* phase begins only when a wildfire breaks out and lasts 48 hours. During this period, communication efforts must be dedicated to protecting the population, guiding them to act safely, increasing risk perception and likely improving compliance with authorities' instructions.

According to Plana et al. (2024), to recover involves restoring or improving the livelihoods and health, as well as the economic, social and environmental assets of communities affected by wildfires, following the principles of sustainable development and the idea of "building back better" (Plana et al., 2024, p. 62) to prevent future risks. Identified with crisis communication itself, the *Recovery* phase begins after the end of the *Rescue* phase and lasts eight weeks. Depending on the month when the *Recovery* phase ends, the loop returns to the phase corresponding to the current month. *Recovery* is a phase related to crisis communication. In this phase, the purpose of communication is to inform affected communities about public policies they can access to restore themselves physically, emotionally and materially.

Regarding the role of public organizations in wildfire management, to evaluate implies accounting for actions and decisions taken before, during and after the crisis and informing about corrections, changes and improvements implemented as a result of lessons learned (Reynolds & Seeger, 2005). Communication plays a crucial role in this process, ensuring transparency and reinforcing public trust in general and particularly among affected audiences. The *Evaluation* phase is the final phase of the model, when communication should serve to examine the risk and crisis management carried out and inform the public about planned actions for the following year, when the loop returns to its initial phase. *Evaluation* is the phase in which public organizations analyze what worked and what did not in managing wildfires. The model prescribes that this reflection be conducted as a way to learn and correct communication strategies before the next loop begins. *Evaluation* phase marks the transition from crisis back to risk.

4. Qualitative Categorization

Each chronological phase of the framework corresponds to specific communication objectives aimed at safeguarding the physical, emotional and material well-being of individuals and communities affected by wildfires, as outlined in Table 1. These objectives develop into a set of informational components (*Table 2*) aligned with the informational perspective of risk and crisis communication. Drawing from an integrative literature review of Diers-Lawson (2020), Wukich (2019, 2016), CDC (2014), and Reich, Bentman and Jackman (2011), the proposed model prescribes 36 key messages that can be disseminated by public organizations before, during and after wildfires or any other type of disaster, with the necessary adaptations.

<i>Communication</i>	<i>Phase</i>	<i>Qualitative categories of messages</i>
Risk	<i>Prevention - Type 1</i>	The message communicates what has happened in previous similar events and what is expected to happen. The message disseminates the contact details of the emergency communication center. The message informs about practical measures to take to minimize risks (do's and don'ts) and self-protection. The message mobilizes citizens to get involved in prevention and/or self-protection activities.
	<i>Preparation - Type 2</i>	The message addresses knowledge, resources and capacities needed by the community and/or individuals to deal with a potential event. The message makes known measures planned or recently carried out to better prepare public organizations and the community itself to respond to the event (i.e. campaigns; awareness-raising; procurement; hiring; training; simulations; and reinforcements). The message, which is educational in nature, presents a kind of guide for preparing the community. The message informs about training opportunities for the community.
	<i>Alert - Type 3</i>	The message warns of penalties for those do not comply with orders and/or cause risk situations. The message calls on the public to take immediate action to prevent wildfires and/or self-protection and/or to consult the probability of risk for the region. The message says exactly what citizens should do between the emergency alert and the start of the crisis. The message provides follow-up on previously sent emergency information. The message tells citizens what to do when a crisis breaks out. The message informs which organizations are available for assistance. The message instructs on how citizens can protect themselves.
Risk -> Crisis	<i>Rescue - Type 4</i>	The message presents cases of citizens who survived the crisis by following the instructions given. The message gives simple and practical self-defense instructions. The message delivers updates on ongoing events. The message clarifies/disproves rumors and misinformation about the event and its management. The message informs about the provision of resources for victims. The message informs about administrative and operational acts directly linked to the event carried out in the last 48 hours.

		The message informs who exactly should be evacuated, by when and at what distance. The message asks society to help the victims (i.e. crowdfunding, donations and volunteering). The message, or the action it refers to, is empathy-oriented and tries to provide emotional support.
Crisis	Recovery - Type 5	The message addresses better preparation for similar future events, while using the freshness of the participants' experiences and memory. The message presents the new post-crisis reality in relation to various aspects of life (i.e. whether another eruption of the crisis is likely, at least in the short term, and how soon they will be able to resume normal life). The message informs that investigations are being carried out/requires clarification about the event. The message informs about public policies from which the affected communities can benefit and/or about the resumption of services that were suspended. The message provides preliminary clarification on the event that occurred. The message promotes a discussion about future preparedness for the occurrence of similar events.
Crisis -> Risk	Evaluation - Type 6	The message reinforces the resilience of society. The message presents a balance sheet or report on the actions taken during the previous phases. The message speaks of broad organizational; structural; legislative; and financial reforms, including investment in large-scale projects that will lead to better preparedness in the future. The message informs about collaboration with journalistic investigations and even the production of documentary films that examine events from an after-the-fact perspective. The message informs about organizational and inter-organizational procedures and their implementation within the organization and with the public. The message shares documents summarizing the conclusions of the investigations carried out.

Table 2 - Qualitative categorization and key messages of the Loop Model. Source: Adapted by the author from Diers-Lawson (2020); Wukich (2019, 2016); CDC (2014); and Reich, Bentman and Jackman (2011).

These categories were grouped according to their pertinence to each chronological phase of the model, following the functional and temporal logic of the communication process. They are not hierarchically ordered, as each category fulfills a distinct and context-dependent purpose within the cycle of risk and crisis communication. Illustrative examples of key messages aligned with the different phases of the framework are presented in Figures 3 to 8. These examples were collected from the official social media profiles of public entities entrusted with risk and crisis management in Portugal, including the Institute for Nature Conservation and Forests (ICNF), the National Authority for Emergency and Civil Protection, the National Republican Guard (GNR), the National Institute of Medical Emergency (INEM), and the Government of Portugal.



Figure 3: Message, issued in 2017, chronologically and qualitatively associated with the Prevention phase (Type 1). The message informs of practical measures to be taken for self-protection or to minimize risks (what to do and what not to do). Source: ICNF Facebook Fanpage³.

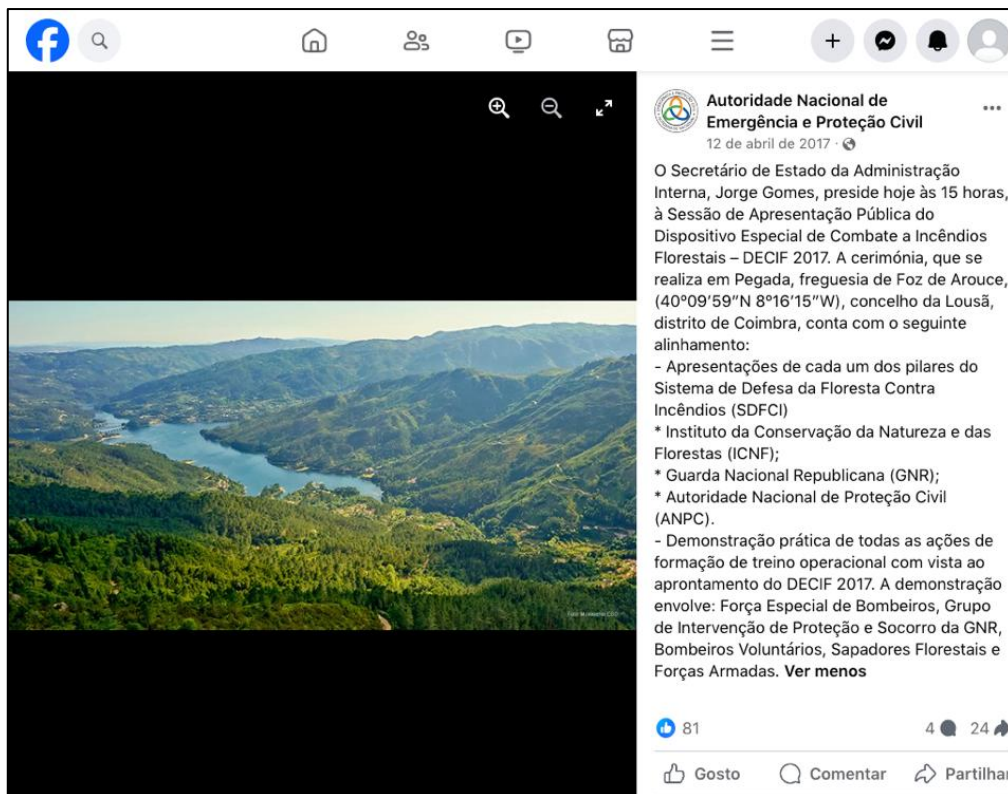


Figure 4: Message, issued in 2017, chronologically and qualitatively associated with the Preparation phase (Type 2). The message publicizes measures planned or recently carried out to better prepare the public administration and the community itself to respond to the event (campaigns, acquisitions, hiring, training, simulations, reinforcements etc). Source: ANEPC Facebook Fanpage⁴.



Figure 5 - Message, issued in 2017, chronologically and qualitatively associated with the Alert phase (Type 3). The message calls on the public to take immediate measures for self-protection and/or consult the probability of risk for the region. Source: GNR Facebook Fanpage⁵.

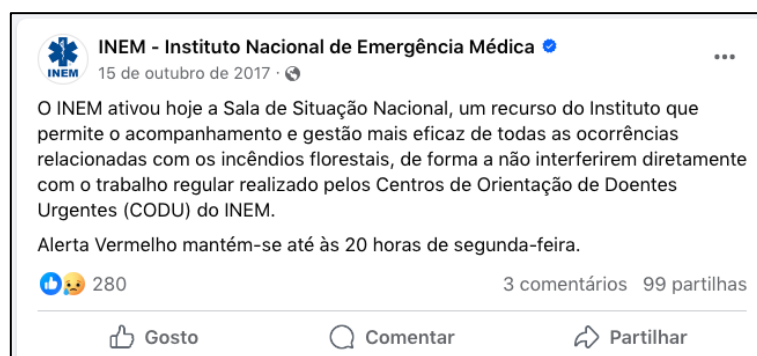


Figure 6 - Message, issued in 2017, chronologically and qualitatively associated with the Rescue phase (Type 4) of the October Fires (15, 16 and 17 October). The message informs about administrative and operational acts directly linked to the event carried out in the last 48 hours. Source: INEM Facebook Fanpage⁶.

HISTÓRICO XXI GOVERNO - REPÚBLICA PORTUGUESA
VOLTAR PARA GOVERNO EM FUNÇÕES


REPÚBLICA PORTUGUESA
XXI GOVERNO CONSTITUCIONAL

PT EN



Notícias

Página Inicial > Comunicação > Notícias

2017-07-04 às 14h33

Programa de saúde de apoio às populações afetadas pelos incêndios «pode durar mais de dois anos»



O Secretário de Estado Adjunto e da Saúde, Fernando Araújo, afirmou que o programa de saúde de apoio às populações afetadas pelo incêndio de 17 de junho em Pedrógão Grande, Figueiró dos Vinhos e Castanheira de Pera «pode durar mais de dois anos».

Em Pedrógão Grande, no final de uma reunião com 14 estruturas do Ministério da Saúde, o Secretário de Estado referiu que «esta é uma resposta dinâmica e flexível, disponível para ir às populações mais remotas, às pessoas mais isoladas, de modo a que ninguém fique para trás» e engloba as vertentes da doença orgânica, saúde mental e saúde pública.

As várias etapas previstas no plano de ação estão «programadas, articuladas e coordenadas» e o número de profissionais será o necessário nas várias áreas e especialidades, sendo que no caso da saúde mental pode estender-se para lá do período de dois anos.

«Temos o exemplo da tragédia de Entre-os-Rios, em que na saúde mental foram precisos mais anos de apoio e, portanto, temos de ter a capacidade de ao longo do tempo ir reconhecendo as necessidades», acrescentou.





Reunião do Secretário de Estado Adjunto e da Saúde, Fernando Araújo, com 14 estruturas do Ministério da Saúde, Pedrógão Grande, 4 julho 2017

Foto: Reunião do Secretário de Estado Adjunto e da Saúde, Fernando Araújo, com 14 estruturas do Ministério da Saúde, Pedrógão Grande, 4 julho 2017

Tags:
[saúde](#), [incêndios](#)

Áreas:
[Saúde](#)

Figure 7 - Message, issued in 2017, chronologically and qualitatively associated with the Recovery (Type 5) of the Pedrógão Grande Fires (20 June to 14 August). The message informs about public policies from which affected communities can benefit and/or about the resumption of services that were suspended. Source: Website of the XXI Portuguese Constitutional Government⁷.

HISTÓRICO XXI GOVERNO - REPÚBLICA PORTUGUESA
VOLTAR PARA GOVERNO EM FUNÇÕES

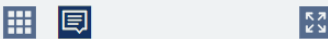

REPÚBLICA PORTUGUESA
XXI GOVERNO CONSTITUCIONAL
PT EN

Notícias

Página Inicial > Comunicação > Notícias
2017-11-28 às 17h08

Relatório do Conselho para a fixação de critérios de Indemnização das Vítimas de Incêndios





Primeiro-Ministro, António Costa, e Ministra da Justiça, Francisca Van Dunem, após reunião com o Comité das Vítimas dos Incêndios, novembro, 28 novembro 2017 (Foto: Clara Azevedo)

Por Resolução do Conselho de Ministros N.º157-C/2017, de 21 de outubro, o Governo assumiu, em nome do Estado, a responsabilidade pelo pagamento das indemnizações decorrentes das mortes das vítimas dos incêndios florestais ocorridos em Portugal Continental, nos dias 17 a 24 de Junho e 15 a 16 de outubro de 2017, e aprovou um mecanismo extrajudicial, de adesão voluntária, destinado à determinação e ao pagamento, de forma ágil e simples, de indemnizações por perdas e danos, não patrimoniais e patrimoniais, por morte das vítimas, aos respetivos familiares, herdeiros e demais titulares do direito à indemnização.

Foi assim constituído um Conselho para fixar, de acordo com o princípio da equidade, os critérios a utilizar no cálculo das indemnizações, bem como os prazos e procedimentos necessários para os titulares do direito poderem exercê-lo, e nomeados como membros do Conselho três juristas de reconhecido mérito e experiência – sendo um indicado pelo Conselho Superior da Magistratura, de entre juizes de tribunal superior, outro indicado pelo Conselho de Reitores das Universidades Portuguesas e o terceiro indicado por associação representativa de titulares do direito de indemnização pela morte das vítimas dos incêndios.

O referido Conselho entregou hoje ao Primeiro-Ministro o Relatório do seu trabalho, com a fixação dos critérios a utilizar para cálculo das indemnizações a pagar pelo Estado aos titulares do direito à indemnização por morte das vítimas dos incêndios que deflagraram em Portugal Continental nos dias 17 de junho e 15 de outubro de 2017.

Caberá agora à Provedora de Justiça a determinação, de acordo com os critérios ora fixados, e mediante um procedimento célere e simples, do montante da indemnização a pagar em cada caso concreto, e o seu respetivo pagamento.

Tags:
[incêndios](#), [justiça](#)

Áreas:
[Primeiro Ministro](#), [Justiça](#)

Figure 8 - Message, issued in 2017, chronologically and qualitatively associated with the Evaluation phase (Type 6). The message informs about organizational and inter-organizational procedures and their implementation in the organization and with the public

Source: Website of the XXI Portuguese Constitutional Government⁸.

When a message about wildfires, issued by a public organization, does not align with any of the categories outlined in Table 2, it indicates that the message is associated with the political perspective rather than the informational perspective of risk and crisis communication. Consequently, it is classified as a strictly political message.

5. Message Design

Regarding message design, we take into account four well-established paradigms in the field of risk and crisis communication, namely the IDEA Model by Sellnow et al. (2019a, 2019b, 2017, 2015); the dimensions of communication in public organizations during disasters by Olsson (2014), combined with thematic content categories by Bytzek (2008); and the SCCT developed by Coombs (2017, 2007), as detailed in *Table 3*.

<i>IDEA Model</i>
Effective risk and crisis communication during the <i>Alert</i> and <i>Rescue</i> phases must contain at least one of the three message elements - “internalization”, “explanation” and/or “action” - advocated in the IDEA Model.
<i>Dimensions of Communication by Public Organizations in Disaster Situations + Thematic Content Categories</i>
The <i>Rescue</i> phase is when all communication efforts, whether operational or strategic, need to be made to guarantee the safety of the population. For this reason, in this phase messages associated with the operational dimension and the strategic dimension can be issued to the same extent.
<i>SCCT</i>
In the <i>Rescue</i> , <i>Recovery</i> and <i>Evaluation</i> phases and/or when the most important political authority is the spokesperson for the message, the predominant response strategies, alone or in combination, are “excuse”; “justification”; “compensation”; “apology”; and “ingratiation”.

Table 3 - Characteristics of message design in the Loop Model. Source: Elaborated by the authors.

The element Distribution, represented by the “D” in the IDEA Model, is not included in the message design, as it pertains to the channels through which the message is disseminated rather than the content of the message itself.

6. Framing of Actions

Risk and crisis communication is characterized by the predominance of the informational perspective when the messages issued during the phases associated with risk – namely *Prevention*, *Preparation* and *Alert* – frame the potential damages that may arise from citizens either adhering to or neglecting the guidance provided by public organizations (Seeger, Sellnow & Ulmer, 2003). This assertion is consistent with Hallahan's (2005, 1999) framing model of actions, which posits that presenting the consequences of compliance or non-compliance with the recommended actions in either positive or negative terms can significantly influence audience behavior.

7. Function of Language

In the *Prevention*, *Alert* and *Rescue* phases, messages utilize the conative (or appellative) function of language (Jakobson, 2003), aiming to connect with the audience through second-person pronouns and verbs (i.e., “you” singular/plural). In contrast, the *Preparation*, *Recovery* and *Evaluation* phases emphasize the referential (or cognitive/denotative) function, resembling news content. The emotive (or expressive) function appears in the *Rescue* and *Recovery* phases, where the spokesperson may adopt a more prominent role, using first-person pronouns and verbs (i.e., “I” and “we”) to enhance rhetorical appeal.

8. Communication Campaign

Risk and crisis communication campaigns —defined as concentrated efforts that adapt language for various media and platforms—are highly effective in preparing the population during the pre-crisis period, specifically during the phases associated with risk. Therefore, we should anticipate finding materials produced as part of communication campaigns in the *Prevention*, *Preparation* and *Alert* phases. This aligns with the original formulation of the CERC model, which associates communication campaigns aimed at raising awareness, informing and educating the population more closely with risk communication (CDC, 2014).

9. Mediation

When risk and crisis communication is aligned with the informational perspective, public organizations engage in both direct communication—through digital social media—and indirect communication via news media. However, during the *Prevention*, *Preparation* and *Alert* phases, direct communication (e.g., social media posts) predominates over mediated communication. In contrast, during the *Recovery* and *Evaluation* phases, mediated communication takes precedence, with messages primarily aimed at journalists (e.g., press releases and news articles published on websites). During the *Rescue* phase, public organizations utilize both mediated and direct communication equally to achieve their objectives (Liu, Fraustino & Jin, 2015; Olsson, 2014; Reich, Bentman & Jackman, 2011).

10. Audience

The way in which the sender addresses the audience—either as a specific or universal audience—can be discerned through the use of pronouns and verb conjugations within the text. When risk and crisis communication adheres to the informational perspective, messages aimed at a specific audience are prevalent during the *Prevention*, *Preparation*

and *Alert* phases. In these instances, verbs are conjugated in the second- and third-person singular (“you”; “the citizen”; “each individual”). Conversely, messages disseminated during the *Rescue*, *Recovery* and *Evaluation* phases target a universal audience, employing second- and third-person plural forms (“we”; “you” plural; “the population”; “everyone”) (Wang et al., 2022; Sellnow et al., 2009; Seeger, Sellnow & Ulmer, 2003), or use an indeterminate subject.

11. Behaviors

Regarding the effects of messages, when risk and crisis communication is associated with the informational perspective, the behavior that the sender intends to elicit from the receiver is compliance with the guidelines provided by public organizations during the risk and emergency-related phases: *Prevention*, *Preparation*, *Alert*, and *Rescue*.

12. Opinions

The *Loop Model* underscores the significance of monitoring the audience's reactions to messages from public organizations, including how these messages are shared and discussed. Interaction occurs through both digital social media and word of mouth, involving peers, influencers, news media and the senders themselves (Austin & Jin, 2017). It is also essential to observe whether the rhetorical arena opened during a crisis reveals multiple voices that may converge, diverge, coexist, or compete, thereby influencing public opinion and adherence to organizational guidelines (Frandsen & Johansen, 2022).

Informational Quality Indicator

As noted by Frandsen and Johansen (2020b), the informational perspective relies on the existence of a rational consensus regarding the perception of risk, whereas the political perspective is characterized by symbolic and framing disputes typical of agonistic models. In scenarios where human lives are at stake, it is evident which perspective should take precedence. While it does not negate the fact that risk and crisis communication developed from an informational perspective can enhance the reputation of public organizations and political authorities (Canel & Luoma-Aho, 2018), it emphasizes the need to prioritize the protection of potentially affected citizens.

To analyze whether the risk and crisis communication from public organizations engaged in wildfire management aligns more closely with the informational or political perspective (Frandsen & Johansen, 2020b), we developed an aggregate indicator derived from the multivariate analysis of the messages, termed the Informational Quality Indicator (IQI). The

ten observed variables in our study yielded thirteen factors, whose presence or absence in the communication from public organizations will be verified. All factors that compose the IQI carry equal weight (1). A summary of these factors is presented in *Table 4*.

<i>Factor 1</i>	Predominance of messages with the framing of actions corresponding to the respective chronological phase.
<i>Factor 2</i>	Predominance of messages with the spokesperson corresponding to the respective chronological phase.
<i>Factor 3</i>	Demonstration of emotions by political authorities during the <i>Recovery</i> chronological phase.
<i>Factor 4</i>	Sending of messages by public organizations with content (qualitative category of messages) corresponding to their respective attributions in wildfire management.
<i>Factor 5</i>	Presence of messages directed at a specific audience in the <i>Prevention</i> , <i>Preparation</i> and <i>Alert</i> chronological phases.
<i>Factor 6</i>	Predominance of messages issued through the channel corresponding to the respective chronological phase.
<i>Factor 7</i>	Presence of communication campaigns in the <i>Prevention</i> , <i>Preparation</i> and <i>Alert</i> chronological phases.
<i>Factor 8</i>	Presence of at least one of the three message elements advocated in the IDEA Model in messages qualitatively identified with the <i>Alert</i> (Type 3) and <i>Rescue</i> (Type 4) phases.
<i>Factor 9</i>	Presence of operational and strategic aspects in messages issued in the <i>Rescue</i> chronological phase.
<i>Factor 10</i>	Predominance of “excuse”; “justification”; “compensation”; “apology”; and “ingratiation” response strategies in messages issued in the <i>Rescue</i> and <i>Recovery</i> chronological phases, as long as they are qualitatively identified with the <i>Rescue</i> (Type 4), <i>Recovery</i> (Type 5), <i>Evaluation</i> (Type 6) phases or strictly political (Type 7) messages; or if the most important political authority is the spokesperson for the message.
<i>Factor 11</i>	Predominance of messages associated with the informational perspective in all chronological phases.
<i>Factor 12</i>	Correlation between the chronological phase and the qualitative category of the message.
<i>Factor 13</i>	Predominance of messages with the function of language corresponding to the respective chronological phase.

Table 4: Risk and crisis communication assessment factors according to the Loop Model. Source: Elaborated by the authors.

The 13 factors must be observed in each phase of the model. The verification of each factor’s presence or absence is assigned values: “-1” (absent), “0” (not applicable or neutral) and “1” (present). The equation that defines the indicator is presented below:

$$IQI = \frac{\sum (f1:f13)}{13 - (n \times \text{value “0”})}$$

The range of variation for the indicator spans from “-1” to “1,” as the factors that comprise it fall within this same range. Consequently, values closer to “-1” indicate lower informational quality, suggesting that risk and crisis communication is more aligned with the political perspective. Conversely, values nearing “1” reflect higher informational quality, indicating a greater alignment with the informational perspective. This equation can be utilized in studies employing content analysis as the empirical research methodology. *Table 5* summarizes the main theoretical constructs and the variables of the framework.

When says		Who says		What says					By which channel says	To who says	With which intention says
Communication	Chronological categorization	Source	Spokesperson	Qualitative categorization	Message design	Framing of actions	Function of language	Communication campaigns	Mediation	Audience	Effects
Risk	Prevention	Executing organizations	Scientific	Type 1	IDEA	Negative	Apellative	Yes	More direct than mediated	Specific	Opinion + Behavior
	Preparation		Technical	Type 2			Referential				
	Alert		Scientific	Type 3			Apellative				
Risk -> Crisis	Rescue	Decision-making + Executing organizations	Technical	Type 4	IDEA + SCTT	Not focused on the consequences of actions	Apeliative + Emotive	No	Direct + Mediated	Universal	Opinion
Crisis	Recuperation	Decision-making organizations	Political	Type 5	SCTT	Positive	Referencial + Emotive				
Crisis -> Risk	Evaluation			Type 6			Referential				
IQI Factor	F11, F12	F4	F2, F3	F12	F8, F9, F10	F1	F13	F7	F6	F5	N.A.

Table 5 - Main theoretical constructs and variables of the Loop Model. Source: Elaborated by the authors.

Discussion and Conclusion

The pursuit of models capable of synthesizing the risk and crisis management process is not new and has grown over the past decades. Sellnow and Seeger (2021) cataloged numerous theories that emphasize communication as a central component of this process. However, the inherent complexity of the concepts of risk and crisis render the development of a definitive and universally applicable model unlikely. This complexity is further compounded in disasters, such as wildfires, which constitute crises with distinct and highly specific characteristics.

To enhance risk and crisis communication regarding wildfires from an informational perspective—prioritizing public protection over political objectives—it is crucial for public organizations to embrace the principles of systematic communication. This entails adopting a strategic, planned, integrated, and structured approach, as exemplified by the proposed model, particularly in phases associated with risk communication, such as *Prevention* and *Preparation*, which are often overlooked (Taylor, Johnston & Ryan, 2022).

The accurate application of the model variables – relating to control, content, media, and audience – is essential for enhancing informational quality indicators and, consequently, safeguarding the population before, during and after wildfires.

We believe that establishing continuous communication about wildfires and actively engaging the public, as well as public and private organizations and civil society, in fostering a risk culture (Plana et al., 2024), could not only help mitigate the immediate impacts of wildfires and promote long-term community resilience. This resilience is crucial for addressing an issue that is becoming increasingly challenging and is expected to intensify in the future.

Community engagement and dialogue are central features in several definitions of risk communication, including the classic formulation by the NRC (1989). According to Sellnow and Sellnow (2023), dialogue in risk communication serves to promote public commitment and individual learning about risks, turning communication into a process of shared reflection and social co-responsibility.

As noted by Sellnow et al. (2009), risk communication is an interactive and dynamic process involving multiple, and often competing, messages. Establishing dialogue, pursuing collaborative conflict resolution, and fostering consensus-building are essential for effective communication and for cultivating mutual trust.

From a theoretical standpoint, this chapter contributes by proposing an original model that integrates several well-tested theories and frameworks. Communication strategies based on the *Loop Model* not only can aid in safeguarding the public but also have the potential to enhance the reputation of public organizations and political actors involved. Practically, the framework is useful for planning and evaluating, from a systematic perspective, the communication of public organizations regarding wildfires or any other type of meteorological disaster, provided the model is adapted properly, such as hurricanes in the Gulf of Mexico, typhoons in Southeast Asia, cyclones in Bangladesh, floods in southern Brazil, and wildfires in California, for example. Nevertheless, this article is limited by its theoretical scope. The empirical validation of both the model and the IQI could not be included within the present study due to space constraints and the conceptual nature of the contribution.

The proposed *model* also encourages participatory and community-based approaches, emphasizing the role of citizen engagement and local knowledge in anticipating communication needs and strengthening public trust.

Thus, the framework is inherently normative, as it serves as a prescriptive framework that guides public organizations how to communicate effectively. However, it can also be regarded as an analytical model, when applied to the examination of messages disseminated during wildfire communication management.

Regarding communication theories, our study advances the rhetorical tradition (Craig, 1999) by exploring the persuasive function of communication in risk and crisis scenarios, such as wildfires. We contend that risk and crisis communication should craft messages tailored to the specific phases outlined in the framework to effectively promote public safety and social engagement. To reaffirm the significance of rhetorical discourse, we proposed an evaluation indicator that assesses the risk and crisis communication practices of public organizations regarding wildfires. The capacity to objectively identify areas for improvement in organizational communication aligns with the principles of the rhetorical tradition, which advocates the continuous improvement of communicative practices.

Future studies should focus on empirically testing the predictive and analytical effectiveness of the model and the IQI through quantitative and qualitative methodologies. Pilot applications in different institutional contexts and in other types of climate- and weather-related disasters with other chronological patterns, such as hurricanes and floods, would be useful to assess its broader validity and adaptability across diverse scenarios.

Acknowledgements

This research was developed with the financial support of the Fundação para a Ciência e a Tecnologia (FCT) — the Portuguese Foundation for Science and Technology — under grant No. 2020.05702.BD.

Biographical notes

Bianca Persici Toniolo - Ph.D. in Communication Sciences, Assistant Invited Professor at the University of Beira Interior (UBI), Researcher at LabCom – Laboratory of Communication and Vice-Chair of the Risk and Crisis Communication Section at ECREA.

Gisela Gonçalves – Ph.D. in Communication Sciences, Associate Professor at the University of Beira Interior (UBI), Coordinator of the Strategic Communication Research Group of LabCom – Laboratory of Communication, Coordinator of the Master Programm in Strategic Communication at UBI, Vice-Chair of the Strategic Communication Working Group at IAMCR, and Vice-President of Sopcom, the Portuguese Association of Communication Sciences.

References

- Almeida, M., Modarres, M. R., & Rodrigues, T. (2024). Proteção das comunidades. In D. X. Viegas (Ed.), *O Que Devo Saber Sobre Tempestades de Fogo?* (pp. 89-107). Lidel.
- An, S.-K., & Cheng, I.-H. (2010). Crisis communication research in public relations journals: Tracking research trends over thirty years. In W. T. Coombs & S. J. Holladay (Eds.), *The Handbook of Crisis Communication* (pp. 65-90). Wiley-Blackwell.
- Austin, L., & Jin, Y. (2017). Social media and crisis communication: Explicating the Social-Mediated Crisis Communication Model. In A. Dudo & L. Kahlor (Eds.), *Strategic Communication: New Agendas in Communication* (pp. 163-186). Routledge.
- Avery, E. J., Lariscy, R. W., Kim, S., & Hocke, T. (2010). A quantitative review of crisis communication research in public relations from 1991 to 2009. *Public Relations Review*, 36(2), 190-192. <https://doi.org/10.1016/j.pubrev.2010.01.001>
- Benoit, W. L. (1995). *Accounts, Excuses, and Apologies: A Theory of Image Restoration Strategies*. State University of New York Press.
- Benoit, W. L. (1997). Image repair discourse and crisis communication. *Public Relations Review*, 23(2), 177-186. [https://doi.org/10.1016/S0363-8111\(97\)90023-0](https://doi.org/10.1016/S0363-8111(97)90023-0)
- Boin, A., 't Hart, P., Stern, E., & Sundelius, B. (2005). *The Politics of Crisis Management: Public Leadership Under Pressure*. Cambridge University Press.
- Bytzek, E. (2008). Flood response and political survival: Gerhard Schröder and the 2002 Elbe flood in Germany. In A. Boin, A. McConnell, P. 't Hart (Eds.), *Governing After Crisis: The Politics of Investigation, Accountability and Learning*, (pp. 85-113). Cambridge University Press.

- Canel, M.-J., & Luoma-aho, V. A. (2018). *Public Sector Communication: Closing Gaps Between Citizens and Public Organizations*. Wiley-Blackwell. 276p.
- Centers for Disease Control and Prevention (CDC). (2014). *Crisis and Emergency Risk Communication*. https://emergency.cdc.gov/cerc/ppt/cerc_2014edition_copy.pdf
- Coombs, W. T. (2007). Protecting organization reputations during a crisis: The development and application of situational crisis communication theory. *Corporate Reputation Review*, 10, 163-176. <https://doi.org/10.1057/palgrave.crr.1550049>
- Coombs, W. T. (2017). Revising Situational Crisis Communication Theory: The influence of social media on crisis communication theory and practice. In Y. Jin & L. Austin (Eds.), *Social Media and Crisis Communication* (pp. 21-37). Routledge.
- Coombs, W. T. (2022). *Ongoing Crisis Communication: Planning, Managing, and Responding*. 6th Edition. Sage.
- Dewey, J. (2007). O desenvolvimento do pragmatismo americano. *Scientia Studia*, 5(2), 227-243. <https://doi.org/10.1590/S1678-31662007000200006>
- Diers-Lawson, A. (2020). *Crisis Communication: Managing Stakeholders Relationships*. Routledge.
- Elgueta-Ruiz, A., & Martínez-Ortiz, J. (2022). Public relations in the new millennium: Towards a synoptic view of fifteen years of research in Public Relations Review in the 21st century (2000-2014). *Public Relations Review*, 48(2), 1-17. <https://doi.org/10.1016/j.pubrev.2022.102175>
- Frandsen, F., & Johansen, W. (2022). Rhetorical Arena Theory: Revisited and expanded. In W. T. Coombs & S. J. Holladay (Eds.), *The Handbook of Crisis Communication* (pp. 170-181). Wiley-Blackwell. <https://doi.org/10.1002/9781119678953.ch12>
- Frandsen, F., & Johansen, W. (2017a). *Organizational Crisis Communication: A Multivocal Approach*. Sage.
- Frandsen, F., & Johansen, W. (2017b). Voices in conflict? The crisis communication of meta-organizations. *Management Communication Quarterly*, 32(1), 90-120. <https://doi.org/10.1177/0893318917705734>
- Frandsen, F., & Johansen, W. (2020a). Reframing the field: Public crisis management, political crisis management, and corporate crisis management. In F. Frandsen & W. Johansen (Eds.), *Crisis Communication* (pp. 59-102). Mouton De Gruyter. <https://doi.org/10.1515/9783110554236-003>
- Frandsen, F., & Johansen, W. (2020b). Public sector communication: Risk and crisis communication. In V. Luoma-aho & M.-J. Canel, *The Handbook of Public Sector Communication* (pp. 229-244). Wiley-Blackwell.
- Hallahan, K. (1999). Seven Models of Framing: Implications for Public Relations. *Journal of Public Relations Research*, 11(3), 205-242. https://doi.org/10.1207/s1532754xjpr1103_02
- Hallahan, K. (2005). Framing theory. In R. L. Heath (Eds.), *The Encyclopedia of Public Relations* (pp. 340-343). Sage.

- Horsley, J. S. (2016). Media framing of disasters: Implications for disaster response communicators. In A. Schwarz, M. W. Seeger & C. Auer (Eds.), *The Handbook of International Crisis Communication Research* (pp. 155-164). Wiley-Blackwell.
- Jakobson, R. (2003). *Essais de Linguistique Générale: Les Fondations du Language*. Volume 1. Les Éditions de Minuit.
- Kuipers, S., & Welsh, N.H. (2017). Taxonomy of the crisis and disaster literature: Themes and types in 34 years of research. *Risk, Hazards & Crisis in Public Policy*, 8(4), 272-283. <https://doi.org/10.1002/rhc3.12123>
- Lasswell, H. (1948). The structure and function of communication in society. In L. Bryson (Eds.), *The Communication of Ideas* (pp. 215-228). Harper & Brothers.
- Lipset, S. M. (1953). Opinion formation in a crisis situation. *Public Opinion Quarterly*, 17(1), 20-46. <https://doi.org/10.1086/266437>
- Liu, B. F., Fraustino, J. D., & Jin, Y. (2015). Social media use during disasters: How information form and source influence intended behavioral responses. *Communication Research*, 43(5), 1-21. <https://doi.org/10.1177/0093650214565917>
- McLean, H., & Ewart, J. (2015). Hindrance or help? A model for the involvement of politicians in communicating with publics during disasters. *International Journal of Mass Emergencies and Disasters*, 33(2), 228-252.
- National Research Council (NRC) (United States). (1989). *Improving Risk Communication*. Committee on Risk Perception and Communication. National Academies Press.
- Novo, I., Moreira, N., Silva, P., Ferreira, J., Rio, J., Pinto, P., & Silva, A. P. (2024). Condições meteorológicas e clima. In D. X. Viegas (Eds.), *O Que Devo Saber Sobre Tempestades de Fogo?* (pp. 9-33). Lidel.
- Olsson, E.-K. (2014). Crisis communication in public organizations. *Crisis Communication*, 22(2), 419-437. <https://doi.org/10.1515/9783110554236-020>
- Palenchar, M. J., & Heath, R. L. (2007). Strategic risk communication: Adding value to society. *Public Relations Review* (33), 120-129. <https://doi.org/10.1016/j.pubrev.2006.11.014>
- Pechta, L. E., Brandenburg, D. C., & Seeger, M. W. (2010). Understanding the dynamics of emergency communication: propositions for a four-channel model. *Journal of Homeland Security and Emergency Management*, 7(1), 1-18. <https://doi.org/10.2202/1547-7355.1671>
- Plana, E., Serra, M., Nebot, S., Smeenk, A., Macri, P., Vendrell, J., Pronto, L., Canaleta, G., Gomes, J., & Alfonso, L. (2024). *Wildfire risk awareness and communication: Analysis of good practices*. Union Civil Protection Knowledge Network. <https://doi.org/10.5281/zenodo.10817549>
- Raupp, J. (2015). Strategizing risk communication. In D. Holtzhausen & A. Zerfass (Eds.), *The Routledge Handbook of Strategic Communication* (pp. 520-532). Routledge.
- Reich, Z., Bentman, M., & Jackman, O. (2011). A crisis communication guide for public organisations. In M. Vos, R. Lund, Z. Reich, & H. Harro-Loit (Eds.), *Developing a Crisis Communication Scorecard* (pp. 265-325). University of Jyväskylä.
- Schnee, C. (2015). *Political Reputation Management: The Strategy Myth*. Routledge.

- Seeger, M. W., Sellnow, T. L., & Ulmer, R. R. (2003). *Communication and Organizational Crisis*. Praeger.
- Sellnow, D. D., Johansson, B., Sellnow, T. L., & Lane, D. R. (2019a). Toward a global understanding of the effects of the IDEA model for designing instructional risk and crisis messages: A food contamination experiment in Sweden. *Journal of Contingencies and Crisis Management*, 27(2), 102-115. <https://doi.org/10.1111/1468-5973.12234>
- Sellnow, D. D., Jones, L. M., Sellnow, T. L., Spence, P., Lane, D. R., & Haarstad, N. (2019b). The IDEA Model as a conceptual framework for designing Earthquake Early Warning (EEW) messages distributed via mobile phone apps. In J. Santos-Reyes, *Earthquakes: Impact, Community Vulnerability and Resilience* (pp. 11-20). IntechOpen. <https://doi.org/10.5772/intechopen.85557>
- Sellnow, D. D., Lane, D. R., Sellnow, T. L., & Littlefield R. S. (2017). The IDEA Model as a Best Practice for Effective Instructional Risk and Crisis Communication. *Communication Studies*, 68(5), 552-567. <https://doi.org/10.1080/10510974.2017.1375535>
- Sellnow, D. D., Limperos, A., Frisby, B. N., Sellnow, T. L., Spence, P. R., & Downs, E. (2015). Expanding the scope of instructional communication research: Looking beyond classroom contexts. *Communication Studies*, 66(4), 417-432. <https://doi.org/10.1080/10510974.2015.1057750>
- Sellnow, T. L., & Seeger, M. W. (2021). *Theorizing Crisis Communication (2nd Edition)*. Wiley.
- Sellnow, T. L., & Sellnow, D. D. (2023). *Before Crisis: The Practice of Effective Risk Communication*. Cognella.
- Sellnow, T. L., Ulmer, R. R., Seeger, M. W., & Littlefield, R. (2009). *Effective Risk Communication: A Message-Centered Approach*. Springer.
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333-339. <https://doi.org/10.1016/J.JBUSRES.2019.07.039>
- Tachkova E. R., & Coombs, W. T. (2022). *Communicating in Extreme Crises: Lessons from the Edge*. Routledge.
- Taylor, M., Johnston, K. A., & Ryan, B. (2022). A community engagement approach to natural hazard communication. In W. T. Coombs & S. J. Holladay (Eds.), *The Handbook of Crisis Communication* (pp. 327-342). Wiley-Blackwell. <https://doi.org/10.1002/9781119678953.ch22>
- Toniolo, B. P. (2025). *As tensões entre as perspectivas informativa e política da comunicação de risco e de crise: A comunicação das organizações públicas portuguesas sobre os incêndios florestais*. University of Beira Interior [Doctoral Thesis]. <http://hdl.handle.net/10400.6/15128>
- United Nations Office for Disaster Risk Reduction (UNDRR) (2022). *GAR Special Report: Our World at Risk*. 256p. <https://www.undrr.org/media/79595/download>

- Viegas, D. X. (2024). Glossário. In D. X. Viegas (Eds.), *O Que Devo Saber Sobre Tempestades de Fogo?* (pp. XI-XIV). Lidel.
- Werder, K. P., Nothhaft, H., Verčič, D., & Zerfass, A. (2018). Strategic communication as an emerging interdisciplinary paradigm. *International Journal of Strategic Communication*, 12(4), 333-351. <https://doi.org/10.1080/1553118X.2018.1494181>
- Wukich, C. (2016). Government social media messages across disaster phases. *Journal of Contingencies and Crisis Management*, 24(4), 230-243. <https://doi.org/10.1111/1468-5973.12119>
- Wukich, C. (2019). Preparing for disaster: Social media use for household, organizational, and community preparedness. *Risk, Hazards & Crisis in Public Policy*, 10(2), 233-260. <https://doi.org/10.1002/rhc3.12161>
- Yamori, K. (2020). Pitfall of disaster information. In K. Yamori (Eds.), *Disaster Risk Communication: A Challenge from a Social Psychological Perspective* (pp. 1-15). Springer. https://doi.org/10.1007/978-981-13-2318-8_1

Notes

¹ Retrieved December, 10, 2025 from <https://www.sgifr.gov.pt/pt/w/diretiva-operacional-nacional-n.%C2%BA-2-decir-2025>.

² Retrieved December, 10, 2025 from <https://prociv.gov.pt/pt/documentacao/diretiva-operacional-nacional-n.%C2%BA-2-decir-2025/>.

³ Retrieved April 23, 2021, from <https://www.facebook.com/InstitutoDaConservacaodaNaturezaedasFlorestas>.

⁴ Retrieved August 19, 2021, from <https://www.facebook.com/AutoridadeNacionalEmergenciaProtecaoCivil>.

⁵ Retrieved August 20, 2021, from <https://www.facebook.com/GuardaNacionalRepublicana>.

⁶ Retrieved August 24, 2021, from <https://www.facebook.com/InstitutoNacionaldeEmergenciaMedica/>.

⁷ Retrieved August 13, 2021, from <https://www.portugal.gov.pt/pt/gc21>.

⁸ Retrieved August 13, 2021, from <https://www.portugal.gov.pt/pt/gc21>.

Allies or Antagonists? The relationship between journalists and public communicators in risks and natural disasters *

Alessandro Lovari**
University of Cagliari

Daniela Pisu***
University of Cagliari

The article investigates the complex relationships between public communicators and journalists in the coverage of risks and natural disasters in Sardinia and Sicily, two Italian islands. Drawing on the theoretical frameworks of public relations and public sector communication, the manuscript presents findings from a qualitative study based on in-depth semi-structured interviews. Findings highlight a hybrid relational model in which institutional procedures intertwine with informal practices during these emergency situations. Although collaboration is perceived necessary by both professionals, it is hindered by bureaucratic constraints, journalistic routines, algorithmic and public pressures. Spectacularization and news sensationalism influence the relationship, raising ethical risks in a contemporary digital ecosystem marked by visibility and disinformation. In this context, journalists and public communicators emerge not as antagonists nor as fully aligned partners, but as imperfect allies jointly responsible for producing reliable and timely information to face risks and natural disasters, essential to supporting vulnerable territories, such as insular contexts.

Keywords: public communicators, journalists, natural disasters, risk communication, insularity

* Article submitted on 25/09/2025. Article accepted on 15/12/2025. The article is the result of the collaboration between the two authors. However, the introduction and the second paragraph were written by Daniela Pisu together with Alessandro Lovari, while all the other paragraphs were written by Alessandro Lovari.

** alessandro.lovari@unica.it

*** daniela.pisu@unica.it

Introduction

Over the past decades, scholarly attention to the concept of risk has grown considerably (Giddens, 1990; Beck, 1992), unfolding in different sectors and disciplines including communication and media research (Covello, Slovic & Winterfeld, 1986; Sellnow & Sellnow, 2024). This development reflects not only this increasing centrality of the concept in academia but also the broadening range of phenomena classified as risks, involving different types of stakeholders, including citizens and the media. The majority of these risks are today deeply mediatized (Hepp, 2019), due to the growing media attention and the role of digital platforms in contemporary communication ecologies. In this scenario public sector organizations play a crucial role in risk prevention, as well as in disaster and crisis communication management (Coman, Gregor & Lilliker, 2025; Coombs, 2020, Comunello & Mulargia, 2018) to inform citizens amid a cacophony of messages often polluted by disinformation (OECD, 2021).

Institutions need to relate with media and journalists who play a pivotal role in shaping public opinion agenda (McCombs & Shaw, 1972) and in the social amplification of risks (Kasperson et al., 1988). Different stakeholders' perception, specific territorial vulnerabilities, and citizens' information seeking practices require multidisciplinary approaches and integrated communication strategies in risk communication (Lovari & Ducci, 2022; Sellnow & Sellnow, 2024).

In this framework, the article investigates the complex relationship between public communicators and journalists in prevention and in the coverage of risks and natural disasters in Sardinia and Sicily, two Italian regions. Drawing on the theoretical frameworks of public relations and public sector communication (Canel & Luoma-aho, 2019), the manuscript presents findings from a qualitative study conducted with in-depth semi-structured interviews¹.

Literature review

The sociology of risk (Beck, 1992; Giddens, 1999) highlights that modernity is characterized by a progressive shift from natural danger to reflexive risk that is, a form of risk generated or amplified by the very social and technological practices designed to control it. Alongside the several dimensions that characterize the concept of risk, communication is being considered as one of the most crucial ones (Heath & O'Hair, 2009). Indeed, the increasing frequency and intensity of natural disasters, exacerbated by climate crises, have made risk communication a crucial dimension of contemporary society, for understanding persistent and emerging problems that occur in different part of the world (Diers-Lawson & Meißner, 2021; Massa & Comunello, 2024). The impacts of climate change, which are distributed unequally across the world, have become clearly visible through floods, droughts, wildfires, heat waves and sea level rise (Reyes et al.,

2021), and are strongly shaping how public sector organizations communicate with citizens and the media and their relationships (World Economic Forum, 2024).

The different forms of citizens' response are closely linked to risk and emergency communication management by government and public sector organizations (i.e., regions, municipalities, civil protections), whose legitimacy depends on trust in institutions (OECD, 2021). Public sector communication (Canel & Luoma-aho, 2019) is the leverage to transparently inform and actively engage multiple stakeholders in local territories also during uncertain situations and critical events that can harm different publics (Lovari, D'Ambrosi & Bowen, 2020). Indeed, in an era characterized by mega crises and mega risks (Sellnow-Richmond & Lukacovic, 2025), as well as by the proliferation of digital platforms, the challenge therefore lies not only in communicating risk, but in nurturing intangible assets, like trust, legitimacy and social capital around its management (Canel & Luoma-aho, 2019). Thus, when risks turn into disasters and emergencies, communicative effectiveness is influenced by the ability of institutions and their communicators to maintain an ongoing dialogue with media and citizens, acknowledging the plurality of knowledge and the narrative dimension of collective safety, while integrating digital communication tools to foster civic engagement and enable digital volunteerism (Comunello, 2014; Lovari & Ducci, 2022; Smith et al., 2021).

Numerous studies have highlighted the important role of media and media coverage in informing citizens about natural risks, playing a central function in shaping public perception of these issues (Boykoff & Boykoff, 2004; Carvalho & Burgess, 2005), in providing information that can prevent disasters (Buoncompagni, 2024; Öhman et al., 2016; Olofsson et al., 2016), or in expanding citizens' opportunities to participate in news production related to local emergencies (Farinosi, 2020). The media primary function is to insert these issues on citizens' agenda (McCombs & Shaw, 1972), to ensure timely and reliable information, while helping to prevent dysfunctional behaviors. At the same time several authors have highlighted how the media convey frames (Entman, 1993; Goffman, 1974) that can influence disasters' news production and the public's perception of these events (Altheide, 1997; Horsley, 2016; Houston, Pfefferbaum & Rosenholtz, 2012). Moreover, they contribute to the emotional dimensions of collective interpretation, often generating moral panic or assigning responsibility without addressing the structural conditions that produce vulnerability (Nogami, 2018; Tierney, Beve & Kuligowsky, 2006).

Depending on the frames used, mass media shape the social perception of risks and disasters and may act as amplification stations within the social imaginary of risk (Coombs, 2013; Luhmann, 1996). For instance, several authors have revealed that media representations of floods tend to privilege the dramatic and event-centred aspects of disaster coverage, while marginalising the systemic and infrastructural weaknesses that amplify such events or their prevention (Miller & Goidel, 2009; Stock, 2007; Zaheer, 2016).

Indeed, journalism is frequently criticized for its tendency toward news sensationalism, spectacularization, or dramatization of news. The terms denote the media's inclination to narrate events in a heightened, emotionally charged, or dramatized manner, emphasizing

the most affective, tragic, emotional or contentious elements to capture and keep public attention (Grabe, Zhuo & Barnett 2010; Sorrentino 2025; Wang & Cohen, 2009). In the context of severe emergencies resulting from natural disasters, the effects of such sensationalism often materialize through high-impact images and catching headlines, the use of myths and heroes (Nogami, 2018), the militarization of journalistic discourse (Horsley 2016), wherein the lexicon of “battle” and “resistance” functions as a means of collective mobilization (Mazzoleni, 2017).

Furthermore, in emergency contexts, the media also play a fundamental role through their watchdog function (Miller & Goidel, 2009; Norris, 2014), monitoring public institutions’ actions and highlighting any inefficiencies in the management of risks and emergencies (Houston et al., 2012). This journalistic oversight is essential to ensure greater transparency and accountability by governments and local authorities (Horsley, 2016). This function is also expressed in the media coverage using a responsibility frame (Kuppuswamy, 2017) when the narrative refers to the importance of the responsibility of politics and local institutions in the event of large-scale natural disasters (Albrecht, 2022; Lovari et al., 2025).

Communicators and journalists: allies or antagonists?

Scholarly literature has long documented the complex and often ambivalent relationship between journalists and communication professionals, particularly those managing media relations, in public and private organizations (Bulicanu, 2022; Supa & Zoch, 2009). While these actors may pursue divergent goals, their interaction remains largely symbiotic and mutually beneficial, grounded in the reciprocal exchange of information (Currah, 2009; Gandy, 1982). Yet, the persistence of journalistic norms of autonomy and independence from external influence (Gans, 1979; Shoemaker & Reese, 2014) often positions organizational communicators as potential constraints on journalistic inquiry and the pursuit of truth. Consequently, several authors have characterized this interaction as a “love-hate relationship” (Graber, 2003; Harcup, 2009), a process that intertwines the organizations’ need to increase their visibility and the newsmaking practices tending to emphasize organizational dysfunctions or crises. This scenario is today complexified by the presence of new digital arenas and by the role of digital influencers in the process of agenda setting (Badham & Luoma-aho, 2025).

Within the public sector, this relationship gains additional complexity due to ongoing processes of mediatization and the disintermediating effects of Web 2.0 technologies (Lovari & Ducci 2022). These developments have significantly reconfigured public sector communication practices and professional boundaries within institutions, as well as the relationships with journalists and citizens (Canel & Luoma-aho, 2019; Solito, Pezzoli & Materassi, 2019; Solito & Sorrentino, 2023). These dynamics extend beyond ordinary

contexts, shaping public information flows and power relations during emergencies and natural disasters giving such events, and their management, an additional digital visibility.

Public relations and communication research have examined the role of media and journalists in their relationships with public communicators - as well as with emergency managers and public information officers - during risk and emergency communication management (Horsley, 2016; Sellnow & Sellnow, 2024; Veil, 2012). In particular during a disaster the goals of journalists and public information professionals should be aligned to favor public outreach and great access to information in these critical situations (Fischer, 2008). The main common goal is to deliver vital information to the public (Veil, 2012). In general, the media cannot always satisfy the public sector communication needs of emergency management or prevention, and communicators are not always able to satisfy the information requests or the pressing deadlines of journalists (Fischer, 2008). Indeed, if public communicators are not ready when journalists contact them, or they are absent from the coverage of an emergency, their institutions may look unresponsive, untransparent and evasive toward the community needs (Veil, 2012). As stated by Miller and Goidel (2009, p. 267), "it is difficult to imagine government effectively responding to a natural disaster in the absence of news organizations reporting on breaking events".

The relationship between these two roles depends on several factors - professional, organizational and cultural ones - related to the development of different routines and models of journalism (Hallin & Mancini, 2012; Sorrentino, 2025), as well as to the level of professionalization of public sector communication (Canel & Luoma-aho, 2019; OECD, 2021) also in risk and crisis management (Lovari, 2025). Scholars pointed out how media still depends on information subsidies from risk and disasters response communicators to provide updates and leads for their ongoing news coverage (Horsley, 2016), even if the growing role of digital platforms has enabled citizens to livestream these events, or to provide information, videos and photos of the disasters directly to the media, bypassing emergency managers or public information officers. This fact can lead to the proliferation of misleading information or unofficial contents that risk being integrated in the news coverage, thus requiring communicators to monitor digital platforms and media for detecting rumors, inaccuracies or false messages that need to be immediately rectified in these critical situations (Lovari & Bowen, 2020).

The adversarial relationship between journalists and public communicators has long been identified as a recurring theme in public relations and emergency management scholarship (Faccioli, 2000; Fischer, 2008; Garner, 1984), highlighting the stereotypical judgments, lack of trust and unrealistic expectations between these two professional roles (Veil, 2012). On one side emergency sources (like governments, public communicators or emergency managers) often criticize the sensationalism and deficiencies of news coverage of disasters; on the other side journalists argue that a deficient media coverage depends mostly on lack of access to information by emergency sources (Grassau, Valenzuela & Puente, 2021). But media frames of disaster are unavoidable, often they are routinely adopted, and communicators should be "cognizant of prevailing frames in a

disaster so that rumors, misinformation and missing information may quickly be remedied” (Horsley, 2016, p.163).

Scholars have argued that one of the most effective strategies to mitigate this tension is to integrate the media throughout all phases of risk and disaster management, from preparedness to recovery and evaluation. Such integration not only enhances media relations but also fosters a mutual understanding of the distinct, yet interdependent roles journalists and communicators play during crises (Reynolds & Seeger, 2005; Seeger, Sellnow & Ulmer, 2003). Moreover, strengthened collaboration with the media may improve the visibility of preparedness, mitigation, and resilience, issues that are frequently underrepresented in disaster journalism (Houston, Pfefferbaum, & Rosenholtz, 2012) but represent strategic themes for risk communication in public sector organizations (Lombardi & Burato, 2017). Furthermore, strategies aimed at improving the relationship and mutual trust between journalists and public communicators should focus on the development of interpersonal relationships, a deeper mutual understanding of respective professional roles, and the adoption of transparent, consistent and ethically grounded communication practices both in ordinary and in critical situations (Lovari & Bowen, 2020; Veil, 2012).

Focusing on the Italian context, a growing body of scholarship has investigated the evolution of risk and emergency communication during natural disasters, examining the roles and interactions of multiple stakeholders (e.g., Baldino et al., 2024; Cerase, 2017; Farinosi & Micalizzi, 2013; Sturloni, 2018). One of the first studies, coordinated by Comunello (2014), revealed the different practices of citizens’ digital volunteerism (Smith et al., 2021), enabled by the increasing prominence of social media, in response to the interventions of local institutions during the 2012 Emilia earthquake and the 2013 flood in Sardinia. Within the scope of social media’s impact, Comunello and Mulargia (2017), focusing on regional and local level administrations, have identified several barriers for an effective use of these platforms during natural disasters: the informal nature of social media collides with institutional bureaucracy, influencing the speed and quality of digital public communication in these critical situations.

Although digital media are growing in importance in citizens’ consumption patterns, traditional media retain a crucial role as trusted intermediaries in the public’s understanding of risks and emergencies. A recent Eurobarometer (2024) reported that more than half of Italian respondents (52%) would rely on national media outlets to get information about disaster-related risks. About 42% prefer to seek information on local or regional media, or via official emergency management bodies such as the *Civil Protection Department, State Police, or Fire and Rescue Services* (31%). Public and governmental institutions follow, with 25% of respondents turning to public communication channels, while 22% reported using social media platforms as a source of information.

Focusing on the role of traditional media, Buoncompagni (2024) analyzed the role of journalists on the heavy floods that impacted Tuscany in 2023. The study highlighted the function of proximity of journalism, and in particular the important role played by local press to promote community resilience during the disaster. Conducting several focus

groups with journalists who covered the event, he pointed out how local media decided to give space to public sector communication - “from the press releases published online by the civil protection services to the contributions of the mayors, from the infographic to the messages published on the websites” (Buoncompagni, 2024, p.5) - to control the noise of information and disinformation. Furthermore, Monaco (2025), investigating the media coverage of newspaper articles published between 2020 and 2024 related to a severe coastal storm surge striking Naples, highlighted limitations of media-driven risk communication, the need for more inclusive narratives, and a general focus on exceptionalism that fails to incorporate forward-looking discussions on mitigation strategies and climate adaptation.

In this context, limited attention - at least within the Italian scholarly landscape - has been devoted to examining the relationship between journalists and public communicators during risk and natural disasters from both points of view, investigating the characteristics of their interactions, practices and mutual expectations. This article aims to partially address this gap through a qualitative study involving journalists and public sector professionals working in insular contexts, an area that remains largely underexplored within international research on risk communication (Diers-Lawson & Meißner, 2021).

Research questions and methodology

This article presents findings from a qualitative study conducted within the national research project PRIN 2022 PNRR INSULANDER that investigates risk communication management in Sardinia and Sicily, two Italian regions and insular territories. Ivčević et al. (2021: 1) noted for the southern Mediterranean regions that different societies “sharing a similar environment may face similar natural hazards but might not necessarily have the same level of risk awareness. They might not share a common understanding of that risk”, may have different levels of anxiety, differ in urbanization levels and consequently may need different communication strategies. Thus, insularity is considered an additional factor of fragility and vulnerability that significantly affects the construction of risk perception and impact on communication strategies and practices (Gelabert, Micallef & Rossello Geli, 2020). Focusing on Italian islands, ISTAT (2024) reports that the maximum number of consecutive dry days (≤ 1 mm of rainfall) reached 44 in Sicily and 52 in Sardinia, compared with a national average of 27. Such climatic conditions exacerbate wildfire risk, as shown by the higher rate of burned forest area per 1,000 km² in the islands in comparison with the national average. The two Italian regions are also severely exposed to the impacts of climate change and extreme weather events, including floods, droughts, and soil aridity.

In this context, this study aims to identify the main elements characterizing the relationship between public communicators and journalists reconstructing, from both

perspectives, the informative routines and the organizational and operational challenges adopted in communicating risks and natural disasters in Sardinia and Sicily.

In particular two research questions guided this study:

RQ1: How is the relationship between journalists and public communication professionals in situations of risk and natural disasters? What are the main dimensions that characterize this relationship?

RQ2: How is the spectacularization of news about risks and natural disasters perceived by journalists and public communicators?

In order to investigate these topic in-depth, semi-structured interviews were carried out with public communicators and journalists operating in these insular contexts. Two semi-structured interview guides (Knott et al., 2022) were created integrating risk and disaster communication, journalism and public sector communication theories. They reported specific sessions related to the characteristics of the two professions and included common issues to compare the different perspectives on specific themes (i.e. risk communication, insularity, newsmaking, mutual relationship, artificial intelligence, etc.). Between April and June 2025, a total of 27 semi-structured interviews were conducted online using Microsoft Teams and Google Meet: 17 in Sardinia (9 communicators, 8 journalists) and 10 in Sicily (6 communicators, 4 journalists).

The sampling procedures differed in relation to the professional profiles, adopting a snowball sampling (Bryman, 2016). For journalists, the selection was based on a mapping of regional media outlets and on a media coverage analysis previously conducted for the INSULANDER PRIN PNRR project activities (Milestone 2, see Baldino et al., 2024). This activity provided a list of journalists working for national, regional and local media who covered natural disasters in these two regions. For public communicators, the selection was carried out through official institutional channels, prioritizing professionals with specific responsibilities in risk communication management at regional and local levels (including Region of Sardinia, Civil Protection, several municipalities).

Interviews were transcribed using Microsoft Teams, but manual text cleaning was performed to ensure full fidelity to the original narratives. The corpus of interviews was subjected to a rigorous qualitative coding process (Braun & Clarke, 2006) by the research team (the authors and two other researchers), aimed at identifying, classifying, and interpreting the main thematic categories emerging from the empirical material. In the first phase, a differentiated codebook was developed for each of the two professional groups under analysis. This methodological choice was motivated by the need to capture the lexical, experiential, and conceptual specificities distinctive of each profile, avoiding an interpretive homogenization between the two groups. At the same time several parts of the coding procedure (with related codes and subcodes) were the same in order to compare specific dimensions: this is the case of the public communicators-journalist's relations that is at the center of this article.

The textual material underwent a process of data reduction and coding assisted by the software NVivo 15, which allowed for a systematic and transparent management of the corpus, ensuring traceability of categories, recurrence of themes, and interconnections

among different codes. This process facilitated the comparison between the two professional groups, enabling a deeper understanding of the points of convergence and divergence emerging from their respective narratives. Interviews' excerpts reported in the next paragraph have been translated into English by the authors. Each quotation is accompanied by the interview number (from 1 to 9, the maximum number for each category), with the abbreviations J for journalists, C for communicators, Sar for Sardinia and Sic for Sicily².

Results and discussion

A multifaceted relationship

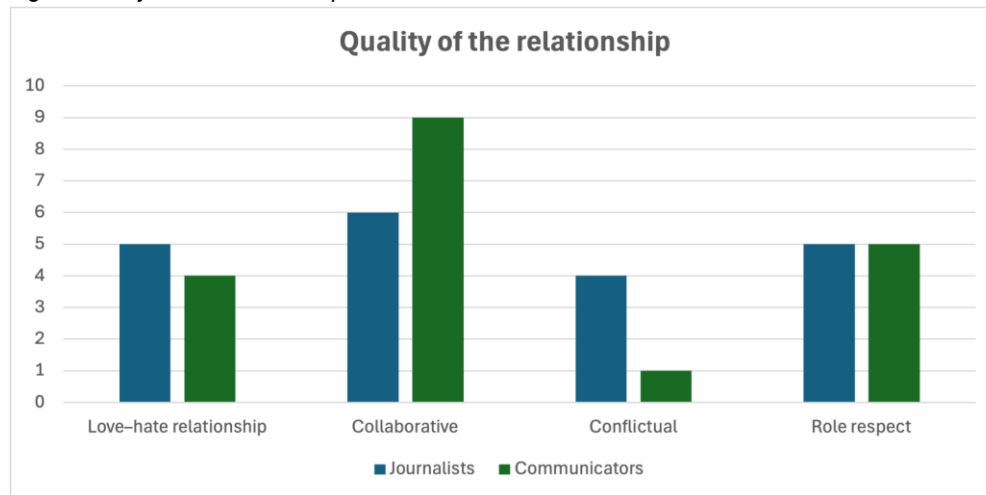
The analysis of the quality of the relationship between journalists and public communicators (RQ1) highlights an overall cooperative yet ambivalent situation (Fig.1). The prevailing perception is a functional collaborative relationship between the two communicative actors, often based on common interest and reciprocal trust (Veil, 2012), although elements of tension and mutual distrust persist and emerge as reported in these interview's excerpts:

We have specific contacts within the world of journalism people with whom, over time, we've built a relationship of trust and collaboration. Very often, when something happens, for example a wildfire, it's the journalist who reaches out to us directly: they might call and ask if any of our volunteers are on site, if they can interview them, or get some additional information. It's a direct, almost personal relationship that has developed and strengthened over the years. (11CSic)

Sometimes I tend to bypass them, but never out of disrespect. Of course, that can cause some tensions. They might say, 'why did you go directly to him instead of through me?' Yes, because sometimes, you know, asking a question directly has a different weight. And one question leads to another, depending on the answer I get, and the story may start one way but end up completely different. (05JSar)

The theme of collaboration emerges as the most recurrent (n.15), explicitly mentioned by six journalists and nine communicators. This indicates a shared representation of the relationship as oriented toward cooperation and dialogue. Both groups acknowledge the need for an interdependent relationship (Bulicanu, 2022) to ensure the timely and accurate circulation of information in situations of risk or natural emergency in the regional territories (Buoncompagni, 2024; Horsley, 2016). However, collaboration is often described as instrumental rather than based on consolidated trust, more an operational necessity dictated by urgency - the incoming risk or the emergency related to a disaster - than a stable, long-term relationship.

Fig.1: Quality of the relationship



Alongside this, the presence of the code “love-hate” relationship, explicitly mentioned by 5 journalists and 4 communicators, highlights the relational ambivalence between media and institutions (Faccioli, 2000; Graber, 2003). Both professionals recognize a dual tension between cooperation and control, transparency and sensationalism, and information openness and strategic communication management. Journalists describe their relationship with press offices as one of selective trust, necessary to quickly access verified information, yet at the same time restrictive in terms of investigative freedom.

The conflictual theme, mentioned by four journalists and only one communicator, points to an asymmetry in perception. Journalists tend to identify sources of friction in the relationship, particularly when public communication appears overly controlled, downplayed or mediated by politicians’ control or internal bureaucracies. For instance:

It’s all about timing. News travels fast, it has a life of its own. I keep saying it: it’s pointless for them to tell us ‘we’re still verifying’ when, by that time, someone might already be dead, or the town might be underwater. When the Montiferru wildfire happened, I was the first to raise the alarm. I was on vacation, but from my house up on the hill I could see the fire front. I called right away, and they kept telling me, ‘we’re checking’. But no, I could see the flames, tall, huge flames, and they were still ‘verifying’. Not verifying the fire, verifying what to tell us. That’s why I say it’s definitely a problem. (04JSar)

On the other side, public communicators tend to moderate these tensions, offering a more conciliatory narrative to have an extensive media coverage in case of forthcoming risks, explaining the need to activate different levels of analysis within their institutions. This divergence reflects the distinct professional logics of the two groups: for journalists, an orientation toward newsworthiness and real-time responsiveness to alert local communities (Grassau, Valenzuela & Puente, 2021); for communicators, a focus on institutional legitimacy and the mitigation of reputational risk in case of natural disasters (Coombs, 2020; La Rocca & Lovari, 2024).

Furthermore, the code “role respect”, mentioned by five participants in each group, represents a point of balance in the mutual recognition of professional competences. Both

professionals emphasize the importance of maintaining clear professional boundaries between journalism and public sector communication, while acknowledging that in emergency situations these boundaries can become more flexible and permeable to prevent an incoming risk or alert the local population in case of unexpected floods or wildfires.

It's clear that we will always aim to communicate in a way that reassures people, while journalists will communicate to sell a story. That's a fundamental difference, and it cannot be changed. What we can do, though, is agree on how we use words. I believe we could develop a sort of shared vocabulary; one where we don't have to call something a catastrophe if it isn't, but rather a smaller-scale event. By fine-tuning our language and setting clearer boundaries between our roles, we could manage to understand each other, especially with the Journalists' Order. (08CSar)

From the interviews emerges how the type of relationship between journalists and public communicators is predominantly collective, although direct one-to-one interactions also persist. Seven communicators and four journalists described the relationship as collective, meaning that it is mediated through group-based online channels and traditional tools such as press conferences, mailing lists, events, or press briefing. This suggests that, in the communicators' narratives, the relationship is perceived as institutionalized and mediated by formal communicative practices, consistent with the role of media relation offices as representatives of public administrations. The collective dimension also aligns with the principles of transparency and equal access to public information - key principles of public communication (Canel & Luoma-aho, 2019; OECD, 2021) - even though communication is often more tightly controlled in cases of emergencies to prevent the spread of panic or misleading information (Lovari & Bowen, 2020).

At the same time, both journalists (3) and communicators (4) acknowledge the presence of personal and direct one-to-one relationships. These interactions, often developed over time, emerge as trust-based practices that enable faster and more flexible information exchange, particularly during emergencies, disasters or crises (Miller & Goidel, 2009). The coexistence of these two relationship types - collective and one-to-one - indicates a dynamic balance between formal roles and personal trust (Veil, 2012). On the one hand, collective relationships ensure uniformity in delivering institutional messages, fostering the traceability in public sector communication's strategies and practices, as reported in this interview:

When we developed the communication plan, we explicitly included this: an institutional channel through which information could be disseminated to everyone promptly and simultaneously. The communication started from the website, which remains the main institutional source, and from there we shared it on social media and through the broadcast list for press. (06CSar)

On the other, individual relationships perform as informal mechanisms of efficiency, grounded in relational networks built over time. In this scenario, journalists often contact

communicators directly to confirm details, and communicators rely on familiar journalists to guarantee responsible media coverage on risks or disasters affecting regional territories. These one-to-one contacts coexist with broader institutional mechanisms, providing flexibility during crises. The overall picture reflects a *structural hybridization in the relationship between journalists and public communicators*. Public communicators tend to privilege the collective dimension, consistent with the organizational culture of administrations and with the practical need to avoid repeating the same information to multiple journalists. Journalists, while participating in these dynamics, tend to value personal relationships more strongly, perceiving them as more effective for obtaining in-depth, timely information and ultimately, for getting the news or the story published.

Analyzing the time-based dimension of the relationship between journalists and public communicators, it emerges that is predominantly structured (10 out of 27), but with a significant discontinuous and intermittent component that reflects the type of expertise and work carried out within the editorial offices (whether journalists are specialized in risk/environment or in news topics), as well as the situational and cyclical nature of their interactions related to the occurrence of risks and natural disasters in the two islands. Most interviewees from both groups (5 journalists, 5 communicators) described the relationship as structured and ongoing, indicating *the presence of stable relational practices and communication channels developed over time*, often grounded in mutual trust or institutionalized procedures (Badham & Luoma-aho, 2025). These relationships are built on previous experiences (or previous disasters and natural emergencies) and maintained through dialogue, a constant mutual flow of information, which becomes particularly crucial in moments of crisis or risk management.

There is constant dialogue with those responsible for public risk communication. We regularly speak with the Regional Authority, the Regional Civil Protection Department - including the department's director - the National Institute of Geophysics and Volcanology, the relevant regional councilors, as well as the Police, Carabinieri, and the Financial Police. It's necessary, because we couldn't communicate anything unless it came from them first. (11JSic)

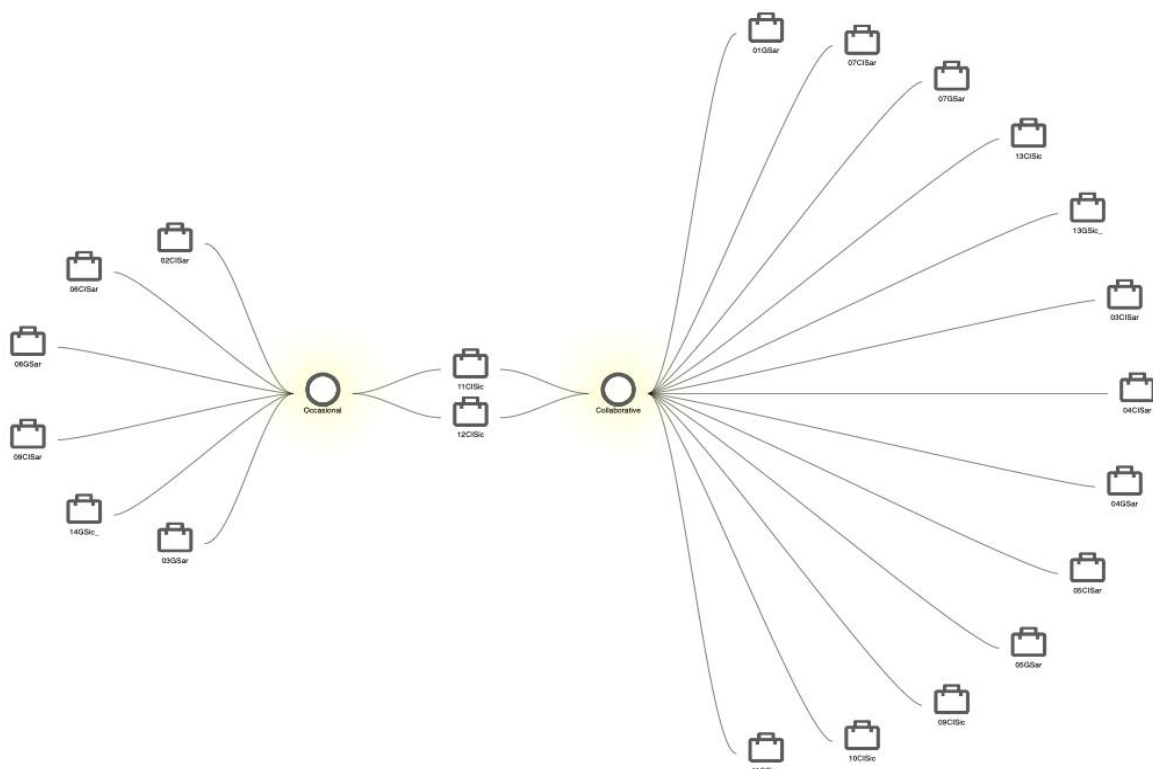
At the same time, three journalists and five communicators reported the relationship as occasional (8/27), activated mainly in specific situations such as public events (i.e., road shows, fairs), press conferences, and tending to diminish once those circumstances end. This suggests that, for part of the interviewees, collaboration is not based on relational continuity but rather on a logic of necessity (Gandy, 1982): communication occurs when needed, rather than as part of a stable or systematic relationship between the two professionals. On the other hand, some communicators specified that their digital channels are constantly updated, so the relationship with journalists is continuous even if mediated, but the manifestation of specific risks can enhance direct and tailored interaction. Finally, the marginal presence of the sub-code "absent" (one journalist and one communicator) indicates that the complete lack of a relationship is quite rare, yet it confirms the existence of cases of communicative isolation, likely associated with poor

disaster journalism in small local contexts (Houston, Pfefferbaum, & Rosenholtz, 2012) or limited professional experience in crisis communication and media relations in the regional institutions. For instance:

No, there isn't an institutional contact person for communication. There are the Civil Protection bulletins, but there isn't a proper communication flow designed for journalism. So, these relationships don't exist as formal channels. Let's say we rely on the bulletins. (10JSic)

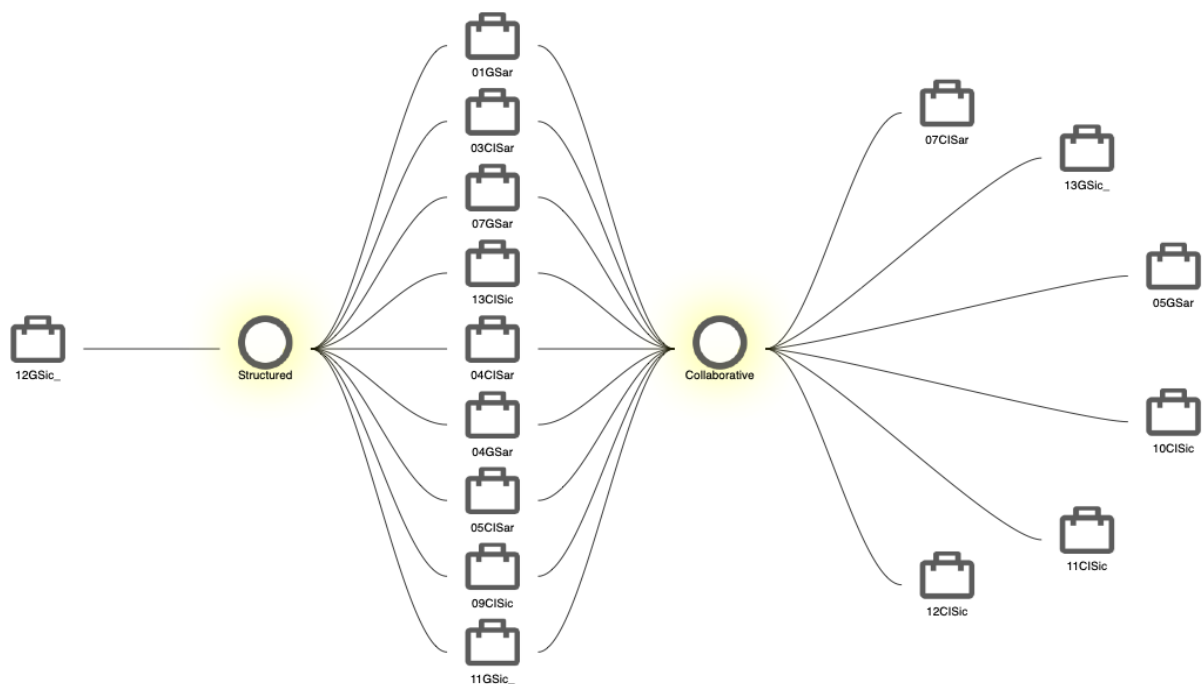
As previously reported, most interviewees describe the relationship between public communicators and journalists as predominantly collaborative, with some exceptions where collaboration is perceived as sporadic, unstructured, or limited to unforeseen risks and emergency situations. Using NVivo 15 to compare subcodes related to temporal dimension (occasional, structured, absent) and the quality of the relationship (collaborative, conflictual, love and hate, without rules, respect of the roles), it can be noticed a partial connection between the two sub-code codes "occasional" and "collaborative" (Fig.2), reflecting a conceptual continuum: collaboration is not binary (present/absent) but exists along a spectrum of stability. From a territorial perspective, the presence of both Sardinian and Sicilian cases in these categories suggests variability is not strongly regional but rather linked to organizational or individual differences.

Fig.2: Relationship between the subcodes "occasional" and "collaborative"



When consolidated, collaboration between journalists and communicators tends to evolve into a structured relationship, supported by routines, frequent contacts, and shared procedures and practices (e.g., mailing lists or joint training activities). This pattern is evident in most cases, as shown in Fig.3, which analyzes the relationship between the codes “structured” and “collaborative”. The temporal dimension and the quality of the relationship appear to be consistent: it is rare to find cases where a collaborative relationship is not also structured or, conversely, occasional. Only in one case was a collaborative tendency mentioned alongside criticism of “non-virtuous mayors” (03CSar) who showed little interest in nurturing media relations during natural disasters.

Fig.3: Relationship between the subcodes “structured” and “collaborative”



Furthermore, focusing the attention on the word “relationships” (in Italian *rapporto*) inside the interviews’ transcripts, a word tree was extracted with NVivo 15 (Fig.4), revealing a multifaceted representation of the relationship between journalists and public communicators in risky situations and natural disasters. The connections branching from the root term show that the concept is almost always contextualized in reference to collaboration, trust, and mutual dependency, nonetheless it also contains traces of conflict, asymmetry, and institutional distance. A large portion of the word tree links “*rapporto*” (relationship) with phrases such as “*di collaborazione*” (for collaboration), “*di fiducia*” (of trust), and “*di collaborazione costante*” (of continuous collaboration), indicating that most professionals view the relationship as functional and necessary for managing journalism and communication in risk and emergency contexts. This reinforces the idea that cooperation between the two actors is not optional or operational but a strategic requirement.

Fig.4: Word Tree “rapporto” (relationship)



Some branches, however, expose the fragility of this collaboration. Mentions such as “rapporto difficile” (difficult relationships), “rapporto che va costruito” (relationship to be built), or “rapporto che dipende dalla fiducia” (relationship depending on trust) reflect a perception of imbalance and the need for continuous negotiation between media and public sector organizations. For example, journalists emphasize their reliance on communicators for access to verified information, while communicators describe the relationship as constrained by organizational hierarchies, the influence of politics, and the need to protect institutional credibility.

Lastly, the word tree highlights an important semantic nuance: the word “rapporto” often co-occurs with verbs like “mantenere” (to keep), “costruire” (to build), and “alimentare” (to nurture), which frame the relationship as an ongoing continuous process rather than a fixed state or singular interaction. This indicates that both sides perceive the

Regarding RQ2, the thematic analysis highlights how news sensationalism (Italian code: spettacolarizzazione) is conceptualized by interviewees not as an intentional practice but as a structural drift within the contemporary media ecosystem. The related word tree (Fig.5) shows a semantic network in which sensationalism intertwines with words such as news, headline, image, story, and urgency, suggesting that the pressure for visibility and timeliness profoundly shapes how news is narrated, especially in contexts of risks and natural disasters.

spettacolarizzazione

- della notizia**
 - «se capita?»: allora questo è molto difficile, purtroppo la mia, la mia esperienza mi dice che sia
 - E secondo lei, quali sono gli esempi di evitare in base anche alla sua esperienza? I: Sì, questi
 - Quindi, può essere capitato, ad esempio, o in un evento, un effetto di questo tipo che abbia riguardato
- delle notizie**
 - e questo può accadere anche quando si trattano eventi legati al dissenso ideologico. Un caso emblematico è stato
 - sono assolutamente presenti. L'informazione, oggi, si scontra con un mare di altre fonti che fanno dello spettacolo
- della spettacolarizzazione**
 - di un evento, di una notizia, come intervenire? O meglio, si può intervenire dal punto di vista istituzionale? I: Ma
 - e viene confuso molto di più del mio colloquio che dire che sono intervenuti 96 operatori. R: così delle differenze
 - è un problema che è sempre più grave. C'è sempre stato, però, con l'avvento del web. L'immediatezza, i debolismi, cioè
 - in questo senso, però mi sento di dire che queste sono in generale le regole del giornalismo, la notiziabilità. Cosa
 - mediana che ricordo con maggiore chiarezza è al periodo della pandemia da Covid - 19. In particolare, durante le prime fasi
 - o comunque una narrazione di quel tipo. Quindi le chiedo: a quel punto, che cosa fare? Cosa si fa per
 - perché alla fine eravamo indotti non da un mese, ma da uno, tre giorni, quattro giorni, non era una cosa ...
- della spettacolarizzazione**
 - avvenire principalmente per titoli o per immagini sparse. Se noi abbiamo seguito un evento all'origine possiamo, abbiamo un'altra
 - forse, serve quando una notizia è debole, quando non ha un impatto emotivo rilevante, ma quando invece si ha
 - sia una realtà, è un dato di fatto e non c'è dubbio che i certi casi l'informazione tende ad accentrare
 - stampa Quindi, come si gestiscono le relazioni con la stampa, che tipo di rapporti vi sono? I: I rapporti inevitabilmente,
 - rende per le questioni sia, la verità è che i cittadini leggono. Vedono quel titolo e allora aprono il link,
 - ... perché però, documentare un evento nel momento in cui sta accadendo prima dell'era digitale, prima dei telefoni era praticamente impossibile,

80 | Alessandro Lovari, Daniela Pisu

the pervasiveness of digital platforms and their algorithms push journalists to simplify complex information and emphasize the emotional and visual dimensions of all the events (Tierney et al., 2006). For instance:

It's clear that the impulse of sensationalism exists, it's real and it's hard to avoid, because it's viscerally connected to the event itself. It makes those who lose their homes, their jobs, or who find themselves living through severe hardship feel angry. In those moments, the newspaper becomes the spokesperson for that anger, for that outrage that's difficult to contain. Sensationalism sometimes stems precisely from this strong, uncontrollable emotion of the people experiencing these tragedies (14JSic)

This process is perceived as a professional constraint, as well as an ethical challenge that can have negative impact on citizens' perceptions and also on public sector communication interventions and credibility (Comunello & Mulargia, 2018; Lovari & Bowen, 2020). As reported in this excerpt:

They wrote a headline that was really dangerous. It said 'Fires between fiction and reality', and I told them to be careful, because Aunt Maria [name of fantasy] reading something like that would get scared, and that's not the message you should be sending. The road ahead is long and difficult, and a lot still needs to be done through training, relationship-building, and direct interaction with the media, because they have the power to convey information however they choose. And this isn't a chess tournament! (03CSar).

Several professionals - both communicators and journalists - explicitly connect this phenomenon to the transformation of journalistic routines: the replacement of traditional press conferences with real-time online updates, live broadcasting via digital platforms, and social media coverage has accelerated the production cycle and amplified the risk of sensationalism and the use of specific narratives (Sorrentino, 2025). These factors intertwine with factors like citizen prosumerism and citizen journalism (Farinosi, 2020), creating a potential competition that influences and accelerates journalistic routines and news production especially in emergency situations.

This issue of sensationalism is becoming increasingly serious. It's always been there, but with the advent of the web, the immediacy, smartphones, and the fact that everyone now has a camera, the whole dynamic has changed. The relationship has been reversed: with digital technology and smartphones, the public has become the witness. So today anyone can see how news events are documented directly by witnesses, who within seconds have already uploaded material that, until recently, would never have been seen. (08JSar)

Furthermore, news sensationalism is not merely a matter of emotions and exaggeration to attract attention (Grabe, Zhuo & Barnett 2010), but reflects a systemic condition of temporal compression, a need to report, interpret, and attract public opinion's attention almost simultaneously, also to increase media visibility and convey online traffic on social media accounts and on their websites. This is clearly reported by some communicators' interviews:

Sensationalism sells, that's the truth since people read it. They see that headline and click on the link, because most of the time they stop at the teaser or the title. But if there's even a chance, they'll open it, and you use something that grabs attention, especially something that appeals to the reader. We know that doing so increases the number of users online. (06CSar)

Yes, journalists...over the years, you learn how to manage them too, because of course they're always looking for sensational stories, for the person who shouts or stirs up controversy. But that's part of their job, and I always try not to overdo it when giving them information (05CSar)

At the same time, the recurrence of words such as “order”, “avoid”, “ethics”, and “rules” highlights a strong normative awareness. Many journalists consciously strive to resist the drift toward dramatization of news in case of risks and natural disasters, recalling professional codes promoting accuracy, moderation, and respect for the public's emotional vulnerability (Buoncompagni, 2024). As some interviewees noted, “no one is immune,” yet “it should be avoided,” emphasizing that news sensationalism represents a form of ethical deviation from professional standards that need to be responsibly respected. Communicators are fully aware of this process: during disasters there is the need of a constant dialogue with journalists to avoid this practice that could create panic in the population and can harm the efficacy of institutions' interventions and their reputation (Lovari, D'Ambrosi & Bowen, 2020).

The thematic analysis reveals a double tension: a) between speed and responsibility, and b) between engagement and restraint. This ambivalence mirrors the dual pressures experienced by journalists, to remain competitive within a fast-paced digital ecosystem while safeguarding credibility and trustworthiness.

Two main interpretive frames of news sensationalism can be distinguished:

- 1) Professional deviation: journalists view it as a professional misstep that undermines credibility, often resulting from emotional involvement, editorial pressure, or the urge to meet audience expectations. It is framed as an ethical issue that needs to be recognized and mediated in the relationship with communicators.
- 2) Systemic effect: rather than an individual choice, sensationalism emerges from structural transformations in communication and journalism production: the acceleration of news cycles, algorithmic visibility and click baiting, as well as the hybridization between public sector communication and social media logics.

Ultimately, these narratives reveal both resistance and resignation. Both professionals demonstrate critical awareness of the ethical and epistemic risks of dramatization of news, while they also acknowledge the difficulty of escaping a system logic that rewards speed, visibility, and emotions, also in public communication (Lovari, 2025). The theme highlights the need for communicators to perform environmental scanning for identification of media frames (Horsley, 2016), as well as the fragilities of journalism in the digital age (Solito & Sorrentino, 2023), suspended between the duty to inform and the spectacularization adopted even during natural disasters.

Conclusion: allies, antagonists or imperfect allies?

This article aimed to investigate the relationship between journalists and public communicators in situations of risk and natural disasters, as well as the perception of the spectacularization of these events in the insular contexts of Sardinia and Sicily. This relation is investigated in the different phases of risk communication, from prevention to emergency management related to natural disasters.

The relationship between journalists and public communicators in these two regions is marked by structural ambivalences (Bulicanu, 2022), exhibiting a hybrid and fluid relational model. This model integrates formalized procedures and institutional channels with direct and informal interactions, crucial for expediting information exchange, especially during emergencies. This hybridization appears to be a structural characteristic of risk communication in insular territories, where geographical proximity amplifies both the necessity for cooperation and the potential for local tensions (La Rocca & Lovari, 2024). Respect of roles and collaboration are recognized as central assets in risk and disaster communication, but they are influenced by bureaucratic procedures that collide with journalist news making routines, framing processes, and citizens' information practices (Comunello, 2014; Fischer, 2008; Smith et al., 2021).

Regarding the perception of news sensationalism, findings indicate a complex and multifaceted phenomenon, exacerbated by the digital media ecosystem's transformations, that today is increasingly problematized by AI, raising concerns of accountability, transparency, and disinformation that can be critical in these situations (Sellnow-Richmond & Lukacovic, 2025; World Economic Forum, 2024). Concurrently, there is an understanding that spectacularization is neither inevitable nor fully accepted; both professional groups appeal to high ethical standards and deontological norms (Bowen, 2024) to counteract sensationalistic tendencies, acknowledging its strong allure, but also its detrimental function. In this context, sensationalism can be understood not merely as an inevitable outcome of journalistic practices driven by platform-oriented business logics, but also as a terrain of mediation in which the relationship between the two professional communities can be constructed, negotiated, and redefined. This issue may offer opportunities for reflexivity - for instance through joint training initiatives as reported in several interviews - that can help establish a more balanced approach toward the public interest, particularly when natural hazards and extreme climate events manifest and should be responsively and timely communicated.

This study presents several limitations, partly influenced by the specific objectives of the INSULANDER research project. First, the decision to focus on selected dimensions of the relationship between journalists and public communicators did not allow for an in-depth exploration of other relevant aspects of the interviews, including the use of different communication tools and digital platforms, the problematic interactions with political actors and public administrators, and the implications of artificial intelligence for innovating risk communication practices. In addition, the unequal number of interviews

conducted in Sicily and Sardinia - largely due to difficulties in identifying and recruiting available contacts - may have affected the comparability of the two insular contexts. Future research will seek to further investigate the interconnections among these factors, with the aim of developing a more comprehensive understanding of the dynamics shaping risk communication, institutional trust, and media–public sector relations in insular settings. Together this study offers relevant insights for its implications for risk and disaster communication, as it investigates points of convergences and tensions between the two professional groups from both sides using in-depth interviews. Greater synergy between journalists and public communicators - supported by respect of roles, shared protocols, continuous training, and conscious use of digital platforms - can reduce the risk of misinformation and strengthen the informational resilience of insular communities.

At the same time, from a theoretical and methodological perspective, the study underscores the importance of contextual and situated analyses that consider risk communication in territorial specificities, such as insularity (Diers-Lawson & Meißner, 2021; Ivčević et al., 2021). Risk communication emerges as a relational, negotiated, and multilayered process whose success depends on the quality of dialogue among communicators and journalists and on the degree of balance achieved among informational, organizational, and ethical dimensions, also built in previous interactions between these professional roles. More than antagonists or allies, journalists and communicators appear to be imperfect allies: called upon to co-construct reliable and timely public information for preventing risk and managing disasters, especially in highly vulnerable insular contexts. Strengthening this alliance, in the mutual respect of their role professional expertise and ethical standards, represents an essential condition for developing responsible and effective risk communication strategies to support the ability of local communities to understand and manage their own vulnerabilities.

Biographical notes

Alessandro Lovari is Associate Professor of Sociology of Communication at the Department of Political and Social Sciences, University of Cagliari (Italy) where he teaches Public Communication, and Institutional Communication, Digital Technologies and Media Relations (graduate program). He is the coordinator of the Phd program in Research and Social Innovation and PI of several national research project on risk and emergency communication in environmental disasters (INSULANDER) as well as on AI impact on communication and citizens (IATI). Lovari's research focuses on public sector communication, public relations, and health communication, investigating digital technologies impact on organizations' practices and citizens' behaviors in ordinary and crisis situations. He was visiting scholar at Purdue University, University of Cincinnati, University of South Carolina, Virginia Commonwealth University (USA), and Jyväskylä University (Finland). He is author of more than hundred publications in monographies, book chapters and journal articles.

Daniela Pisu, PhD in Social Sciences, is a research fellow in the Sociology of Cultural and Communication Processes at the Department of Political and Social Sciences, University of Cagliari (Italy). A specialist social worker and family mediator, she has worked for over ten years in public and private services in the field of planning and coordination of personal social services. Her main research areas include Sociology, Social Policy, Social Work, and Risk Communication in emergency conditions.

References

- Altheide, D. L. (1997). The news media, the problem frame, and the production of fear. *The Sociological Quarterly*, 38(4), 647–668. <https://doi.org/10.1111/j.1533-8525.1997.tb00758.x>
- Albrecht, M. M. (2022). *Trumping the media: Politics and democracy in the post-truth era*. Bloomsbury Publishing.
- Baldino, N., Atzeni, C., & Lovari, A. (2024). La rappresentazione mediatica degli incendi in Sicilia e Sardegna. In G. La Rocca & A. Lovari (Eds.), *Comunicazione del rischio insulare* (pp. 40–62). FrancoAngeli.
- Badham, M., & Luoma-aho, V. (2025). *Digital communication management: Theories and practices for a global and volatile world*. Taylor & Francis.
- Beck, U. (1992). From industrial society to the risk society: Questions of survival, social structure and ecological enlightenment. *Theory, Culture & Society*, 9(1), 97–123. <https://doi.org/10.1177/026327692009001006>
- Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77–101. <https://doi.org/10.1191/1478088706qp063oa>
- Bryman, A. (2016). *Social research methods* (5th ed.). Oxford University Press.
- Boykoff, M. T., & Boykoff, J. M. (2004). Balance as bias: Global warming and the US prestige press. *Global Environmental Change*, 14(2), 125–136. <https://doi.org/10.1016/j.gloenvcha.2003.10.001>
- Bowen, S. A. (2024). “If it can be done, it will be done”: AI ethical standards and a dual role of public relations. *Public Relations Review*, 50(1), 1–13. <https://doi.org/10.1016/j.pubrev.2024.102513>
- Bulicanu, V. (2022). The relational framework of journalists and communicators. *International Journal of Communication Research*, 12(4), 269–271.
- Buoncompagni, G. (2024). Media and natural disasters: Organising storytelling in the age of climate change. *Journalism and Media*, 5(2), 614–625. <https://doi.org/10.3390/journalmedia5020041>
- Canel, M., & Luoma-aho, V. (2019). *Public sector communication: Closing gaps between public sector organizations and citizens*. Wiley-Blackwell.

- Carvalho, A., & Burgess, J. (2005). Cultural circuits of climate change in UK broadsheet newspapers, 1985–2003. *Risk Analysis*, 25(6), 1457–1469. <https://doi.org/10.1111/j.1539-6924.2005.00692.x>
- Cerese, A. (2017). *Rischio e comunicazione*. Egea.
- Coombs, W. T. (2013). *Applied crisis communication and crisis management: Cases and exercises*. Sage.
- Coombs, W. T. (2020). Public sector crises: Realizations from Covid-19 for crisis communication. *Partecipazione e Conflitto*, 13(2), 990–1001. <https://doi.org/10.1285/i20356609v13i2p990>
- Coman, I. A., Gregor, M., & Lilleker, D. (2025). *Risk communication and COVID-19 governmental communication and management of pandemic*. Routledge.
- Comunello, F. (Ed.). (2014). *Social media e comunicazione d'emergenza*. Guerini & Associati.
- Comunello, F., & Mulargia, S. (2017). Tra risposte protocollate e «social sensing». L'uso dei social media per la comunicazione d'emergenza nelle istituzioni locali italiane. *Sociologia e Ricerca Sociale*, 112, 111–137. <https://doi.org/10.3280/SR2017-112006>
- Comunello, F., & Mulargia, S. (2018). *ShakeNetworks: Social media in earthquake-related communication*. Emerald Publishing.
- Corsale, A., Perelli, C., & Sistu, G. (2020). Large island, big issues. In M. Grimalt Gelabert, A. Micallef, & J. Rossello Geli (Eds.), *The Anthropocene and islands* (pp. 59–77). Il Sileno Edizioni.
- Covello, V. T., Slovic, P., & Von Winterfeldt, D. (1986). Risk communication: A review of the literature. *Risk Abstract*, 3, 171–182.
- Currah, A. (2009). *What's happening to our news*. Reuters Institute for the Study of Journalism, University of Oxford.
- Diers-Lawson, A., & Meißner, F. (2021). Editor's essay: Moving beyond Western corporate perspectives: On the need to increase the diversity of risk and crisis communication research. *Journal of International Crisis and Risk Communication Research*, 4(1), 165–176. <https://doi.org/10.30658/jicrcr.4.1.6>
- Ducci, G., & Lovari, A. (2021). The challenges of public sector communication in the face of the pandemic crisis: Professional roles, competencies and platformization. *Sociologia della Comunicazione*, 61, 9–19. <https://doi.org/10.3280/SC2021-061002>
- Entman, R. M. (1993). Framing: Towards clarification of a fractured paradigm. *Journal of Communication*, 43(4), 51–58. <https://doi.org/10.1111/j.1460-2466.1993.tb01304.x>
- Eurobarometer (2024). *Disaster risk awareness and preparedness of the EU population*. <https://europa.eu/eurobarometer/surveys/detail/3228>
- Faccioli, F. (2000). *Comunicazione pubblica e cultura del servizio*. Carocci.

- Farinosi, M., & Micalizzi, A. (Eds.). (2013). *NetQuake: Media digitali e disastri naturali. Dieci ricerche empiriche sul ruolo della Rete nel terremoto dell'Aquila*. FrancoAngeli.
- Farinosi, M. (2020). Informazione locale e citizen journalism nei contesti di emergenza. La (contro)narrazione del terremoto dell'Aquila. *Problemi dell'Informazione*, 3, 451–478. <https://doi.org/10.1445/99261>
- Fischer, H. W. (2008). *Response to disaster: Fact versus fiction and its perpetuation: The sociology of disaster*. University Press of America.
- Gelabert, M., Micallef, A., & Rossello Geli, J. (Eds.). (2020). *The Anthropocene and islands: Vulnerability adaptation and resilience to natural hazards and climate change*. Il Sileno Edizioni.
- Gans, H. J. (1979). The messages behind the news. *Columbia Journalism Review*, 17(5), 40–45.
- Garner, G. W. (1984). *The police meet the press*. C. C. Thomas.
- Giddens, A. (1990). *The consequences of modernity*. Polity Press.
- Giddens, A. (1999). Risk and responsibility. *Modern Law Review*, 62, 1–10. <https://doi.org/10.1111/1468-2230.00188>
- Gandy, O. H. (1982). *Beyond agenda setting: Information subsidies and public policy*. Alex Publishing Company.
- Goffman, E. (1974). *Frame analysis*. Harper.
- Grabe, M. E., Zhuo, S., & Barnett, B. (2010). Explicating sensationalism in television news: Content and the bells and whistles of form. *Journal of Broadcasting & Electronic Media*, 45(4), 635–655. https://doi.org/10.1207/s15506878jobem4504_6
- Graber, D. A. (2003). *The power of communication: Managing information in public organizations*. CQ Press.
- Grassau, D., Valenzuela, S., & Puente, S. (2021). What “emergency sources” expect from journalists: Applying the hierarchy of influences model to disaster news coverage. *International Journal of Communication*, 15, 1349–1371.
- Hallin, D. C., & Mancini, P. (Eds.). (2012). *Comparing media systems beyond the Western world*. Cambridge University Press.
- Heath, R. L., & O'Hair, H. D. (2009). The significance of crisis and risk communication. In R. L. Heath & H. D. O'Hair (Eds.), *Handbook of risk and crisis communication* (pp. 5–31). Routledge.
- Harcup, T. (2009). *Journalism: Principles and practices* (2nd ed.). Sage.
- Hepp, A. (2019). *Deep mediatization*. Routledge.
- Houston, J. B., Pfefferbaum, B., & Rosenholtz, C. E. (2012). Disaster news: Framing and frame changing in coverage of major US natural disasters, 2000–2010. *Journalism & Mass Communication Quarterly*, 89(4), 606–623. <https://doi.org/10.1177/1077699012456022>

- Horsley, J. S. (2016). Media framing of disasters: Implications for disaster response communicators. In A. Schwarz, M. W. Seeger, & C. Auer (Eds.), *The handbook of international crisis communication research* (pp. 155–164). John Wiley & Sons.
- Kasperson, R. E., Renn, O., Slovic, P., Brown, H. S., Emel, J., Goble, R., Kasperson, J. X., & Ratick, S. (1988). The social amplification of risk: A conceptual framework. *Risk Analysis*, 8, 177–187. <https://doi.org/10.1111/j.1539-6924.1988.tb01168.x>
- Knott, E., Rao, A. H., Summers, K., & Teeger, C. (2022). Interviews in the social sciences. *Nature Reviews Methods Primers*, 2, 73. <https://doi.org/10.1038/s43586-022-00150-6>
- Kuppuswamy, S. (2017). A study on the print media coverage of disasters. *Global Media Journal – Indian Edition*, 8(2), 1–22.
- Ivčević, A., Mazurek, H., Siame, L., Bertoldo, R., Statzu, V., Agharroud, K., & Bellier, O. (2021). The importance of raising risk awareness: Lessons learned about risk awareness sessions from the Mediterranean region (North Morocco and West Sardinia, Italy). *Natural Hazards and Earth System Sciences Discussion*, 1–25. <https://doi.org/10.5194/nhess-21-3749-2021>
- La Rocca, G., & Lovari, A. (Eds.). (2024). *Comunicazione del rischio insulare: Prospettive in comunicazione, politiche pubbliche e analisi dei contesti*. FrancoAngeli.
- Lombardi, M., & Burato, A. (2017). Crisis Management della comunicazione: istituzioni e social media. *Sistemi intelligenti*, 29(2), 371–392.
- Lovari, A., & Ducci, G. (2022). *Comunicazione pubblica, istituzioni, pratiche, piattaforme*. Mondadori Università.
- Lovari, A., & Bowen, S. A. (2020). Social media in disaster communication: A case study of strategies, barriers, and ethical implications. *Journal of Public Affairs*, 1–9. <https://doi.org/10.1002/pa.1967>
- Lovari, A., D'Ambrosi, L., & Bowen, S. A. (2020). Re-connecting voices: The (new) strategic role of public sector communication after the Covid-19 crisis. *Partecipazione e Conflitto*, 13(2), 970–989. <https://doi.org/10.1285/i20356609v13i2p970>
- Lovari, A. (2025). The future of public sector communication: Algorithmic challenges, generative platforms, and new responsibilities. In S. Bowen (Ed.), *Handbook of innovations in strategic communication: AI, futurism and directions* (pp. 215–224). Edward Elgar Publishing.
- Lovari, A., Atzeni, C., Baldino, N., & La Rocca, G. (2025). Media narratives and resilience: Analysing disaster risk communication in Sardinia and Sicily. *Proceedings of the 2025 International Crisis and Risk Communication Conference*, 13(1), 20–23. <https://doi.org/10.69931/001c.142825>
- Luhmann, N. (1996). On the scientific context of the concept of communication. *Social Science Information*, 35(2), 257–267. <https://doi.org/10.1177/053901896035002005>

- Mazzoleni, G. (2017). Changes in contemporary communication ecosystems ask for a “new look” at the concept of mediatization. *Javnost—The Public*, 24(2), 136–145. <https://doi.org/10.1080/13183222.2017.1290743>
- Massa, A., & Comunello, F. (2024). La comunicazione del rischio “Made in Italy”. Riflessioni a partire da un literature review. *Rivista Trimestrale di Scienza dell’Amministrazione*, 3, 1–30. <https://doi.org/10.32049/RTSA.2024.3.09>
- McCombs, M. E., & Shaw, D. L. (1972). The agenda-setting function of mass media. *The Public Opinion Quarterly*, 36(2), 176–187. <https://doi.org/10.1086/267990>
- Miller, A., & Goidel, R. (2009). News organizations and information gathering during a natural disaster: Lessons from Hurricane Katrina. *Journal of Contingencies and Crisis Management*, 17(4), 266–273. <https://doi.org/10.1111/j.1468-5973.2009.00586.x>
- Monaco, S. (2025). Risk communication in coastal cities: The case of Naples, Italy. *Land*, 14(6), 1288, 1–17. <https://doi.org/10.3390/land14061288>
- Nogami, T. (2018). Disaster myths among disaster response professionals and the source of such misconceptions. *Journal of Contingencies and Crisis Management*, 26(4), 491–498.
- Norris, P. (2014). Watchdog journalism. In M. Bovens, R. E. Goodin, & T. Schillemans (Eds.), *The Oxford handbook of public accountability* (pp. 525–542). Oxford University Press.
- Öhman, S., Giritli Nygren, K., & Olofsson, A. (2016). The (un)intended consequences of crisis communication in news media: A critical analysis. *Critical Discourse Studies*, 13(5), 515–530. <https://doi.org/10.1080/17405904.2016.1174138>
- Olofsson, A., Öhman, S., & Nygren, K. G. (2016). An intersectional risk approach for environmental sociology. *Environmental Sociology*, 2(4), 346–354. <https://doi.org/10.1080/23251042.2016.1246086>
- OECD. (2021). *Report on public communication: The global context and the way forward*. <https://doi.org/10.1787/22f8031c-en>
- Reynolds, B., & Seeger, M. W. (2005). Crisis and emergency risk communication as an integrative model. *Journal of Health Communication*, 10(1), 43–55. <https://doi.org/10.1080/10810730590904571>
- Reyes, R. S., Nguyen, V. M., Schott, S., Berseth, V., Hutchen, J., Taylor, J., & Klenk, N. (2021). A research agenda for affective dimensions in climate change risk perception and risk communication. *Frontiers in Climate*, 3. <https://doi.org/10.3389/fclim.2021.751310>
- Sellnow, T. L., & Sellnow, D. D. (2024). *Before crisis: The practice of effective risk communication*. Cognella Academic Publishing.
- Sellnow-Richmond, D., & Lukacovic, M. (2025). *Communication amid mega-crises*. Routledge.
- Seeger, W. M., Sellnow, T. L., & Ulmer, R. R. (2003). *Communication and organizational crisis*. Praeger.

- Shoemaker, P. J., & Reese, S. D. (2014). *Mediating the message in the 21st century: A media sociology perspective*. Routledge.
- Smith, W. R., Robertson, B. W., & Stephens, K. K. (2021). A different type of disaster response digital volunteer: Looking behind the scenes to reveal coordinating actions of disaster knowledge workers. *Journal of Contingencies and Crisis Management*, 29(2), 116–130. <https://doi.org/10.1111/1468-5973.12352>
- Solito, L., Pezzoli, S., & Materassi, L. (2019). Prove di «maturità». Social media e bisogni informativi della cittadinanza. *Problemi dell'Informazione*, 44(2), 339–364. <https://doi.org/10.1445/94254>
- Solito, L., & Sorrentino, C. (2023). An interweaving of influences: How the digital environment redefines journalism. In M. C. Negreira-Rey, J. Vázquez-Herrero, J. Sixto-García, & X. López-García (Eds.), *Blurring boundaries of journalism in digital media* (pp. 39–52). Springer.
- Sorrentino, C. (2025). *Il giornalismo ha un futuro*. Il Mulino.
- Stock, P. V. (2007). Katrina and anarchy: A content analysis of a new disaster myth. *Sociological Spectrum*, 27(6), 705–726. <https://doi.org/10.1080/02732170701534218>
- Sturloni, G. (2018). *La comunicazione del rischio per la salute e per l'ambiente*. Mondadori Università.
- Supa, D. W., & Zoch, L. M. (2009). Maximizing media relations through a better understanding of the public relations-journalist relationship: A quantitative analysis of changes over the past 23 years. *Public Relations Journal*, 3(4), 1-28. <https://doi.org/10.4324/9781003131700-17>
- Tierney, K., Bevc, C., & Kuligowski, E. (2006). Metaphors matter: Disaster myths, media frames, and their consequences in Hurricane Katrina. *The Annals of the American Academy of Political and Social Science*, 604(1), 57–81. <https://doi.org/10.1177/0002716205285589>
- Veil, S. R. (2012). Clearing the air: Journalists and emergency managers discuss disaster response. *Journal of Applied Communication Research*, 40(3), 289-306. <https://doi.org/10.1080/00909882.2012.679672>
- Wang, L., & Cohen, A. (2009). Factors affecting viewers' perceptions of sensationalism in television news: A survey study in Taiwan. *Issues & Studies*, 45, 125–157.
- World Economic Forum. (2024). *Global risks report*. <https://www.weforum.org/publications/global-risks-report-2024/>
- Zaheer, L. (2016). Natural catastrophes and role of Pakistani mass media. *Journal of Political Studies*, 23, 207–232.

Notes

¹ This article reports selected empirical findings from the research project PRIN 2022 PNRR INSULANDER (INvestigating the Strategic role of commUnication for resiLient islAnds copiNg DisastErs Risk management), funded by the European Union – Next Generation EU, PNRR, Mission 4, Component 2, Project Code: P2022A4YAL.

² The authors wish to acknowledge Norma Baldino, Cinzia Atzeni, Alessio Genovese, as members of the INSULANDER research team, for their contribution to the interviews' guides, the administration of the interviews and the transcription of the empirical material.

Smart Alerts for Complex Risks: A Dual-Risk Framework for Understanding AI-Driven Crisis Communication in the Computational Age*

Kulsawasd Jitkajornwanich**
Texas Tech University

Kerk F. Kee***
Texas Tech University

This paper develops a framework for “smart alerts” that explains how artificial intelligence (AI) and computational social media analytics are reshaping crisis communication, amplifying detection speed and message personalization while introducing new categories of technological risk. We theorize a dual-risk structure: (1) primary hazards (e.g., floods, wildfires) that alerts aim to mitigate, and (2) secondary risks embedded in AI-mediated communication systems (false positives, bias, privacy, deepfakes). Using a speculative design approach and an illustrative technical case study of Twitter-based flood detection in Thailand, we show how human–AI collaboration models (AI-assisted, human-supervised, and parallel processing) can be operationalized from data ingestion and geocoding to visualization and verification. We propose three cross-cutting design and governance mechanisms: graduated confidence communication, multi-source verification, and adaptive governance architectures. They jointly balance the speed–accuracy dilemma while safeguarding equity and democratic accountability. The framework advances crisis and strategic communication by (a) reframing time in predictive messaging (from reactive to anticipatory communication), (b) specifying organizational design patterns for decision rights and oversight in AI-enabled warning systems, and (c) articulating implementable practices that can sustain public trust. We conclude with implications for empirical evaluation and policy design.

Keywords: Smart alerts, AI-enabled crisis communication, human–AI collaboration, predictive risk communication, multi-source verification, adaptive governance architectures

* Article submitted on 22/09/2025. Article accepted on 20/12/2025.

** kulsawasd.jitkajornwanich@ttu.edu

*** kerk.kee@ttu.edu

Introduction

The perfect storm of artificial intelligence, social media platforms, and computational communication methods is transforming how societies detect, communicate, and respond to disasters and emergencies. As climate change intensifies the frequency and severity of natural disasters, traditional emergency management systems face unexpected challenges in providing timely, accurate, and actionable crisis communication (Alexander, 2014; Vogler & Meissner, 2024). The limitations of traditional disaster response mechanisms, which typically require hours of data processing and verification before issuing public warnings, have become apparent in contexts where minutes can determine life-or-death situations.

This paper presents a framework for understanding how AI-enabled social media analytics can revolutionize crisis communication while simultaneously introducing new categories of risk and ethical complexity. Drawing from computational communication research and disaster management literature, we explore the dual nature of risk communication in the digital/computational age: how emerging technologies both enhance our capacity to detect and communicate about disasters while creating novel forms of risk embedded within the communication processes themselves (Beck, 2009; Klinga & Lundgren, 2024). The framework addresses a critical gap in current literature by examining not only how AI and social media can improve disaster response, but also how over-reliance on these technologies might create new vulnerabilities in emergency management systems (Meißner & Diers-Lawson, 2022).

The research questions we are exploring include: How can AI-enabled social media analytics transform early warning systems while managing risks of mis- and dis-information and false positives? What speculative frameworks best capture the relationship between human decision-making and automated systems in crisis communication? How might future dependencies on AI-driven disaster analytics create new forms of systemic risk that communicators and policymakers have to anticipate? These questions are urgent given recent technological developments in generative AI, which expand analytical capabilities and at the same time make it harder to verify the analytical outcomes.

Our analysis employs a speculative design approach, examining both promising applications and concerning implications of AI-enabled crisis communication through multiple perspectives. We propose a theoretical framework that positions crisis communication at the intersection of technological capability, institutional authority, and public trust, while acknowledging the uncertainties in predicting how these relationships will develop. The paper contributes to strategic communication scholarship by offering an analysis of how emerging technologies might transform fundamental assumptions/theories about risk communication, institutional legitimacy, and public engagement during emergencies. We next synthesize prior research, delineate the systemic risks introduced by AI-mediated communication, and then advance a dual-risk framework operationalized through a Thailand flood-detection case and implementation guidance.

To guide the readers, the paper proceeds in five stages. First, the literature review synthesizes research on social media, AI-enabled detection, and human–AI collaboration to establish the conceptual foundations of smart-alert systems. Second, the analytical approach clarifies how the paper integrates a secondary re-analysis of the Thailand case with a speculative design–driven framework to surface generalizable design principles. Third, the conceptual sections develop the dual-risk structure and examine human–AI collaboration dynamics that shape the detection, verification, and communication of emerging hazards. Fourth, the Thailand case analysis and the subsequent Context and Transferability section show how these dynamics materialize in practice and how they must be adapted across sociotechnical, institutional, and linguistic domains. Fifth, the Strategic Frameworks for Implementation and the Adaptive Governance Architectures for Smart-Alert Systems sections articulate operational and institutional mechanisms (e.g., graduated confidence communication, multi-source verification, and adaptive governance architectures) that help organizations balance primary and secondary risks in real-world deployments.

Literature Review

Social Media as Crisis Information Infrastructure

This section synthesizes research across social media analytics, AI-enabled detection, and human–AI interaction to establish the conceptual foundations for the dual-risk framework developed later in the paper. Scholarship in crisis informatics has increasingly shown that social media now operates as both an informational sensor network and a dynamic communication arena during emergencies. Rather than functioning solely as downstream channels for official alerts, platforms such as X (formerly known as Twitter), Facebook, TikTok, and LINE enable publics to co-produce situational awareness through real-time observations, problem reporting, emotional expression, and community coordination (Mirbabaie et al., 2020). This collective behavior establishes a distributed early-warning system in which crisis-relevant signals frequently surface before formal authorities issue updates or confirm events (Reuter & Kaufhold, 2018).

Yet the same affordances that enable rapid signal emergence (e.g., low barriers to posting, high visibility, and algorithmic amplification) also introduce volatility. Studies show that during uncertain, high-stress events, information flows mutate quickly, generating mixtures of verified facts, speculation, emotional narratives, and misleading content (Gu et al., 2022). Public attention can swing dramatically in response to dramatic imagery, sensational framing, or rapid-fire sharing, creating feedback loops that heighten pressure on institutions to respond. This dual role of social media as both accelerator and distorter complicates institutional efforts to maintain authoritative communication in real time.

AI-Enabled Detection and Predictive Analytics

Concurrently, advancements in AI-driven detection systems have attempted to systematize the use of social metadata for disaster monitoring by leveraging natural language processing, image processing, computer vision, anomaly detection, and geospatial inference to identify hazards earlier than traditional infrastructures. Recent reviews (Reuter et al., 2023) indicate that these systems often succeed at surfacing meaningful crisis signals (e.g., localized flooding, wildfire spread, infrastructure disruption) faster than human analysts or official sensors.

However, the incorporation of AI also expands the crisis communication risk landscape. Machine-learning models depend on probabilistic inference, incomplete data, and historically biased training sets, generating error distributions that are difficult for non-experts to interpret. Research in human–AI interaction demonstrates that individuals alternately over-trust or over-scrutinize automated outputs, depending on prior expectations, the visual presentation of AI results, and the sociotechnical context in which decisions are made (Wu et al., 2024). These dynamics create vulnerabilities not only in detection accuracy but also in public confidence, as erroneous alerts or false negatives can cascade into legitimacy crises for institutions.

Synthetic Media and Misinformation

A rapidly evolving layer of complexity stems from the emergence of synthetic media and deepfakes, driven by generative AI. High-fidelity deepfakes can convincingly simulate political leaders, emergency managers, meteorologists, or even eyewitnesses, mimicking voices, facial expressions, and emotional cues with increasing realism. Recent empirical work demonstrates that deepfake audio and video can meaningfully distort risk perception, hinder institutional verification processes, and trigger public confusion, particularly during fast-moving disasters when information needs are acute and time pressure is high (Vaccari & Chadwick, 2020).

Further, comparative analyses show that AI-generated misinformation spreads faster than human-crafted content, elicits stronger emotional and threat responses, and is harder for users to evaluate heuristically in crisis contexts (Vaccari & Chadwick, 2020; Vosoughi et al., 2018). This transforms misinformation management from a downstream message-correction task into an upstream authenticity and provenance challenge, heightening the secondary risks associated with premature alerts, adversarial manipulation, and cross-platform spillover.

Human-AI Collaboration and Crisis Decision-Making Under Uncertainty

Large-scale crisis communication systems increasingly depend on hybrid human–AI workflows, in which AI provides interpretive or predictive support for human-led decisions. Research consistently shows that hybrid models outperform either human-only or automation-only systems in high-uncertainty environments. However, collaboration introduces new epistemic and operational challenges. Human decision-makers may exhibit automation bias (accepting AI outputs with insufficient scrutiny) or algorithm aversion (discounting AI outputs after observing errors), both of which can distort decisions during emergencies.

The Alsmosis framework (Bozdog, 2023) offers an important conceptual complement here. Rather than seeing humans and AI as discrete actors that exchange information in linear stages, Alsmosis describes how responsibilities, interpretations, and decision heuristics gradually “seep” between human and algorithmic components. This highlights why errors or blind spots in one part of the system can migrate into others, producing compound failures that are difficult to trace. In crisis communication systems, this seepage means that thresholds, confidence scores, and signal classifications can implicitly shape human judgment, even when officials believe they are exercising independent oversight.

Synthesis: Toward a Dual-Risk Understanding of AI-Driven Crisis Communication

Across these literatures, a consistent insight emerges: the same technological advances that promise earlier detection and more adaptive communication simultaneously introduce new forms of secondary risk. Social media accelerates informational visibility but magnifies noise; AI enhances detection but embeds probabilistic uncertainty; synthetic media expands expressive richness but destabilizes authenticity; and hybrid human–AI collaboration boosts capacity but complicates responsibility and trust.

Together, these developments establish the analytical foundation for a dual-risk perspective: modern crisis communication systems must manage not only the primary hazard (floods, fires, storms, public health threats) but also the risks produced by the detection, interpretation, and communication processes themselves. This synthesis sets the stage for the next section, which presents a framework for understanding how AI-enabled crisis systems generate, mediate, and potentially mitigate dual-risk dynamics through governance, verification pathways, and adaptive communication strategies.

Analytical Approach and Methodological Positioning

This section clarifies how the paper integrates a secondary re-analysis of the Thailand case with a speculative design-driven conceptual approach, forming the methodological bridge between the literature review and the theoretical framework that follows. This paper integrates two complementary methodological components: a secondary case analysis and a speculative design-driven conceptual framework. First, we conduct a targeted re-analysis of an existing technical case, the Thailand Twitter-based flood detection system by Jitkajornwanich et al. (2018), focusing not on reproducing its engineering details but on extracting design principles, governance implications, and risk dynamics relevant to contemporary smart-alert systems. This constitutes a situated analytical reinterpretation, where previously published technical outcomes are examined through the lenses of crisis communication theory, human–AI collaboration, and uncertainty management.

Second, we adapt methods from speculative design to articulate a forward-looking dual-risk framework, using conceptual constructs (e.g., multi-source validation, graduated confidence communication) not as labels but as scaffolds for imagining plausible, near-future communication architectures. Together, these components provide both an empirical anchor and a normative, anticipatory framework for evaluating AI-enabled alert systems.

Role of Speculative Design in the Framework

Speculative design in this paper does not aim to forecast technological futures exactly but to clarify the design space of emerging AI-driven warning systems. It allows us to reason about conditions that do not yet exist at scale (e.g., automated multi-source verification, confidence-scaled public alerts) by formalizing the trade-offs, uncertainties, and governance challenges that such systems would introduce. The speculative component therefore functions as a methodological bridge that connects what we observe in the Thailand case (the challenges of noisy signals, variable precision, and verification pressure) to what institutions will face as they adopt increasingly automated infrastructures. This approach elevates the framework from description to conceptual intervention, helping identify where risks accumulate and how they might be mitigated.

Positioning the Paper as a Hybrid Conceptual–Empirical Contribution

This methodological configuration positions the paper between empirical and conceptual category. The re-analysis of the Thailand case provides an empirical grounding that demonstrates how early-warning systems operate within real sociotechnical constraints, while the speculative design component offers an anticipatory lens for identifying risks and governance requirements that are only partially visible in current deployments. This hybrid approach is appropriate given the rapidly evolving nature of AI-enabled crisis

communication, where the empirical record remains incomplete and conceptual tools are needed to articulate future design requirements and institutional consequences.

Conceptualizing Dual Risk Structures

Building on the analytical approach above, this section introduces the paper's central conceptual contribution: the dual-risk structure that characterizes AI-enabled crisis communication systems and shapes the design tensions that follow. We define dual-risk structures in AI-enabled crisis communication as the need to manage, in tandem, (a) primary risks (the hazard itself: wildfire, flood, hurricane) and (b) secondary risks embedded in the communication system (verification errors, privacy–surveillance, bias/equity, ethical/cultural fit). The framework treats these as interdependent: interventions that lower primary risk (e.g., faster alerts) may raise secondary risk (e.g., false alarms, legitimacy loss). Accordingly, our design problem is to balance these risks through human–AI role design, confidence-matched messaging, verification across heterogeneous data sources, and adaptive oversight. The dual-risk framework builds from the insights surfaced in the Thai case but extends beyond them through speculative design, allowing us to articulate risk structures that are only partially visible in current implementations.

This dual-risk framing also clarifies how AI-enabled alert systems transform crisis communication into a multi-layered, interdependent process. Primary risks unfold in physical time and space, whereas secondary risks accumulate within sociotechnical infrastructures, such as algorithms, platform governance, institutional protocols, and public trust dynamics. Secondary risks include, but are not limited to, verification failures, bias and inequitable outcomes, privacy and surveillance concerns, and erosion of institutional trust, as summarized in Table 1 below. By conceptualizing these as mutually constitutive rather than separate domains, the framework makes visible why speed and accuracy cannot be optimized independently, why error tolerance becomes a normative rather than technical decision, and why institutions must design for uncertainty rather than attempting to eliminate it. This perspective provides the conceptual foundation for the subsequent sections on human–AI collaboration, verification, confidence-matched messaging, and governance.

Figure 1 and *Table 1* summarize the dual-risk structure and visually situate the design and governance components that the subsequent sections unpack in detail. Graduated Confidence Communication (GCC) refers to aligning public alert messages with model confidence; Multi-Source Verification (MSV) addresses the validation of AI-generated signals across data sources; and Adaptive Governance Architectures (AGA) captures institutional oversight and accountability mechanisms. The categories summarized in Table 1 represent analytically separable manifestations of secondary risk within the broader dual-risk structure illustrated in Figure 1. These components are intentionally cross-cutting: while each is shown as addressing specific secondary risks, they function jointly and reinforce one another across the system.

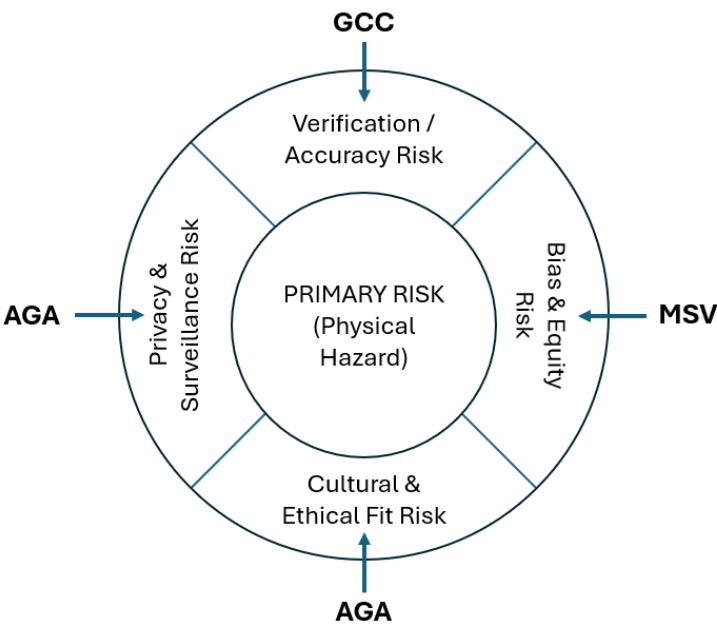


Figure 1 - Dual-Risk Structure of AI-Enabled Smart-Alert Systems. Primary and secondary risks are shown as analytically distinct but dynamically interdependent, with feedback effects that can amplify crisis outcomes. Arrows indicate primary mitigation pathways rather than exclusive or ranked solutions; all three mechanisms operate jointly across multiple secondary risks.

Secondary Risk Category	Risk Description	Primary Mitigation via GCC	Primary Mitigation via MSV	Primary Mitigation via AGA
Verification / Accuracy	Signal noise, inconsistent data, false positives/negatives	Communicates uncertainty with confidence bands	Cross-checks with heterogeneous models and data sources	Establishes thresholds and oversight
Bias & Equity	Uneven platform adoption, representational bias	Reduces overconfidence bias	Includes multiple validation channels and incorporates human in the loop	Ensures equitable error distribution
Privacy & Surveillance	Data sensitivity, monitoring concerns	Limits unnecessary alerts and escalation	Minimizes intrusive and maximizes transparent data collection	Provides institutional safeguards
Cultural & Ethical Fit	Misaligned messaging, legitimacy loss	Tailors message confidence to appropriate context	Uses local knowledge in verification	Ensures culturally grounded governance

Table 1 - Mapping Secondary Risks to Framework Components

Human-AI Collaboration Models

Having established the dual-risk structure, this section examines how humans and AI systems interact within that structure, highlighting the collaboration dynamics that influence real-time detection, verification, and messaging. Rather than positioning AI and human decision-makers in opposition, the framework emphasizes complementary capabilities and collaborative decision-making structures. AI systems excel at rapid data processing, pattern recognition across large datasets, and consistent application of predefined criteria. Human decision-makers contribute contextual understanding, ethical reasoning, communication skills, and accountability mechanisms that AI systems currently lack. Effective crisis communication systems should leverage these complementary strengths while maintaining clear lines of responsibility and oversight.

The framework identifies three primary collaboration models: AI-assisted human decision-making, where AI provides analytical support for human-led communication decisions; human-supervised AI communication, where AI systems generate initial responses subject to human review and approval; and parallel processing systems, where AI and human analysis proceed simultaneously with integration occurring at the communication stage. Each model involves different trade-offs between speed, accuracy, accountability, and resource requirements.

An additional conceptual lens that enriches this section is the *Alsmosis* framework proposed by Bozdag (2023), which describes how algorithmic processes and human practices increasingly “seep into” one another, producing hybrid assemblages where the boundaries of agency, accountability, and authorship become blurred. *Alsmosis* is useful for crisis communication because it shows that collaboration between humans and AI is rarely discrete or cleanly partitioned into separate stages; instead, human judgment, algorithmic recommendation, automated filtering, and institutional protocols continuously shape one another.

Incorporating this perspective clarifies how our three collaboration modes (i.e., AI-assisted decision-making, human-supervised AI communication, and parallel human-AI processing) operate not as rigid categories but as fluid configurations that shift depending on data quality, urgency, and risk tolerance. It also reinforces our dual-risk argument: as human and AI roles intermingle, errors, biases, or gaps in one domain can migrate into another, creating new secondary risks. Applying the *Alsmosis* lens therefore underscores the necessity of explicit governance structures, audit trails, and decision-rights clarity to prevent invisible drift of responsibility and to maintain legitimacy in AI-mediated emergency messaging.

Critical to all collaboration models is the maintenance of human agency and accountability in final communication decisions. While AI systems may process information and generate recommendations faster than human analysts, the framework insists that human decision-makers must retain ultimate authority over public communication during emergencies. This principle reflects both practical concerns about AI reliability and ethical commitments to democratic accountability in government emergency response. The

framework recognizes that maintaining human oversight while leveraging AI capabilities requires careful institutional design and clear protocols for when and how human judgment should override AI recommendations.

Temporal Complexity and Predictive Communication

The framework addresses the unique challenges posed by AI's capacity for predictive analysis and speculative warnings about potential future events. Traditional crisis communication has focused primarily on communicating about events that are currently occurring or imminent. AI analytics enable communication about events that may occur based on probabilistic assessments of current conditions and historical patterns. This predictive capability creates new opportunities for proactive emergency preparation while raising complex questions about appropriate communication strategies for uncertain risks.

The framework distinguishes between categorical predictions (event will occur), probabilistic predictions (event has X% likelihood of occurring), and speculative scenarios (event could occur under certain conditions). Each category requires different communication approaches, with categorical predictions warranting immediate action recommendations, probabilistic predictions requiring risk-benefit analysis, and speculative scenarios focusing on preparation and monitoring. The framework emphasizes that communication strategies must be matched to prediction certainty levels to maintain public trust and encourage appropriate responses.

Temporal considerations also extend to the evolving nature of AI capabilities and the need for communication systems that can adapt to technological change. The framework anticipates that AI analytical capabilities will continue expanding rapidly, potentially enabling prediction of events or conditions that are currently unforeseeable. Crisis communication systems must therefore be designed with sufficient flexibility to incorporate new AI capabilities while maintaining consistent public understanding and trust.

Emerging Opportunities and Applications

Multi-Modal Data Integration. Contemporary AI systems demonstrate unprecedented capabilities for integrating diverse data sources to create comprehensive situational awareness during developing emergencies. Social media platforms generate textual content, images, video, audio, geolocation data, engagement metrics, and temporal patterns that can be analyzed simultaneously to identify and characterize emerging situations (Alam et al., 2018). When combined with traditional data sources including satellite imagery, meteorological sensors, seismic monitoring, and institutional reporting, AI systems can potentially detect disasters faster and more accurately than any single data source alone.

The integration of emotional and sentiment analysis adds another dimension to disaster detection capabilities. Social media users experiencing or witnessing emergency situations often express fear, urgency, surprise, or other emotional responses that can provide additional signals about developing events. Engagement metrics such as sharing rates, reaction types, and comment patterns may indicate public perception of threat levels and the effectiveness of official communication. These behavioral indicators can inform both detection algorithms and communication strategy optimization.

Geospatial analysis capabilities enable AI systems to map disaster progression in real-time using crowdsourced location data extracted from social media posts, even when users have not explicitly enabled location sharing. The Thailand flood detection system demonstrated techniques for extracting location information from textual content, enabling detailed mapping of affected areas based on user reports (Jitkajornwanich et al., 2018). Advanced geocoding algorithms can identify specific streets, neighborhoods, or landmarks mentioned in social media posts and translate these references into precise coordinates for mapping and visualization.

Personalized Risk Communication. AI analytics create possibilities for personalized crisis communication that adapts message content, timing, and delivery channels to individual recipient characteristics and circumstances. Machine learning algorithms can analyze individual social media behavior, location patterns, demographic information, and communication preferences to optimize emergency message design and delivery. This personalization capability could significantly improve message effectiveness while raising complex privacy and equity concerns.

Individual risk tolerance preferences could be incorporated into personalized warning systems, allowing some users to receive alerts about low-probability events while others receive notifications only for high-confidence predictions. This approach addresses the fundamental tension between false positives and false negatives by allowing individual users to calibrate their personal alert thresholds. However, implementing such systems requires sophisticated user interface design and clear communication about the implications of different threshold settings.

Personalized communication could also adapt message design to reduce panic and encourage appropriate responses based on individual psychological profiles and historical response patterns. AI systems might learn which message framings, visual elements, or information structures are most effective for different user groups or populations. This capability could potentially reduce the amplification of fear and anxiety that sometimes accompanies emergency communication while ensuring that critical safety information reaches its intended audience effectively.

Proactive Communication and Scenario Planning. AI-enabled predictive analytics create opportunities for proactive communication about potential future risks before they materialize into actual emergencies. Weather pattern analysis, social media trend

monitoring, and environmental sensor data can identify conditions that historically correlate with specific types of disasters, enabling early warning communication that help populations prepare for possible events. This predictive capability extends the temporal scope of crisis communication from reactive response to proactive preparation.

Speculative scenario communication could help populations understand and prepare for novel or unprecedented risk combinations that AI analysis identifies as potentially emerging. Climate change creates new combinations of environmental conditions that may produce disaster scenarios outside historical experience. AI systems capable of analyzing complex interactions between multiple risk factors could identify potential scenarios that human analysts might not anticipate, enabling communication about preparation strategies for previously unimaginable events.

Case Study Analysis: Twitter-Based Flood Detection in Thailand

Based on the early flood warning system developed by Jitkajornwanich et al. (2018) in Thailand, the following case analysis illustrates how the dual-risk dynamics and collaboration patterns described above materialize in a real-world sociotechnical context, showing both the strengths and limitations of social-media-based early detection. We treat the Thailand system not as a technical prototype to be described in detail but as an empirical illustration that, when reinterpreted, reveals generalizable principles and tensions relevant to AI-enabled crisis detection.

Technical Implementation

The Thailand flood-detection prototype illustrates how social media can serve as an early indicator of emerging hazards in contexts where official monitoring infrastructure is uneven or delayed. The system continuously collected Thai-language tweets that referenced flood-related terms and used a language-aware location extraction process to approximate where users were reporting problems. Because geotagged tweets are rare in Thailand, the system inferred locations by matching textual references to known administrative units (e.g., neighborhoods, districts, provinces) and broadening the search when finer-grain data were unavailable. This hierarchical approach balanced geospatial precision with computational efficiency, demonstrating a key design trade-off for smart-alert systems: finer-grain geocoding increases specificity but slows detection, while coarser matching accelerates alerts but may expand uncertainty zones.

The system also compared real-time tweet volumes against historically observed baselines to estimate whether emerging floods were statistically meaningful anomalies. This calibration tied directly to local communication patterns allowed the system to differentiate

between routine chatter and genuine hazard-related spikes. Importantly, the goal was not to produce a fully automated warning tool but to show how platform-native signals, when processed thoughtfully, can complement official monitoring by detecting disruptions earlier than traditional channels.

User Behavior Analysis

Analysis of user activity revealed the distinctive communicative role that Twitter played during flood events in Thailand. Users frequently posted hyperlocal observations such as rising water levels, blocked roads, or drainage failures. These are information that rarely appears in formal reports but is crucial for situational awareness. Many users also engaged in “collective problem reporting,” tagging local authorities, journalists, or municipal accounts, which created organic feedback loops between residents and institutions. These interactions reinforced the notion that social media serves not only as a sensor network but also as a participatory crisis communication ecosystem, where communities help surface risks that institutions may not yet have verified.

At the same time, user behavior exhibited variability that affects early-warning reliability. During severe monsoon periods, for instance, high posting volumes reflected both genuine hazard conditions and heightened public anxiety, making it harder to distinguish meaningful signals from ambient noise. This underscores the importance of integrating behavioral baselines into smart-alert pipelines, a requirement relevant across cultural and linguistic contexts.

Visualization of Flood Signals

The visualization module went beyond standard Google Maps API implementation by aggregating geocoded tweet densities into simple heatmaps. The prototype translated social media activity into simple, interpretable visual outputs that indicated the approximate intensity and location of flood signals. Rather than providing precise maps or predictive modeling, the visualizations offered coarse but actionable indicators of where attention was increasing. These displays illustrated how even minimal geospatial inference, when aggregated across thousands of posts, can reveal patterns that would otherwise be difficult to perceive.

The visualizations served a conceptual purpose: they demonstrated how smart-alert systems must communicate degrees of confidence, not absolute certainty. Color scales, thresholds, and signal categories allowed the system to present emerging risks while making uncertainty explicit. This aligns directly with the design principles discussed in later sections, especially *Graduated Confidence Communication (GCC)*.

Validation and Implications for Early Detection

Comparison with official data from Thailand's Royal Irrigation Department and the Bangkok Metropolitan Administration showed that social-media-derived signals frequently appeared minutes to hours before formal alerts or press releases. Although the system was not designed for operational forecasting, it demonstrated the potential for social signals to function as an anticipatory layer in multimodal early-warning infrastructures. This confirmatory evidence supports the broader argument of this paper: social media can accelerate hazard detection, but it also adds ambiguity that must be managed through structured verification and careful communication.

Importantly, the Thailand case revealed the operational limits of using social media for early detection. Linguistic ambiguity, uneven platform adoption, misinformation, and high-volume "chatter spikes" can all create false positives. These limitations highlight why smart-alert systems require multi-source validation (MSV), where social-data signals are examined alongside hydrological sensors, municipal reports, satellite imagery, or local community networks.

Context and Transferability

While the Thailand case grounds the framework empirically, this section clarifies the contextual limits of those insights and outlines how the underlying design principles can be adapted across different linguistic, institutional, and technological environments. Although this paper articulates its framework in broad terms, the operational realities of AI-enabled warning systems remain deeply context dependent. Factors such as platform adoption, linguistic diversity, institutional capacity, governance cultures, and public trust norms shape how dual-risk dynamics manifest in practice. As a result, the framework should be understood as a transferable analytic model, not a universal prescription. Its components require local calibration, cultural grounding, and institutional adaptation. Making these contextual limits explicit strengthens the framework by clarifying where its principles hold and where they must be modified.

While the Thailand case demonstrates the feasibility of social-media-based early warning systems, its implementation also reveals a series of procedural steps that can guide adaptation in other national or organizational contexts. At minimum, any deployment requires (1) access to platform data streams with sufficient volume and geographic relevance; (2) a mapping layer capable of translating user-generated content into approximate locations; and (3) a verification pathway that connects social signals to authoritative data sources (e.g., sensors, official monitoring agencies, or trusted community partners). The Thai system operationalized these components through hashtag- or keyword-based data collection, language-specific tokenization, multi-level geocoding using internal administrative boundaries, and rate-based thresholds calibrated to local tweeting behaviors.

In other countries or hazard settings, each of these steps would need to be contextualized rather than duplicated.

Practically, adaptation begins with establishing platform baselines. That includes understanding normal posting volumes, linguistic cues, dialects, and platform adoption rates within the population of interest. Because geocoding precision varies widely across languages and countries, implementers must assess which location cues are realistically extractable (e.g., street names, landmarks, neighborhood terms) and whether supplemental data sources (e.g., local place-name databases, crowd-curated gazetteers, or government-provided shapefiles) are available. Threshold-setting likewise must be recalibrated using local behavioral patterns, taking into account that a 0.05% spike in Thailand corresponds to very different absolute numbers in regions with lower social media activity.

Moreover, cross-context adoption requires embedding the system within existing institutional workflows. This includes identifying which agencies have authority to issue warnings, how AI-generated signals will enter decision protocols, and what verification steps are required before public communication. Even in lightweight implementations, smart-alert systems benefit from clear governance design: where data are stored, who validates alerts, how uncertainty is communicated, and what safeguards are in place to mitigate privacy, bias, and misclassification risks. These procedural considerations allow organizations to adapt the underlying model, even with very different languages, platforms, or hazards, while preserving the integrity and legitimacy of the alerts. Taken together, these contextual considerations point to the need for structured design approaches capable of guiding the operational deployment of AI-enabled alert systems.

Finally, thresholds, acceptable false-positive rates, and the prioritization of hazard types are not purely technical parameters; they are decisions made by institutions, communities, or cross-agency bodies that reflect local risk cultures, governance traditions, and social expectations. What counts as an “acceptable” error rate in one country or community may be unacceptable (or even harmful) in another. Recognizing these differences is essential for adapting smart-alert systems responsibly across diverse sociotechnical and institutional contexts.

Strategic Frameworks for Implementation

Building on the conceptual and empirical foundations developed in earlier sections, this part of the paper consolidates the operational design principles necessary for implementing AI-enabled smart-alert systems. These principles (i.e., Graduated Confidence Communication/GCC and operational Multi-Source Verification/MSV) translate the dual-risk framework into concrete mechanisms that guide how emerging signals are interpreted, validated, and communicated under conditions of uncertainty. Design patterns map directly to specific classes of secondary risk. In operational terms, GCC (messages scaled to model confidence) primarily mitigates verification and accuracy risks and secondarily supports

ethical fit by tempering language, while MSV (cross-validation across social, sensor, and administrative data) mitigates verification/accuracy and bias/equity risks. The institutional role of adaptive governance architectures, such as addressing privacy–surveillance risks and sustaining legitimacy, is developed in the subsequent section.

Graduated Confidence Communication (GCC)

GCC serves as an operational mechanism to manage uncertainty and prevent premature escalation in early-warning contexts. Rather than issuing binary warnings, GCC communicates *graded levels of confidence* that reflect the evolving strength of the underlying signals. This structure provides institutional actors and publics with clearer expectations about how early indicators should be interpreted, and it helps prevent the overconfidence, misinterpretation, or alarm fatigue that often result from premature or overly definitive alerts.

GCC functions by translating probabilistic outputs in AI systems into structured tiers of communication (e.g., “signal of interest,” “signal strengthening,” “high-confidence alert”). This approach aligns directly with the dual-risk framework: it reduces secondary risks related to legitimacy loss, miscommunication, and trust erosion, while enabling earlier and more nuanced communication that mitigates primary risk (i.e., hazard impact). In addition to calibrating message timing and intensity, GCC also mitigates interpretive errors by reducing the risk of overconfidence in early or ambiguous signals. By presenting confidence bands rather than binary alerts, GCC helps prevent premature escalation and protects against the legitimacy losses associated with false alarms, especially in communities with past experiences of alert fatigue.

Multi-Source Verification

Operational MSV focuses on *how* detection systems cross-check and validate early-warning signals during the initial stages of hazard emergence. Because early signals extracted from social media often suffer from noise, demographic imbalance, linguistic ambiguity, or platform-specific biases, MSV strengthens detection pipelines by integrating information from diverse, independent, transparent, and heterogeneous data sources.

Operational examples include:

- cross-analyzing keyword surges with sensor readings
- validating geocoded posts with rainfall or hydrological data
- checking unexpected spikes against historical baselines
- using cross-platform confirmation (e.g., Twitter + local forums)
- evaluate and cross-validate prediction outputs across different models

In the dual-risk framework, MSV reduces secondary risks such as inaccurate classification, false positives, and representational bias while supporting faster detection of primary risks. MSV not only strengthens accuracy by cross-checking heterogeneous data sources but also reduces representational and equity-related risks. By incorporating multiple sources, including institutional datasets, community signals, environmental sensors, and geospatial information, MSV ensures that no single demographic group or platform artifact disproportionately shapes the alert output. While GCC and MSV provide essential operational mechanisms for early detection and communication, these design principles alone cannot guarantee equitable, accountable, or culturally legitimate alert outcomes. These broader concerns require institutional governance structures, addressed in the next section.

Adaptive Governance Architectures for Smart-Alert Systems

This section extends the framework from operational design into the institutional domain. Adaptive governance architectures (AGA) establish the oversight structures, decision protocols, and normative commitments that determine how smart-alert systems function in practice, ensuring that they are legitimate, accountable, and aligned with community expectations. AGA integrates institutional MSV, oversight, ethical safeguards, and the explicit management of value-laden choices such as acceptable false-positive rates and error burdens.

AGA defines the institutional actors responsible for validating alerts, escalating signals, approving public communication, and managing uncertainty. Even robust detection algorithms require human-centered review processes to determine when preliminary signals merit action. Institutions must therefore designate:

- validation authorities
- escalation thresholds
- interagency coordination pathways
- documentation standards
- transparency and audit requirements

These structures address the secondary risks of legitimacy, authority confusion, and inconsistent decision-making. AGA also supports public trust by creating clear, accountable lines of responsibility. Because thresholds inevitably encode normative decisions about what constitutes “acceptable risk,” GCC and MSV offer ways to surface and communicate these value-laden judgments rather than hiding them within technical parameters.

Normative Choices in Smart-Alert System Design and the Distribution of Uncertainty

Although AI-enabled alert systems are often framed as technical tools, their configuration depends on a series of explicit normative choices about what levels of error, delay, and uncertainty are socially acceptable. Decisions such as setting allowable false-positive or false-negative rates are not merely statistical thresholds; they reflect underlying judgments about whose safety is prioritized, who bears the burden of unnecessary alerts, and how institutions balance precaution against public fatigue.

For instance, a lower false-positive tolerance may reduce public annoyance but increases the likelihood that emerging risks go undetected, disproportionately affecting communities with less access to alternative information sources. Conversely, a higher tolerance for false positives may improve early detection but impose uneven costs on groups who cannot easily absorb disruptions caused by frequent alerts. Making these normative dimensions explicit is essential for designing equitable, transparent, and democratically legitimate smart-alert systems. These choices are not technical; they are ethical and political.

Examples include deciding:

- whether false positives are more harmful than false negatives
- whether alerts should prioritize vulnerable communities
- how cultural expectations shape acceptable levels of uncertainty
- how to avoid disproportionate burden on marginalized groups

Governance clarifies the normative choices embedded in early-warning systems by establishing criteria for equitable error distribution, culturally grounded review mechanisms, privacy safeguards, and limits on institutional overreach.

Privacy, Cultural Fit, and Sociotechnical Safeguards

Governance must ensure that alert systems adhere to societal expectations around privacy, data protection, cultural appropriateness, and community legitimacy. This includes:

- data minimization policies
- appropriate use of location data
- community consultation
- culturally appropriate messaging standards
- institutional transparency

These safeguards reduce secondary risks associated with surveillance, cultural mismatch, and trust erosion.

Multi-Source Verification as Governance Practice

Whereas the earlier section on Multi-Source Verification focused on MSV as a detection mechanism, here MSV serves as an institutional review process that determines who validates alerts, how validation decisions are justified, and how cross-agency verification prevents institutional bias or procedural opacity.

Governance-oriented MSV includes:

- cross-institutional validation committees
- human-in-the-loop review
- documentation of verification pathways
- accountability tracking

This version of MSV maps directly onto the governance layer of the figure and table, linking accuracy and equity concerns to institutional practices. Together, the operational mechanisms and the governance structures complete the system-level framework for AI-enabled smart-alert systems, positioning the paper to reflect more broadly on future research and practical implications in the next section.

Discussion and Future Directions

The preceding sections form the basis for a broader reflection on how AI-enabled warning systems reshape crisis communication, highlighting implications for research, emergency management practice, public trust, and long-term societal resilience. Our framework underscores that alert-system performance cannot be evaluated solely in accuracy terms; it must also be assessed in relation to the underlying normative choices that determine which errors are minimized, for whom, and at what social cost.

Implications for Strategic Communication

The integration of AI technologies into crisis communication challenges fundamental assumptions in strategic communication about message control, audience targeting, and communication effectiveness. The speed and scale of AI analytics create possibilities for real-time message optimization and audience analysis that exceed traditional communication planning capabilities while introducing new uncertainties about message interpretation and response. The shift from reactive to predictive communication fundamentally alters the temporal dynamics of strategic communication, extending it into speculative territory where communicators must address uncertain futures rather than known present conditions.

The democratization of information production through social media platforms combined with AI analytical capabilities creates new power dynamics in crisis communication that strategic communication theory has not adequately addressed. Traditional models assume institutional control over authoritative information sources, but AI systems can potentially identify and amplify non-institutional voices that provide valuable insights about developing situations. This distributed intelligence model challenges institutional gatekeeping roles while potentially improving information quality and response speed.

Resilience-Focused Approaches

The speculative framework advanced in this paper aligns with resilience-focused approaches to crisis communication. Research has shown that communicating mental health resources during COVID-19, particularly for vulnerable groups, can enhance community resilience (Akhther & Islam, 2022). Analysis of Historically Black Colleges and Universities' (HBCUs) website-based communication showed that mental health received minimal importance in pandemic response, with larger and advanced degree-granting institutions providing relatively greater mental health resources than smaller institutions.

Organizational memory, through after-action reports, plays a critical role in shaping resilience and renewal after crises. Research analyzing multiple reports created after the 2017 Las Vegas shooting found that reports from different professional fields commemorate crises in disparate ways that select and deflect memories of trauma, with these reports playing important emotional roles in making sense of organizational trauma (Rice & Bloomfield, 2022). Ultimately, AI-enabled systems must be embedded within collaborative emergency management structures that emphasize coordination, trust, and technological integration. Research investigating collaborative strategies to enhance emergency management identifies the importance of interagency cooperation, community engagement, technological integration, and policy development (Alkhouzaie et al., 2024).

Testable Implications

To translate the framework into evaluable claims, we derive testable propositions that link each design pattern—Graduated Confidence Communication (GCC), Multi-Source Verification (MSV), and adaptive governance architectures (AGA)—to measurable outcomes under real operational constraints. This step shifts the paper from theoretical contribution to empirical program design by specifying what success would look like (e.g., compliance, false-positive control, equitable coverage, legitimacy) and how to compare alternatives. The propositions below are therefore written to be fielded in drills, simulations, or quasi-experiments with agencies and platforms.

We propose four testable propositions:

- P1 (GCC efficacy). Confidence-matched alerts (vs. single-tone alerts) will yield higher compliance and lower alert fatigue under equal model accuracy.
- P2 (MSV accuracy). Multi-source verification will reduce false positives without lengthening time-to-alert beyond operational thresholds.
- P3 (Equity). Language-aware pipelines will increase detection/alert coverage in linguistic-minority areas relative to baseline models.
- P4 (AGA/trust). Transparent audit and role-clarity protocols will increase perceived legitimacy and willingness to comply with AI-mediated alerts.

Operationalization: P1 can be tested via A/B trials of alert templates in agency drills or controlled, IRB-approved online experiments measuring comprehension, intended compliance, and fatigue. P2 can use retrospective event logs to compare MSV vs. single-source pipelines on false-positive rate and time-to-alert. P3 requires stratified analyses by language/locale (coverage rate, precision/recall by subgroup). P4 can be evaluated with pre/post or difference-in-differences designs around policy rollouts (audit transparency, role charters), measuring perceived legitimacy and compliance intention.

Conclusion

The convergence of artificial intelligence, social media analytics, and crisis communication represents both unprecedented opportunity and substantial risk for emergency management systems. AI-enabled technologies offer the potential to detect disasters faster, communicate warnings more effectively, and tailor emergency responses to specific community needs and individual circumstances. The Thailand flood detection case study demonstrates that these capabilities are technically feasible and can provide valuable supplements to traditional emergency monitoring systems.

However, the implementation of AI-enabled crisis communication also introduces new categories of risk that require careful management and ongoing attention. The speed—accuracy dilemma inherent in emergency response becomes more complex when AI systems can process information faster than human verification systems can confirm its reliability. Privacy and surveillance concerns arise from the comprehensive data collection required for effective AI disaster detection. Algorithmic bias and equity issues may systematically exclude certain communities from AI-enabled early warning benefits.

The speculative framework presented here emphasizes the dual risk structure of AI-enabled crisis communication, recognizing that technologies designed to manage uncertainty may themselves become sources of uncertainty. Effective implementation requires explicit acknowledgment and management of both primary risks (the disasters we seek to detect and communicate about) and secondary risks (embedded within our technological communication systems). This dual risk structure creates complex interdependencies that traditional crisis communication models have not adequately addressed.

The framework's emphasis on graduated confidence communication, multi-source verification systems, and adaptive governance mechanisms provides practical approaches for managing these complexities while leveraging AI capabilities for improved emergency response. However, successful implementation requires sustained attention to ethical considerations, ongoing public education about AI system capabilities and limitations, and continuous adaptation to rapidly evolving technological capabilities.

Closing the loop, we argue that, framed as a dual-risk problem, smart-alert systems must show how GCC, MSV, and AGA jointly lower hazard harm while containing system-embedded risks. By naming the trade-offs, mapping mitigations to risk classes, and specifying testable propositions, we provide a path from speculative design to auditable, equitable practice in AI-enabled emergency messaging.

The stakes for getting this right are substantial. Climate change and technological advancement are creating new categories of risk that exceed the capabilities of traditional emergency management systems. AI-enabled crisis communication offers tools for addressing these challenges, but only if implemented thoughtfully with appropriate attention to both opportunities and risks. The framework presented here provides a foundation for that implementation, but its ultimate value will depend on continued research, development, and refinement through practical application.

Biographical Notes

Kulsawasd "Bo" Jitkajornwanich (Ph.D. in Computer Science, University of Texas at Arlington) is an Assistant Professor in the College of Media & Communication at Texas Tech University. His research focuses on computational communication, artificial intelligence, natural language processing (NLP), social media analytics, spatio-temporal database systems, and big spatial data, among other topics. His work has appeared in communication journals such as *Journal of Communication* (Oxford University Press), computer science journals such as *Applied Artificial Intelligence* (Taylor & Francis), and interdisciplinary journals such as *JMIR Formative Research* (JMIR Publications). His work has also been featured in key research handbooks, such as the *Handbook of Innovations in Strategic Communication*, and the *Routledge Handbook of Employee Communication and Organizational Processes*. He is the 2019 recipient of the National Dissertation Award in the "Information Technology and Communication Arts" category from the National Research Council of Thailand.

Kerk F. Kee (Ph.D. in Communication Studies, University of Texas at Austin), is the Virginia & Choc Hutcheson Professor in Mass Communication in the College of Media & Communication at Texas Tech University. His scholarship centers on information diffusion and technology adoption in organizational, health, science, environmental, risk, and crisis communication contexts. He has published research in communication journals such as *New Media & Society*, and *Communication Research*; computer science journals including *ACM Computer Surveys*, and *IEEE Transactions on Human-Machine Systems*; as well as interdisciplinary journals like *AI & Society*, and *International Journal of Information*

Management. He received a prestigious CAREER award from the US National Science Foundation's Computer & Information Science & Engineering (CISE) Directorate in 2015.

References

- Akhther, N., & Islam, K. (2022). Communicating mental health coping resources among college students of color: A resilience approach to COVID-19 response. *Journal of International Crisis and Risk Communication Research*, 5(2). <https://doi.org/10.63278/jicrcr.v5i2.122>
- Alam, F., Ofli, F., & Imran, M. (2018). CrisisMMD: Multimodal Twitter datasets from natural disasters. *Proceedings of the International AAAI Conference on Web and Social Media*, 12(1), 465–473. <https://doi.org/10.48550/arXiv.1805.00713>
- Alexander, D. E. (2014). Social media in disaster risk reduction and crisis management. *Science and Engineering Ethics*, 20(3), 717–733. <https://doi.org/10.1007/s11948-013-9502-z>
- Alkhouzaie, H., Mutawam, M., Dallak, H., Alahmari, S., Alsawat, A., Ogdi, A., & Ogdi, Q. (2024). Strengthening emergency management: Collaborative strategies for effective crisis response. *Journal of International Crisis and Risk Communication Research*, 7(1). <https://doi.org/10.63278/jicrcr.v7i1.89>
- Ansah, P. O. (2022). COVID-19 dialogue on Facebook: Crisis communication relationship between Ghanaian authorities and citizens. *Journal of International Crisis and Risk Communication Research*, 5(1). <https://doi.org/10.63278/jicrcr.v5i1.101>
- Beck, U. (1992). *Risk society: Towards a new modernity*. Sage Publications.
- Beck, U. (2009). *World at risk*. Polity Press.
- Bozdag, A. A. (2023). Almsosis and the pas de deux of human-AI interaction: Exploring the communicative dance between society and artificial intelligence. *Online Journal of Communication and Media Technologies*, 13(4). <https://doi.org/10.30935/ojcmr/13414>
- Coombs, W. T. (2014). *Ongoing crisis communication: Planning, managing, and responding*. Sage Publications.
- Diers-Lawson, A., & Meißner, F. (2021). Editor's essay: Moving beyond Western corporate perspectives: On the need to increase the diversity of risk and crisis communication research. *Journal of International Crisis and Risk Communication Research*, 4(1). <https://doi.org/10.63278/jicrcr.v4i1.31>
- Fire Safety Research Institute. (2024, April 17). *Phase One: Lahaina Fire Comprehensive Timeline Report*. UL Research Institutes & State of Hawai'i Office of the Attorney General. <https://fsri.org/resource/phase-one-lahaina-fire-comprehensive-timeline-report>
- Gu, M., Guo, H., Zhuang, J., Du, Y., & Qian, L. (2022). Social media user behavior and emotions during crisis events. *International Journal of Environmental Research and Public Health*, 19(9), 5197. <https://doi.org/10.3390/ijerph19095197>

- Jitkajornwanich, K., Kongthong, C., Khongsoontornjaroen, N., Kaiyasuan, J., Lawawirojwong, S., Srestasathiern, P., Srisonphan, S., & Vateekul, P. (2018). Utilizing Twitter data for early flood warning in Thailand. *2018 IEEE International Conference on Big Data*, 5165–5169. <https://doi.org/10.1109/bigdata.2018.8621961>
- Johansson, S. (2024). The anxiety threshold: Exploring the relationship between discrete emotions and information-seeking repertoires during a societal crisis. *Journal of International Crisis and Risk Communication Research*, 7(1). <https://doi.org/10.63278/jicrcr.v7i1.135>
- Jong, W., & Brataas, K. (2021). Victims as stakeholders: Insights from the intersection of psychosocial, ethical, and crisis communication paths. *Journal of International Crisis and Risk Communication Research*, 4(1). <https://doi.org/10.63278/jicrcr.v4i1.28>
- Klinga, M., & Lundgren, M. (2024). Making sense of disinformation in the Swedish heterogeneous society: Understandings, experiences, and vulnerabilities. *Journal of International Crisis and Risk Communication Research*, 7(1). <https://doi.org/10.63278/jicrcr.v7i1.138>
- Kryvasheyev, Y., Chen, H., Obradovich, N., Moro, E., Van Hentenryck, P., Fowler, J., & Cebrian, M. (2016). Rapid assessment of disaster damage using social media activity. *Science Advances*, 2(3), e1500779. <https://doi.org/10.1126/sciadv.1500779>
- LaCour, M., Serra, M. J., Duvall, M., & Hislop, C. (2023). Getting lost in the “realm of possibility”: Common phrases used to communicate rare events have substantially different effects on decision-making. *Journal of International Crisis and Risk Communication Research*, 6(2). <https://doi.org/10.63278/jicrcr.v6i2.132>
- Lee, Y.-I., Lu, X., Voges, T., & Jin, Y. (2023). Fending off unverified accusation with narratives: The role of primary and secondary narratives in organization’s response effectiveness in an ongoing crisis. *Journal of International Crisis and Risk Communication Research*, 6(1). <https://doi.org/10.63278/jicrcr.v6i1.126>
- Madden, S., Eng, N., & Myrick, J. G. (2023). Emotional responses to wireless emergency alerts for COVID-19 and predictors of public health compliance. *Journal of International Crisis and Risk Communication Research*, 6(1). <https://doi.org/10.63278/jicrcr.v6i1.128>
- Maui Emergency Management Agency. (2025). *Maui Wildfires 2023 After-Action Report*. County of Maui. <https://www.mauicounty.gov/DocumentCenter/View/151355/MEMA-2023-Wildfire-After-Action-Report>
- Meißner, F., & Diers-Lawson, A. (2022). Editorial essay: Innovation in risk and crisis communication: Toward new topics, theories, and methods. *Journal of International Crisis and Risk Communication Research*, 5(2). <https://doi.org/10.63278/jicrcr.v5i2.118>
- Miller, A. N., Collins, C., Neuberger, L., Todd, A., Sellnow, T. L., & Boutemen, L. (2021). Being first, being right, and being credible since 2002: A systematic review of crisis and emergency risk communication (CERC) research. *Journal of International Crisis and Risk Communication Research*, 4(1). <https://doi.org/10.63278/jicrcr.v4i1.26>

- Mirbabaie, M., Bunker, D., Stieglitz, S., Marx, J., & Ehnis, C. (2020). Social media in times of crisis: Learning from Hurricane Harvey for the coronavirus disease 2019 pandemic response. *Journal of Information Technology*, 35(3), 195–213. <https://doi.org/10.1177/0268396220929258>
- Page, T. G., & Clementson, D. E. (2023). The power of style: Sincerity's influence on reputation. *Journal of International Crisis and Risk Communication Research*, 6(2). <https://doi.org/10.63278/jicrcr.v6i2.131>
- Prasad, M. (2022). Templated crisis communication for people with disabilities, access and functional needs. *Journal of International Crisis and Risk Communication Research*, 5(2). <https://doi.org/10.63278/jicrcr.v5i2.123>
- Reuter, C., & Kaufhold, M. A. (2018). Fifteen years of social media in emergencies: A retrospective review and future directions for crisis informatics. *Journal of Contingencies and Crisis Management*, 26(1), 41–57. <https://doi.org/10.1111/1468-5973.12196>
- Reuter, C., Hughes, A. L., & Kaufhold, M.-A. (2023). Social media in crises: A systematic literature review and future agenda. *Computers in Human Behavior*, 140, 107568.
- Rice, R. M., & Bloomfield, E. F. (2022). Commemorating disorder in after-action reports: Rhetorics of organizational trauma after the Las Vegas shooting. *Journal of International Crisis and Risk Communication Research*, 5(1). <https://doi.org/10.63278/jicrcr.v5i1.102>
- Rodríguez-Díaz, C. E. (2018). Maria in Puerto Rico: Natural disaster in a colonial archipelago. *American Journal of Public Health*, 108(1), 30–32. <https://doi.org/10.2105/AJPH.2017.304198>
- Rogers, D., & Tsirkunov, V. (2011). *Costs and benefits of early warning systems*. World Bank Publications.
- Schwarz, A., Sellnow, D. D., Sellnow, T. D., & Taylor, L. E. (2024). Instructional risk and crisis communication at higher education institutions during COVID-19: Insights from practitioners in the Global South and North. *Journal of International Crisis and Risk Communication Research*, 7(1). <https://doi.org/10.63278/jicrcr.v7i1.134>
- Starbird, K., Maddock, J., Orand, M., Achterman, P., & Mason, R. M. (2014). Rumors, false flags, and digital vigilantes: Misinformation on Twitter after the 2013 Boston Marathon bombing. *iConference 2014 Proceedings*, 654–662. <https://doi.org/10.9776/14308>
- Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1). <https://doi.org/10.1177/2056305120903408>
- Vogler, D., & Meißner, F. (2024). The mediated construction of crises—Combining automated and qualitative content analysis to investigate the use of crisis labels in headlines of Swiss news media between 1998 and 2020. *Journal of International Crisis and Risk Communication Research*, 7(1). <https://doi.org/10.63278/jicrcr.v7i1.136>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>

- Wang, Y., & Chen, J. (2022). What motivates information seeking and sharing during a public health crisis? A combined perspective from the uses and gratifications theory and the social-mediated crisis communication model. *Journal of International Crisis and Risk Communication Research*, 5(2). <https://doi.org/10.63278/jicrcr.v5i2.120>
- Wu, Y., Cong, Y., Zhang, H., Wang, H., & Hu, P. (2024, September 5). *Emotional Contagion and Public Risk Perception in Crisis Events on Social Media: Insights from the Chinese Context during COVID-19*. <https://doi.org/10.31234/osf.io/s3utx>

Processing Fabricated Foreign Risks with AI Fact-Checker: The role of Third-Person Perception, Nationalism and Conspiratorial Orientation*

Qi Chen**
KU Leuven

Yicheng Zhu***
Beijing Normal University

The proliferation of fabricated information presents a global challenge, particularly when infused with conspiratorial narratives. As AI technologies become integrated into news production and verification, interfaces increasingly display labels such as “verified by a human” or “verified by AI,” altering how individuals evaluate risk-related content. Personality traits—especially blind nationalism and conspiratorial orientation—shape responses to such information, while perceptions of social norms and susceptibility (third-person perception, TPP) mediate these effects. This study investigates how verification source (human vs. AI) and personality traits interact to influence third-person perception and subsequent information-seeking intentions. In a between-subjects experiment (N = 506), participants were randomly assigned to read misinformation labeled as verified either by a human or by AI, then completed measures assessing perceived social norms, third-person perception (TPP), and information-seeking intentions. Results show that H1 was supported: blind nationalism consistently predicted higher perceived social norms of the foreign-risk misinformation in both conditions (human: $\beta = 0.465$, $p < .001$; AI: $\beta = 0.371$, $p < .001$). H3 was supported: higher perceived norms were associated with lower third-person perception in both conditions (human: $\beta = -0.244$, $p = .003$; AI: $\beta = -0.364$, $p < .001$). H4 was supported: lower TPP predicted stronger intentions to seek additional information (human: $\beta = -0.211$, $p < .001$; AI: $\beta = -0.208$, $p < .001$). The indirect pathway from nationalism to information-seeking via perceived norms and TPP was significant in both groups, with comparable indirect effects across conditions. In contrast, H2 was not supported: conspiratorial orientation had no significant effect on perceived norms or on any downstream outcome in either condition. These findings underscore that verification interface moderates the psychological pathways through which personal predispositions shape engagement with misinformation. The third-person effect emerges as a key mediator, translating normative perceptions into behavioral intentions. This study contributes to cross-contextual understanding of AI-mediated news credibility and offers practical insights for designing verification systems that enhance public resilience to misinformation.

Keywords: Conspiratorial Orientation, Nationalism, Fact-Checker, Third-Person Perception, Information Seeking

* Article submitted on 22/09/2025. Article accepted on 12/12/2025.

** qi.chen@student.kuleuven.be

*** yichengresearch@gmail.com

Introduction

Foreign-risk misinformation—false or misleading claims that portray multinational companies or foreign actors as threats—has become especially consequential. These narratives can inflame nationalist sentiments, undermine transnational trust, disrupt overseas operations of multinational corporations, and, more broadly, threaten the sociopolitical foundations of economic globalization.

In response, fact-checking has become an increasingly important safeguard in global information flows. Verification practices are themselves evolving: platforms now experiment with AI-supported tools such as Grok’s AI-driven fact-checking features and X’s (Twitter’s) Community Notes, while others, such as Meta, have recently discontinued their U.S. Third-Party Fact-Checking program. These developments signal a shifting verification ecosystem where traditional journalist-driven review coexists—and at times competes—with computational approaches. As a result, fact-checking interfaces—whether human-labeled or AI-labeled—have become central components of contemporary misinformation governance, especially for contested foreign-risk claims.

Understanding public responses to these verification efforts requires attention to individual-level psychological mechanisms. Research shows that conspiracy beliefs vary across political and cultural contexts (Armaly et al., 2022; Lorenzo Posocco & Jarfhlaith Watson, 2024; Malešević et al., 2024; Jia & Luo, 2023; Mir & Siddiqui, 2024) and exert systematic influence on how people interpret risk-related narratives.

Misinformation acceptance is shaped not only by accuracy judgments but also by trust, identity, and perceived norms (Lan et al., 2025). During the COVID-19 “infodemic,” audiences often accepted both consensus accounts and misinformed narratives simultaneously (Zhu et al., 2024), with lower trust in science predicting higher endorsement of misinformation (Kahan, 2017). These findings suggest that correcting foreign-risk misinformation involves more than supplying accurate information; it also requires understanding the cognitive heuristics and trust dispositions that guide public interpretation of fact-checks.

Third-person perception (TPP) offers a complementary lens for understanding how individuals evaluate foreign-risk misinformation. TPP holds that people typically judge others as more susceptible to misinformation than themselves (Davison, 1983; Gunther, 1992). In environments shaped by nationalism or conspiratorial thinking, perceived descriptive norms—beliefs about what others generally accept—can compress or expand this “self–other” gap.

At the same time, the RISP (Risk Information Seeking and Processing) model highlights the role of perceived social norms as motivational antecedents that shape how individuals process risk information and decide whether to seek additional information (Kahlor, 2007). Prior research shows that specific conspiracy schemas (e.g., GILC; Kahan et al., 2012; Lan et al., 2025; Zhu et al., 2024) can become socially legible and normatively communicated, binding identity to belief.

Integrating these perspectives, the present study pursues two analytical objectives:

Model how nationalism, conspiratorial orientation, and perceived social norms jointly shape TPP judgments and subsequent information-seeking intentions when individuals encounter foreign-risk misinformation. Compare model path strength and viability across two fact-checking interface conditions (AI-labeled vs. human-labeled).

By embedding these comparisons within broader dynamics of nationalism, conspiratorial orientation, and subjective norm perception, the study advances understanding of how fact-checking interfaces interact with cognitive heuristics derived from both the RISP and TPP traditions. The findings contribute to research on foreign risk misinformation governance and offer practical implications for platform design, fact-checking strategies, and media literacy initiatives, with particular relevance for the Chinese information environment.

Literature Review

Nationalism and Perceived Social Norms of Foreign-Risk Misinformation

Nationalism has long been theorized as a double-edged orientation in mass communication research. On one hand, it fosters solidarity, civic engagement, and symbolic attachment to the nation-state (Billings et al., 2013). On the other, it can cultivate exclusionary attitudes, xenophobia, and susceptibility to narratives that position outgroups as threats (Martikainen & Sakki, 2021). In the domain of misinformation, these exclusionary aspects are particularly consequential: when misinformation invokes foreign adversaries, nationalist orientations provide a ready-made interpretive frame that legitimizes and amplifies such content (Flew & Iosifidis, 2020).

A robust body of scholarship documents how national identity interacts with risk communication. Guan's (2020) analysis of conspiratorial beliefs in China illustrates how nationalist rhetoric frequently intersects with populist and conspiratorial discourses, generating a "foreign-plot" schema in which outside powers are cast as malevolent actors. This schema is not merely an elite-driven rhetorical device; rather, it becomes a recognizable narrative structure that ordinary citizens draw upon when processing news and rumors about international issues. By attributing risk to foreign malfeasance, nationalist frames increase the plausibility of otherwise questionable claims and feed into a broader communicative climate that regards suspicion of outsiders as normative (Jaspal & Cinnirella, 2010). In this sense, nationalism both supplies the lens through which misinformation is judged credible and reinforces a perception that "most people" endorse such judgments (Rowling et al., 2015).

The mechanism linking nationalism to perceived norms can be understood through the concept of descriptive norm perception: individuals estimate what most others believe based on repeated exposure to cues in their environment. Nationalist discourses are often amplified in mainstream media and echoed by social elites, especially in times of diplomatic conflict or crisis (Whitmeyer, 2002). When these discourses align with misinformation narratives—for instance, blaming a foreign country for economic decline, environmental

hazards, or health crises—the perceived consensus around such claims intensifies (Shayegh et al., 2022). Repeated exposure fosters pluralistic ignorance, where individuals who may personally doubt the claim nevertheless assume that “everyone else” accepts it. Over time, this dynamic transforms isolated rumors into perceived social norms, especially in digitally networked spaces where algorithmic curation favors emotionally charged, identity-reinforcing content (Mihelj & Jiménez-Martínez, 2021).

Importantly, nationalism’s effect on norm perception is not simply a matter of persuasion but of social identity signaling. Communication scholars have shown that when political or cultural identities are salient, individuals often align their expressed beliefs with in-group norms, regardless of their private attitudes (Kahan, 2015; Kunda, 1990). This identity-protective cognition means that people who identify strongly as nationalists may publicly endorse foreign-risk misinformation as a way of affirming loyalty to the nation, even if privately ambivalent. Such public alignment then becomes visible evidence for others, reinforcing the cycle of perceived consensus. The result is a self-reinforcing loop: nationalist identity increases openness to foreign-risk misinformation, public endorsement signals group loyalty, and the accumulation of such signals makes the norm perception of misinformation acceptance even stronger.

It is important to consider the cultural context: Chinese nationalism and public trust in AI technology likely shape how fact-checking cues are received. Chinese society exhibits remarkably higher trust in AI compared to Western countries. An Edelman survey found that 87% of Chinese respondents express confidence in AI, versus roughly one-third or less in the US and Europe. This trust gap reflects differing media environments – in China, state-led narratives portray AI as a pillar of national progress, fostering optimism and acceptance of AI-driven solutions (AI Jazeera, 2025). Nationalistic sentiment can amplify this effect: citizens proud of China’s tech advancements may view an “AI-verified” label as credible and domestically trustworthy, aligning with their national pride. By contrast, Western publics, often more skeptical of AI in media roles, might question an algorithm’s bias or transparency. Indeed, only 17% of Americans welcome greater AI use in daily life, compared to over half of Chinese respondents. These differences suggest that in China’s media landscape, an AI fact-checker might be afforded a level of public trust – and perhaps perceived impartiality – that would not be assumed elsewhere. Chinese nationalism’s influence further means that fact-checks congruent with national narratives (e.g., debunking foreign threats) could resonate strongly (Semafor, 2025). In sum, the Chinese public’s high AI trust and nationalist information norms create a distinctive context, one where an AI fact-checking interface may enjoy legitimacy and uptake in ways not mirrored in Western settings.

In sum, prior research suggests that nationalism, particularly in its exclusionary or “blind” form, is positively associated with the perception that misinformation about foreign risks is widely accepted by society. Nationalist frames legitimize threat-based misinformation, encourage public alignment with group narratives, and elevate the perceived descriptive norm that others share these beliefs. On this basis, we propose the following hypothesis:

H1: Nationalism is positively associated with perceived social norm of foreign-risk misinformation.

Conspiratorial Orientation and Perceived Social Norms of Foreign-Risk Misinformation

If nationalism supplies a collective identity lens through which foreign-risk misinformation is rendered plausible, conspiratorial orientation provides the cognitive style that makes such misinformation compelling (Kim & Lee, 2024; Romer & Jamieson, 2022). Communication and psychology literatures converge in defining conspiratorial orientation as a relatively stable tendency to interpret events as the result of secret plots orchestrated by powerful actors (Douglas et al., 2017). This disposition extends beyond specific conspiracy theories; it reflects a generalized mindset characterized by suspicion toward official accounts, heightened sensitivity to perceived threats, and a search for hidden explanations.

Several mechanisms explain why conspiratorial orientation should foster the perception that misinformation about foreign risks is not only believable but socially normative. First is the epistemic function. Conspiracy thinking thrives in environments of uncertainty and crisis, offering seemingly coherent narratives that simplify complexity and restore a sense of causal order (Sutton & Douglas, 2020). When individuals high in conspiratorial orientation encounter misinformation implicating foreign actors—whether about disease origins, cyberattacks, or economic sabotage—they are primed to accept such accounts because they fit the broader script of hidden manipulation by outsiders. Crucially, these individuals also tend to believe that others share their suspicions. Studies of conspiracy endorsement during the COVID-19 pandemic, for example, show that mistrust in science and institutions strongly predicts belonging to “misinformed” or “both consensus-and-misinformed” belief profiles (van Mulukom et al., 2020). Perceiving oneself in the company of likeminded skeptics reinforces the sense that conspiracy-based misinformation is a widely held norm.

Second is the social identity function. Conspiracy beliefs often provide group-based cohesion, delineating “truth-seekers” from the gullible mainstream. In such cases, conspiratorial orientation not only sustains belief in foreign-risk misinformation but also frames adherence as a patriotic stance (Daunt et al., 2023). The more people see conspiracy narratives being circulated in their online networks, the more they interpret this circulation as evidence of social consensus. The communicative environment—particularly on social media platforms that privilege sensational or counter-hegemonic content—magnifies this perception by algorithmically exposing users to like-minded posts, producing what appears to be widespread agreement.

Third, conspiratorial orientation is linked to a pervasive distrust of authority and mainstream media, which weakens the corrective impact of fact-checking or expert refutations (Ternullo, 2022; Van Prooijen et al., 2022). Individuals with high conspiracy orientation often reject “official” corrections as further evidence of cover-ups, thereby reducing the perceived legitimacy of alternative viewpoints. When mainstream challenges are dismissed, the visible survival of misinformation in the communicative ecosystem is interpreted as tacit validation: if the narrative persists despite denial, it must resonate with what many others believe. This logic fuels the perception of descriptive norms even in the absence of accurate indicators of public opinion.

Taken together, these mechanisms underscore why conspiratorial orientation is more than an individual cognitive bias; it is a social force that actively shapes perceptions of what others believe. The epistemic need for certainty, the identity-protective benefits of “insider knowledge,” and the distrust of corrective institutions converge to transform foreign-risk misinformation into a perceived normative stance. Accordingly, we advance the second hypothesis:

H2: Conspiratorial orientation is positively associated with perceived social norm of foreign-risk misinformation.

Perceived Social Norms and the Third-Person Perception

The third-person perception (TPP) describes the tendency for individuals to judge media content as exerting stronger effects on others than on themselves (Davison, 1983). Its behavioral corollary, the third-person effect (TPE), has been documented across diverse issues—from violent media and pornography to political advertising and misinformation—and often manifests in support for restrictive policies aimed at “protecting” others (Perloff, 2009). The canonical explanations for TPP emphasize two psychological mechanisms: ego-enhancement, whereby individuals protect their self-image as rational and immune to influence, and social desirability, whereby they distance themselves from “undesirable” content presumed to affect less discerning audiences (Gunther, 1992). Yet this line of research has only rarely incorporated the role of perceived descriptive norms—that is, one’s sense of what most others believe—when evaluating self/other media effects.

Why should perceived social norms of foreign-risk misinformation matter for TPP? At its core, TPP relies on a comparative judgment: I am less influenced than they are. This comparative gap becomes meaningful when the audience believes that others are vulnerable and credulous. If, however, an individual perceives that foreign-risk misinformation is widely accepted—that it represents not a marginal belief but a mainstream norm—the basis for exaggerating others’ susceptibility weakens. Under strong perceived norms, individuals may view acceptance of misinformation as socially typical rather than deviant. As a result, the distinction between self and others collapses, attenuating the TPP gap.

This logic aligns with social-norms theory in communication, which argues that perceptions of collective endorsement guide individual attitudes and behaviors. When misinformation about foreign risks is framed as common sense or patriotic truth, distancing oneself from its influence may no longer serve the ego-enhancement function. Instead, acknowledging shared susceptibility—or at least refraining from asserting immunity—becomes a way to maintain group alignment. In other words, strong perceived norms convert what might otherwise be a stigmatized belief into a socially desirable one. If accepting foreign-risk misinformation is seen as the normative stance of “people like us,” then claiming that others are more affected than oneself no longer yields reputational benefits.

Empirical evidence supports this reasoning. For example, Koo et al. (2021) found that third-person perceptions and perceived norms jointly shape how people respond to misinformation, implying that normative beliefs can modulate the usual self–other bias in perceived influence. In the context of public health myths, individuals who perceive a false claim as widely accepted tend to exhibit a smaller third-person perception gap – rather than viewing “others” as more gullible, they judge the influence on self and others to be more comparable (Oktavianus & Bautista, 2023). Likewise, in polarized political environments where false claims circulate within like-minded networks, the TPP effect is often attenuated because misinformation becomes viewed as a normative group belief rather than a deviant opinion. This echo chamber dynamic, akin to a false consensus effect, leads individuals to assume that many others share the same misinformation, reducing the incentive to assert one’s own immunity (Luzsa & Mayr, 2021). In sum, strong descriptive norms appear to erode the self–other disparity central to TPP. Although prior studies did not focus specifically on foreign-risk narratives, the underlying mechanism should be parallel: perceived norms shape the desirability and diagnosticity of claiming immunity from influence.

Thus, we contend that perceived social norms of foreign-risk misinformation are negatively associated with TPP. As the perception grows that misinformation acceptance is typical and widely shared, the motivational basis for asserting greater immunity than others erodes. Instead of accentuating the self–other difference, individuals may converge in their judgments of susceptibility. This reasoning leads us to propose the third hypothesis:

H3: Perceived social norm of foreign-risk misinformation is negatively associated with third-person perception (TPP).

Third-Person Perception and Information-Seeking Intention

The final link in our conceptual chain connects third-person perception (TPP) to information-seeking behaviors. Decades of TPP research has established that the “self–other gap” in perceived media effects is not a purely cognitive quirk; it carries behavioral implications. When individuals believe that others are more vulnerable to harmful media messages, they often endorse interventions to mitigate those effects—supporting censorship, regulations, or corrective campaigns (Perloff, 2009). Yet the same comparative logic can also suppress proactive engagement at the individual level: if one perceives oneself as relatively immune, the incentive to seek out additional information diminishes.

The mechanism here is straightforward. Information-seeking is often motivated by personal relevance and perceived vulnerability. In risk communication, individuals turn to additional sources when they feel uncertain, exposed, or at risk (Griffin, Dunwoody, & Neuwirth, 1999). But TPP reinforces the opposite inference: “others are more misled than I am”. By externalizing vulnerability, individuals reduce their sense of personal need to investigate further. Instead, they may assume that corrective responsibility lies with institutions or that the greater danger resides in others’ susceptibility, not their own. This logic yields a negative association between TPP and future information-seeking intentions.

Decades of research suggest that people who think others are more affected by harmful media content often support remedies for those others while neglecting to change their own behavior. Recent misinformation studies confirm this pattern. Chung and Kim (2021) showed that when individuals were presented with a fact-check debunking a false news story, their third-person perception increased, which in turn significantly reduced their willingness to share the misinformation on social media. Similarly, in a health communication context, a heightened third-person perception has been linked to lower personal intentions to correct or verify false information – even as those same individuals remain supportive of corrective interventions aimed at “gullible” others (Oktavianus & Bautista, 2023). This outward-focused response reflects an illusory superiority bias: people tend to overestimate their own immunity to misinformation and thus see less need to seek out additional information or fact-check for themselves (Corbu et al., 2025). In other words, the protective impulse operates outwardly rather than inwardly, directing efforts toward safeguarding others while one’s own proactive information-seeking is curtailed.

The communication environment further reinforces this displacement of responsibility. Algorithmic personalization and online echo chambers filter exposure in ways that rarely challenge users’ existing beliefs, fostering a false sense of consensus (Luzsa & Mayr, 2021). Within these insular networks, debunking efforts often fail to penetrate. For instance, Zollo et al. (2017) found that on Facebook, fact-checking posts were predominantly consumed by users in pro-science communities, whereas members of conspiracy-driven communities scarcely engaged with those corrections. This means that those most susceptible to the misinformation often do not see the debunking at all, blunting its corrective potential. In such settings, individuals confident in their own judgment remain surrounded by like-minded content, which only reinforces their perspective and further diminishes their perceived need to verify information or seek out additional sources. TPP thus dovetails with these structural features of digital media: by keeping people in information silos where their lack of personal vigilance is never challenged, it ultimately dampens the likelihood of future information-seeking even when foreign-risk misinformation carries real social or political consequences. In sum, the literature suggests that TPP attenuates the perceived necessity of personal information acquisition. By framing oneself as less vulnerable than others, individuals reduce their motivation to pursue additional knowledge about foreign risks. Thus, we propose the fourth hypothesis:

H4: Third-person perception is negatively associated with future information-seeking intention about the foreign risk

Integrating the abovementioned mechanisms, the study also tends to address the following research question:

RQ: How do individual dispositions—specifically nationalism and conspiratorial orientation—interact with fact-checking cues to shape perceived social norms of foreign-risk misinformation, and how do these perceived norms, in turn, influence third-person perception and intentions to seek additional information about foreign risks?

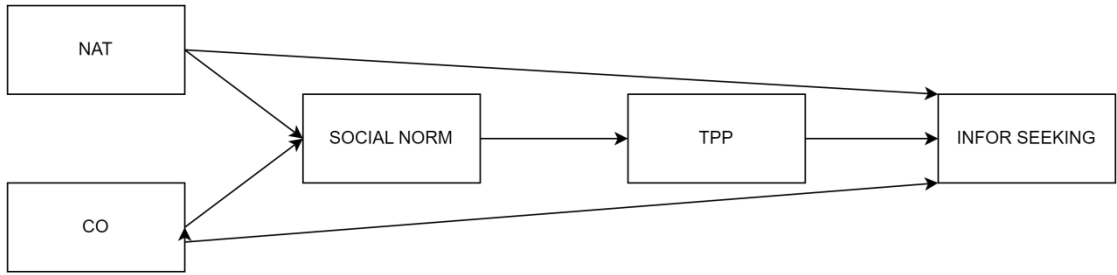


Figure1 - Main model

Method

Sample and Sampling Method

The sample was recruited through a major Chinese online panel provider (Wenjuanxing, www.wjx.cn) using quota sampling based on the sixth national population census of China (National Bureau of Statistics of China, 2018) Quotas were set to match the adult population in Beijing in terms of age, gender, and education level. Survey invitations were distributed to a representative online panel, and participants were randomly assigned to one of two experimental conditions.

For the purpose of the current study, we utilized data from two conditions: the AI fact-checker group (n = 252) and the human fact-checker group (n = 254). These two groups viewed the same misinformation relating to foreign risks, differing only in the label indicating the identity of the fact-checker. Participants who failed the manipulation check (i.e., could not correctly identify who had fact-checked the news article) were retained for robustness purposes but flagged for sensitivity analysis.

The experiment was conducted in two waves, from September 9 to 28, 2024, and from January 13 to 19, 2025. Participants completed a series of questions before and after the stimulus exposure, including demographic information, personality measures (nationalism and conspiratorial orientation), and outcome variables such as third-person perception and information-seeking intention.

Experimental Design

This study employed a between-subjects experimental design with a single manipulated factor: the identity of the fact-checker (AI vs. human). Participants were randomly assigned to view a conspiracy-themed news article that included either an AI fact-checker label or a human fact-checker label, indicating who had verified the content.

The stimulus materials consisted of a news screenshot adapted from real-world misinformation, covering topics related to foreign risk conspiracy narratives (e.g., Pfizer’s alleged role in mutating the Covid-19 virus). The core content of the article—including the

headline, body text, and visual layout—remained identical across conditions. Only the verification label was varied: one version displayed the message “This article has been verified by our AI system,” while the other showed “This article has been verified by professional human fact-checkers”. The screenshot was designed to simulate a typical mobile news app interface familiar to Chinese users (e.g., Tencent News or Toutiao). The verification label appeared in simplified Chinese and was positioned between the article headline and body text. It was rendered in bold font, slightly larger than the body text, to ensure prominence and visibility within the layout.

Participants were instructed to carefully read the news content, after which they completed a structured questionnaire measuring their personality traits (nationalism, conspiratorial orientation), perceived social norm, third-person perception, and intention to seek further information. Prior to exposure, all participants answered baseline demographic questions and control variables (e.g., prior knowledge of the topic).

Core Variable Measurement (Independent Variables)

This study's independent variables primarily include the fact-checker identity and two core personality traits: conspiratorial orientation and blind nationalism.

Fact-Checker Identity

This is the core independent variable of this study, operationalized as a dichotomous variable. For the purpose of this study, we primarily utilized data from the AI fact-checker group and the human fact-checker group. This variable was manipulated through distinct verification labels clearly displayed within the news articles, such as "This article has been verified by our AI system" or "This article has been verified by professional human fact-checkers". In data coding, the AI fact-checker group was coded as 0, and the human fact-checker group was coded as 1.

Conspiratorial Orientation (CO)

The measurement of this personality trait aligns with the conceptualization and operationalization proposed by Kim and Lee (2024). The scale comprises three dimensions (i.e., conspiratorial realism, susceptibility to popular folklore, and workplace conspiratorial realism), totaling 11 items. Each item was measured on a 7-point Likert scale (1 = absolutely disagree, 7 = absolutely agree). An illustrative item is: "Those people in power will use shadowy means to gain profit or advantage rather than lose it" ($M = 4.62$, $SD = 1.18$).

Blind Nationalism

This variable was measured based on the "nationalism" scale adapted from Kosterman and Feshbach (1989) which consists of four items. They were measured on a 5-point Likert scale (1 = absolutely disagree, 5 = absolutely agree), as asking for agreement on statements including "People should support their country even if the country is in the wrong", "I would rather be Chinese than a citizen of any other country in the world", "The world would be a better place if people from other countries were more like the Chinese" and "Overall, China is better than most other countries". directly reflecting a tendency towards unconditional and "blind" national pride and affiliation ($M = 4.44$, $SD = 0.65$).

Endogenous Variable

The endogenous variables in this study include third-person perception and future information-seeking intention, incorporating perceived social norms as a related or mediating variable.

Perceived Social Norm of Foreign Risk Misinformation

These variable measures participants' perceived subjective norms regarding whether they should stay informed about misinformation related to foreign risks. Following the RISP framework (Kahlor, 2007), we operationalized norms using injunctive items that capture perceived social approval and expectations. Four items were rated on a 5-point Likert scale (1 = Strongly disagree, 5 = Strongly agree), including "People whose opinions I value would approve of my staying on top of information about Boeing's potential role in Flight 5735 crash" and "It is expected of me that I seek information about Boeing's potential role in Flight 5735 crash" ($M = 3.65$, $SD = 0.81$). Although social norms can include descriptive elements, our focus on injunctive norms aligns with established RISP operationalization and the construct's theorized role as a motivational antecedent of information-seeking intention.

Third-Person Perception (TPP)

This serves as a core mediating variable in our study. TPP quantifies an individual's tendency to perceive media messages as having a lesser impact on themselves compared to others, particularly those with differing viewpoints. It is measured by comparing two components:

Perceived News Effect on Self: Comprises 1 item, measured on a 5-point Likert scale (from "None" to "A Great Deal"). An example is: "How much influence do you think the news article has on yourself?" ($M = 3.19$, $SD = 1.09$).

Perceived News Effect on Others: Comprises 2 items, also measured on a 5-point Likert scale (from "None" to "A Great Deal"). Examples include: "How much influence do you think the news article has on your friends?" and "How much influence do you think the news article has on other people?" ($M = 3.64$, $SD = 0.92$).

Future Information Seeking Intention

This variable assesses participants' willingness to actively seek additional information after reading the news article. It includes 3 items, measured on a 5-point Likert scale (from "Strongly disagree" to "Strongly agree"). Illustrative items are: "I plan to seek more information about it in the future," "I intend to find out more about it," and "In the future, I will try to seek as much information as I can about it" ($M = 3.38$, $SD = 1.03$).

Reliability and Validity Assessment

All multi-item scales demonstrated good internal consistency reliability (see Table 1). The conspiratorial orientation scale yielded a Cronbach's α of 0.87, composite reliability CR = 0.88, and average variance extracted AVE = 0.52; the nationalism scale showed $\alpha = 0.85$, CR = 0.86, and AVE = 0.55; the perceived social norms scale had $\alpha = 0.81$, CR = 0.83, and AVE = 0.51; the third person perspective scale had $\alpha = 0.84$, CR = 0.86, and AVE = 0.63 and the information seeking intention scale had $\alpha = 0.89$, CR = 0.90, and AVE = 0.63. All AVE values exceeded 0.50 and $CR > 0.70$, indicating adequate convergent validity for each construct.

Table 1- Inter-Construct Correlation Matrix

Construct	NAT	CO	SOCIAL NORM	TPP	INFOR SEEKING
NAT	$\alpha = .85$ CR = .86 AVE = .55	-.018	.294***	-.098*	.291***
CO		$\alpha = .87$ CR = .88 AVE = .52	.121*	-.058	.079
SOCIAL NORM			$\alpha = .81$ CR = .83 AVE = .51	-.234**	.583***
TPP				$\alpha = .84$ CR = .86 AVE = .63	-.216**
INFOR SEEKING					$\alpha = .89$ CR = .90 AVE = .63

*Note : Values in parentheses are Cronbach's α reliability coefficients; * $p < .05$, ** $p < .01$.*

Analytical Strategy

Given the complexity of the theoretical framework—encompassing core personality traits, psychological mechanisms (specifically the third-person effect), and behavioral intentions—and the need to compare paths across different fact-checker identities (AI vs. human), the core analytical approach of this study will utilize path analysis. Data will be processed and analyzed using Multi-Group Analysis in IBM SPSS AMOS software. This method enables simultaneous examination of complex direct and indirect effects and facilitates the comparison of model parameters across groups, thereby effectively addressing the central research question: how do different types of information verification interfaces change the pathways through which personality traits influence the third-person effect and information-related behaviors?

To assess whether the identity of the fact-checker (AI vs. human) changes these pathways, this study will compare the strength and significance of key paths between groups—such as the influence of blind nationalism on perceived social norms, and the effect of conspiratorial thinking on both third-person perception and information-seeking behavior. By examining the statistical significance and magnitude of these path coefficients, we aim to compare how these influences change as the identity of the fact-checker shifts.

Moreover, the study will investigate the mediating role of the third-person effect in the relationship between personality traits and information-seeking intentions, employing the Bootstrap method to test the significance of both direct and indirect effects. Demographic variables—including age, gender, education level, and household income—as well as participants' prior knowledge, will be included as covariates to control for their potential influence on endogenous variables. Blind nationalism and conspiratorial thinking, as key personality constructs, will be central to the model.

Results

Structural Equation Model Fit Results

The structural equation analysis of the proposed model indicated that the default model yielded a chi-square value of 8.689 ($df = 4$, $p = .013$), with a chi-square/degrees of freedom ratio of 4.344, slightly above conventional recommendations but still within an acceptable range. The root mean square residual ($RMR = .034$) and overall fit indices ($GFI = .993$, $AGFI = .899$) suggest good correspondence between the model and the observed data.

The human-checker structural model showed excellent fit to the data: $\chi^2(4) = 6.389$, $p = 0.173$ ($CMIN/DF = 1.60$), $RMSEA = 0.036$ (90% CI [0.000, 0.089], $PCLOSE = 0.72$), $SRMR = 0.022$, $GFI = 0.997$, $AGFI = 0.991$, $NFI = 0.993$, $IFI = 0.995$, $CFI = 0.995$, $PNFI = 0.782$, $PCFI = 0.796$, and $RMR = 0.034$. Hoelter's critical N was 523 at $\alpha = 0.05$ and 749 at $\alpha = 0.01$. All indices exceed conventional thresholds (e.g. $RMSEA < 0.08$, $CFI \approx 1.0$, $SRMR < 0.08$), indicating a very good model fit.

The AI-checker model also fit well by standard criteria. For example, $\chi^2(4) = 8.51$ ($p = .074$; CMIN/DF = 2.13), RMSEA = 0.058 (90% CI [.000, .127], PCLOSE = .22), SRMR = 0.025, GFI = 0.991, AGFI = 0.904, NFI = 0.969, IFI = 0.975, CFI = 0.970, PNFI = 0.720, PCFI = 0.735, RMR = 0.035. Hoelter's $N = 505$ at $\alpha = 0.05$ and 685 at $\alpha = 0.01$. These values meet conventional fit criteria ($CFI \geq .97$, $RMSEA < .08$, etc.), indicating acceptable to good fit.

Overall, the structural equation model demonstrates adequate absolute fit, incremental fit, and parsimony, providing a reliable basis for subsequent path analyses and hypothesis testing.

Test of the Main Model

Path analysis was conducted to examine the hypothesized main model (see Figure 1). Maximum likelihood estimation with robust standard errors and 5,000 bootstrap resamples was employed to account for potential violations of the normality assumption, ensuring that the parameter estimates were not biased by non-normal distributions of the observed variables (Finney & DiStefano, 2006). Error variance for each observed variable was specified based on its sample variance and reliability estimate, following established procedures to control for measurement error (Schumacker & Lomax, 2010). Participants' demographic variables, including age and gender, as well as social desirability and prior exposure to foreign-risk information, were included as covariates for all paths. Correlations among key variables are summarized in Table 1. The model demonstrated acceptable fit across indices, with $\chi^2(4) = 6.389$, $p = .173$, CMIN/DF = 1.597, RMSEA = 0.036, 90% CI [.000, .089], SRMR = 0.022, CFI = 0.995, and TLI = 0.978, indicating that the hypothesized structure adequately represented the observed data.

In the human-checker condition, nationalism (NAT) was a significant predictor of perceived social norms of foreign-risk misinformation (SocNorm), with a standardized coefficient of $b = 0.465$, $SE = 0.084$, $p < .001$, 95% CI [0.299, 0.631], thereby providing strong support for H1. This finding aligns with theoretical expectations that heightened nationalistic sentiments, particularly blind nationalism, increase individuals' alignment with and perception of socially endorsed misinformation concerning foreign risks. In contrast, conspiratorial orientation (CO) did not significantly predict SocNorm in this group, $b = 0.035$, $SE = 0.035$, $p = .315$, 95% CI [-0.034, 0.104], indicating that the hypothesized effect of a general conspiratorial mindset on social norm perception (H2) was not supported. This suggests that, while conspiratorial tendencies may shape broader information processing, they do not directly translate into heightened perceptions of normative acceptance for misinformation when evaluated in the context of human-mediated fact-checking.

Perceived social norms (SocNorm) negatively influenced third-person perception (TPP) in the human-checker condition, $b = -0.244$, $SE = 0.083$, $p = .003$, 95% CI [-0.406, -0.082], confirming H3. Participants who believed that others endorsed misinformation were less likely to perceive themselves as uniquely susceptible, consistent with established social-

cognitive mechanisms of third-person perception (Davison, 1983; Gunther, 1992). This effect highlights the psychological process whereby normative beliefs about misinformation create a perceived social consensus that paradoxically decreases individuals' estimation of their own susceptibility. Third-person perception in turn negatively predicted future information seeking intentions regarding foreign risks (InfoSeek), $b = -0.211$, $SE = 0.048$, $p < .001$, 95% CI $[-0.307, -0.115]$, confirming H4. These results collectively suggest a mediational pathway in which nationalism increases perceived social norms, which then influence TPP and ultimately shape information-seeking behaviors.

The direct effect of nationalism on information seeking intention was also significant, $b = 0.495$, $SE = 0.088$, $p < .001$, 95% CI $[0.323, 0.667]$. The total effect of nationalism on information seeking, integrating both direct and indirect paths through SocNorm and TPP, was positive and significant, $b = 0.519$, $SE = 0.099$, 95% CI $[0.324, 0.711]$, indicating that participants with stronger nationalistic sentiments exhibited increased intention to further seek information about foreign risks, despite the mitigating influence of third-person perception. In contrast, CO did not show significant direct or indirect effects on information seeking, emphasizing the specific salience of nationalism over generalized conspiratorial orientation in shaping normative and behavioral responses in this context.

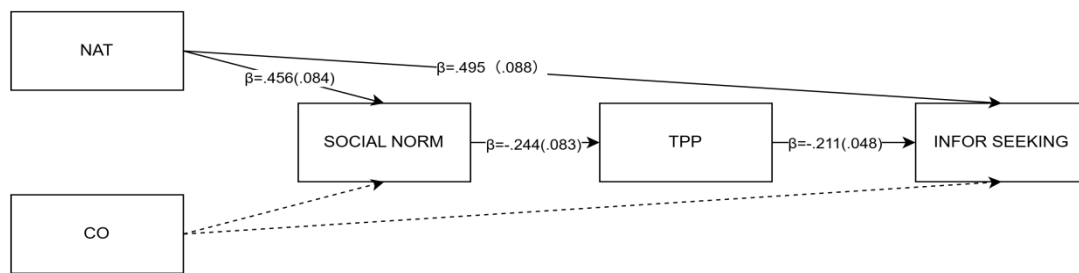


Figure 2 - Model results (Human-checker condition)

In the AI-checker condition, similar patterns emerged. Nationalism remained a significant predictor of perceived social norms, $b = 0.371$, $SE = 0.085$, $p < .001$, 95% CI $[0.192, 0.537]$, providing further support for H1 across both checker types. Again, CO did not significantly affect SocNorm, $b = 0.028$, $SE = 0.035$, $p = .419$, 95% CI $[-0.050, 0.117]$, suggesting that conspiratorial orientation has limited influence when information is evaluated through AI-based verification. SocNorm negatively predicted TPP, $b = -0.364$, $SE = 0.077$, $p < .001$, 95% CI $[-0.516, -0.209]$, supporting H3, and TPP negatively predicted InfoSeek, $b = -0.208$, $SE = 0.051$, $p < .001$, 95% CI $[-0.326, -0.094]$, supporting H4. The direct effect of nationalism on information seeking intention was significant, $b = 0.30$, $SE = 0.073$, $p < .001$, 95% CI $[0.16, 0.44]$, indicating that even after accounting for the indirect pathway through social norms and third-person perception, stronger nationalistic sentiments independently predicted greater intention to seek further information about foreign risks. The total effect of nationalism on information seeking intention in the AI-checker group was $b = 0.329$, $SE = 0.098$, 95% CI $[0.101, 0.521]$, lower than in the human-checker group, which suggests that

AI-mediated fact-checking may attenuate, but not eliminate, the influence of nationalist sentiments on subsequent information-seeking behaviors.

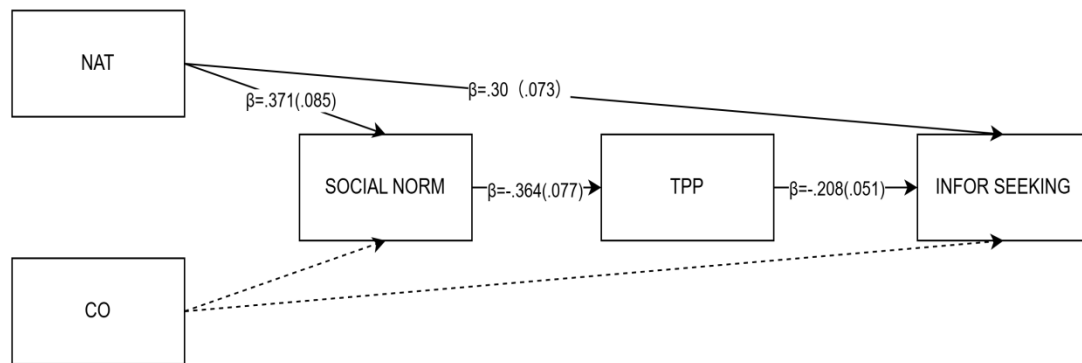


Figure 3 - Model Results (AI-checker condition)

The model explained 19.3% of the variance in SocNorm, 11.2% in TPP, and 28.6% in InfoSeek for the human-checker group, and 16.7%, 14.5%, and 24.1%, respectively, for the AI-checker group. The consistent support for H1, H3, and H4 across both checker types, combined with the lack of support for H2, indicates that blind nationalism reliably elevates perceived social norms regarding foreign-risk misinformation, which subsequently shapes third-person perception and information-seeking intentions. The stronger indirect effect of social norm on TPP observed in the AI-checker condition suggests that dehumanized verification sources may heighten reliance on perceived social consensus, potentially altering how normative beliefs translate into personal susceptibility judgments and behavior. These results carry important implications for fact-checking strategies and policy interventions. In environments characterized by rising nationalism, particularly blind nationalism, fact-checking efforts may face heightened challenges, as individuals are more likely to seek further information despite corrective messages. Multinational organizations and policymakers should consider the influence of socio-political attitudes on misinformation dynamics and anticipate that normative and perceptual mechanisms can amplify attention to foreign-risk misinformation.

Table 2 - Summary of Hypothesis Test Results Across Checker Conditions

Hypothesis	Path	Human Condition (β)	AI Condition (β)	p
H1	Nationalism → SocNorm	0.465	0.371	p < .001
H2	Conspiratorial Orientation → SocNorm	0.035 (ns)	0.028 (ns)	p > .3
H3	SocNorm → TPP	-0.244	-0.364	p < .01
H4	TPP → InfoSeek	-0.211	-0.208	p < .001
—	Nationalism → InfoSeek (Direct)	0.495	0.300	p < .001

Note: ns = not significant; all coefficients are standardized.

Overall, the data provide robust evidence for the mediational pathway connecting nationalism to information seeking through perceived social norms and third-person perception, highlighting the nuanced psychological processes that underlie engagement with foreign-risk misinformation. The divergence in path strengths between human- and AI-checked misinformation underscores the importance of the checker type in conditioning these effects, pointing to future research directions on technological mediation in misinformation correction.

Discussion

The present study set out to investigate the mechanisms through which nationalism and conspiratorial orientation influence perceptions of foreign-risk misinformation and subsequent information behaviors, with particular attention to whether these mechanisms differ when misinformation is fact-checked by AI versus human checkers. Through multi-group path analysis, the study examined hypothesized relationships among nationalism (NAT), conspiratorial orientation (CO), perceived social norm of misinformation (SocNorm), third-person perception (TPP), and information-seeking intention (InfoSeek). The findings not only provide support for the hypothesized paths (H1 - H4) but also yield important insights into the conditioning role of fact-checker identity.

Summary of Findings

H1 predicted that nationalism would positively predict perceived social norms of foreign-risk misinformation. This hypothesis was supported in the human-checker condition, where nationalism significantly increased SocNorm, suggesting that stronger nationalistic sentiment fosters a heightened perception that foreign-risk misinformation is widely accepted by others. This resonates with prior research showing that nationalist orientations amplify in-group bias and collective endorsement. Under AI the effect was smaller but still significant, and importantly nationalism also directly predicted information-seeking in both groups.

H2 was not supported. Coefficients were near zero in both conditions ($\beta \approx 0.03$, ns), contrary to expectations. Thus, conspiratorial worldviews did not significantly influence perceived norms of misinformation, even under human checking. One plausible explanation is that conspiratorial worldviews rely heavily on interpersonal and socially grounded cues to judge consensus. When the corrective message is framed as algorithmic, individuals may perceive it as detached from human communicative contexts, thereby weakening its normative salience. Another possibility is that AI corrections are interpreted as “technical” rather than “social,” which blunts the role of conspiratorial predispositions. This asymmetry suggests that conspiracy thinking does not uniformly shape normative judgments across communicative environments.

H3 proposed that perceived social norms would negatively predict TPP. Supporting the third-person perception framework (Davison, 1983), results showed that higher levels of perceived consensus were associated with a smaller self-other discrepancy, thus lowering TPP. The effect was stronger under AI ($\beta \approx -0.364$) than human ($\beta \approx -0.244$).

H4 predicted that TPP would negatively predict information-seeking intention. This hypothesis was supported: lower levels of TPP were linked to stronger motivation to seek additional information. The finding highlights TPP not merely as a perceptual bias but as a psychologically consequential mechanism that translates into active behaviors.

Overall, human checkers triggered a normative path driven by nationalism (rather than conspiratorial orientation), elevating social norms, reducing third-person perception, and ultimately promoting higher information-seeking intention. By contrast, AI checkers elicited a more direct route from nationalism to information-seeking. Fact-checker identity thus shifts whether individuals rely on social norms or heuristic shortcuts in processing misinformation.

Theoretical Implications

These findings contribute to several strands of communication and media psychology literature.

First, they refine understanding of nationalism and misinformation processing. Prior research suggests nationalism heightens in-group bias and fosters resistance to threatening information (Mummendey et al., 2001). This study shows that under human correction, nationalism exerts influence through perceived consensus, consistent with the idea that nationalistic predispositions are socially embedded. Under AI correction, however, nationalism motivated direct information-seeking, suggesting that in algorithmic contexts individuals no longer rely on consensus judgments but instead turn their predispositions into behavioral engagement. This conditional role of nationalism underscores the importance of communicative context in shaping how predispositions manifest.

Second, The null effect of conspiracy worldview on normative perceptions challenges models that treat conspiracy orientation as a uniform lens on social reality. Conspiracy-minded individuals often distrust authorities and peers alike, which may blunt any single social cue's influence. Interestingly, our findings - together with recent research - imply that an AI label may even neutralize conspiracists' bias against correction. Lan et al. (2025) showed that since conspiracists view humans as suspect, they can paradoxically place more trust in an AI fact-checker ("positive machine heuristic"), leading them to be more willing to accept the debunking. Thus, under an AI verifier, high-CO participants did not need to rely on perceived consensus at all; their conspiratorial skepticism was effectively "turned off" by the impersonal label, which made them more open to fact-checking. This suggests new theory: automation cues can override individual differences in skepticism.

Third, the study advances research on third-person perception. The negative effect of SocNorm on TPP aligns with the proposition that prevalence perceptions constrain self-

other differentials. Importantly, the study demonstrates that TPP shapes risk-aversion behaviors (Zhu et al., 2021), such as risk information-seeking, reinforcing its value as a mediator linking cognitive perceptions to behavioral outcomes.

Finally, the findings expand theorization of fact-checking as a communicative act. Previous research has largely focused on AI vs. human agency's effect on perceptual and behavioral outcomes (DeVerna et al., 2024). This study highlights such agency's impact on the psychometric mechanism of risk information seeking. The results suggest that corrections should not be understood solely in terms of their informational content but also in terms of their relational positioning as human or algorithmic communicators. As recent work on AI communication cautions (Lovari & De Rosa, 2025), the rise of generative AI complicates traditional models of trust, credibility, and relationality, making fact-checker identity a central concern for future theory.

It is noteworthy that while we interpret the AI fact-checker effect as invoking a more heuristic cue-based route (a 'shortcut' as described earlier), we did not directly measure participants' depth of processing or need for cognition in this experiment. The absence of such measures means we infer the cognitive pathway indirectly from the pattern of results. It is possible that both AI and human fact-check labels primarily engaged heuristic processing, but via different cues - e.g., human fact-checks triggered a social consensus heuristic (relying on normative perceptions), whereas AI fact-checks triggered a machine credibility heuristic (relying on the perceived objectivity of technology). This would explain why we did not observe any overt difference in deliberative cognitive effort: participants in both conditions may not have engaged in extensive systematic scrutiny, instead trusting the cue (human or AI) presented. Prior research on source cues supports this notion that fact-check flags often work through peripheral credibility cues rather than deep analytical processing (Sundar, 2008; Metzger & Flanagin, 2013). Future studies should incorporate direct measures of cognitive elaboration - for instance, need for cognition scales or thought-listing tasks - to verify whether an AI fact-check truly reduces analytic processing or if audiences simply trust it as an authoritative cue. Our findings suggest a scenario of heuristic acceptance (relying on consensus or automation cues) in both conditions, which is an important nuance: the AI checker reinforced reliance on a credibility heuristic without necessarily increasing critical thinking effort. A deeper probe into users' cognitive engagement can illuminate whether AI-based corrections risk encouraging passive consensus-taking over active inquiry.

Practical Implications

The findings carry several implications for policymakers, platforms, and organizations.

For fact-checking organizations, the study suggests that human and AI checkers are not interchangeable. AI-driven corrections may encourage nationalistic audiences to actively seek more information, which could aid vigilance but also risk amplifying the salience of

contested issues. Human-led corrections, meanwhile, are more deeply entangled with social norms: while they may enhance perceived legitimacy, they also allow predispositions such as nationalism and conspiracy thinking to shape perceptions of consensus, potentially undermining corrective intent.

For governments and communication professionals, the results underscore the importance of embedding fact-checking in strategies that emphasize transparency, ethical responsibility, and dialogic engagement. As Lovari (2020) reminds us, government interventions in misinformation crises must balance authority with public trust, while Lovari and Bowen (2020) stress the ethical imperative of two-way communication and community participation in disaster contexts. In the case of foreign-risk misinformation, this means that fact-checking cannot be treated merely as a technical add-on but must be integrated into broader communication strategies that respect citizens' trust and moral expectations.

For multinational corporations, the evidence highlights the risks of over-relying on one mode of correction. In nationalist contexts, AI corrections may inadvertently fuel further attention to sensitive issues, while in environments where conspiracy thinking is widespread, human corrections may unintentionally reinforce perceived consensus around falsehoods. Tailoring the identity of the fact-checker to the audience's socio-political context is thus a key strategic choice.

Given the above context, our results yield actionable insights for implementing fact-checking interventions in China. First, timing is crucial. In a fast-moving rumor cycle about foreign threats, deploying AI-driven fact-checks early - but not so early as to seem irrelevant - can leverage the speed of AI to correct falsehoods before they fully take root. Chinese platforms might integrate AI fact-check labels as soon as a misleading foreign-risk story starts trending, to intercept misinformation while public interest is high. However, care must be taken that such early intervention does not inadvertently pique curiosity in fringe claims; monitoring user reactions can inform the optimal timing.

Second, message design should be culturally resonant. Fact-check messages verified by AI should be framed in a way that respects Chinese audience values. For example, emphasizing that a claim has been "AI-verified for accuracy to protect the public interest" can appeal to collective welfare. Tying accuracy to patriotic duty could also be effective - recent research shows that nationalist framing can motivate public-minded behaviors in China (e.g., wearing masks during COVID-19). Thus, positioning fact-checks as part of safeguarding national stability and truth may enhance acceptance.

Third, implement safeguards to maintain trust in AI fact-checking. This includes combining AI with human oversight: an ideal system might use AI to flag and correct most claims, with human editors auditing a sample for quality and handling ambiguous cases. Explicitly communicating this hybrid approach (for instance, a note that "AI findings are reviewed by independent experts") can strengthen credibility and alleviate concerns about algorithmic errors. Similarly, providing users with transparency - e.g. brief explanations or source links showing why the AI marked something as false - will help prevent misunderstandings and bolster trust.

In practice, Chinese platforms could pilot these strategies during periods of heightened foreign-related misinformation (for example, around major geopolitical events), gathering feedback to refine the approach. By thoughtfully tuning the timing, framing, and oversight of AI fact-checkers to China's media culture, stakeholders can harness the efficiency of AI while reinforcing public trust and minimizing unintended effects. These recommendations serve as a roadmap for applying our findings in real-world Chinese contexts, ensuring that the promise of AI fact-checking is realized in a safe and culturally attuned manner.

Limitations and Future Research

Several limitations merit consideration. First, the data were cross-sectional, limiting causal inference. Future research should adopt longitudinal or experimental designs to establish temporal order. Additionally, because all measures were collected in a single survey at one point in time, the study is susceptible to common-method bias (Podsakoff et al., 2003). The use of self-report questionnaires for both predictors and outcomes can inflate observed relationships due to shared method variance. We did take procedural steps to mitigate this (e.g., assuring respondents of anonymity and varying question order), but we cannot fully rule out this bias. The cross-sectional nature of the data also limits causal inference - while our structural model is theoretically grounded, longitudinal or experimental research is needed to verify the temporal ordering of effects. In future studies, adopting multi-wave designs or incorporating objective behavioral measures would help address these common-method and causality concerns. Second, the study focused on foreign-risk misinformation, which may not generalize to other misinformation domains (e.g., health, politics) or socio-political and cultural contexts. Third, perceptions of AI versus human fact-checkers were based on manipulated framing rather than actual exposure to different correction interfaces. Real-world experiments using platform-integrated AI or human corrections could yield more ecologically valid insights. Fourth, the current model focused on a limited set of psychological mechanisms (NAT, CO, SocNorm, TPP). Additional factors such as trust in institutions, media literacy, or affective polarization may further moderate the observed relationships.

Conclusion

In conclusion, the present study demonstrates that the perceived identity of the fact-checker - AI versus human - meaningfully alters the mechanisms through which predispositions shape perceptions and behaviors in response to foreign-risk misinformation. Human checkers activate elaborate normative pathways where nationalism shapes perceptions of consensus and susceptibility, whereas AI checkers trigger heuristic shortcuts whereby nationalism directly drives further engagement. These findings enrich theoretical understanding of misinformation processing and offer concrete guidance for practitioners

grappling with the challenges of misinformation in an era where both human and algorithmic corrections coexist.

Biographical note

Qi Chen is a master's student in Quantitative Analysis for the Social Sciences at KU Leuven, Belgium. His research interests include media effects, media psychology, and human–AI interaction.

Yicheng Zhu (Ph. D. University of South Carolina) is associate professor at School of Journalism and Communication, Beijing Normal University. His research focuses on transnational communication, international public relations and media effects.

Bibliography

- Al Jazeera. (2025, November 19). "Trust in AI far higher in China than West, poll shows". Al Jazeera. <https://www.aljazeera.com/economy/2025/11/19/trust-in-ai-far-higher-in-china-than-west-poll-shows>
- Armaly, M. T., Buckley, D. T., & Enders, A. M. (2022). Christian Nationalism and Political Violence: Victimhood, Racial Identity, Conspiracy, and Support for the Capitol Attacks. *Political Behavior*, 44(2), 937–960. <https://doi.org/10.1007/s11109-021-09758-y>
- Ashley, S., Craft, S., Maksl, A., Tully, M., & Vraga, E. K. (2023). Can news literacy help reduce belief in COVID misinformation? *Mass Communication and Society*, 26(4), 695–719. <https://doi.org/10.1080/15205436.2022.2137040>
- Billings, A. C., Brown, N. A., Brown, K. A., Guoqing, Leeman, M. A., Ličen, S., Novak, D. R., & Rowe, D. (2013). From Pride to Smugness and the Nationalism Between: Olympic Media Consumption Effects on Nationalism Across the Globe. *Mass Communication and Society*, 16(6), 910–932. <https://doi.org/10.1080/15205436.2013.822519>
- Bryanov, K., Kliegl, R., Koltsova, O., Miltsov, A., Pashakhin, S., Porshnev, A., Sinyavskaya, Y., Terpilovskii, M., & Vziatysheva, V. (2023). What drives perceptions of foreign news coverage credibility? A cross-national experiment including Kazakhstan, Russia, and Ukraine. *Political Communication*, 40(2), 115–146. <https://doi.org/10.1080/10584609.2023.2172492>
- Chung, M., & Kim, N. (2021). When I learn the news is false: How fact-checking information stems the spread of fake news via third-person perception. *Human Communication Research*, 47(1), 1–24. <https://doi.org/10.1093/hcr/hqaa010>
- Corbu, N., Halagiera, D., Jin, S., Stanyer, J., Strömbäck, J., Matthes, J., Aalberg, T. (2025). Illusory superiority about misinformation detection and its relationship to knowledge and fact-checking intentions: Evidence from 18 countries. *Mass Communication and Society*. Advance online publication. <https://doi.org/10.1080/15205436.2025.2495206>

- Das, R., & Ahmed, W. (2022). Rethinking fake news, disinformation, and ideology during the time of COVID-19 global pandemic. *IIM Kozhikode Society & Management Review*, 11(1), 146–159. <https://doi.org/10.1177/22779752211027382>
- Daunt, K. L., Greer, D. A., Jin, H. S., & Orpen, I. (2023). Who believes political fake news? The role of conspiracy mentality, patriotism, perceived threat to freedom, media literacy and concern for disinformation. *Internet Research*, 33(5), 1849–1870. <https://doi.org/10.1108/INTR-07-2022-0565>
- Davison, W. P. (1983). The third-person effect in communication. *Public Opinion Quarterly*, 47(1), 1–15. <https://doi.org/10.1086/268763>
- Davison, W. P. (1996). The third-person effect revisited. *International Journal of Public Opinion Research*, 8(2), 113–119.
- DeVerna, M. R., Yan, H. Y., Yang, K.-C., & Menczer, F. (2024). Fact-checking information from large language models can decrease headline discernment. *Proceedings of the National Academy of Sciences*, 121(50), e2322823121.
- Enders, A. M., et al. (2019). Who are conspiracy theorists? A comprehensive approach to explaining conspiracy. *Social Science Quarterly*, 100(6), 2019–2032. <https://doi.org/10.1111/ssqu.12711>
- Finney, S. J., & DiStefano, C. (2006). Nonnormal and categorical data in structural equation modeling. In G. R. Hancock & R. O. Mueller (Eds.), *Structural Equation Modeling: A Second Course* (pp. 269–314). Information Age Publishing.
- Flew, T., & Iosifidis, P. (2020). Populism, globalisation and social media. *International Communication Gazette*, 82(1), 7–25. <https://doi.org/10.1177/1748048519880721>
- Griffin, R. J., Dunwoody, S., & Neuwirth, K. (1999). Proposed model of the relationship of risk information seeking and processing to the development of preventive behaviors. *Environmental Research*, 80(2), S230–S245. <https://doi.org/10.1006/enrs.1998.3940>
- Guan, S. (2020). Diversifying conspiracy beliefs and populist ideologies in the Chinese context. *Social Science Quarterly*, 101(2), 461–472. <https://doi.org/10.1111/ssqu.12766>
- Gutierrez-Huerter O, G. (2024). Responding to institutional plurality: Micro-politics in the rollout of a global corporate social responsibility norm in a multi-national enterprise. *Human Relations*, 77(7), 1037–1068. <https://doi.org/10.1177/00187267231176671>
- Gunther, A. C. (1992). Biased press and public perception: The role of perceived media influence in the third-person effect. *Journal of Communication*, 42(2), 17–31. <https://doi.org/10.1111/j.1460-2466.1992.tb00739.x>
- Jaspal, R., & Cinnirella, M. (2010). Media representations of British Muslims and hybridised threats to identity. *Contemporary Islam*, 4(3), 289–310. <https://doi.org/10.1007/s11562-010-0126-7>
- Jia, H., & Luo, X. (2023). I Wear a Mask for My Country: Conspiracy Theories, Nationalism, and Intention to Adopt Covid-19 Prevention Behaviors at the Later Stage of Pandemic Control in China. *Health Communication*, 38(3), 543–551. <https://doi.org/10.1080/10410236.2021.1958982>

- Kahan, D. M. (2016). The politically motivated reasoning paradigm, part 1: What politically motivated reasoning is and how to measure it. In R. Scott & S. Kosslyn (Eds.), *Emerging Trends in the Social and Behavioral Sciences: An Interdisciplinary, Searchable, and Linkable Resource* (pp. 1 - 16). Wiley. <https://doi.org/10.1002/9781118900772.etrds0417>
- Kahan, D. M. (2017). 'Ordinary science intelligence' : A science-comprehension measure for study of risk and science communication, with notes on evolution and climate change. *Journal of Risk Research*, 20(8), 995 - 1016. <https://doi.org/10.1080/13669877.2016.1148067>
- Kahan, D. M., Peters, E., Wittlin, M., Slovic, P., Ouellette, L. L., Braman, D., & Mandel, G. (2012). The polarizing impact of science literacy and numeracy on perceived climate change risks. *Nature Climate Change*, 2(10), 732 - 735. <https://doi.org/10.1038/nclimate1547>
- Kahlor, L. A. (2007). An Augmented Risk Information Seeking Model: The Case of Global Warming. *Media Psychology*, 10(3), 414 - 435. <https://doi.org/10.1080/15213260701532971>
- Kim, J.-N., & Lee, S. (2024). Conceptualizing Conspiratorial Thinking: Explicating Public Conspiracism for Effective Debiasing Strategy. *American Behavioral Scientist*, 68(10), 1366-1394. <https://doi.org/10.1177/00027642231175637>
- Koo, A. Z. X., Su, M. H., Lee, S., Ahn, S. Y., & Rojas, H. (2021). What motivates people to correct misinformation? Examining the effects of third-person perceptions and perceived norms. *Journal of Broadcasting & Electronic Media*, 65(1), 111 - 134. <https://doi.org/10.1080/08838151.2021.1903896>
- Kosterman, R., & Feshbach, S. (1989). Toward a measure of patriotic and nationalistic attitudes. *Political Psychology*, 10(2), 257-274. <https://doi.org/10.2307/3791647>
- Kunda, Z. (1990). The case for motivated reasoning. *Psychological Bulletin*, 108(3), 480-498. <https://doi.org/10.1037/0033-2909.108.3.480>
- Kragh, M., Andermo, E., & Makashova, L. (2022). Conspiracy theories in Russian security thinking. *Journal of Strategic Studies*, 45(3), 334 - 368. <https://doi.org/10.1080/01402390.2020.1717954>
- Lan, D., Zhu, Y., Liu, M., & He, C. (2025). AI Agency in Fact-Checking: Role-Based Machine Heuristics and Publics' Conspiratorial Orientation. *Media and Communication*, 13. <https://www.ssoar.info/ssoar/handle/document/102711>
- Lovari, A. (2020). Spreading (Dis)Trust: Covid-19 misinformation and government intervention in Italy. *Media and Communication*, 8(2), 458 - 461. <https://doi.org/10.17645/mac.v8i2.3219>
- Lovari, A., & Bowen, S. A. (2020). Social media in disaster communication: A case study of strategies, barriers, and ethical implications. *Journal of Public Affairs*, 20(3), e1967. <https://doi.org/10.1002/pa.1967>

- Lovari, A., & De Rosa, F. (2025). Exploring the challenges of generative AI on public sector communication in Europe. *Media and Communication*, 13. <https://doi.org/10.17645/mac.9644>
- Luzsa, R., & Mayr, S. (2021). False consensus in the echo chamber: Exposure to favorably biased social media news feeds leads to increased perception of public support for own opinions. *Cyberpsychology: Journal of Psychosocial Research on Cyberspace*, 15(1), Article 3. <https://doi.org/10.5817/CP2021-1-3>
- Malešević, S., Uzelac, G., Carol, S., & David, L. (2024). Plotting against our nation: COVID-19, nationalisms, and conspiracy theories in five European countries. *National Identities*, 26(2), 141-171. <https://doi.org/10.1080/14608944.2023.2288970>
- Martikainen, J., & Sakki, I. (2021). Boosting nationalism through COVID-19 images: Multimodal construction of the failure of the ‘dear enemy’ with COVID-19 in the national press. *Discourse & Communication*, 15(4), 388 - 414. <https://doi.org/10.1177/17504813211002039>
- Metzger, M. J., & Flanagin, A. J. (2013). Credibility and trust of information in online environments: The use of cognitive heuristics. *Journal of Pragmatics*, 59, 210-220. <https://doi.org/10.1016/j.pragma.2013.03.012>
- Mihelj, S., & Jiménez-Martínez, C. (2021). Digital nationalism: Understanding the role of digital media in the rise of ‘new’ nationalism. *Nations and Nationalism*, 27(2), 331-346. <https://doi.org/10.1111/nana.12685>
- Mir, A., & Siddiqui, N. (2024). Nationalism, status, and conspiracy theories: Evidence from Pakistan. *British Journal of Political Science*, 54(4), 1159 - 1179. <https://doi.org/10.1017/S0007123424000140>
- Mummendey, A., Klink, A., & Brown, R. (2001). Nationalism and patriotism: National identification and out - group rejection. *British Journal of Social Psychology*, 40(2), 159-172. <https://doi.org/10.1348/014466601164740>
- Oktavianus, J., & Bautista, J. R. (2023). Motivating healthcare professionals to correct online health misinformation: The roles of subjective norm, third-person perception, and channel differences. *Computers in Human Behavior*, 147, 107839. <https://doi.org/10.1016/j.chb.2023.107839>
- Perloff, R. M. (2009). Mass media, social perception, and the third-person effect. In A. A. Raney & J. Bryant (Eds.), *Media effects: Advances in theory and research* (pp. 503-522). Routledge.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology*, 88(5), 879-903. <https://doi.org/10.1037/0021-9010.88.5.879>
- Posocco, L. & Watson, I. (2024). Conspiracy theories and nationalism. Exploring the intersections and implications. *Politica & Società*, 1, 89 - 122. <https://doi.org/10.4476/115027>

- Romer, D., & Jamieson, K. H. (2022). Conspiratorial thinking as a precursor to opposition to COVID-19 vaccination in the US: A multi-year study from 2018 to 2021. *Scientific Reports*, 12(1), 18632. <https://doi.org/10.1038/s41598-022-22870-y>
- Rowling, C. M., Gilmore, J., & Sheets, P. (2015). When Threats Come from Within: National Identity, Cascading Frames, and the U.S. War in Afghanistan. *The International Journal of Press/Politics*, 20(4), 478 - 497. <https://doi.org/10.1177/1940161215599384>
- Salvati, M., Pellegrini, V., De Cristofaro, V., & Giacomantonio, M. (2024). What is hiding behind the rainbow plot? The gender ideology and LGBTQ+ lobby conspiracies (GILC) scale. *British Journal of Social Psychology*, 63, 295 - 318. <https://doi.org/10.1111/bjso.12678>
- Semafor. (2025, November 18). "China's trust of AI is higher than the US". Semafor. <https://www.semafor.com/article/11/18/2025/chinas-trust-of-ai-is-higher-than-the-us>
- Shayegh, J., Storey, L., Turner, R. N., & Barry, J. (2022). A Social Identity Approach to How Elite Outgroups Are Invoked by Politicians and the Media in Nativist Populism. *Political Psychology*, 43(6), 1009-1025. <https://doi.org/10.1111/pops.12798>
- Schumacker, R. E., & Lomax, R. G. (2010). *A beginner's guide to structural equation modeling* (3rd ed). Routledge.
- Sun, B., Jin, L., Ren, Z., & Yue, G. et al. (2023). Is a punisher always trustworthy? In-group punishment reduces trust. *Current Psychology*, 42, 22965 - 22975. <https://doi.org/10.1007/s12144-022-03395-2>
- Sundar, S. S. (2008). *The MAIN model: A heuristic approach to understanding technology effects on credibility* (pp. 73-100). MacArthur Foundation Digital Media and Learning Initiative.
- Sutton, R. M., & Douglas, K. M. (2020). Conspiracy theories and the conspiracy mindset: Implications for political ideology. *Current Opinion in Behavioral Sciences*, 34, 118-122. <https://doi.org/10.1016/j.cobeha.2020.02.015>
- Ternullo, S. (2022). "I'm not sure what to believe": Media distrust and opinion formation during the COVID-19 pandemic. *American Political Science Review*, 116(3), 1096-1109. <https://doi.org/10.1017/S0003055422000334>
- Turner, M.M., Jang, Y., Wade, R. et al. (2024). The effects of moral norms and anticipated guilt on COVID19 prevention behaviors. *Current Psychology*, 43, 16767-16779 (2024). <https://doi.org/10.1007/s12144-023-04477-5>
- Van Prooijen, J.-W., Spadaro, G., & Wang, H. (2022). Suspicion of institutions: How distrust and conspiracy theories deteriorate social relationships. *Current Opinion in Psychology*, 43, 65-69.
- Whitmeyer, J. M. (2002). Elites and popular nationalism. *The British Journal of Sociology*, 53(3), 321-341. <https://doi.org/10.1080/0007131022000000536>
- Zhang, Z., Li, M., Liu, Q., Chen, C., & Qi, C. (2023). Group membership and adolescents' third-party punishment: A moderated chain mediation model. *Front. Psychol.*, 14, 1251276. <https://doi.org/10.3389/fpsyg.2023.1251276>

- Zhu, Y., Fitzpatrick, M. A., & Bowen, S. A. (2024). Factors Related to Compliance with CDC COVID-19 Guidelines: Media Use, Partisan Identity, Science Knowledge, and Risk Assessment. *Western Journal of Communication*, 88(3), 567 - 594. <https://doi.org/10.1080/10570314.2023.2219239>
- Zhu, Y., Wei, R., Lo, V.-H., Zhang, M., & Li, Z. (2021). Collectivism and Altruistic Behavior: A Third-Person Effect Study of COVID-19 News Among Wuhan Residents. *Global Media and China*, 6(4), 476-491.
- Zollo, F., Bessi, A., Del Vicario, M., Scala, A., Caldarelli, G., Shekhtman, L., Quattrociocchi, W. (2017). Debunking in a world of tribes. *PLoS ONE*, 12(7), e0181821. <https://doi.org/10.1371/journal.pone.0181821>

Responding to Online Hoaxes: The Role of Contextual Priming, Crisis Narrative Type and Communication Strategy*

Anli Xiao**
University of South Carolina

Yang Cheng***
North Carolina State University

This study investigates the interaction between contextual priming, crisis narrative type and crisis communication strategy in the context of a geopolitical conflict. Specifically, it examines how an international company, when its country of origin is being caught in geopolitical tension and being portrayed positively vs. negatively by the media, should respond to misinformation paracrisis on social media using different crisis narrative types and crisis communication strategies. An online experiment (N = 506) found that a narrative response might be more effective; people's judgment of the crisis response is partially influenced by the contextual priming of how the company's country of origin is portrayed. This study enriched communication literature by examining strategic crisis narratives in the context of geopolitical conflicts, and exploring the roles of different response types, response strategies and contextual priming into crisis situations. This study yields important implications on how international organizations should respond to threats amid geopolitical uncertainties. Media scanning may be critical when geopolitical tensions are present. International organizations may benefit from adopting informative, accommodative responses rather than informative, defensive responses when handling misinformation-driven paracrisis. Alternatively, a narrative response might work better than an informative response.

Keywords: contextual priming, crisis communication, narrative message, social media hoax, international organizations

*Article submitted on 22/09/2025. Article accepted on 15/12/2025.

** axiao@mailbox.sc.edu

*** ycheng20@ncsu.edu

Introduction

Hoaxes are fabricated acts (Merriam-Webster, 2018) that present detrimental threats to organizations. Those “deceptive alerts designed to undermine the public’s confidence in an organization, product, service or person” (Veil *et al.*, 2012, p. 328) have been increasingly common in the social media space, widely adopted by groups or individuals to gain engagement and attention (Park and Rim, 2019; Veil *et al.*, 2015). Social media hoaxes can be understood as a form of paracrisis—publicly visible reputational threats that emerge rapidly online and do not necessarily involve confirmed organizational wrongdoing (Coombs, 2019; Coombs & Holladay, 2012). In this study, paracrisis is treated as a contextual condition rather than a distinct crisis category, characterized by low responsibility attribution, high uncertainty, and rapid amplification through networked media. As a result, viral hoaxes can still cause substantial reputational harm (Veil *et al.*, 2012), as publics may not immediately recognize the lack of veracity in unsubstantiated claims. Without timely and strategically appropriate responses, organizations may experience financial loss and erosion of brand reputation, legitimacy, and trust (Veil *et al.*, 2012). Strategically responding to social media hoaxes therefore remains a critical challenge for organizations, particularly international ones.

Hoaxes and organizational crisis response messages do not appear in a vacuum (Yi, 1993). They are embedded in complex media environments in which prior contextual information can activate cognitive schemas that shape how subsequent messages are interpreted (Yi, 1990, 1993). This contextual priming effect has been well documented in advertising and persuasion research (e.g., Shen & Chen, 2007; Yi, 1990) and in studies of social issue interpretation (e.g., Abraham & Appiah, 2006; Dixon, 2007). Such priming effects may be especially consequential for international organizations, whose crisis responses may be evaluated through the lens of geopolitical tensions and country-of-origin perceptions.

Recent geopolitical conflicts illustrate how international companies can become symbolic targets. For example, Chinese consumers boycotted American brands during the U.S.–China tariff war (Feng & Xiao, 2025), and Moroccans boycotted IKEA following Sweden’s foreign policy decisions (Taylor, 2015). These reactions suggest that media portrayals of geopolitical relations may prime audiences to evaluate organizations differently before any crisis response is even encountered. Despite these developments, no research has systematically examined how geopolitical contextual priming shapes audience evaluations of organizational responses to misinformation-driven paracrises.

Addressing this gap, the present study investigates how crisis response format (narrative vs. informative) and crisis communication strategy (accommodative vs. defensive) interact with geopolitical contextual priming to influence audience evaluations of organizational responses to social media hoaxes. Using an experimental design, the study manipulated media portrayals of China as a geopolitical friend versus threat and examined how U.S. participants evaluated a Chinese company’s response to a social media hoax. By focusing on paracrisis as a contextual condition shaped by media and geopolitical environments, this

research offers a focused, theory-driven extension of crisis communication research rather than a broad theoretical expansion.

Literature Review

Hoaxes and misinformation paracrisis on social media

The rise of social media has posed numerous challenges for crisis communication. One particular challenge is the ease and speed with which inaccurate information about companies and organizations can spread (Oukropec, 2024). A social media hoax is a type of paracrisis (Oukropec, 2024; Chen & Holladay, 2018) because social media exposes potential reputational risks to the entire online public, creating the need for these risks to be managed publicly rather than solely among stakeholders (Oukropec, 2024). Chen and Holladay (2018) categorized six types of paracrisis, including challenge, faux pas, guilt by association, misinformation, social media misuse, and social media account hacking. A social media hoax is a form of misinformation paracrisis characterized by the “circulation of messages about the organization that lack veracity” (Chen & Holladay, 2018, p. 109).

Rather than conceptualizing paracrisis as a distinct theoretical category, this study treats misinformation paracrisis as a contextual condition characterized by ambiguity, uncertainty, and low initial responsibility attribution. These features distinguish hoaxes from conventional crises and make them particularly sensitive to external contextual cues, including media framing and geopolitical narratives.

Prior research on paracrisis responses has yielded mixed findings. Research on responses to paracrisis on social media generally recommends a more defensive approach (Lewandowsky et al., 2012; Coombs & Holladay, 2012). However, closer examinations of paracrisis responses have revealed inconsistent results. Humor and mockery may be effective in some cases (Kim et al., 2016); however, Honisch and Manchon (2019) found the opposite effect and instead recommended a reform strategy. Overall, Coombs (2017) suggested that responses to paracrisis may take the form of variations of traditional crisis response strategies, which were later specified as refusal, refutation, repression, recognition, reform, and revision (Coombs & Holladay, 2015).

While the lack of consensus regarding effective response strategies in paracrisis warrants further investigation, it is also important to examine the contexts in which hoaxes and misinformation paracrisis occur. Most existing studies on paracrisis or online hoaxes examine these phenomena in relatively isolated environments, testing audience reactions to response strategies without accounting for broader media or international contexts. Yet, in an era of globalization, crises and misinformation travel both physically and symbolically across national boundaries, making crisis communication increasingly complex for organizations (Schwarz, Seeger, & Auer, 2016). International organizations may be particularly disadvantaged when local media facilitate othering or favor certain national or ethnic groups (Nohrstedt, 2011). In crisis communication, geopolitical tensions play an

important role in shaping crises themselves as well as public perceptions and judgments. For example, Claessen (2022) argued that mutual frustrations among European countries contributed to the portrayal of “geopolitical othering” of Ukraine in Russian narratives during the Ukraine crisis. Similarly, Taher, El Kolaly, and Tarek (2025) employed the micro-influencer impact model to examine crisis communication during geopolitical conflicts and found that perceived authenticity, expertise, narrative personalization, and digital literacy play significant roles in shaping public opinion and crisis responses.

Despite growing scholarly interest in crises occurring within geopolitical conflicts, existing research has not examined how geopolitical framing interacts with response strategies and response formats in misinformation paracrises. This gap motivates the present study’s focus on contextual priming and response effectiveness.

Contextual priming and U.S.- China relations

Messages “don’t appear in a vacuum” (Yi, 1993, p. 1), and audience interpretations of messages are shaped by the broader media environment in which those messages are embedded. Accordingly, people’s perceptions of a social media hoax or paracrisis can be influenced not only by message content, but also by contextual information encountered prior to or alongside the crisis message. Examining the effectiveness of crisis communication across different information environments is therefore critical.

Contextual priming refers to the influence of surrounding media content that activates cognitive schemas used to interpret a focal message (Shen & Chen, 2007). Most people read messages in an environment filled with various other information, such as page background, price, information for another product, etc. (Mandel and Johnson, 2002). Contextual priming has received much attention in advertising and persuasion research as product information can be ambiguous and interpreted in many ways: an ad that describes a piece of luggage as being light can be interpreted as the luggage being easy to carry around or being less durable (Yi, 1990).

Contextual priming is effective because people interpret information using recent, active knowledge structures for schemas (Higgins and King 1981; Wyer and Srull 1981) activated by the last stimuli or message. They are adopted to interpret current information because of its accessibility, even though the schema might not be directly related to the new information (Fiske and Taylor, 1991). Two different effects, assimilation effect and contrast effect, might occur (Sherif and Hovland 1961). Assimilation effect describes people’s tendency to adopt the primed context and judge the subsequent message accordingly (Sherif and Hovland 1961; Shen and Chen, 2007). However, people can also bias away from the context and reject the primes when people are highly involved in the topic and engaged in elaborated thinking on the message (Meyers-Levy and Sternthal, 1993; Shen and Chen, 2007).

Priming certain attributes of products can activate consumers’ interpretation in a subsequent ad, and consequently, enhance people’s judgment of the brand (Yi, 1990). Schmitt (1994) found that although participants said they were not aware of the visual

priming, they were still affected by it unconsciously. Contextual information on websites also influences people's evaluation of a subsequent ad and their purchase intention (Jeong and King, 2010). This trend remains consistent across different studies examining magazine ads, TV commercials, online display or banner ads (Jeong and King, 2010), job and recruitment ads (Walker *et al.*, 2011), and ads in blogs (Segev *et al.*, 2015). Contextual priming can also influence attitudes and behaviors such as president approval (Althaus and Kim, 2006) and stereotyped judgment on crime situations (Abraham and Appiah, 2006; Dixon, 2007; Dixon & Azocar, 2007). These findings suggest that contextual priming shapes evaluative judgments even when audiences are not consciously aware of its influence (Schmitt, 1994).

In the context of international crisis communication, contextual priming related to a company's country of origin may be especially consequential. When organizations operate across national boundaries, audiences may rely on geopolitical cues and media portrayals of a country as heuristics for evaluating organizational credibility and intent (Borden, 2016). For instance, when Chinese people boycotted American brands as a response to the tariff war between the U.S. and China (Feng and Xiao, 2025) and when Moroccans boycotted IKEA to respond to Sweden's foreign policy (Taylor, 2015), these companies being caught in geopolitical conflicts, partly due to contextual priming effects induced by political or economic tensions between countries (Taylor, 2015). These reactions illustrate how geopolitical narratives can prime audiences to evaluate organizations through a national lens, independent of organizational behavior.

Such contextual priming may be particularly influential in misinformation paracrises, where uncertainty is high and organizations are already in a disadvantaged position. In these situations, contextual cues may shape acceptance of the crisis response, defined as the perceived degree to which audiences accept an organizational response (Yang *et al.*, 2010), as well as organizational reputation (Kim, 2001) and corporate trust (Shen, 2017).

Although only a limited number of studies have examined geopolitical contextual priming in crisis communication, existing research suggests that news framing can significantly affect how crises and organizational responses are perceived (Mason, 2019). However, few studies have examined how geopolitical contextual priming interacts with specific crisis response strategies and response formats in misinformation-driven paracrises, particularly for international organizations.

This study, in particular, utilizes the geopolitical relationship between the United States and China as the media context. The two countries have "one of the world's most important and complex bilateral relationships" (Council on Foreign Relations, 2025). Since 1949, their relationship has experienced fluctuations, with periods of tension and cooperation in areas such as trade, military affairs, and national security (U.S. Government Accountability Office, 2024). Data from the Pew Research Center indicate that in 2024, 42% of participants viewed China as an enemy, approximately 50% viewed China as a competitor, and only 6% viewed China as a friend (Huang, Silver, & Clancy, 2024). In 2025, 77% of U.S. respondents reported unfavorable views of China, and 42% believed that China posed the greatest threat to the United States (Huang, Silver, & Clancy, 2025).

However, it is important to note that the U.S.–China relationship has not always been negative. Between 2009 and 2011, Americans generally held positive views of China, and

in 2017, public opinion in the United States was nearly evenly divided (Silver, 2022). In contrast, data from 2021 revealed a sharp 28% decline in positive attitudes toward China (Pew Research Center, 2021). These negative perceptions were attributed to bilateral economic relations, the COVID-19 outbreak, and human rights issues (Pew Research Center, 2021). Despite ongoing tensions and fluctuations, the United States and China remain critical partners. As the world's two largest economies, China is one of the largest markets for U.S. exports, while the United States is China's top export market (Council on Foreign Relations, 2025). The status of the relationship between the two countries, along with their close economic interdependence, provides a valuable opportunity to examine how fluctuations reflected in the media environment may influence public perceptions of a Chinese company during a misinformation-driven paracrisis. Any misinformation-driven paracrisis or social media hoax has the potential to cause significant damage to companies operating in both countries.

It is important to distinguish between country-of-origin effects and contextual priming of a country. Country of origin is typically treated as a static organizational attribute, whereas contextual priming captures dynamic media portrayals of a country as positive or negative, which may fluctuate with international relations and geopolitical uncertainty. By focusing on contextual priming rather than static country-of-origin effects, this study examines how shifting geopolitical narratives shape audience evaluations of organizational crisis responses. H1 is posited below.

H1: After exposure to a positive (H1a) vs. negative (H1b) story about China, participants who read a company's response to a misinformation paracrisis from China will report (1) greater acceptance of the response (2) higher company reputation and (3) higher trust toward the company.

Crisis narrative type: Narratives vs. informative response

Narratives are chains of causally connected events unfolding across time and space (Bordwell & Thompson, 2010) that include plots and sequences designed to immerse audiences in a coherent story (Green, 2006; Shen et al., 2014). Compared with informative messages, which present discrete facts and arguments, narrative messages organize information into meaningful structures that facilitate sensemaking under uncertainty. Prior research shows that narrative persuasion can shape attitudes toward social, health, and political issues more effectively than informative messages (Shen et al., 2014). Through the strategic use of narrative, Gabbay & Leenders (2003) found that recruiters could influence the recruit's decision to trust the company. The narrative construction of events could also increase the narrative's credibility (van Krieken and Sanders, 2016). Narrative effectiveness is often explained through cognitive and affective mechanisms, such as transportation and identification. Narrative messages tend to transport readers into the story world, reduce self-awareness, and foster emotional and cognitive engagement with characters and events (Green, 2006; Cohen, 2001; Shen et al., 2014). In contrast, informative responses rely more

heavily on analytic processing and factual scrutiny, which may heighten skepticism when audiences face ambiguity or uncertainty.

Crisis communication has long been described as inherently narrative in nature. As Yang et al. (2010) argued, crisis communication “is essentially narratives” (p. 475), and Heath (1997) described it as the enactment of a narrative of control under conditions of high uncertainty. Accordingly, narrative responses may be particularly effective during crises by providing causal explanations and restoring a sense of order, thereby reducing resistance to organizational messages (Yang et al., 2010). Prior studies have shown that narrative crisis responses can increase perceived interactivity, positive organizational attitudes, and message acceptance (Yang et al., 2010).

Despite these advantages, the effectiveness of narrative responses in misinformation-driven paracrises remains underexplored. Social media hoaxes often lack clear sources for verification and are designed to be emotionally engaging and easily shareable (Itti & Baldi, 2009; Vosoughi et al., 2018). In such contexts, audiences may engage in “post-fact” reasoning and rely less on factual correction alone (Berthon & Pitt, 2018). Informative responses that focus solely on refutation or denial may therefore be insufficient or even counterproductive, particularly when perceived as unfair or overly aggressive (Johar et al., 2010; Mills & Robson, 2019).

These features suggest that narrative responses may offer a distinct advantage in social media hoaxes by contextualizing events, explaining organizational actions, and guiding audience sensemaking, rather than merely correcting facts. However, boundary conditions remain unclear, particularly regarding whether narrative responses are more effective than informative responses when audiences are exposed to negative contextual cues, such as unfavorable geopolitical framing.

Building on prior findings that narrative messaging positively influences response acceptance (Yang et al., 2010), organizational trust (Gabbay & Leenders, 2003), and attitudes toward organizations (Yang et al., 2010), this study examines whether narrative responses are more effective than informative responses in the context of social media hoaxes.

H2: When compared with an informative response, a narrative response can generate higher levels of (1) acceptance of the response (2) company reputation and (3) trust toward the company.

Crisis communication strategies

In paracrisis situations, Coombs and Holladay (2015) identified six response strategies—refusal, refutation, repression, recognition/reception, reform, and revision—which can be positioned along a continuum from defensive to accommodative responses (Coombs, 2019). Defensive strategies primarily seek to deny the existence of a crisis, minimize organizational responsibility, or refute allegations, whereas accommodative strategies acknowledge concerns and emphasize corrective actions, apologies, or compensation. Based on the

situational crisis communication theory (SCCT), stakeholders would attribute high, low, or medium crisis responsiveness to organizations when facing different crises. If the assumed crisis responsiveness is high, communicators might adopt accommodative strategies to achieve satisfactory outcomes; if the attributed responsibilities are low, they adopt defensive strategies (Coombs, 2007). However, SCCT was largely developed in the context of conventional crises and assumes relatively stable attributions of responsibility, an assumption that may not hold in social media-driven paracrises.

Prior research has extensively examined the effectiveness of crisis communication strategies across multiple outcome variables, including response acceptance, organizational attitudes, supportive word-of-mouth intentions, corporate reputation, and trust (Cheng, 2018; Jin, 2010; Yang et al., 2010). For example, Yang et al. (2010) demonstrated that accommodative strategies, such as corrective actions, enhance response acceptance, particularly when audiences are actively engaged in crisis communication. Ngai and Jin (2016) further showed that hybrid strategies combining accommodative and defensive elements may prevent escalation in social-mediated crises, while Ma (2020) found that compensation strategies can reduce consumer reactance.

Corporate reputation is a key indicator of crisis communication effectiveness (Claeys et al., 2010). Empirical studies consistently show that accommodative strategies, especially apologies are more effective in reputation restoration than purely defensive responses (Coombs & Holladay, 2008; Claeys et al., 2010). Similarly, trust is a central dimension of organization–public relationships and a critical outcome in crisis contexts (Kang & Park, 2017; Shen, 2017; Yang et al., 2015). Research suggests that accommodative strategies can foster trust when stakeholders attribute low responsibility to the organization (e.g., natural disasters), but may fail to restore trust when responsibility is perceived as high and public skepticism dominates (Coombs, 2019; Yang, 2016).

Importantly, social media hoaxes differ from traditional crises in that responsibility is often ambiguous, evidence is contested, and misinformation spreads rapidly through emotionally charged networks. In such paracrisis contexts, stakeholder attributions may be shaped less by objective responsibility and more by contextual cues and media framing, raising questions about whether standard SCCT prescriptions apply. As a result, it remains unclear whether accommodative or defensive strategies are more effective when organizations respond to social media hoaxes, particularly under varying geopolitical and contextual priming conditions.

Building on this gap, the present study examines how defensive and accommodative response strategies operate within misinformation-driven paracrises, and how their effectiveness depends on response format (narrative vs. informative) and contextual priming. Based on prior research and the unique features of social media hoaxes, we advance the following hypotheses:

H3: When compared with a defensive crisis communication strategy, an accommodative crisis response can generate higher levels of 1) acceptance of the response, 2) company reputation, and 3) trust toward the company.

This article also investigates how crisis narrative type and crisis communication strategy together influence crisis communication outcomes. Past research has identified a narrative

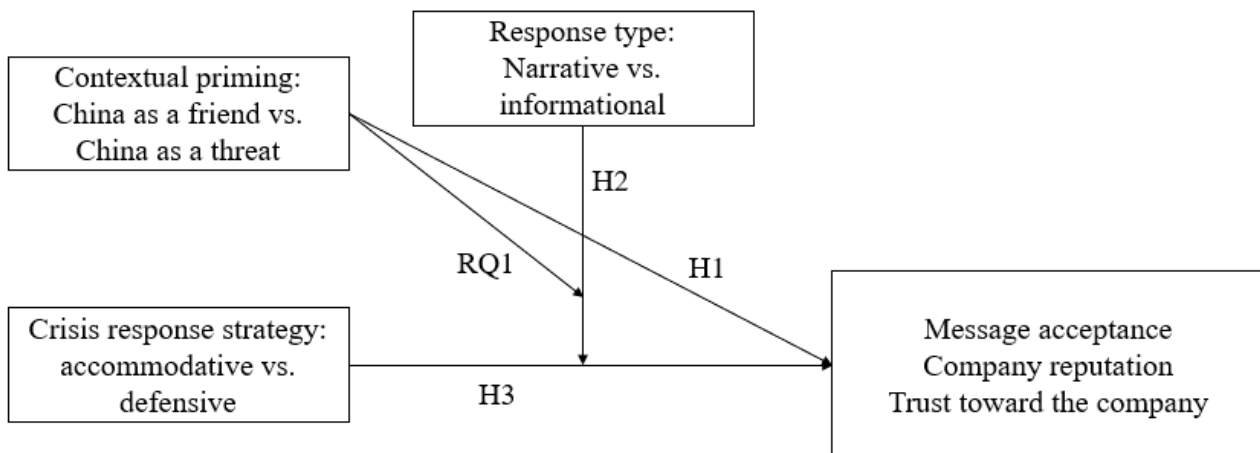
response to be more effective in crisis communication (Yang *et al.* 2010; Gabbay and Leenders, 2003) and an accommodative response to be more effective than a defensive response (Coombs, 2015). This study hypothesizes the following interactional effect:

H4: A narrative, accommodative response will generate higher levels of 1) acceptance of the response, 2) company reputation, and 3) trust toward the company than other types of responses.

The interactions between the three proposed independent variables are also examined. As indicated above, messages do not appear in a vacuum (Yi, 1993). Investigating how the complex media environment (geopolitical uncertainties between two countries, in this study) might affect the effectiveness of crisis response is important as it enhances our understanding of crisis responses effectiveness. Exposure to a previous message could affect how they judge the subsequent information, and such impact remains under-investigated in responses to social media hoax. Moreover, the examination would explain how and to what extent the reports on geopolitical conflicts affect responses to social media hoaxes is still unknown. Scholars have argued that the priming effect and news frames could affect the effectiveness of crisis response (Yang *et al.*, 2010; Mason, 2019), but these studies did not examine strategic crisis responses from an international and geopolitical perspective. Figure 1 displays the hypotheses in this article.

RQ1: How does contextual priming (China as a friend vs. China as a threat) moderate the relationship between response type and crisis communication strategy?

Figure 1- Hypothesized model



Method

A 2 (contextual priming: China as a friend vs. as a threat) x 2(response type: narrative vs. informative) x 2 (crisis communication strategy: accommodative vs. defensive) between-group experiment was conducted online. This specific context was chosen because of the recent geopolitical conflict between the U.S. and China, and that the tensions were well-documented by the media (Gandel, 2016).

The experiment was launched on Qualtrics.

Sample

Participants were recruited via Lucid, an online vendor. $N = 506$ participants from the United States joined the online experiment. Participants had a mean average age of 45.81 ($SD = 17.03$), ranging from 18-year-old to 81-year-old. 244 of them were males (48.2%) and 261 of them were females (51.6%). Most of the participants were white (73.5%), followed by Black/African American (non-Hispanic) (10.9%), and Latino/Hispanic (6.1%). 29 of the participants were Asian/Pacific Islander (5.7%), 6 were Native American (1.2%), and 2.6% of them reported their race as "other." Most of them received bachelor's degrees (25.1%), 23.7% of them received high school degrees, 23.1% received some college education. In our sample, 11.7% received associate degrees, 10.7% received master's degrees, 4% received less than a high school diploma and 1.8% of them received doctorate degrees. 25.1% of the participants reported income between \$20,000-40,000, 21.5% of them reported income under \$20,000, and 16.6% reported income between \$40,000-\$60,000. About 13.8% said \$100,000 and higher, 12.1% of them reported \$60,000 - \$80,000, 7.5% reported \$80,000-\$100,000, and 3.4% did not report their income.

Procedure

Researchers conducted a pre-test using participants from Amazon Mechanical Turk to check the experiment flow. For formal data collection, participants were recruited by an online vendor, Lucid. Participants clicked the link to participate in the study. Upon consent, they answered questions relating to the control variables and then were randomly exposed to an article depicting either China as a friend or China as a threat. After answering the manipulation check questions, participants read information about a Chinese food company and an incident at a U.S. location – Two employees posted a video accusing the company of violating food safety measures. This scenario was chosen as an adaptation of a real social media crisis that happened to Domino's Pizza when two employees posted videos of their food contaminating behaviors in the kitchen. The two employees were eventually arrested and charged with food tampering. Then, participants were randomly assigned to read a response from the company, which argued that the online video was a hoax, and the video

was created as a prank by the two employees. Participants then proceed with the questionnaire and were debriefed at the end of the questionnaire. Upon completion, participants received compensation from the vendor.

Stimulus

Contextual priming was manipulated through an online article created based on current coverage regarding China and U.S. relations and online news articles describing how China facilitated/threatened the U.S. economy. Specifically, one story described China as a friend, discussing how China's investment in the U.S. stimulated the local economy. In the "China as a threat" context, the article discussed how China's investment in the U.S. took away local jobs and endangered the local economy. All other parts of the story remained similar, and the two versions are in similar length.

Response types (narrative vs. informative) were manipulated via two versions of a Facebook post. The manipulation was adopted from Shen *et al.*'s study (2014) and followed their message manipulation format. The design of the Facebook post mimics the design of a real Facebook post. In both conditions, the response was from the company's CEO's Facebook account. In the narrative condition, the response was formatted as a story using narrative elements: the initiation of the event (CEO received the hoax information), exposition (CEO realized actions needed to be taken), complication (CEO realized the complexity of the hoax and started an investigation), and event culmination (CEO concluded the investigation and offered a resolution) (Shen *et al.*, 2014). The informative condition was crafted using a typical news story format, starting with the summarization of the event, followed by a series of actions taken by the CEO and the company. Then the CEO quoted the investigators and stated the company's resolution and its stand on the social media hoax. The information and CEO's presence in both conditions were similar.

Two strategies were created at the end of the story to manipulate crisis communication strategy in the company's response. In the accommodative condition, the response emphasized that the company was sorry about what happened and similar events would never happen again in the future. In the defensive condition, the company defended itself and stressed it was a victim as well, and the two employees who made the video were to blame and should be held responsible. Figure 2 and figure 3 showcase two of the stimuli used.

Figure 2 - Response stimuli example (narrative, accommodative response)

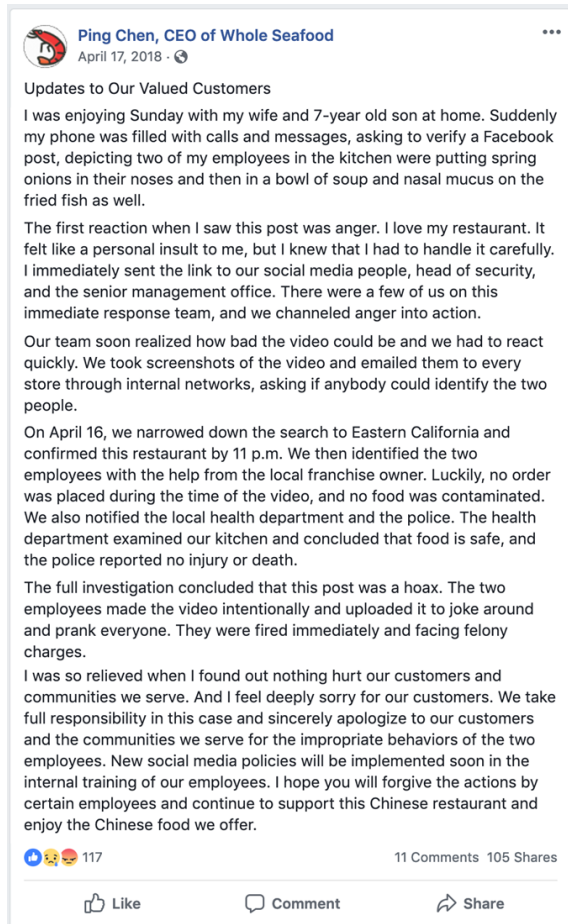


Figure 3 - Priming stimuli example (China as a threat)

Despite its economic growth, China poses threat to the U.S.

Within about 15 years, China's economy will surpass America's and become the largest in the world. As of this moment, a consensus has formed that China poses a significant threat to the U.S. economy. Ever since 2001, when the U.S. agreed to allow China into the World Trade Organization, U.S. workers have been nervous.

Small towns across America have suffered just as workers have. After a factory closes, unemployed residents move away for new work, shops close, and mayors lacking tax revenue cut services. The result is the towns becoming hallowed out shells with decaying prospects. The loss was not concentrated in just one state or region, like Michigan or the Midwest. Every congressional district in every state bled jobs.

A new study shows the growth in the U.S. trade deficit with China between 2001 and 2017 cost 3.4 million American workers their jobs. Among the 3.4 million who lost jobs since 2001 is Cathy Coreman from Pennsylvania. She was struck down twice, Cathy said, "The 1st time it happened was in 2012 when the mirror manufacturing company I worked for closed its factory and transferred its business to China. The second time my employer laid me off as it prepared to permanently close its glass factory and move half of the work to China. That is the way U.S. factory workers live now: nervous. Any day could be their last on the job. Any day, their work could be shipped to China. Finding new work that pays a living wage is difficult in many places.

China has threatened the American economy by using low-wage labor, providing government subsidies, free or cut-rate raw materials, and discounted utility bills. China keeps people employed by subsidizing overproduction of materials like steel and aluminum, then dumps the excess at below-production-cost prices on the international market. These market-distorting practices kill industries in market economies such as in the U.S. China is a threat, not a friend.

Measurements

All items were measured on 1-7 point scales.

Acceptance of response messages was measured by adapting five items from Blumstein *et al.* (1974). Participants were asked to rate the response message. A sample item was “Not acceptable” to “acceptable”. Cronbach $\alpha = .96$, $M = 5.18$, $SD = 1.64$.

Corporate reputation was measured by adopting 7 items from Tao and Wilson (2016) to measure the perceived reputation of the company. Participants were asked to rate their levels of agreement on statements. A sample statement read “This company delivers high-quality products and services”. Cronbach $\alpha = .94$, $M = 4.77$, $SD = 1.22$.

Trust toward the company was measured by using five items from Cheng, Shen, and Jiang (2020). Participants were asked to rate their levels of agreement on statements. A sample statement is “I believe that the company treats stakeholders like me fairly and justly”. Cronbach $\alpha = .94$, $M = 4.61$, $SD = 1.31$.

Demographic variables including age, gender, race, education, and income were also measured. Control variables including social media use and frequency of reading online hoaxes were also measured.

Three manipulation check questions were used. The first manipulation check question was to examine whether the contextual priming (China as a friend vs. as a threat) was successful by asking participants’ levels of agreement on the item “the article I just read described China as a friend of the U.S.”. The second manipulation check question examined the manipulation of response type by asking the participants to rate their levels of agreement on 5 items, with an example reads like “I think this response tells a story about this incident rather than purely listing information”. The third manipulation was to check the manipulation of crisis communication strategy by asking levels of agreement on 4 items such as “the response apologized for the online hoax” and “the response took responsibility for this online hoax on its own”.

A 2 (contextual priming: China as a friend vs. as a threat) x 2(response type: narrative vs. informative) x 2 (crisis communication strategy: accommodative vs. defensive) between-group experiment was conducted online.

Results

Manipulation check

Independent t-tests were performed to check the manipulations. Manipulation of contextual information was successful $t(504) = 2.13$, $p < .05$. Those exposed to the news depicting China as a friend reported that China was described as a friend rather than a threat (friend: $M = 4.19$, $SD = 2.28$ vs. threat: $M = 3.76$, $SD = 2.28$). Response type (narrative vs. informative) was also manipulated successfully, $t(504) = 5.82$, $p < .00$. Those exposed to narrative information reported a significantly higher level of story narration ($M = 5.16$, SD

= .85) than those exposed to an informative message ($M = 4.67$, $SD = 1.02$). Those exposed to an accommodative response reported that the message was more accommodative ($M = 5.50$, $SD = 1.26$) than those exposed to a defensive message ($M = 4.64$, $SD = 1.27$).

Hypotheses testing

To test H1, H2, H3 and H4 and answer the research questions, a MANOVA test was employed, controlling for time spent on social media and frequency of reading online hoaxes. The more time people spend on social media and the more frequently people read about an online hoax, the more likely they were familiar with the phenomenon of online hoaxes. Therefore, they might be more likely to judge a hoax more critically and to evaluate the affected company differently. Correlations between the dependent variables were all significant and acceptable, with the lowest correlation being $r = .52$ and the highest being $r = .85$. In the MANOVA test, we first examined how the independent variables affected the dependent variables as a group. Upon identifying the significant results, the main effects of the independent variables were then investigated. For significant two-way interactions, we employed pairwise post-hoc tests. For the significant three-way interaction, researchers split the data by the different conditions of the moderating variable (contextual priming) and then probed the two-way interactions under each condition.

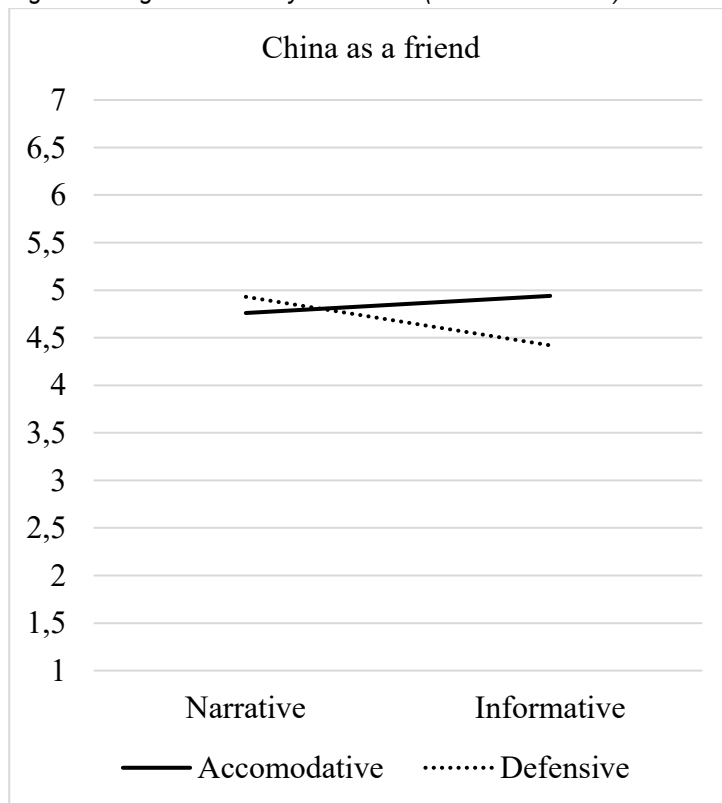
Results showed that response type (narrative vs. informative) had a significant main impact, Wilks' $\Lambda = .98$, $F(3, 494) = 3.36$, $p < .05$, partial $\eta^2 = .02$. No significant difference was found on contextual priming (Wilks' $\Lambda = .10$, $F(3, 494) = .017$, $p > .1$, partial $\eta^2 = .02$) and on crisis communication strategy (Wilks' $\Lambda = .10$, $F(3, 494) = .057$, $p > .1$, partial $\eta^2 = .00$). H1(a), H1(b) and H3 were not supported.

To further investigate H2 (response type, narrative vs. informative), univariate analysis for response type showed that response type was significant on all three dependent variables. A narrative response was more acceptable ($M = 5.38$, $SD = .10$) than an informative response ($M = 5.00$, $SD = .10$), $F(1, 496) = 6.84$, $p < .01$, partial $\eta^2 = .01$. Those who read a narrative response also reported the company to have better reputation ($M = 4.88$, $SD = .074$) than those who read an informative response ($M = 4.67$, $SD = .08$), $F(1, 496) = 3.79$, $p = .05$, partial $\eta^2 = .01$. A narrative response brought higher level of trust towards the company ($M = 4.77$, $SD = .08$) than an informative response ($M = 4.46$, $SD = .08$), $F(1, 496) = 7.56$, $p < .01$, partial $\eta^2 = .02$. Therefore, H2 was partially supported.

A significant two-way interaction was found between response type and crisis communication strategy, Wilks' $\Lambda = .98$, $F(3, 494) = 10.98$, $p < .01$, partial $\eta^2 = .02$, and the difference was found on message acceptance, $F(1, 496) = 7.56$, $p < .01$, partial $\eta^2 = .02$. Pairwise post hoc comparison showed that an informative and accommodative response was significantly more acceptable ($M = 5.34$, $SD = .15$) than an informative, defensive response ($M = 4.63$, $SD = .14$). There was no significant difference among narrative responses or on other dependent variables. Therefore, H4 was not supported.

RQ1 probes how contextual priming moderates the relationship between crisis response strategy and response type. A significant three-way interaction was found, Wilks' $\Lambda = .99$, $F(3, 494) = 4.19$, $p < .1$, partial $\eta^2 = .02$. The significant impact was found on company reputation, $F(1, 496) = 6.32$, $p < .05$, partial $\eta^2 = .01$. After separating the file by how China was portrayed, the ANOVA test showed that a significant difference was found only when China was portrayed as a friend (Figure 4). When people were primed with news portraying China as a friend, a significant two-way interaction was found for response type and crisis communication strategy on company reputation. $F(1, 252) = 2.81$, $p < .05$, partial $\eta^2 = .02$. Pairwise post hoc test revealed that an informative, accommodative response led to significantly higher perceived company reputation ($M = 4.94$, $SD = .15$) than an informative, defensive response ($M = 4.42$, $SD = .14$). After reading an article about China being a friend of U.S. economy, an informative, accommodative response was more effective than an informative, defensive response. No significant differences were found on other independent variables.

Figure 4 - Significant 3-way interaction (China as a friend) and its effect on company reputation.



Discussion

This study examined people's perceptions of the corporate messages on a misinformation paracrisis on social media under different contextual priming effects featuring different geopolitical relations. Results of this study highlights the strategic complexity corporations

need to navigate while responding to hoax on social media amid geopolitical uncertainties. Rather than positioning any single response strategy as universally effective, the findings demonstrate that crisis response effectiveness is contingent upon response format, strategy, and the broader geopolitical media context in which the response is embedded. Results suggest that a narrative response might be more effective than an informative response in generating higher message acceptability perceived company reputation. In addition, an informative, accommodative response was more acceptable than an informative, defensive response. Importantly, these effects were not uniform across conditions. Results from the three-way interaction indicate that people processed crisis response messages differently depending on the surrounding media environment. Only when participants were primed with a geopolitical context portraying China as a friend of the U.S. did an informative, accommodative response outperform an informative, defensive response in terms of company reputation.

This study enriched the crisis communication literature by examining the impact of narrative response format on message acceptability, corporate trust, and reputation. As previous research indicated, narrative messages could be more persuasive (van Krieken and Sanders, 2016; Gabbay and Leenders, 2003; Yang *et al.*, 2010) and could help reduce counter-arguing and increase message effectiveness (Moyer-Guse & Nabi, 2010). Our findings extend this work by demonstrating that narrative advantages persist even in paracrises characterized by low organizational responsibility and high informational ambiguity, such as social media hoaxes. Results of this study supported the argument that a narrative response in a paracrisis is more effective than an informative message in increasing the perceived message acceptability and company reputation. When responding to a paracrisis, using strategies such as explaining the chain of events that lead to the paracrisis might help the company to gain better message evaluation and restore its reputation. Perhaps a narrative response gives the audiences a more immersive context to understand the situation (Green, 2006; Shen *et al.*, 2014); in addition to explaining the paracrisis, a narrative response might also provide entertainment values (Shen *et al.*, 2014) and hence, is more persuasive. Results also reflected Heath's (1997) argument of crisis communication being the enactment of narrative control during uncertain times. Moreover, results extended previous research on the effect of strategic narratives during uncertain times (i.e., Yang *et al.*, 2010) and indicated that narrative storytelling countering a paracrisis could be effective in terms of its influence on both message acceptance and reputation. Narrative responses may provide audiences with a coherent interpretive framework under uncertainty, particularly when misinformation disrupts causal clarity.

The significant two-way interaction suggest that an informative, accommodative message is more effective in increasing message acceptance than an informative, defensive message. When messages are presented in a non-narrative format, participants prefer to see the organization taking a softer and more accommodative stand when refuting hoaxes on social media. Such response was perceived to be more acceptable. A firmer, defensive stand might make the participants perceive the message to be less acceptable. This result is in line with previous research that a more accommodative stand might be more effective for corporations to restore their reputation (Claey *et al.*, 2010; Coombs and Holladay, 2008).

This result suggests that when a narrative response is not available or feasible, and when the company has to respond using an informative format, taking a softer stand and be apologetic might make the response more acceptable than taking a firm, unapologetic stand. However, rather than presenting this as a novel main effect, we interpret it as context-dependent evidence that even low-responsibility misinformation paracrises may activate expectations for accommodation when audiences lack confidence in factual verification.

The most theoretically revealing finding is the significant three-way interaction on company reputation. When predicting reputation, only after exposure to a geopolitical prime portraying China as a friend of the U.S. did participants perceive an informative, accommodative response as more effective than an informative, defensive response. This suggests that geopolitical contextual priming conditions the applicability of crisis response strategies, rather than merely amplifying their effects. When China was portrayed as a friend, participants appeared more receptive to accommodative signals; when China was portrayed as a threat, response strategies had limited influence on evaluations.

This pattern invites consideration of the long-standing and often negative U.S.–China relationship as an interpretive backdrop. Even when primed with a positive portrayal, participants may still hold chronically accessible geopolitical schemas that lead them to expect accommodative behavior from Chinese companies regardless of crisis responsibility or misinformation status. Rather than undermining the findings, this interpretation underscores the importance of distinguishing between *country-of-origin as a static attribute* and *geopolitical contextual priming as a dynamic media condition*. While entrenched geopolitical attitudes may set baseline expectations, the results show that short-term contextual cues can still meaningfully shape how crisis responses are evaluated.

Notably, when China was portrayed as a threat, no response strategy emerged as clearly effective. This suggests that in hostile geopolitical media environments, organizational crisis communication efforts may face structural constraints, limiting the persuasive power of both accommodative and defensive strategies. This finding advances crisis communication theory by showing that response strategies may only function as intended within relatively favorable geopolitical contexts, emphasizing the need for strategic narratives that help organizations navigate compounded reputational risks under geopolitical tension.

The impact of contextual priming deepens our understanding of paracrisis communication. Although Coombs and Holladay conceptualized paracrises as publicly visible but responsibility-ambiguous threats, empirical research has rarely examined how broader media and geopolitical contexts shape paracrisis evaluations. While prior reviews have highlighted the role of political and cultural environments (Cheng et al., 2020), this study provides causal evidence that geopolitical contextual priming interacts with response format and strategy to influence crisis outcomes. Although no main effect of contextual priming was observed, the interaction effects demonstrate that its influence is conditional and nonlinear.

Finally, this study did not find a significant main effect of crisis communication strategy alone, diverging from some prior research (Claeys et al., 2010; Coombs, 2019; van Zoonen & van der Meer, 2015). This suggests that in misinformation paracrises, response strategy may be overshadowed by message format and contextual interpretation. Social media

hoaxes may therefore constitute a distinct crisis type in which traditional strategy prescriptions are insufficient, warranting further theoretical and empirical attention.

Theoretical Implications

This study contributes to corporate and crisis communication theories. It examined how companies should respond to social media hoaxes most effectively and extended prior research on hoaxes, paracrisis, and social media crisis communication in a digital age. It also enriched crisis communication literature by integrating different response types with contextual priming and examining their impact on response message acceptance, corporate trust, and organizational reputation.

Although contextual priming did not show a significant main effect, the observed 3-way interaction provides nuanced insights into how cognitive processing and activated schemas influence audience reactions in complex crisis environments. These findings advance theoretical understanding of paracrisis communication, particularly under conditions of cross-cultural and global organizational contexts (Coombs, 2019; Coombs & Holladay, 2012). In addition, the results highlight how crisis communication outcomes may be shaped by geopolitical tensions and international contexts, offering guidance for future research on country-specific framing effects in crisis scenarios.

Practical Implications

The results indicate that audiences prefer narrative-based responses from companies when addressing social media paracrisis. Companies' responses should include storytelling that outlines the development of the paracrisis and the sequence of organizational actions taken to manage it. When a narrative response is not feasible, an informative, accommodative response is generally more acceptable than an informative, defensive response, even in low-responsibility situations, as it fosters greater trust and message acceptance.

This study also demonstrates that audience judgments are influenced by contextual cues, including geopolitical and country-of-origin factors. Organizations, particularly international ones, should conduct careful media screening and consider the geopolitical framing of their country of origin before issuing responses, as negative portrayals can undermine the effectiveness of informative responses. Narrative responses may be more resilient across varying contexts, supporting higher trust and perceived credibility. Companies should clearly explain the chain of events leading to their response actions and craft messages that immerse audiences in the rationale and context of the crisis response, while remaining attentive to the limitations posed by the study's focus on country framing and the specific experimental context.

Limitations and Future Studies

This study has a few limitations. This research used only one country and one scenario to reflect the media environment and geopolitical conflicts, and the results might not be generalizable to other scenarios and other international companies. It only focused on one type of social media. Other media sources could be manipulated and examined as well. Second, hoaxes are only one type of social-mediated crisis; other types, such as online misconduct and information mismanagement, can be investigated in future studies. Future studies should investigate different types of contextual priming and other social media crisis on various social media platforms.

Conclusion

This study investigated how contextual information, crisis response type, and crisis communication strategies together influence how companies should best respond to paracrises such as social media hoaxes. The experiment showed that a narrative response might be more effective, and informative, accommodative response is more effective than an informative, defensive response. The study found that participants did adopt an active, accessible cognitive scheme primed by contextual information when judging a subsequent crisis response message; however, the results are complicated by other messaging factors.

Biographical notes

Anli Xiao (Ph.D.) is an assistant professor at the School of Journalism and Mass Communications, University of South Carolina. Her research concentrations include corporate social advocacy, global and international public relations, and nonprofit public relations. Her research covers important topics such as the roles of organizations in public discourse, the impact of cultures and values on global public relations practices and people's perceptions and evaluations of corporate social responsibility, how organizations can use digital media to strengthen organization-public relationships and encourage prosocial behaviors.

Yang Cheng (Ph.D., MBA) is a tenured Professor in the Department of Communication at North Carolina State University, a University Faculty Scholar, and an associate editor of the *Journal of Contingencies and crisis management*. Recognized among Stanford University's top 2% most-cited scholars, her research focuses on artificial intelligence, crisis communication, relationship management, and global public relations. She has published over 75 peer-reviewed journal articles in leading outlets such as *New Media & Society* and *Public Relations Review*. Dr. Cheng has secured more than \$2 million in competitive research funding from agencies including the National Science Foundation (NSF) and the National Endowment for the Humanities (NEH). Her work has earned numerous international awards, and she is frequently invited to deliver lectures worldwide. She teaches courses in strategic management, crisis communication, and research methods, and is

widely respected for advancing interdisciplinary research at the intersection of AI, communication, and public relations.

References

- Abraham, L., & Appiah, O. (2006). Framing news stories: The role of visual imagery in priming racial stereotypes. *Howard Journal of Communications*, 17(3), 183–203. <https://doi.org/10.1080/10646170600829584>
- Althaus, S. L., & Kim, Y. M. (2006). Priming effects in complex information environments: Reassessing the impact of news discourse on presidential approval. *The Journal of Politics*, 68(4), 960–976. <https://doi.org/10.1111/j.1468-2508.2006.00483.x>
- Berthon, P., Treen, E., & Pitt, L. (2018). How truthiness, fake news and post-fact endanger brands and what to do about it. *GfK Marketing Intelligence Review*, 10(1), 18–23. <https://doi.org/10.2478/gfkmir-2018-0003>
- Berthon, P. R., & Pitt, L. F. (2018). Brands, truthiness and post-fact. *Journal of Macromarketing*, 38(2), 218–227. <https://doi.org/10.1177/0276146718755869>
- Blumstein, P. W., Carssow, K. G., Hall, J., Hawkins, B., Hoffman, R., & Ishem, E. (1974). The honoring of account. *American Sociological Review*, 40, 551–566. <https://doi.org/10.2307/2094421>
- Borden, J. (2016). Effects of national identity in transnational crises: Implications of social identity theory for attribution and crisis communications. *International Journal of Communication*, 10, 377–397.
- Bordwell, D., & Thompson, K. (2010). *Film art: An introduction*. McGraw-Hill.
- Chen, F., & Holladay, S. J. (2018). Identifying and responding to social media risks: Towards an organizational paracrisis communication framework. *International Journal of Strategic Communication*, 12, 59–75. <https://doi.org/10.1080/1553118X.2017.1378220>
- Cheng, Y. (2018). How social media is changing crisis communication strategies: Evidence from the updated literature. *Journal of Contingencies and Crisis Management*, 26(1), 58–68. <https://doi.org/10.1111/1468-5973.12130>
- Cheng, Y., Shen, H., & Jiang, Q. (2020). Corporate dialogue in crises of China: Examining dialogic strategies and communicative outcomes in a child abuse scandal. *Public Relations Review*, 46(1), 101816. <https://doi.org/10.1016/j.pubrev.2019.101816>
- Claessen, E. (2022). The making of a narrative: The use of geopolitical othering in Russian strategic narratives during the Ukraine crisis. *International Communication Gazette*, 16(1). <https://doi.org/10.1177/17506352211029529>
- Claeys, A.-S., Caubergh, V., & Vyncke, P. (2010). Restoring reputations in times of crisis: An experimental study of the Situational Crisis Communication Theory and the moderating effects of locus of control. *Public Relations Review*, 36(3), 256–262. <https://doi.org/10.1016/j.pubrev.2010.05.004>

- Cohen, J. (2001). Defining identification: A theoretical look at the identification of audiences with media characters. *Mass Communication and Society*, 4(3), 245–264. https://doi.org/10.1207/S15327825MCS0403_01
- Coombs, T. (2007). Protecting organization reputations during a crisis: The development and application of Situational Crisis Communication Theory. *Corporate Reputation Review*, 10(3), 163–176. <https://doi.org/10.1057/palgrave.crr.1550049>
- Coombs, W. T. (2017). Digital naturals and the rise of paracrises: The shape of modern crisis communication. In S. C. Duhe (Ed.), *New media and public relations* (2nd ed., pp. 281–290). Peter Lang.
- Coombs, W. T., & Holladay, S. J. (2008). Comparing apology to equivalent crisis response strategies: Clarifying apology's role and value in crisis communication. *Public Relations Review*, 34(3), 252–257. <https://doi.org/10.1016/j.pubrev.2008.04.001>
- Coombs, W. T., & Holladay, S. J. (2012). The paracrisis: The challenges created by publicly managing crisis prevention. *Public Relations Review*, 38(3), 408–415. <https://doi.org/10.1016/j.pubrev.2012.04.004>
- Coombs, W. T., & Holladay, S. J. (2015). CSR as crisis risk: Expanding how we conceptualize the relationship. *Corporate Communications: An International Journal*, 20(2), 144–162. <https://doi.org/10.1108/CCIJ-11-2013-0078>
- Coombs, T. (2019). *Ongoing crisis communication: Planning, managing, and responding*. Sage Publications.
- Council on Foreign Relations. (2025). *Timeline: U.S.–China relations 1949–2025*. <https://www.cfr.org/timeline/us-china-relations>
- Dixon, T. L. (2007). Psychological reactions to crime news portrayals of Black criminals: Understanding the moderating roles of prior news viewing and stereotype endorsement. *Communication Monographs*, 73(2), 162–187. <https://doi.org/10.1080/03637750600690643>
- Dixon, T. L., & Azocar, C. L. (2007). Priming crime and activating Blackness: Understanding the psychological impact of the overrepresentation of Blacks as lawbreakers on television news. *Journal of Communication*, 57(2), 229–253. <https://doi.org/10.1111/j.1460-2466.2007.00341.x>
- Feng, K., & Xiao, B. (2025, April 11). This is China's attitude: Chinese netizens mock US tariffs with memes. *ABC News*. <https://www.abc.net.au/news/2025-04-11/chinese-boycott-of-trump-tariffs-in-us-trade-war/105166500>
- Fiske, S. T., & Taylor, S. E. (1991). *Social cognition: From brains to culture*. McGraw-Hill Higher Education.
- Fortune. (n.d.). The biggest American companies now owned by the Chinese. Retrieved July 28, 2023, from <https://fortune.com/2016/03/18/the-biggest-american-companies-now-owned-by-the-chinese/>
- Fransen, M. L., Smit, E. G., & Verlegh, P. W. J. (2015). Strategies and motives for resistance to persuasion: An integrative framework. *Frontiers in Psychology*, 6, 1201. <https://doi.org/10.3389/fpsyg.2015.01201>
- Gabbay, S. M., & Leenders, R. Th. A. J. (2003). Creating trust through narrative strategy. *Rationality and Society*, 15(4), 509–539. <https://doi.org/10.1177/1043463103154004>

- Green, M. C. (2006). Narratives and cancer communication. *Journal of Communication*, 56(Suppl 1), S163–S183. <https://doi.org/10.1111/j.1460-2466.2006.00288.x>
- Heath, R. L. (1997). *Strategic issues management: Organizations and public policy challenges*. SAGE Publications.
- Higgins, E. T., & King, G. (1981). Accessibility of social constructs: Information processing consequences of individual and contextual variability. In N. Cantor & J. Kihlstrom (Eds.), *Personality, cognition, and social interaction* (pp. 69–122). Lawrence Erlbaum Associates Publishers.
- Honisch, S. V., & Manchon, L. M. (2019). The effects of paracrisis origin and response strategy on Facebook audience's perceived organizational reputation and behavioral intentions. *Corporate Reputation Review*, 23(3), 133–144. <https://doi.org/10.1057/s41299-019-00073-7>
- Huang, C., Silver, L., & Clancy, L. (2024). *China's relationship with the U.S.* Pew Research Center Global Attitudes. <https://www.pewresearch.org/global/2024/05/01/chinas-relationship-with-the-u-s/>
- Huang, C., Silver, L., & Clancy, L. (2025, April 17). *Negative views of China have softened slightly among Americans*. Pew Research Center. <https://www.pewresearch.org/global/2025/04/17/negative-views-of-china-have-softened-slightly-among-americans/>
- Itti, L., & Baldi, P. (2009). Bayesian surprise attracts human attention. *Vision Research*, 49(10), 1295–1306. <https://doi.org/10.1016/j.visres.2008.09.007>
- Jack Walker, H., Feild, H. S., Giles, W. F., Bernerth, J. B., & Short, J. C. (2011). So what do you think of the organization? A contextual priming explanation for recruitment Web site characteristics as antecedents of job seekers' organizational image perceptions. *Organizational Behavior and Human Decision Processes*, 114(2), 165–178. <https://doi.org/10.1016/j.obhdp.2010.11.004>
- Jeong, Y., & King, C. M. (2010). Impacts of website context relevance on banner advertisement effectiveness. *Journal of Promotion Management*, 16(3), 247–264. <https://doi.org/10.1080/10496490903281395>
- Jin, Y. (2010). Making sense sensibly in crisis communication: How publics' crisis appraisals influence their negative emotions, coping strategy preferences, and crisis response acceptance. *Communication Research*, 37(4), 522–552. <https://doi.org/10.1177/0093650210368256>
- Johar, G. V., Birk, M. M., & Einwiller, S. A. (2010). How to save your brand in the face of crisis. *MIT Sloan Management Review*, 51(4), 57. <https://ssrn.com/abstract=2896848>
- Kang, M., & Park, Y. E. (2017). Exploring trust and distrust as conceptually and empirically distinct constructs: Association with symmetrical communication and public engagement across four pairings of trust and distrust. *Journal of Public Relations Research*, 29(2–3), 114–135. <https://doi.org/10.1080/1062726X.2017.1337579>
- Kim, Y. (2001). Measuring the economic value of public relations. *Journal of Public Relations Research*, 13(1), 3–26. https://doi.org/10.1207/S1532754XJPRR1301_2

- Kim, S., Zhang, X. A., & Zhang, B. W. (2016). Self-mocking crisis strategy on social media: Focusing on Alibaba chairman Jack Ma in China. *Public Relations Review*, 42(5), 903–912. <https://doi.org/10.1016/j.pubrev.2016.09.006>
- Lewandowsky, S., Ecker, U. K., Seifert, C. M., Schwarz, N., & Cook, J. (2012). Misinformation and its correction: Continued influence and successful debiasing. *Psychological Science in the Public Interest*, 13(3), 106–131. <https://doi.org/10.1177/1529100612451018>
- Ma, L. (2023). How the interplay of consumer-brand identification and crises influences the effectiveness of corporate response strategies. *International Journal of Business Communication*, 60(1), 84–104. <https://doi.org/10.1177/2329488419898222>
- Mandel, N., & Johnson, E. J. (2002). When web pages influence choice: Effects of visual primes on experts and novices. *Journal of Consumer Research*, 29(2), 235–245. <https://doi.org/10.1086/341573>
- Mason, A. (2019). Media frames and crisis events: Understanding the impact on corporate reputations, responsibility attributions, and negative affect. *International Journal of Business Communication*, 56(3), 414–431. <https://doi.org/10.1177/2329488416648951>
- Merriam-webster.com. (2018). Definition of hoax. <https://www.merriam-webster.com/dictionary/hoax>
- Meyers-Levy, J., & Sternthal, B. (1993). A two-factor explanation of assimilation and contrast effects. *Journal of Marketing Research*, 30(3), 359–368. <https://doi.org/10.1177/002224379303000307>
- Mills, A. J., & Robson, K. (2020). Brand management in the era of fake news: Narrative response as a strategy to insulate brand value. *Journal of Product & Brand Management*, 29(2), 159–167. <https://doi.org/10.1108/JPBM-12-2018-2150>
- Moncus, J. J., & Silver, L. (2021, April 12). Americans' views of Asia-Pacific nations have not changed since 2018 – with the exception of China. *Pew Research Center*. <https://www.pewresearch.org/fact-tank/2021/04/12/americans-views-of-asia-pacific-nations-have-not-changed-since-2018-with-the-exception-of-china/>
- Moyer-Gusé, E., & Nabi, R. L. (2010). Explaining the effects of narrative in an entertainment television program: Overcoming resistance to persuasion. *Human Communication Research*, 36(1), 26–52. <https://doi.org/10.1111/j.1468-2958.2009.01367.x>
- Ngai, C. S. B., & Jin, Y. (2016). The effectiveness of crisis communication strategies on Sina Weibo in relation to Chinese publics' acceptance of these strategies. *Journal of Business and Technical Communication*, 30(4), 451–494. <https://doi.org/10.1177/1050651916651907>
- Oukropec, J. (2025). Strategic Defense Against Misinformation: Empirical Findings on Paracrisis Responses. *International Journal of Strategic Communication*, 19(1), 54–85. <https://doi.org/10.1080/1553118X.2024.2382984>
- Park, K., & Rim, H. (2019). Social media hoaxes, political ideology, and the role of issue confidence. *Telematics and Informatics*, 36, 1–11. <https://doi.org/10.1016/j.tele.2018.11.001>

- Schmitt, B. H. (1994). Contextual priming of visual information in advertisements. *Psychology and Marketing*, 11(1), 1–14. <https://doi.org/10.1002/mar.4220110103>
- Schwarz, A., Seeger, M. W., & Auer, C. (2016). Significance and structure of international risk and crisis communication research: Toward an integrative approach. In A. Schwarz, M. W. Seeger, & C. Auer (Eds.), *The handbook of international crisis communication research* (pp. 1–10). Wiley Blackwell.
- Segev, S., Wang, W., & Fernandes, J. (2014). The effects of ad–context congruency on responses to advertising in blogs. *International Journal of Advertising*, 33(1), 17–36. <https://doi.org/10.2501/IJA-33-1-017-036>
- Shen, F., & Chen, Q. (2007). Contextual priming and applicability: Implications for ad attitude and brand evaluations. *Journal of Advertising*, 36(1), 69–80. <https://doi.org/10.2753/JOA0091-3367360105>
- Shen, F., Ahern, L., & Baker, M. (2014). Stories that count: Influence of news narratives on issue attitudes. *Journalism & Mass Communication Quarterly*, 91(1), 98–117. <https://doi.org/10.1177/1077699013514414>
- Shen, F., Sheer, V. C., & Li, R. (2015). Impact of narratives on persuasion in health communication: A meta-analysis. *Journal of Advertising*, 44(2), 105–113. <https://doi.org/10.1080/00913367.2015.1018467>
- Shen, H. (2016). Refining organization–public relationship quality measurement in student and employee samples. *Journalism & Mass Communication Quarterly*, 94(4), 994–1010. <https://doi.org/10.1177/1077699016674186>
- Sherif, M., & Hovland, C. I. (1961). *Social judgment*. Yale University Press.
- Silver, L., Devlin, K., & Huang, C. (2020). *Unfavorable views of China reach historic highs in many countries*. Pew Research Center. <https://www.pewresearch.org/global/2020/10/06/unfavorable-views-of-china-reach-historic-highs-in-many-countries/>
- Slater, M. D. (2002). Entertainment education and the persuasive impact of narratives. In M. C. Green, J. J. Strange, & T. C. Brock (Eds.), *Narrative impact: Social and cognitive foundations* (pp. 157–181). Lawrence Erlbaum Associates Publishers.
- Taher, A., El Kolaly, H., & Tarek, N. (2025). Examining crisis communication in geopolitical conflicts: The micro-influencer impact model. *Journalism and Media*, 6, 116. <https://doi.org/10.3390/journalmedia6030116>
- Tao, W., & Wilson, C. (2016). The impact of corporate communication on company evaluation: Examining the message effects of CSR, corporate ability, and hybrid strategies. *International Journal of Strategic Communication*, 10(5), 426–444. <https://doi.org/10.1080/1553118X.2016.1223671>
- Taylor, A. (2015, October 7). Sweden’s subtly radical “feminist” foreign policy is causing a stir. *The Washington Post*. <https://www.washingtonpost.com/news/worldviews/wp/2015/10/07/swedens-subtly-radical-feminist-foreign-policy-is-causing-a-stir/>
- U.S. Government Accountability Office. (2024). *U.S.–China relations*. <https://www.gao.gov/u.s.-china-relations>

- Utz, S., Schultz, F., & Glocka, S. (2013). Crisis communication online: How medium, crisis type and emotions affected public reactions in the Fukushima Daiichi nuclear disaster. *Public Relations Review*, 39(1), 40–46. <https://doi.org/10.1016/j.pubrev.2012.09.010>
- van Krieken, K., & Sanders, J. (2016). Diachronic changes in forms and functions of reported discourse in news narratives. *Journal of Pragmatics*, 91, 45–59. <https://doi.org/10.1016/j.pragma.2015.11.002>
- van Zoonen, W., & van der Meer, T. (2015). The importance of source and credibility perception in times of crisis: Crisis communication in a socially mediated era. *Journal of Public Relations Research*, 27(5), 371–388. <https://doi.org/10.1080/1062726X.2015.1062382>
- Veil, S. R., Reno, J., Freihaut, R., & Oldham, J. (2015). Online activists vs. Kraft foods: A case of social media hijacking. *Public Relations Review*, 41(1), 103–108. <https://doi.org/10.1016/j.pubrev.2014.11.017>
- Veil, S. R., Sellnow, T. L., & Petrun, E. L. (2012). Hoaxes and the paradoxical challenges of restoring legitimacy. *Management Communication Quarterly*, 26(2), 322–345. <https://doi.org/10.1177/0893318911426685>
- Vosoughi, S., Roy, D., & Aral, S. (2018). The spread of true and false news online. *Science*, 359(6380), 1146–1151. <https://doi.org/10.1126/science.aap9559>
- Wyer, R. S., & Srull, T. K. (1981). Category accessibility: Some theoretical and empirical issues concerning the processing of social stimulus information. In N. Cantor & J. Kihlstrom (Eds.), *Personality, cognition, and social interaction* (pp. 161–197). Lawrence Erlbaum Associates Publishers.
- Xu, J., & Wu, Y. (2015). Using Twitter in crisis management for organizations bearing different country-of-origin perceptions. *Journal of Communication Management*, 19(3), 239–253. <https://doi.org/10.1108/JCOM-06-2013-0050>
- Yang, S.-U., Kang, M., & Johnson, P. (2010). Effects of narratives, openness to dialogic communication, and credibility on engagement in crisis communication through organizational blogs. *Communication Research*, 37(4), 473–497. <https://doi.org/10.1177/0093650210362682>
- Yang, S.-U., Kang, M., & Cha, H. (2015). A study on dialogic communication, trust, and distrust: Testing a scale for measuring organization–public dialogic communication (OPDC). *Journal of Public Relations Research*, 27(2), 175–192. <https://doi.org/10.1080/1062726X.2015.1007998>
- Yi, Y. (1990). The effects of contextual priming in print advertisements. *Journal of Consumer Research*, 17(2), 215. <https://doi.org/10.1086/208551>
- Yi, Y. (1993). Contextual priming effects in print advertisements: The moderating role of prior knowledge. *Journal of Advertising*, 22(1), 1–10. <https://doi.org/10.1080/00913367.1993.10673391>

Beyond Interpersonal Abuse: Conceptualizing Organizational Gaslighting as a Strategic Communication Risk*

Elina Erzikova**

Central Michigan University

Kerk F. Kee*****

Texas Tech University

Shannon A. Bowen***

University of South Carolina

Mary Beth West*****

Independent Researcher

Lana Ivanitskaya****

Central Michigan University

Gaslighting, traditionally examined as a psychological phenomenon within intimate relationships, has recently gained interdisciplinary attention that extends into institutional and organizational contexts. In this paper, we define gaslighting as a risky strategic communication practice in which individuals or organizations deliberately obscure truth, discredit critics, manipulate narratives, and normalize doubt in order to maintain power and control. Normative ethical theory is used to explain the unethical nature of gaslighting as well as offer deontological (principled) ethical analysis tools and recommendations. This bibliometric study analyzed 235 academic publications indexed in the Web of Science to map the disciplinary distribution and thematic clusters of gaslighting research. Keyword co-occurrence mapping reveals a highly fragmented field—85% of journals (n = 195) published only a single article—yet concentrated scholarly activity exists in feminist philosophy, interpersonal violence, and education. Four thematic clusters emerged: (1) structural inequities and institutional contexts, (2) identity, power, and organizational life, (3) psychological traits, emotional harm, and interpersonal relations, and (4) abuse, control, and help-seeking. This study proposes Organizational Gaslighting as a Strategic Communication Risk (OGSCR) framework for understanding organizational gaslighting, encompassing four interrelated dimensions: power and hierarchy, narrative control, discrediting and delegitimization, normalization through internal culture, and stakeholder responses to issue or policy change. By reframing gaslighting as a risky communication practice, the study underscores how organizations may not only communicate about risks but also generate risks through their communication strategies. This dual nature—communication of risk and risk of communication—highlights the ethical, strategic, and societal stakes of gaslighting, making it a critical phenomenon for advancing debates in risk and strategic communication. This framework serves as a foundation for theoretical development and encourages cross-disciplinary collaboration.

Keywords: Gaslighting, Organizational gaslighting; Strategic communication; Bibliometric analysis; VOSViewer.

* Article submitted on 22/09/2025. Article accepted on 15/12/2025.

** erzikle@cmich.edu

*** sbowen@sc.edu

**** ivani11@cmich.edu

***** kerk.kee@ttu.edu

***** marybethwest@mymail.com

“You’re not going out of your mind, you’re slowly and systematically being driven out of your mind” – *Gaslight*, 1944 film by Metro-Goldwyn-Mayer starring Ingrid Bergman as an heiress gaslit for her fortune, starting the popular use of the term “gaslighting.”

Introduction

Gaslighting, once primarily associated with interpersonal relationships and psychological abuse, has in recent years gained attention as a broader sociocultural and organizational phenomenon (Sweet, 2019). Scholars have documented its manifestations in politics, healthcare, education, gender relations, and media, with a growing recognition that gaslighting can also occur at the institutional and corporate levels. In these contexts, gaslighting may involve the strategic distortion of facts, denial of lived experiences, or manipulation of narratives to protect organizational interests, often at the expense of stakeholders’ trust and well-being. This shift in scope—from a private, relational dynamics to a public, systemic issue—has opened new avenues for interdisciplinary inquiry.

Despite its expanding relevance, gaslighting scholarship remains conceptually fragmented and disciplinarily concentrated. Research is scattered across psychology, sociology, gender studies, medical ethics, and legal studies, with limited integration between fields (Erzikova et al., 2025). Moreover, the term “gaslighting” is often used without a consistent theoretical foundation, leading to varied definitions and conceptual ambiguities. While some studies focus on its rhetorical and communication dimensions, others examine its psychological impacts or sociopolitical implications. This lack of cohesion complicates the task of mapping the field’s development and identifying opportunities for theoretical and applied contributions.

For strategic communication—and particularly public relations—the study of gaslighting presents both a challenge and an imperative. Strategic communication is a discipline concerned with ethics, issues management, research and data-driven information management, stakeholder perception, relationships, media, and reputation management. The public relations function and its communication and message design activities occupy a central role in either perpetuating or resisting manipulative practices. Understanding how gaslighting is conceptualized and studied across disciplines can therefore inform the profession’s ethical boundaries, accountability mechanisms, and potential for societal responsibility and leadership.

For the purposes of this study, we define gaslighting as a strategic communication practice in which individuals or organizations abuse their power and position and violate ethical norms deliberately to obscure truth, discredit critics, manipulate narratives, and normalize doubt in order to maintain power and control. Such practices can set organizations on divergent paths: toward reform and ethical accountability, toward stakeholder disengagement and moral outrage, or toward a continued “dark” gaslighting strategy—an intentionally manipulative and high-risk approach that, while reinforcing short-term control,

ultimately heightens ethical and reputational risks and perpetuates cycles of institutional manipulation, creating myriad ethical concerns.

Ethical Theory and Concerns

Conceptualizing gaslighting as a communication practice involving some degree of specified yet veiled intent to deceive—often in ways that might be premeditated in advance by the communicator—also invites framing it as a form of high-risk communication that operates in direct contradiction to ethics and the public good. In this sense, gaslighting does not only distort reality for individuals or groups, but also generates reputational, relational, and ethical risks for organizations. By strategically manipulating uncertainty, organizations that engage in gaslighting create new hazards for public trust, stakeholder relationships, and long-term legitimacy. Rather than mitigating risks through transparent communication, gaslighting functions as a risk-amplifying strategy on multiple concurrent levels: it increases uncertainty through the purposeful obfuscation and/or confusion of facts; solidifies disinformational narratives; silences dissenting voices; and undermines the credibility of both institutions and their publics. This risk-laden dimension underscores why gaslighting must be studied as part of the broader field of risk and strategic communication.

Heath (2025) conceptualized strategic communication as the intent to minimize lies, mis/disinformation by distilling truth and seeking what he termed fact agency. He explained, “democracy requires shared issue contests of fact, value, policy, identity, identification, and place’ (p. 14). Ethical theory would add that we must have a shared understanding of the truth or an assumption of veracity. The principle of veracity, rather than being based in a single philosophy, is the foundation of most moral philosophy; it is based on rational engagement with fact (De George, 2010). Without an assumption of veracity, communication would break down and social organization would become impossible (Bowen, 2010; Kant, 1997). Veracity, conceived of as full disclosure, candor, honesty, and rationally examining the merit of multiple perspectives (Bowen, 2016). Obviously, gaslighting is antithetical to the principle of veracity.

The present study focuses on the normative moral philosophy of Kantian deontology as a theoretical basis for critiquing the practice of gaslighting and offering an analytical framework for examining its substance and recommending correction. A deontological framework, based on moral principle gained through reason holds absolute congruence with the principle of veracity (Scruton, 1982). Additionally, Kantian deontology (Kant, 1997/1799) is based on one’s duty to uphold the moral law, dignity and respect for all involved stakeholders, and acting from a basis of goodwill or good intent alone (as opposed to selfishness, bias, convention, or prejudice). These tenants of deontology, further explicated in the categorical imperative, are applied to gaslighting in the ethical implications section of this article.

Purpose of Bibliometric Study: Conceptualization

Our working definition draws from the Organizational Gaslighting as a Strategic Communication Risk (OGSCR) framework developed in this paper. We offer this brief definition at the outset to set the stage. The remainder of the paper elaborates this conceptualization in greater detail and situates it within broader scholarly debates and empirical findings.

This study addresses multi-disciplinary gaps in the literature through a bibliometric and thematic analysis of scholarly literature on gaslighting. Bibliometric analysis is regarded as one of the major methods for conducting literature reviews (Donthu et al., 2021). By mapping conceptual clusters, identifying temporal trends, and assessing the disciplinary distribution of publications, it offers a comprehensive overview of how the gaslighting concept has evolved and where it is heading. Four research questions guide the inquiry:

RQ1: What are the key conceptual clusters structuring scholarly discourse on gaslighting across disciplines, and how do these clusters inform strategic communication theory and public relations practice?

RQ2: How is *gaslighting* positioned within the conceptual network, and what thematic connections does it reveal across different domains of scholarship?

RQ3: How do the frequency of occurrence, average publication year, and average citation count of key terms enhance our understanding of the development and influence of gaslighting scholarship?

RQ4: How is the academic discourse on gaslighting distributed across different journals, and what does the concentration of publications reveal about the disciplinary focus and fragmentation of the field?

By answering these questions, the study seeks to advance both scholarly understanding and practical awareness of gaslighting, positioning it as a critical area for future research in communication ethics and organizational practice.

Method

This study employed a mapping approach utilizing bibliometric analysis and co-occurrence network visualization to explore the scholarly discourse on gaslighting and its implications for strategic communication. This method facilitated a systematic, data-driven examination of how key concepts related to gaslighting have emerged, clustered, and influenced academic literature across various disciplines (Ivanitskaya & Erzikova, 2023).

Data Collection

Data were retrieved from the Web of Science (WoS) Core Collection database using the search term *gaslighting*. WoS databases are frequently used in VOSviewer-based studies

as WoS provides valuable analytical metrics (Cabeza et al., 2020). Using a query included in Appendix 1, we identified studies that mentioned gaslighting in their titles or abstracts or used it as a keyword. Covering the full timespan available in the database up to August 2025, the search included English-language publications across all fields, excluding Construction Building Technology, which contained technical (engineering) uses of the search term.

The final dataset comprised 235 publications, including keywords, citation counts, and publication years (see Appendixes 2 and 3). Keywords that appeared in at least three publications in the dataset were included in the network analysis to ensure both relevance and interpretive clarity, resulting in a final total of 116 keywords included in the map (Fig. 1).

Data Analysis

The bibliometric data were analyzed using VOSviewer, a widely used tool for visualizing scientific fields (van Eck & Waltman, 2022; Visser et al., 2021). A co-occurrence analysis of keywords (both author keywords and index terms) was conducted to identify conceptual linkages within the dataset. Each node (circle) represented a keyword found in academic literature, with the size of the node indicating the keyword's frequency (how often it appears), and the lines between nodes showing co-occurrence relationships (how often the keywords appear together in documents). The colors represented clusters—groups of keywords that are thematically related.

Additionally, VOSviewer provided quantitative indicators for each keyword, including occurrences (the number of studies indexed with a keyword), average publication year (the mean year of publication for documents containing the keyword), and average citations (the average number of citations received by documents associated with each keyword).

Results

The results are presented in accordance with the study's research questions.

RQ1: What are the key conceptual clusters structuring scholarly discourse on gaslighting across disciplines, and how do these clusters inform strategic communication theory public relations practice?

The keyword co-occurrence analysis revealed four distinct thematic clusters, each representing a different but interrelated dimension of gaslighting scholarship. The four Clusters are revealed separately below. While the boundaries between clusters are porous,

each presents a coherent set of concepts, contexts, and methodological orientations (see Fig.1).

Cluster 1 (Red): Structural Inequities and Institutional Contexts

Cluster 1, the largest node with 69 keywords, centers on gaslighting as a phenomenon embedded in structural and institutional inequities. Keywords such as *access*, *barriers*, *disparities*, *structural racism*, *discrimination*, and *equity* seem to point to systemic conditions that shape individual experiences. Health-related terms (*health*, *health-care*, *long covid*, *pain*, *patient*, *diagnosis*, *medical gaslighting*, *mental-health*) suggest a strong emphasis on the medical domain, where marginalized groups encounter dismissive or delegitimizing treatment.

The presence of *critical race theory*, *intersectionality*, *racial gaslighting*, and *testimonial injustice* underscores scholarly interest in epistemic harm and the racialized dimensions of disbelief (Zembylas, 2024). Demographic identifiers (*adolescents*, *adults*, *children*, *youth*, *lgbtq plus*, *transgender*, *gay*, *women*) indicate the wide range of populations studied, while *social media* and *tiktok* suggest contemporary communication channels through which both gaslighting and counter-narratives emerged.

Overall, this cluster frames gaslighting as a sociostructural problem linked to systemic injustice, with particular attention to healthcare, institutional betrayal, and the compounding effects of intersecting identities.

Cluster 2 (Green): Identity, Power, and Organizational Life

Cluster 2, consisting of 20 keywords, focuses on the interplay between identity construction, power dynamics, and workplace or organizational environments. Terms such as *authentic leadership*, *management*, *autonomy*, *agency*, and *work* point toward organizational behavior and leadership studies. Identity-related keywords (*gender*, *men*, *self*, *authenticity*, *social epistemology*, *epistemic injustice*) might reflect management literature on the ways in which gaslighting undermines personal and professional credibility.

Workplace-specific harms are captured in *workplace gaslighting*, *harassment*, *sexual harassment*, *bullying*, and *sexism*, possibly illustrating sociological research on how the phenomenon manifests in professional hierarchies. The inclusion of *betrayal* and *construction* suggests an emphasis on relational breakdown and the social construction of meaning in organizational contexts. Scholarly approaches such as *autoethnography* highlight the role of lived experience and reflexivity in studying these dynamics. Collectively, this cluster seems to conceptualize gaslighting as a form of identity-based harm situated within organizational cultures and leadership practices.

Cluster 3 (Blue): Psychological Traits, Emotional Harm, and Interpersonal Relations

Cluster 3, with 17 keywords, captures the psychological and affective dimensions of gaslighting, linking it to harmful personality traits and emotional outcomes. The presence of *dark triad*, *narcissism*, *machiavellianism*, and *psychopathy* connects gaslighting to manipulative interpersonal strategies and personality disorders. Emotional and mental health terms (*anger*, *anxiety*, *depression*, *trauma*, *emotion*, *emotions*) indicate the profound affective consequences for targets. *Adolescent mental health* and *community* suggest that these impacts are studied not only at the individual level but also within broader social environments.

Concepts such as *feminism* and *politics* seem to highlight the ideological and advocacy contexts in which emotional harm is contested and addressed. Interpersonal relationships (*relationships*, *aggression*, *abuse*) remain a central focus, positioning gaslighting as both a personal and social pathology.

Cluster 4 (Yellow): Abuse, Control, and Help-Seeking

Cluster 4, the smallest with 10 keywords, is characterized by the language of abuse and coercion, positioning gaslighting within a broader continuum of interpersonal violence. Keywords such as *abuse*, *psychological abuse*, *sexual violence*, *domestic violence*, *coercive control*, and *intimate partner violence* firmly anchor this cluster in gender-based violence research largely associated with psychology. The presence of *help-seeking* indicates an applied orientation toward support systems and intervention strategies, while *epistemology* and *qualitative* point to theoretical and methodological engagement with how knowledge and meaning about harm are produced and validated. Terms like *transphobia* highlight the intersection of violence with marginalized identities. This cluster seems to present gaslighting as both a tactic of domination and a barrier to recognition and escape from abusive situations, with qualitative research approaches often employed to capture survivor perspectives.

Overall, these clusters indicate that gaslighting is not conceptualized as a singular or isolated behavior but as a complex, multidimensional phenomenon operating simultaneously at structural, organizational, interpersonal, and psychological levels. Cluster 1 (red), representing 58% of all keywords in the dataset, situates gaslighting within systemic inequities and institutional contexts, particularly in healthcare and social justice domains. Cluster 2 (green) extends the analysis into organizational life, where leadership, identity, and credibility are contested. Cluster 3 (blue) highlights the psychological and emotional dimensions, linking gaslighting to harmful personality traits and the affective harm experienced by targets. Cluster 4 (yellow) frames gaslighting within broader patterns of abuse and coercive control, drawing attention to barriers to help-seeking and epistemic struggles. The thematic interconnections across clusters suggest that effective interventions

and theoretical frameworks must integrate structural analysis, identity-based perspectives, psychological insights, and survivor-centered approaches.

The network visualization illustrates the interconnectedness among the four thematic clusters. Key terms such as *women* (65 links), *gender* (54), *health* (41), *intimate partner violence* (33), *intersectionality* (25), *racism* (24), *politics* (23), and *coercive control* (21), among others, serve as crucial links that unify all clusters. This interconnectedness highlights the complexity of gaslighting as a multifaceted topic by emphasizing the interplay among diverse perspectives and the potential for collaborative discourse among various disciplines, methodological perspectives, and stakeholders.

Importantly, terms directly related to strategic communication, public relations, or media discourse are absent across all clusters, indicating that gaslighting has yet to be systematically theorized as a deliberate communication or reputational strategy employed by organizations. The disciplinary emphasis remains strongly rooted in psychology, sociology, and feminist theory, with relatively limited engagement from organizational communication or communication ethics. This absence poses a risk for understanding the deceptive practice of gaslighting and the dire ethical implications of its use in the communication field.

From a strategic communication perspective, these findings highlight an important opportunity for the field to examine how communication strategies may perpetuate, legitimize, or disrupt gaslighting across contexts. Recognizing gaslighting as both discourse and power (as both a strategy and a tactic) broadens the ethical responsibilities of practitioners, requiring careful discernment of narrative control and truth-telling, as well as a balance between organizational reputation management and stakeholder well-being. The cross-cluster connections point to the desperate need for ethical communication strategies that actively counter epistemic injustice, strengthen credibility for marginalized voices, and foster honest, responsible communication and accountable dialogue within and beyond organizational boundaries. In conclusion, RQ1 finds that numerous fields discuss gaslighting in four clustered conceptual ways, yet the communication fields (strategic, organizational, and public) have little academic perspective on the topic, despite the overwhelming risk of unethical and propagandistic uses of gaslighting.

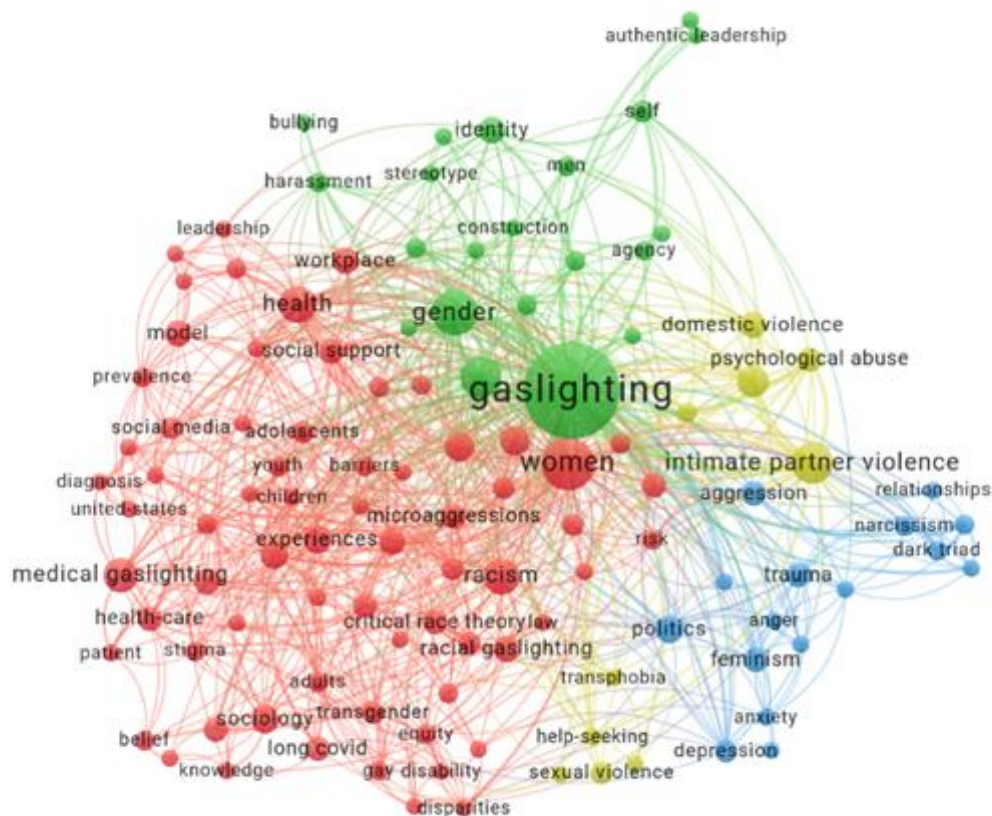


Figure 1- A keyword co-occurrence map for 235 articles from 195 academic journals included into the study. Available online: <https://tinyurl.com/29tckt7t> as an interactive map with selectable clusters.

RQ2: How is gaslighting positioned within the conceptual network, and what thematic connections does it reveal across different domains of scholarship?

The bibliometric network visualization places *gaslighting* (98 links) at the center of a dense web of thematic connections, confirming its role as a cross-disciplinary concept. The term exhibits strong co-occurrence with *women*, *gender*, *workplace*, *intimate partner violence*, *medical gaslighting*, and *racism*, indicating that scholarly attention extends well beyond its original interpersonal abuse context.

Its linkage to *women* and *gender* points to the sustained association between gaslighting and gendered power dynamics. The frequent co-occurrence with *workplace*, *bullying*, and *harassment* suggests that the concept is increasingly applied to organizational and professional environments. The connection with *medical gaslighting*, *health*, and *health care* underscores its relevance in patient-provider relationships, particularly in contexts where patient credibility is undermined.

Other prominent links—such as *intimate partner violence*, *psychological abuse*, *narcissism*, *trauma*, and *depression*—indicate continued scholarly engagement with

gaslighting's psychological and relational harm. Simultaneously, associations with *racism*, *racial gaslighting*, *politics*, and *equity* reveal that researchers are examining it as a systemic and sociopolitical phenomenon.

Only a few terms strongly associated with strategic communication ethics appeared in the gaslighting – *epistemic injustice* (green) cluster. Those terms were *autonomy*, *agency*, *authenticity*, and *authentic leadership*. As these terms are well conceptualized within the communication literature, those studies would offer a fruitful ground for conceptualizing gaslighting studies in our own field.

RQ3: How do the frequency of occurrence, average publication year, and average citation count of key terms enhance our understanding of the development and influence of gaslighting scholarship?

The results reveal key temporal and citation-based patterns that improve our understanding of the development and influence of gaslighting-related scholarship. By examining frequency of occurrence, average publication year, and citation metrics (both raw and normalized), we can trace how the conceptualization of the term has evolved, identify mature versus emerging areas of inquiry, and assess the scholarly impact of specific conceptual domains.

Frequency Signals Conceptual Centrality

The most frequently occurring term is *gaslighting* itself, with 93 mentions, affirming its centrality in the literature. High-frequency terms also include *women* (28), *gender* (21), and *epistemic injustice* (17), indicating that gender-based, philosophical, and identity-related perspectives have become dominant frames in recent discourse (see Appendix 2).

Additionally, *intimate partner violence* (17), *medical gaslighting* (12), *racism* (13), and *coercive control* (10) rank highly in frequency, suggesting a growing body of work that links gaslighting to systemic abuse and health disparities. These clusters bridge personal experience with institutional critique—highlighting the concept's expansion from interpersonal abuse to structural injustice.

Recent Publication Dates Reflect a Rapidly Emerging Field

The average publication by year across keywords increased between 2022–2024, signaling a field that is both new and rapidly expanding. For example, high-frequency and high-impact concepts such as *epistemic injustice* (avg. year: 2022.6), *intimate partner violence* (2023.3), and *medical gaslighting* (2023.7) have all gained momentum very recently.

Moreover, newer terms like *workplace gaslighting* (2025), *nurses* (2025), and *adults* (2024.8) point to future directions in professional contexts and adult health systems, underscoring how the research is diffusing into applied and occupational domains. In contrast, some foundational or critical terms such as *autoethnography* (2021), *corruption* (2020.7) and *transphobia* (2020.3) date earlier, suggesting that they may have catalyzed early academic interest in structural gaslighting and intersectional analyses.

Citation Metrics Highlight Both High-Impact and Emerging Areas

Citation averages and normalized citation counts reveal the scholarly influence of specific subtopics. For example, *help-seeking* (avg. citations: 113.7; normalized: 2.11) and *autonomy* (avg. citations: 65.3; normalized: 2.04) have exceptionally high citation averages, suggesting that foundational literature on personal or moral agency and resistance in abusive contexts has significantly influenced subsequent research. In addition, *domestic violence* (avg. citations: 51.3; normalized: 2.11) and *coercive control* (36.4; normalized: 1.61) reflect strong, ongoing interest in gaslighting's role within abusive relationships.

Furthermore, emerging terms with high normalized citation scores—such as *nurses* (normalized: 4.52), *betrayal* (4.85), *workplace gaslighting* (3.10), and *dark triad* (3.24) [narcissism, selfishness, and psychopathy]—signal new but already impactful directions in applied research, particularly at the intersection of personality psychology, labor environments, and professional ethics. Finally, terms with low average citation counts but recent publication dates, such as *social media* (0.31), *performance* (0.13), and *leadership* (0.11), may represent nascent areas of study that require time to accrue impact.

Normalized Citation Counts Help Identify Punch-Above-Weight Topics

Normalized citations adjust for recency bias and spotlight concepts that are disproportionately influential relative to their age. For instance, *knowledge* (normalized: 2.69), *care* (2.60), *trauma* (2.42), and *psychological abuse* (2.41) all carry high normalized scores, indicating that despite being recent, these terms are conceptually rich and highly cited relative to peers. Additionally, terms like *machiavellianism* (3.98) and *psychopathy* (3.24) reflect growing interest in linking gaslighting to dark personality traits, further broadening the field's interdisciplinary relevance to psychology and leadership studies.

By examining frequency, recency, and citation impact, this analysis suggests that gaslighting scholarship is experiencing an expansion phase, both theoretically and empirically. High-frequency and high-citation terms reveal the field's conceptual anchors, particularly within the contexts of gender, health, and harm. Meanwhile, newer and increasingly cited terms—such as *workplace gaslighting*, *autonomy*, *epistemic injustice*, and *betrayal*—suggest that gaslighting has become a strategic concept, instrumental for

understanding contemporary power, identity, and resistance in both interpersonal and institutional communication.

RQ4: How is the academic discourse on gaslighting distributed across different journals, and what does the concentration of publications reveal about the disciplinary focus and fragmentation of the field?

The analysis encompassed 195 distinct journals that published scholarly articles with *gaslighting* as a keyword (see Appendix 4). Among these, a substantial majority—166 journals, or 85%—contributed only a single article, indicating a wide but shallow distribution of gaslighting-related research across academic outlets. A small subset of journals demonstrated higher publication activity on this topic. For instance, *Hypatia: A Journal of Feminist Philosophy* and the *Journal of Interpersonal Violence* each published five articles, representing the highest counts at approximately 2.1% of the total corpus. Other journals with notable contributions included the *International Journal of Qualitative Studies in Education* with four articles, and several others such as *Journal of Family Violence*, *Leadership*, and *SSM Qualitative Research in Health* with three articles each.

This pattern suggests that while gaslighting is a topic of interest in diverse fields, it has particular traction in feminist philosophy, interpersonal violence, education, and leadership research. Beyond these few journals with multiple contributions, the majority of publications are dispersed singly across a broad spectrum of journals spanning disciplines such as philosophy, psychology, health communication, sociology, gender studies, and political science, among others. This dispersion reflects an interdisciplinary engagement with gaslighting but also signals that the field remains relatively fragmented, without a primary disciplinary home. The term is notably absent as a keyword for articles in journals representing public relations, strategic communication, business ethics, public diplomacy, propaganda, journalism, and media ethics.

Discussion

The bibliometric mapping shows that gaslighting occupies a central and integrative position in contemporary scholarship, bridging psychological, organizational, medical, and sociopolitical domains. Its strongest links—with terms such as *women*, *gender*, *workplace*, *medical gaslighting*, *intimate partner violence*, and *racism*—reflect both the breadth of its application and the persistence of foundational themes. While research retains the concept's roots in interpersonal abuse, it now extends into systemic contexts where power, control, and credibility are contested.

The prominence of *women* and *gender* underscores the gendered nature of gaslighting, often framed as an abuse tactic embedded in patriarchal structures to undermine agency,

silence dissent, and delegitimize lived experiences. This framing aligns with early psychological studies and feminist theoretical contributions that situate gaslighting within larger systems of oppression. It should also be noted that the public relations field – both domestically in the United States and internationally – is female-dominated in lower and mid-level position by some 60-70%, while senior and executive ranks of the industry skew male (Grunig, L., Grunig, J., & Dozier, 2002). How gender-driven factors influence organizational gaslighting occurrence, particularly in various sectors of the industry, is due for in-depth study and analysis, given the field's gender skew.

Workplace-related associations—such as *harassment*, *bullying*, and *leadership*—signal growing recognition of gaslighting in professional contexts. Here, it may function as a managerial or peer-to-peer strategy to control narratives, obscure accountability, or discredit employee perspectives. This adaptation of the concept beyond intimate relationships suggests fertile ground for research in organizational communication and leadership ethics.

Connections to *medical gaslighting* highlight a significant shift toward examining the phenomenon as an epistemic injustice in healthcare. Patients' accounts are dismissed, reframed, or pathologized—often resulting in delayed diagnoses, inadequate care, or further marginalization, particularly for women, racial minorities, and those with contested or chronic illnesses.

The clustering of terms such as *racism*, *racial gaslighting*, *politics*, and *equity* further positions gaslighting as a systemic and sociopolitical tool. Racial gaslighting, for example, can deny racism's existence, invalidate discrimination claims, or manipulate public discourse to maintain dominant narratives. Despite these expansions, psychological terms such as *intimate partner violence*, *psychological abuse*, *narcissism*, *trauma*, *anxiety*, and *depression* remain closely linked, underscoring the concept's enduring resonance in mental health research.

However, gaslighting research remains fragmented across journals, often appearing as single-article contributions. While outlets like *Hypatia* and *Journal of Interpersonal Violence* anchor the concept in feminist theory and abuse studies, its interdisciplinary spread lacks a cohesive scholarly community. This dispersion mirrors gaslighting's multifaceted nature but also hampers theoretical integration. Greater cross-disciplinary dialogue could strengthen conceptual clarity, advance methodological approaches, and establish firmer disciplinary foundations.

Building on these findings, we advance a framework that reframes organizational gaslighting not simply as a set of communication tactics but as a series of risk dynamics within strategic communication. This risk-oriented model highlights how gaslighting practices actively generate ethical, relational, and legitimacy hazards for organizations and their stakeholders.

From Interpersonal Pathology to Institutional Strategy

While institutional contexts appear in the network—e.g., *leadership*, *bullying*, *trust*—there is a gap in literature on how organizations themselves enact gaslighting to preserve legitimacy, deflect criticism, and suppress dissent. The lack of exploration limits our understanding of the communication strategies institutions use to shape perception and maintain authority—a phenomenon with direct implications for strategic communication ethics.

To address this gap, we propose the Organizational Gaslighting as a Strategic Communication Risk (OGSCR) framework, which synthesizes insights from organizational communication, moral philosophy, public relations ethics, psychology, and critical management studies. Traditionally understood in psychological contexts as interpersonal manipulation that causes individuals to doubt their perceptions, memory, or sanity, gaslighting can also be systemic—embedded in errant or absent policies, leadership behaviors of misconduct, ethically compromised strategic messaging, and low-accountability with high-dysfunction organizational cultures. In this organizational form, it involves the deliberate use of communication to obscure truths, shift blame, manipulate policy, and protect reputational power or position.

The OGSCR Framework: Organizational Gaslighting as Communication Risk

To address the communication and ethical hazards revealed in our analysis, we propose the Organizational Gaslighting as a Strategic Communication Risk (OGSCR) framework, which reconceptualizes gaslighting as a set of interrelated risk dynamics rather than a legitimate strategic tool. The OGSCR Framework identifies several interrelated dimensions through which organizational gaslighting may manifest, offering a lens to analyze both internal and external communication practices (Fig. 2). Importantly, we recognize that labeling gaslighting as “strategic communication” may appear to frame a harmful and unethical practice in terms that are often associated with purposeful, legitimate, or even positive organizational action. Instead, our intention is to critically locate it within the scholarly field of strategic communication, where planned communication choices—ethical or unethical—shape organizational relationships, legitimacy, and risk. Therefore, we deliberately frame this model a strategic communication risk to avoid implying that gaslighting is a legitimate or constructive form of strategy.

Power and Hierarchy: Gaslighting often originates from individuals or groups in positions of authority who exploit hierarchical structures to suppress opposition, limit alternative viewpoints, and maintain institutional control. For example, a senior executive may override committee recommendations without explanation, signaling that dissent is unwelcome.

Narrative Control: Organizations strategically reshape or obscure narratives through selective disclosure, reframing, or omission, protecting institutional legitimacy at the expense of honesty. Annual reports, press releases, or public statements may highlight successes while minimizing or omitting failures, shaping stakeholder perceptions.

Discrediting and Delegitimization: Those who challenge the status quo may be portrayed as emotional, irrational, or incompetent, diminishing their credibility and discouraging further opposition. Whistleblowers, for instance, may be labeled as troublemakers, overly sensitive, or even as participants in misconduct themselves under false accusations, undermining their influence within the organization (Greenwood, 2025).

Normalization through Internal Culture: Gaslighting behaviors become routine and culturally embedded when repeated, justified, or ignored. Over time, employees and stakeholders may internalize doubt, self-censor, or rationalize unethical practices, creating a self-reinforcing cycle that – over time – manifests a deeply engrained culture of complicity, resignation, and worst of all, normalized corruption.

Issue/Policy Change/Ethical Analyses: When issues, policies, or ethical dilemmas arise, they enter a stage where stakeholders engage in interpretation, redefinition, questioning, experience moral objection, or make decisions about acceptance or exit. This phase functions as a critical juncture, determining whether the organization pursues corrective action, experiences stakeholder withdrawal, or escalates into further manipulation.

Reform, Regulatory, or Ethical Sanctions/Leadership: One possible outcome is corrective intervention. Organizations may adopt reforms, face regulatory oversight, or undergo leadership changes to address shortcomings and restore legitimacy. This path emphasizes accountability and the potential for ethical renewal.

Potential Exit from Relationship (Moral Outrage): Another outcome involves stakeholders disengaging from the organization due to moral or ethical concerns. Such exits may be driven by outrage, disillusionment, or loss of trust, resulting in reputational damage and weakened relational ties.

Gaslighting Strategy (Dark): A third trajectory is the deliberate pursuit of a high-risk, manipulative strategy. This *Gaslighting Strategy (Dark)* seeks to reinforce short-term control but carries profound ethical and reputational risks. The model highlights that this path feeds back—represented by a dotted line—into *Power and Hierarchy*, perpetuating cycles of institutional control and manipulation.

The OGSCR framework serves both scholarly and practical purposes. Researchers can use it to identify and categorize instances of organizational gaslighting, while practitioners may apply it as an audit tool to detect harmful communication patterns, guide ethics training programs, and promote accountability. The framework draws on organizational

communication, social psychology, and ethics literature to conceptualize how systemic gaslighting functions as a mechanism of control within and beyond organizations.

Organizational Gaslighting in Strategic Communication Risk (OGSCR) Framework

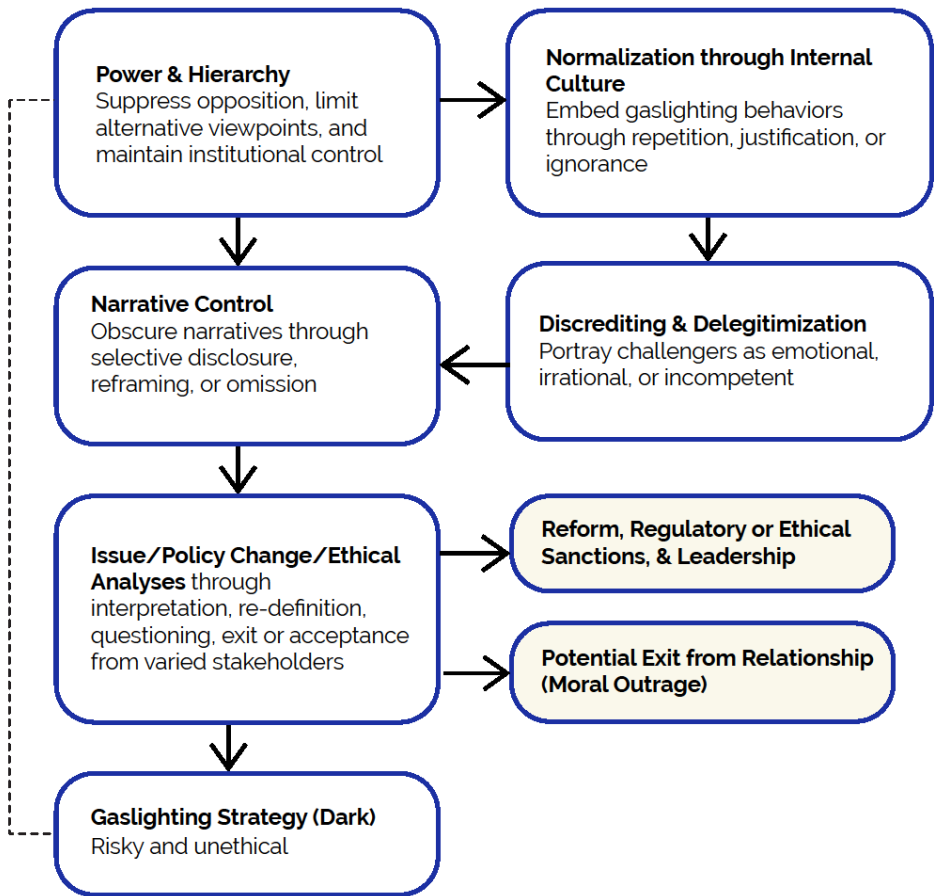


Figure 2 - The Organizational Gaslighting as a Strategic Communication Risk (OGSCR) Framework.

Ethical Implications for Strategic Communication

Recognizing gaslighting as an institutional strategy reframes ethical debates in strategic communication, public relations management, and organizational communication. For example, narrative control, while central to crisis management, crosses ethical boundaries when it intentionally and with premeditation denies responsibility, distorts stakeholder experiences, or conceals systemic harm. Suppression of dissent can be disguised as miscommunication, while performative CSR or DEI messaging may function as gaslighting when progressive rhetoric is used to deflect criticism or marginalize whistleblowers.

AI-assisted communication compounds these risks, as persuasive automation can blur the line between perception and deception (Bowen, 2024; Bowen & Erzikova, 2025; Bowen & Pashentsev, 2025). These practices threaten moral responsibility, accountability, and mutual understanding—core values of ethical strategic communication.

To further guide the implementation of ethical correction when the OGSCR Framework detects gaslighting, the Kantian ethical consideration triangle (Bowen, 2004) can be used to guide issues management and symmetrical problem resolution, with both internal and external stakeholders. In this ethical consideration triangle, the categorical imperative's three forms designed by philosopher Immanuel Kant (1785/1964) offer tests that can be considered with each of the publics and stakeholders with whom an organization bears responsibility. The triangle asks decision makers to consider:

1. Duty to moral principle – what is the primary principle to be upheld by rationality?
2. Dignity and respect – what option would maintain these for all involved publics and stakeholders, in a reversible and non-exploitative manner?
3. Intention or good will – is the action taken from these motivations alone, as incorruptible? (Kant, 1785/1964 as adapted from Bowen, 2004).

These moral guidelines, or categorical requirements for ethical behavior, can then be used to create decision options for symmetrical discussion with internal and external stakeholders, creating an ethical approach (Bowen, 2004). Over time, the use of this ethical consideration triangle can help to ingrain ethics in the organizational culture and to enact its use by all levels of employees (Bowen, 2025). Practically, the OGSCR framework can again be used to assess the organizational culture to see the change resulting from ethical analyses and to determine if further training and enculturation is needed to keep gaslighting from re-emerging. One would expect stakeholders with suspicion or moral objections to exit the relationship, or perhaps engage in hearings or regulatory actions, more risks of the unethical gaslighting strategy. Governmental action, regulation, media attention, activist communications, and other negative repercussions may spur a change in leadership or for more ethical accountability. Yet, when negative repercussions of gaslighting are absent, it can become a repeated and unethical dark strategy used in applied settings due to its effectiveness. This use of gaslighting as a strategy is not normative and remains unethical.

Expanding the Strategic Communication Research Agenda

Integrating organizational gaslighting into strategic communication research enables more thorough analysis of communication power, particularly regarding epistemic injustice, organizational silence, and manipulative discourse. This perspective challenges the assumption that strategic communication is inherently constructive, positioning it instead as a contested terrain where strategy and messages can both foster trust and distort reality. A more ethically attuned scholarship must acknowledge that communication strategies can be

enabling or oppressive—and that unexamined practices risk reinforcing the very power asymmetries they claim to address.

In this context, leadership becomes pivotal, as Erzikova and Martinelli (2024) argued: leaders who act as moral entrepreneurs can set ethical precedents, champion transparency, and resist manipulative organizational norms. Key moral entrepreneurship traits—such as moral courage to challenge unethical practices, a visionary commitment to societal well-being, and the ability to mobilize others around shared ethical values—are essential for counteracting gaslighting dynamics. By openly naming harmful communication patterns, reframing organizational narratives toward honesty, and developing cultures of accountability, such leaders ensure that strategic communication serves the public good rather than entrenched power. Using a deontological ethical approach invests leaders with humility and the ability to listen to others (Place & Bowen, 2025).

Limitations and Future Research Directions

This study presents a new conceptual framework for organizational gaslighting in strategic communication but has several limitations. It is based on bibliometric analysis, which may overlook emerging or marginalized perspectives, particularly those outside mainstream academic databases. Furthermore, reliance on English-language sources narrows the cultural perspective, potentially missing variations across geopolitical or linguistic contexts. The framework remains conceptual and has not been empirically tested in real organizational contexts, limiting its direct practical application. While keyword co-occurrence analysis provides a broad overview of scholarship, it does not capture the lived experiences of individuals subjected to gaslighting or the complex, evolving nature of institutional rhetoric.

To advance research on organizational gaslighting, future studies should conduct in-depth case studies examining how gaslighting manifests in specific contexts such as crisis communication, workplace culture, or brand management, and how it affects stakeholder relationships over time. Comparative research across sectors—including corporate, nonprofit, and governmental organizations—could clarify how institutional logics or accountability structures influence gaslighting behaviors. Methodologically, integrating critical discourse analysis and media framing analysis will help uncover organizational rhetorical strategies that distort reality and maintain control. Research should also explore intersections with social identities—race, gender, ability—to understand how gaslighting contributes to structural inequities within communication systems. Finally, investigating long-term reputational risks such as erosion of public trust, diminished employee morale, and threats to brand legitimacy will offer important insights for ethical strategic communication practice. In other words, future research should build on this foundation by explicitly examining gaslighting as a risk communication problem.

Conclusion

Taken together, the bibliometric mapping and the OGSCR framework reveal that while gaslighting scholarship is highly fragmented, its most significant theoretical advance lies in recognizing gaslighting as a risky communication practice. The bibliometric analysis demonstrates how existing research emphasizes psychological and interpersonal harm but largely omits the communication risks posed by organizational gaslighting. The OGSCR framework addresses this omission by theorizing gaslighting as a communication practice that generates ethical, relational, and reputational risks for organizations and their stakeholders. This integration provides a clearer throughline from empirical patterns to conceptual development, aligning the study with ongoing debates in risk and strategic communication.

The scholarly discourse on gaslighting highlights its harmful effects across intimate relationships, health care, and racial injustice. However, this keyword co-occurrence analysis reveals that gaslighting remains predominantly framed as a psychological or interpersonal issue, with little attention to strategic communication, organizational behavior, or institutional narrative control—exposing a critical theoretical and practical gap.

This study addresses that gap by proposing a new conceptual framework: Organizational Gaslighting as a Strategic Communication Risk. Grounded in interdisciplinary theory, the framework redefines gaslighting as a deliberate organizational practice to obscure truth, discredit critics, and uphold reputational power. It outlines four dimensions—Power and Hierarchy, Narrative Control, Discrediting and Delegitimization, and Cultural Normalization—that apply to both internal dynamics (e.g., employee relations) and external messaging (e.g., crisis communication). Corrective action can be taken to institutionalize ethics using a Kantian framework with the ethical consideration triangle applying normative ethics of duty, dignity and respect, and good intention to stakeholder relationships, both internally and externally, and engaging in symmetrical dialogue.

By expanding the scope of gaslighting scholarship, this framework opens new avenues for research in communication, public relations, ethics, and media studies. Future research can use this framework to analyze organizational gaslighting across various sectors and media, enhancing our understanding of how manipulation operates in institutional contexts, including both internal dynamics and external interactions.

Biographical Notes

Elina Erzikova, Ph.D., is a Professor of Public Relations at Central Michigan University. She received her Ph.D. from the University of Alabama and is a Fellow of the Plank Center for Leadership in Public Relations at the same institution. Her primary areas of research interest include the relationships between media and power, as well as public relations ethics and leadership. She has published in journals such as *Political Communication*, *Mass Communication and Society*, *Journalism Studies*, *Journalism Practice*, *Public Relations Review*, *Journal of Strategic Communication*, and *International Communication Gazette*. In 2020, she co-authored the book *Russian Regional Journalism: Struggle and Survival in the Heartland*, published by Peter Lang Publishing. Additionally, Erzikova co-edited the *Handbook of Innovations in Strategic Communication: AI, Futurism and Directions* with Shannon A. Bowen, which brings together leading international scholars to explore how technological advances are transforming communication, ethics, and society.

Shannon A. Bowen, Ph.D., (University of Maryland) is a Professor at the University of South Carolina where she founded, along with Dr. Erzikova, the Global Strategic Communication Consortium to study ethics and futurism in the strategic communication, resulting in the *Handbook of Innovations in Strategic Communication: AI, Futurism and Directions* (2025). Bowen won numerous awards and top papers, and served as principle investigator on grants from the National Science Foundation, the International Association of Business Communicators, Arthur W. Page Center, and others. Her research areas are applied ethics, artificial intelligence (AI), internal communication, and issues management. She has several books and more than one hundred publications, many catalogued under ORCID: 0000-0001-7675-5002

Dr. Lana Ivanitskaya is an Industrial and Organizational Psychologist at Central Michigan University who applies rigor and methods of psychology to the scientific study of the workplace, specializing in healthcare organizations and systems. As a methodologist, she studies applications of custom network visualizations to bibliometric and social media analyses. In addition, her research involves healthcare workforce, health planning, and health system study from the patient perspective. She has published in journals such as *Journal of Occupational and Organizational Psychology*, *Journal of Organizational Behavior*, *Informatics*, *Gerontologist*, *Health Information Management Journal*, and *Journal of Health Care for the Poor and Underserved*.

Kerk F. Kee, Ph.D., (University of Texas at Austin), is the Virginia & Choc Hutcheson Professor in Mass Communication at Texas Tech University. His work centers on the diffusion of innovation and the dissemination of information in organizational, science, health, and environmental contexts. He has published in journals such as *Communication Research*, *Journal of Computer-Mediated Communication*, *New Media & Society*, *International Journal of Information Management*, and *International Journal of Communication*, to name a few. His research has been funded by the US National Science Foundation, Academy of Sciences, Bill & Melinda Gates Foundation, and the Robert Wood Johnson Foundation.

Mary Beth West is an independent researcher with 30 years of PR experience. She is the founder of the #PRethics Community on LinkedIn. The London, U.K.-based Public Relations and Communications Association (PRCA) named Mary Beth as a Fellow (FPRCA) in summer 2022, following her prior appointment as a founding member and co-chair of the PRCA's Global Ethics Council and as a member of the PRCA's Global Advisory Council. Mary Beth's commentary has been included in the *Wall Street Journal*, *Harvard Business Review*, the *USA Today* Network, *Business Insider*, *PRweek*, *PR News*, *Strategic*, and other business publications. She has spoken at international conferences and forums on public relations ethics and other timely topics in cities around the world, including in Belgrade, Davos, Dubai, Paris, Vienna and Warsaw.

References

- Bowen, S. A. (2004). Expansion of ethics as the tenth generic principle of public relations excellence: A Kantian theory and model for managing ethical issues. *Journal of Public Relations Research*, 16(1), 65-92. https://doi.org/10.1207/s1532754xjpr1601_3
- Bowen, S. A. (2010). The nature of good in public relations: What should be its normative ethic? In R. L. Heath (Ed.), *The Sage handbook of public relations* (pp. 569-583). Sage. <https://doi.org/10.4135/9781452276236.n167>
- Bowen, S. A. (2016). Clarifying ethics terms in public relations from A to V, authenticity to virtue. *Public Relations Review*, 42, 564-572. <https://doi.org/10.1016/j.pubrev.2016.03.012>
- Bowen, S. A. (2024). "If it can be done, it will be done:" AI ethical standards and a dual role for public relations. *Public Relations Review*, Special issue on Artificial Intelligence (AI), 50(5), 1-13. <https://doi.org/10.1016/j.pubrev.2024.102513>
- Bowen, S. A. (2025). Enacting ethics beyond institutionalization: Creating ethical culture, leadership, and employee communication. In S. Kim, P. M. Buzzanell, A. Mazzei, & J. N. Kim (Eds.), *The Routledge handbook of employee communication and organizational processes* (pp. 164 – 173). Routledge. <https://doi.org/10.4324/9781003415619>
- Bowen, S. A., & Erzikova, E. (2025). Devolving ethical standards of 2050: International and disciplinary failure. In A. V. Laskin & K. Freberg (Eds.), *Public relations and strategic communication in 2050: Trends shaping the future of the profession* (pp. 49 - 60). Routledge. <https://doi.org/10.4324/9781003426653>
- Bowen, S. A., & Pashentsev, E. (2025). AI, biases, and cultures: Values in a machine world. In Y. Cheng & D. Verčič (Eds.), *AI and strategic communication: Navigating the future* (pp. 251-269). Wiley. <https://doi.org/10.1002/9781394275434.ch12>
- Cabeza, L. F., Chàfer, M., & Mata, É. (2020). Comparative Analysis of Web of Science and Scopus on the Energy Efficiency and Climate Impact of Buildings. *Energies*, 13(2), 409. <https://doi.org/10.3390/en13020409>
- De George, R. T. (2010). *Business ethics* (7th ed.). Prentice Hall.
- Donthu, N., Kumar, S., Mukherjee, D., Pandey, N., & Lim, W. M. (2021). How to conduct a bibliometric analysis: An overview and guidelines. *Journal of business research*, 133, 285-296. <https://doi.org/10.1016/j.jbusres.2021.04.070>
- Erzikova, E., & Martinelli, D. (2024). Public relations professionals' views of moral entrepreneurship: An exploration of the concept, its benefits and barriers. *Corporate Communications: An International Journal*, 29(4), 607-625. <https://doi.org/10.1108/CCIJ-02-2023-0020>
- Erzikova, E., Bowen, S. A., Ivanitskaya, L., Kee, K., & West, M. B. (2025, May 13). *Gaslighting in the workplace: Ethical implications for public relations professionals*. Paper presented at the Global Strategic Communication Consortium Conclave, Sardinia, Italy.

- Greenwood, C. A. (2025). Employee whistleblowing. In S. Kim, P. M. Buzzanell, A. Mazzei, & J. N. Kim (Eds.). *The Routledge handbook of employee communication and organizational processes* (pp. 140 – 150). Routledge.
- Grunig, L., Grunig, J. & Dozier, D. (2002). *Excellent public relations and effective organizations: A study of communication management in three countries*. Erlbaum.
- Heath, R. L. (2025). Strategic communication beyond the amphitheater: Future strategic issues management. In S. A. Bowen & E. V. Erzikova (Eds.), *Handbook of innovations in strategic communication: AI, futurism and directions* (pp. 13 - 26). Edward Elgar.
- Ivanitskaya, L. & Erzikova, E. (2023). Social construction of public relations knowledge by academic institutions worldwide: A bibliometric network analysis. *Public Relations Review*, 49(4). <https://doi.org/10.1016/j.pubrev.2023.102361>
- Kant, I. (1964). *Groundwork of the metaphysic of morals* (H. J. Paton, Trans.). Harper & Row. (Original work published 1785)
- Kant, I. (1997). *Critique of practical reason* (M. Gregor, Trans.). Cambridge University Press. (Original work published 1799)
- Kant, I. (1997). *Groundwork of the metaphysic of morals* (T. E. M. Gregor, Trans.). Cambridge. (Original work published 1785)
- Place, K., & Bowen, S. A. (2025). Ethics in strategic communication: Humility as a Deontological cornerstone of future communication management. In S. A. Bowen & E. Erzikova (Eds.), *Handbook of innovations in strategic communication AI, futurism, and directions* (pp. 53-76). Elgar.
- Scruton, R. (1982). *Kant*. Oxford University Press.
- Sweet, P. L. (2019). The sociology of gaslighting. *American Sociological Review*, 84(5), 851–875. <https://doi.org/10.1177/0003122419874843>
- van Eck, N.J., & Waltman, L. (January, 2022). VOSviewer Manual 1.6.18. CWTS, University of Leiden. Retrieved from https://www.vosviewer.com/documentation/Manual_VOSviewer_1.6.18.pdf.
- Visser, M., van Eck, N.J., & Waltman, L. (2021). Large-scale comparison of bibliographic data sources: Scopus, Web of Science, Dimensions, Crossref, and Microsoft Academic. *Quantitative Science Studies*, 2(1), 20–41. https://doi.org/10.1162/qss_a_00112
- Zembylas, M. (2024). Racial gaslighting as affective injustice: a conceptual framework for education. *Race Ethnicity and Education*, 27(4), 417–431. <https://doi.org/10.1080/13613324.2024.2306680>

Appendix 1. Search query

Gaslighting (Topic) and Construction Building Technology (Exclude – Web of Science Categories) and Letter or Art Exhibit Review or News Item or Review Article or Meeting Abstract or Book Review or Editorial Material (Exclude – Document Types) and English (Languages)

Appendix 2. Gaslighting articles by year

<i>Year</i>	<i>Number of Articles</i>	<i>Percentage</i>
1981	1	0.4
1988	1	0.4
2003	1	0.4
2007	1	0.4
2014	2	0.9
2017	2	0.9
2018	6	2.6
2019	5	2.1
2020	13	5.5
2021	17	7.2
2022	31	13.2
2023	39	16.6
2024	56	23.8
2025	60	25.5
Total	235	100.0

Appendix 3. Descriptive statistics for top keywords (n=116) in research publications indexed in Web of Science

<i>Label</i>	<i>Cluster</i>	<i>Occurrences</i>	<i>Avg. pub. year</i>	<i>Avg. citations</i>	<i>Avg. norm. citations</i>
1 access	1	4	2023	5.25	0.9278
2 adolescents	1	5	2023.2	13.2	1.7107
3 adults	1	4	2024.75	0.25	0.0791
4 barriers	1	4	2023.25	4	0.5964
5 belief	1	4	2023	4.75	0.723
6 care	1	5	2023.2	13.4	2.5964
7 children	1	3	2024.3333	6.6667	2.1092
8 consequences	1	6	2022.5	9.6667	0.6691
9 corruption	1	3	2020.6667	22	1.5479
10 critical race theory	1	6	2022.1667	13.3333	0.8301

11	diagnosis	1	3	2024.3333	4	1.2655
12	disability	1	3	2024.6667	0.6667	0.2109
13	discourse	1	3	2023.6667	8	0.929
14	discrimination	1	7	2022.8571	16.7143	1.9648
15	disparities	1	3	2024	10.6667	0.9612
16	distress	1	3	2023	2	0.2575
17	education	1	5	2023.2	5	0.6626
18	endometriosis	1	3	2024.3333	5.3333	1.6874
19	equity	1	4	2024.5	1.25	0.3955
20	experiences	1	8	2024.625	2.75	1.6161
21	family	1	3	2022.3333	15.6667	0.8474
22	gay	1	4	2024.25	4.5	1.0804
23	health	1	14	2023.3571	11.3571	1.228
24	health-care	1	6	2023.6667	9.5	1.4962
25	impact	1	7	2023.2857	1.8571	0.3298
26	institutional betrayal	1	5	2021	14.8	1.1828
27	intersectionality	1	9	2022.4444	39.1111	0.9894
28	knowledge	1	3	2021.6667	34.3333	2.6894
29	law	1	3	2023	3	0.147
30	leadership	1	3	2024.3333	0.3333	0.1055
31	lgbtq plus	1	3	2024.6667	0.6667	0.3436
32	long covid	1	6	2023.8333	24.8333	3.2665
33	manipulation	1	6	2022.8333	2.1667	0.375
34	medical gaslighting	1	12	2023.6667	11.5833	1.568
35	mental-health	1	8	2023.125	17.25	1.0859
36	microaggressions	1	6	2023.5	13.8333	0.7428
37	misinformation	1	3	2023.3333	0	0
38	model	1	7	2023.4286	8	0.762
39	nurses	1	3	2025	6.3333	4.5238
40	outcomes	1	3	2024	4.6667	0.886
41	pain	1	3	2023.6667	6.6667	1.9402
42	participation	1	3	2023.6667	4	0.9638
43	patient	1	3	2024	6	1.1603
44	performance	1	4	2024.25	0.75	0.1266
45	perspectives	1	3	2024.3333	3.6667	1.1601
46	power	1	5	2022.6	19.6	1.4863
47	prevalence	1	4	2024.25	3.25	0.6961
48	qualitative research	1	4	2024.25	0.75	0.2373
49	race	1	6	2023	19.3333	1.3516
50	racial gaslighting	1	7	2023.1429	20.4286	0.7282
51	racial microaggressions	1	4	2021.75	11.75	0.9314

52	racism	1	13	2022.8462	21.9231	1.2042
53	resilience	1	4	2024	3.25	1.0282
54	risk	1	4	2021	18.25	0.9283
55	social media	1	5	2024.4	1	0.3074
56	social support	1	6	2023.6667	5.1667	0.6096
57	sociology	1	8	2023.625	18.75	1.6897
58	stigma	1	4	2023	4.75	0.7786
59	stress	1	3	2023	11.6667	1.0507
60	structural racism	1	3	2023	1	0.0889
61	testimonial injustice	1	4	2022.5	11	1.4019
62	tiktok	1	3	2024.3333	1	0.3164
63	transgender	1	5	2022.8	7.4	1.719
64	united-states	1	3	2024.3333	4.6667	1.4765
65	violence	1	9	2022.3333	20.2222	1.2068
66	women	1	28	2023.6429	8.1429	1.4387
67	workplace	1	7	2022.8571	6.4286	0.9536
68	workplace gaslighting	1	3	2025	4.3333	3.0952
69	youth	1	4	2022.75	12	0.7146
70	agency	2	4	2024.25	0	0
71	authentic leadership	2	3	2022	9.6667	0.8711
72	authenticity	2	3	2022	9.6667	0.8711
73	autoethnography	2	3	2021	15.3333	1.6389
74	autonomy	2	3	2020.6667	65.3333	2.0377
75	betrayal	2	3	2024.3333	9.3333	4.8483
76	bullying	2	3	2022.6667	6.3333	0.6668
77	construction	2	3	2024	6.3333	2.0038
78	epistemic injustice	2	17	2022.6471	17.7059	1.6943
79	gaslighting	2	93	2022.5484	11.6774	1.2667
80	gender	2	21	2023	17.2381	1.126
81	harassment	2	4	2022.25	10.5	0.6711
82	identity	2	7	2024.1429	2	0.4219
83	management	2	3	2023	11.6667	2.6651
84	men	2	4	2023.5	7	0.3853
85	self	2	5	2024.2	3	0.4061
86	sexism	2	4	2023.5	4.5	1.7514
87	sexual harassment	2	4	2023.75	4.25	1.1735
88	social epistemology	2	3	2022.3333	5.6667	0.4283
89	stereotype	2	3	2024	4.3333	2.2845
90	work	2	4	2024	2.75	0.8701
91	adolescent mental health	3	3	2024.6667	0.6667	0.2109
92	aggression	3	6	2023	5.8333	2.3314

93	anger	3	3	2023.6667	4	1.1179
94	anxiety	3	3	2024.6667	3	0.9492
95	community	3	3	2022.6667	6	1.1478
96	dark triad	3	4	2023.75	6.25	3.2369
97	depression	3	5	2024.4	1.2	0.3797
98	emotion	3	3	2024.3333	3	0.9492
99	emotions	3	3	2023	3	0.2703
100	feminism	3	6	2022.5	3.3333	0.6952
101	machiavellianism	3	3	2024	6.3333	3.9783
102	narcissism	3	4	2024	5.75	3.1894
103	politics	3	8	2023.375	3.875	0.7466
104	psychopathy	3	4	2023.75	6.25	3.2369
105	relationships	3	3	2023.3333	8.6667	1.4627
106	trauma	3	6	2022.8333	6.5	2.4211
107	abuse	4	4	2024.75	2	0.9312
108	coercive control	4	10	2023	36.4	1.6056
109	domestic violence	4	7	2022.5714	51.2857	2.1092
110	epistemology	4	3	2022	10	0.9012
111	help-seeking intimate partner	4	3	2021	113.6667	2.114
112	violence	4	17	2023.2941	25.7647	2.0934
113	psychological abuse	4	6	2023.1667	19.8333	2.4126
114	qualitative	4	3	2024.6667	1	0.3164
115	sexual violence	4	5	2024	4.8	0.6932
116	transphobia	4	3	2020.3333	29	1.4609

Appendix 4. List of journals (n=195) included into analysis.

	<i>Publication Title</i>	<i>Record Count</i>	<i>% of 235</i>
1	HYPATIA A JOURNAL OF FEMINIST PHILOSOPHY	5	2.119
2	JOURNAL OF INTERPERSONAL VIOLENCE	5	2.119
3	INTERNATIONAL JOURNAL OF QUALITATIVE STUDIES IN EDUCATION	4	1.695
4	JOURNAL OF FAMILY VIOLENCE	3	1.271
5	LEADERSHIP	3	1.271
6	SSM QUALITATIVE RESEARCH IN HEALTH	3	1.271
7	TOPOI AN INTERNATIONAL REVIEW OF PHILOSOPHY	3	1.271
8	CRITICAL PERSPECTIVES ON ACCOUNTING	2	0.847
9	ENVIRONMENT AND PLANNING E NATURE AND SPACE	2	0.847
10	EPISTEME A JOURNAL OF INDIVIDUAL AND SOCIAL EPISTEMOLOGY	2	0.847
11	FRONTIERS IN PSYCHOLOGY	2	0.847
12	FRONTIERS IN SOCIOLOGY	2	0.847
13	GENDER WORK AND ORGANIZATION	2	0.847

14	HEALTH COMMUNICATION	2	0.847
15	HEALTHCARE	2	0.847
16	HUMAN RELATIONS	2	0.847
17	HUMANITIES BASEL	2	0.847
18	INFORMATION	2	0.847
19	JOURNAL OF SOCIAL AND PERSONAL RELATIONSHIPS	2	0.847
20	NURSING ETHICS	2	0.847
21	PATIENT EDUCATION AND COUNSELING	2	0.847
22	PHILOSOPHICAL QUARTERLY	2	0.847
23	PHILOSOPHICAL STUDIES	2	0.847
24	PROFESSIONAL CASE MANAGEMENT	2	0.847
25	PSYCHOANALYTIC QUARTERLY	2	0.847
26	RAE REVISTA DE ADMINISTRACAO DE EMPRESAS	2	0.847
27	SOCIAL EPISTEMOLOGY	2	0.847
28	SOCIAL SCIENCE MEDICINE	2	0.847
29	WOMEN A CULTURAL REVIEW	2	0.847
30	AFRICAN STUDIES	1	0.424
31	AGENDA EMPOWERING WOMEN FOR GENDER EQUITY	1	0.424
32	AMERICAN JOURNAL OF POLITICAL SCIENCE	1	0.424
33	AMERICAN SOCIOLOGICAL REVIEW	1	0.424
34	ANGLIA ZEITSCHRIFT FUR ENGLISCHE PHILOLOGIE	1	0.424
35	ANTIPODE	1	0.424
36	APPLIED ACOUSTICS	1	0.424
37	ASIAN WOMEN	1	0.424
38	AUSTRALASIAN JOURNAL OF PHILOSOPHY	1	0.424
39	AUSTRALIAN JOURNAL OF POLITICAL SCIENCE	1	0.424
40	AUSTRALIAN JOURNAL OF POLITICS AND HISTORY	1	0.424
41	BIBLICAL INTERPRETATION A JOURNAL OF CONTEMPORARY APPROACHES	1	0.424
42	BIG DATA SOCIETY	1	0.424
43	BMC NURSING	1	0.424
44	BMC PRIMARY CARE	1	0.424
45	BMC PSYCHIATRY	1	0.424
46	BMC WOMENS HEALTH	1	0.424
47	BMJ OPEN	1	0.424
48	BRAIN BROAD RESEARCH IN ARTIFICIAL INTELLIGENCE AND NEUROSCIENCE	1	0.424
49	BRITISH JOURNAL OF SOCIOLOGY OF EDUCATION	1	0.424
50	BUFFALO LAW REVIEW	1	0.424
51	CANADIAN JOURNAL OF NURSING RESEARCH	1	0.424
52	CANADIAN MEDICAL ASSOCIATION JOURNAL	1	0.424
53	CANADIAN THEATRE REVIEW	1	0.424
54	COLLEGE COMPOSITION AND COMMUNICATION	1	0.424
55	COMMUNICATION RESEARCH AND PRACTICE	1	0.424
56	COMMUNICATION STUDIES	1	0.424
57	COMMUNICATION THEORY	1	0.424

58	COMMUNIST AND POST COMMUNIST STUDIES	1	0.424
59	COMMUNITY COLLEGE JOURNAL OF RESEARCH AND PRACTICE	1	0.424
60	CONTEMPORARY FAMILY THERAPY	1	0.424
61	CONTEMPORARY FRENCH AND FRANCOPHONE STUDIES	1	0.424
62	CONTEMPORARY POLITICAL THEORY	1	0.424
63	COOPERATION AND CONFLICT	1	0.424
64	COUNSELLING PSYCHOTHERAPY RESEARCH	1	0.424
65	CRITICAL ARTS SOUTH NORTH CULTURAL AND MEDIA STUDIES	1	0.424
66	CRITICAL SOCIOLOGY	1	0.424
67	CURRENT PSYCHOLOGY	1	0.424
68	CYPRUS TURKISH JOURNAL OF PSYCHIATRY AND PSYCHOLOGY	1	0.424
69	DIALOGUE CANADIAN PHILOSOPHICAL REVIEW	1	0.424
70	DISCOURSE STUDIES	1	0.424
71	EIGHTEENTH CENTURY FICTION	1	0.424
72	ENERGY RESEARCH SOCIAL SCIENCE	1	0.424
73	ENGAGING SCIENCE TECHNOLOGY AND SOCIETY	1	0.424
74	ENVIRONMENT AND PLANNING A ECONOMY AND SPACE	1	0.424
75	ENVIRONMENTAL POLICY AND GOVERNANCE	1	0.424
76	EQUALITY DIVERSITY AND INCLUSION	1	0.424
77	EQUITY EXCELLENCE IN EDUCATION	1	0.424
78	ERGO AN OPEN ACCESS JOURNAL OF PHILOSOPHY	1	0.424
79	ETHICAL THEORY AND MORAL PRACTICE	1	0.424
80	ETHICS	1	0.424
81	EUROPEAN JOURNAL OF MARKETING	1	0.424
82	FEMINIST ANTHROPOLOGY	1	0.424
83	FRONTIERS IN EDUCATION	1	0.424
84	GENDER AND EDUCATION	1	0.424
85	GENERATIONS	1	0.424
86	GRIFFITH LAW REVIEW	1	0.424
87	GROUP	1	0.424
88	HEALTH SOCIAL CARE IN THE COMMUNITY	1	0.424
89	HISPANIC HEALTH CARE INTERNATIONAL	1	0.424
90	HUMAN RESOURCES FOR HEALTH	1	0.424
91	INFORMATION COMMUNICATION SOCIETY	1	0.424
92	INFORMATION WISSENSCHAFT UND PRAXIS	1	0.424
93	INQUIRY AN INTERDISCIPLINARY JOURNAL OF PHILOSOPHY	1	0.424
94	INTERNATIONAL JOURNAL OF APPLIED PSYCHOANALYTIC STUDIES	1	0.424
95	INTERNATIONAL JOURNAL OF BUSINESS COMMUNICATION	1	0.424
96	INTERNATIONAL JOURNAL OF COMMUNICATION	1	0.424
97	INTERNATIONAL JOURNAL OF GROUP PSYCHOTHERAPY	1	0.424
98	INTERNATIONAL JOURNAL OF GYNECOLOGY OBSTETRICS	1	0.424
99	INTERNATIONAL JOURNAL OF HUMAN RIGHTS IN HEALTH CARE	1	0.424
100	INTERNATIONAL JOURNAL OF LAW CRIME AND JUSTICE	1	0.424
101	INTERNATIONAL JOURNAL OF PHILOSOPHICAL STUDIES	1	0.424
102	INTERNATIONAL JOURNAL OF PSYCHOLOGICAL RESEARCH	1	0.424

103	INTERNATIONAL JOURNAL OF TRANSGENDER HEALTH	1	0.424
104	INTERPRETATION A JOURNAL OF BIBLE AND THEOLOGY	1	0.424
105	JAMA NETWORK OPEN	1	0.424
106	JMIR INFODEMIOLOGY	1	0.424
107	JOURNAL OF ADVANCED NURSING	1	0.424
108	JOURNAL OF ASIAN PUBLIC POLICY	1	0.424
109	JOURNAL OF COMMUNICATION	1	0.424
110	JOURNAL OF COMMUNITY APPLIED SOCIAL PSYCHOLOGY	1	0.424
111	JOURNAL OF CREATIVITY IN MENTAL HEALTH	1	0.424
112	JOURNAL OF CRIME JUSTICE	1	0.424
113	JOURNAL OF CRIMINOLOGICAL RESEARCH POLICY AND PRACTICE	1	0.424
114	JOURNAL OF CRITICAL REALISM	1	0.424
115	JOURNAL OF HEALTH CARE FOR THE POOR AND UNDERSERVED	1	0.424
116	JOURNAL OF HUMANISTIC PSYCHOLOGY	1	0.424
117	JOURNAL OF INDUSTRIAL RELATIONS	1	0.424
118	JOURNAL OF LESBIAN STUDIES	1	0.424
119	JOURNAL OF LGBT YOUTH	1	0.424
120	JOURNAL OF LITERARY CULTURAL DISABILITY STUDIES	1	0.424
121	JOURNAL OF MANAGEMENT ORGANIZATION	1	0.424
122	JOURNAL OF MIND AND MEDICAL SCIENCES	1	0.424
123	JOURNAL OF PAIN	1	0.424
124	JOURNAL OF PATIENT EXPERIENCE	1	0.424
125	JOURNAL OF PEDIATRIC AND ADOLESCENT GYNECOLOGY	1	0.424
126	JOURNAL OF PERINATAL NEONATAL NURSING	1	0.424
127	JOURNAL OF PREVENTION INTERVENTION IN THE COMMUNITY	1	0.424
128	JOURNAL OF PUBLIC HEALTH DENTISTRY	1	0.424
129	JOURNAL OF SEXUAL AGGRESSION	1	0.424
130	JOURNAL OF SOCIAL WORK EDUCATION	1	0.424
131	JOURNAL OF THE AMERICAN MUSICOLOGICAL SOCIETY	1	0.424
132	JOURNAL OF THE SOCIETY FOR SOCIAL WORK AND RESEARCH	1	0.424
133	JOURNAL OF WOMEN AND MINORITIES IN SCIENCE AND ENGINEERING	1	0.424
134	LANGUAGES	1	0.424
135	LEGAL MEDICINE	1	0.424
136	LITERARY VOICE	1	0.424
137	MEDICAL EDUCATION	1	0.424
138	MEMORY	1	0.424
139	MERIDIANS FEMINISM RACE TRANSNATIONALISM	1	0.424
140	MIDWEST STUDIES IN PHILOSOPHY	1	0.424
141	MONIST	1	0.424
142	NURSING FORUM	1	0.424
143	NURSING INQUIRY	1	0.424
144	PASTORAL PSYCHOLOGY	1	0.424
145	PERFORMANCE RESEARCH	1	0.424
146	PERSONAL RELATIONSHIPS	1	0.424
147	PERSONALITY AND INDIVIDUAL DIFFERENCES	1	0.424

148	PERSONALITY AND SOCIAL PSYCHOLOGY REVIEW	1	0.424
149	PERSPECTIVES ON PSYCHOLOGICAL SCIENCE	1	0.424
150	PHILOSOPHICAL PERSPECTIVES	1	0.424
151	PHILOSOPHICAL PSYCHOLOGY	1	0.424
152	PHILOSOPHY AND SOCIETY FILOZOFIJA I DRUSTVO	1	0.424
153	PHILOSOPHY PSYCHIATRY PSYCHOLOGY	1	0.424
154	PHILOSOPHY SOCIAL CRITICISM	1	0.424
155	PLASTIC AND RECONSTRUCTIVE SURGERY GLOBAL OPEN	1	0.424
156	POLITICAL QUARTERLY	1	0.424
157	POLITICS GROUPS AND IDENTITIES	1	0.424
158	POPULAR COMMUNICATION	1	0.424
159	PSYCHOLOGY OF LEADERS AND LEADERSHIP	1	0.424
160	PSYCHOLOGY OF MEN MASCULINITIES	1	0.424
161	PSYCHOLOGY PUBLIC POLICY AND LAW	1	0.424
162	QUALITATIVE HEALTH RESEARCH	1	0.424
163	QUALITATIVE PSYCHOLOGY	1	0.424
164	QUALITY OF LIFE RESEARCH	1	0.424
165	RACE AND JUSTICE	1	0.424
166	RACE ETHNICITY AND EDUCATION	1	0.424
167	RBGN REVISTA BRASILEIRA DE GESTAO DE NEGOCIOS	1	0.424
168	REPRESENTATIONS	1	0.424
169	REVIEW OF EDUCATION	1	0.424
170	REVIEW OF HIGHER EDUCATION	1	0.424
171	SEXUAL AND RELATIONSHIP THERAPY	1	0.424
172	SHOFAR AN INTERDISCIPLINARY JOURNAL OF JEWISH STUDIES	1	0.424
173	SIGNS AND SOCIETY	1	0.424
174	SOCIAL FORCES	1	0.424
175	SOCIAL POLITICS	1	0.424
176	SOCIAL PROBLEMS	1	0.424
177	SOCIAL SCIENCES BASEL	1	0.424
178	SOCIETY AND MENTAL HEALTH	1	0.424
179	SOCIOLOGICAL INQUIRY	1	0.424
180	SOCIOLOGICAL PERSPECTIVES	1	0.424
181	SOCIOLOGY OF HEALTH ILLNESS	1	0.424
182	SOUTHERN COMMUNICATION JOURNAL	1	0.424
183	SUSTAINABILITY	1	0.424
184	SYNTHESE	1	0.424
185	TEACHING AND LEARNING IN MEDICINE	1	0.424
186	TEACHING AND LEARNING IN NURSING	1	0.424
187	TRANSLATIONAL GASTROENTEROLOGY AND HEPATOLOGY	1	0.424
188	TYDSKRIF VIR LETTERKUNDE	1	0.424
189	UNIVERSITY OF TORONTO LAW JOURNAL	1	0.424
190	VICTORIAN STUDIES	1	0.424
191	VIOLENCE AGAINST WOMEN	1	0.424
192	VISUAL ANTHROPOLOGY REVIEW	1	0.424

193	WESTERN JOURNAL OF COMMUNICATION	1	0.424
194	WOMENS HISTORY REVIEW	1	0.424
195	WOMENS STUDIES INTERNATIONAL FORUM	1	0.424

Governing Harms from Deepfakes in Crisis Situations: Comparing Legal and Regulatory Frameworks of G7 Countries and the EU*

Katerina Tsetsura**
University of Oklahoma

H M Murtuza***
University of Oklahoma

Mark Raymond****
University of Oklahoma

Typhaine Joffe*****
University of Oklahoma

This study analyses deepfake-related initiatives of the Group of Seven (G7) countries—Canada, France, Germany, Italy, Japan, the United Kingdom, the United States—and the United Nations and the European Union from a comparative perspective to examine in what ways, if any, AI-generated inaccurate content, generated in times of crises such as natural disasters, is regulated in these countries. Using the Social Amplification of Risk Framework (SARF), a theory that explains how risk perceptions and communication create ripple effects, we demonstrate why the potentially detrimental risks that might come from deepfakes, which aim to distort societies in times of crises, should be accounted for in national and global initiatives to regulate the AI-generated content. We collected and thematically analysed documents using a qualitative open coding approach. The findings demonstrated that while existing and proposed country-specific laws and regulations reviewed offer useful principles, they were not designed to address the kinds of digital harms arising from the use of deepfakes in crises such as natural disasters. Global initiatives also contained the same limitations: despite encouraging responsible innovation and digital transparency, existing frameworks did not address the kinds of harms associated with deepfake use in disaster scenarios. Overall, these initiatives failed to provide concrete strategies for crisis management or harm mitigation from deepfakes deployed to mislead the public in natural disasters or to initiate or escalate violent conflict. Based on the analysis, the article offers implications and recommendations for policymakers and for future studies.

Keywords: risk mitigation, AI legislation, natural disasters, violent conflict, AI regulation

* Article submitted on 14/10/2025. Article accepted on 20/12/2025.

** tsetsura@ou.edu

*** hmmmurtuza@ou.edu

**** mraymond@ou.edu

***** typhaine.e.joffe-1@ou.edu

In today's world of information, our ability to detect, identify, and correctly pinpoint true and false information is greatly challenged. Various types of inaccurate, untrue, or fake information overwhelm our everyday information consumption, creating multiple risks to our well-being, security, and the overall functioning of society. Deepfakes create a particular challenge to national and global efforts to resist misinformation and increase security risks associated with misinformation consumption, particularly in times of risks associated with natural disasters and violent conflicts, including hybrid warfare (HybridCoE, 2020).

In 2019, AI firm Deeptext identified 15,000 deepfake videos online, and this number almost doubled in just nine months (Sample, 2020). Some experts anticipate that as much as 90 percent of digital content could be synthetically generated within just a few years (Lawson, 2023). Currently, the European Union is leading the way with its Artificial Intelligence (AI) Act, the first and most comprehensive legislation to regulate AI content (EU, 2025). In 2024, the U.S. states alone were introducing around 25 deepfake bills weekly, according to Business Software Alliance (2024).

So far, comparative studies on AI-related regulatory frameworks have been limited to examining the overall approach to AI regulation (Cajueiro & Rezende Celestino, 2026) or focusing on specific regulatory acts as individual case studies, such as the EU AI Act (Dasharathraj et al., 2025). No one study has examined and compared the actual or proposed regulations or acts related to the production and dissemination of deepfakes associated with national crises, such as natural disasters and violent conflicts. This study, the first of its kind, collected and analysed actual and proposed policy documents to identify whether states appear to be anticipating these concerning use cases for deepfakes. The analysis will allow policymakers, communication professionals, and other stakeholders to pay particular attention to how the regulation of deepfakes, which aim to destabilize society, can be created and implemented.

This study analyses the 32 deepfake-related legal initiatives of the Group of Seven (G7) countries, Canada, France, Germany, Italy, Japan, the United Kingdom, and the United States; as well as efforts by the European Union, the United Nations, and the Council of Europe, from a comparative perspective. The goal is to identify the commonalities, differences, and regulatory gaps, especially in the context of potential harms deepfakes can impose in natural disasters and violent conflicts. We use the Social Amplification of Risk Framework (SARF) (Kasperson et al., 2022) as a theoretical basis to establish the need for regulating deepfakes related to natural disasters and potential societal crises. SARF explains how risk amplification and attenuation happen through the social system. We argue that the far-reaching consequences of ill-motivated deepfakes might have a *ripple effect* (Kasperson & Kasperson, 1996) from the individual to society if they remain unchecked and unregulated.

This study contributes to previous research in three ways. First, it conceptualizes the harmful, ill-motivated deepfakes as a risk rather than a mere content-level problem, using SARF. Second, it presents a thematic analysis of the existing regulatory frameworks of G7 countries and major transnational entities and identifies the commonalities,

divergences, and gaps in dealing with deepfakes in the context of natural disasters or violent conflict. Third, the study discusses how the SARF-informed risk management approach can minimize the harmful effects of deepfakes in times of natural disasters or violent conflicts. Understanding the G7 regulatory frameworks is crucial as it may assist other government agencies and AI platform service providers in the process of adopting policies and regulations related to deepfakes.

We first define deepfakes and discuss the crisis-specific harms associated with them. Next, we discuss SARF and introduce our research questions and methodology. Finally, we present our empirical findings and offer implications for policymakers as well as recommendations for future research.

Defining Deepfakes

Increasingly, deepfakes become part of everyday media interactions. Westerlund (2019) wrote that “deepfakes are hyper-realistic videos digitally manipulated to depict people saying and doing things that never actually happened” (p. 40). Another definition offered by Van der Sloot and Wagenveld (2022) treats deepfakes as “content (video, audio or otherwise) that is wholly or partially fabricated or existing content (video, audio or otherwise) that has been manipulated” (p. 1). Despite having negative connotative and denotative meanings, deepfakes can also be used for more benign or unobjectionable purposes (Westerlund, 2019). Still, most of the communication studies are concerned with deepfakes’ negative consequences (Lee, 2019; Stewart, 2019; Van der Sloot & Wagenveld, 2022; Westerlund, 2019).

The term “deepfake” first gained prominence after the 2016 United States presidential election, when researchers examined various features of fake news (Lee, 2019). The rapid improvement and proliferation of deepfake technology led international organizations to view deepfakes as a significant future threat (**HybridCoE, 2020**). Deepfakes can be misused in a variety of ways, including for the generation of misinformation and fake news; political manipulation; cyberbullying and harassment; privacy-related concerns; and fraud through impersonation (Vig, 2024). In addition, deepfakes could compromise national security as malign actors use them to spread disinformation, to interfere in elections, and deepen distrust between citizens and authorities (Westerlund, 2019). These various use cases can harm society, social relationships, democracy, and the rule of law (Van der Sloot & Wagenveld, 2022). Thus, both academics and practitioners pointed out an increasing need to regulate deepfakes (HybridCoE, 2020; Quirk, 2023).

Harms Associated with Deepfakes in Crises

Previous studies focused mostly on describing the personal risks and privacy concerns associated with deepfakes (e.g., Cochran & Napshin, 2021; Kugler & Pace, 2021). However, existing research generally overlooks at least two other potential use cases for deepfake technology that we believe merit attention, involving different kinds of potential harms. First, recently, experts warned that deepfakes have the potential to create chaos or provide inaccurate information in times of natural disasters (Kerbage, 2025). In some cases, deepfakes mimicked responses from public authorities and law enforcement (Dauer, 2022). Such deepfakes have the potential to lead to injury or death, or to prevent individuals from accessing services and benefits essential to their well-being at times of heightened vulnerability (Chapagain et al., 2024).

Second, the possibility exists for deepfakes to be used to either initiate or escalate violent conflicts (BBC, 2019; Kerbage, 2025). Technology could also be used to provoke violence at the community level, but we focus here on the possibility of interstate conflict. Potential examples could include alerts about a fraudulent incoming attack through civil defence systems, or a deepfake video of a national leader making a declaration of war or similar political announcement.

These potential applications of deepfake technology differ from the scenarios discussed in academia and the media, which are typically concerned with financial (de Rancourt-Raymond & Smaili, 2023) or electoral disruptions (Barari, Lucas, & Munger, 2025). These risks may or may not be financially motivated and do not involve individuals' private information; rather, they might attack state and non-state actors, organizations, and networks. Such attacks are not motivated by a desire to bully or intimidate specific individuals but rather to create chaos and disorder (Hameleers, 2023) in times of natural disasters (Hilberts et al., 2025) or contribute to inciting conflict and violence in the most unstable and fluid environments (Topor, 2024). To better understand the mechanisms behind acting upon deepfakes, we propose using the social amplification of risk framework as a theoretical basis for this study.

Social Amplification of Risk Framework (SARF)

Social Amplification of Risk Framework, or SARF, aims to understand why a risk ripples or amplifies in a certain way to impact society (Kasperson et al., 2022). Kasperson and Kasperson (1996) framed risk as "threats of harm to people and nature but also to other things or ends that people value, such as community or political freedom" (p. 98). We argue that such societal threats can also emerge from deepfakes to amplify the risks. Hence, governments and societies need to be fully prepared to minimize risks.

SARF also states that a ripple effect of risks happens in three major stages (Kasperson & Kasperson, 1996). First, the information about a risk-related event is shared and

reshaped as it moves through different communication channels, and its interpretation may change the risk perception. In the second stage, these modified perceptions influence how the individual or organization behaves. Finally, these behaviours shape the experience of risk, and people's behavioural responses can prompt new rounds of communication, as communicators adjust their messages in response to audience reactions to reinforce or discourage certain behaviours.

We argue that harmful deepfakes behave like risk signals, especially in the context of natural disasters or potentially violent conflicts. First, deepfakes may appear on digital platforms and spread through social media, messaging apps, or other channels. In the process, deepfakes can be reshaped, forwarded, re-captioned, and reframed. Then, deepfakes can alter perceptions and influence behaviour, which, in turn, can be based on deepfake messages that can be potentially life-threatening in the context of natural disasters (e.g., evaluation in an opposite direction, toward tornado or hurricane paths) or violent conflicts (e.g., incitement of violence).

Hence, the SARF-guided approach highlights the importance of examining the existing regulatory frameworks to address deepfake-induced risks at the governmental level, as they may cause greater harm in crisis contexts.

Research Questions

Because the use of deepfakes to exacerbate crises entails significant and specific potential harms, we were interested in determining whether states have enacted or proposed laws, regulations, or policies that specifically address such scenarios.

Specifically, we asked:

RQ1: What are the similarities in addressing the topic of deepfakes in the proposed or enacted laws, policies, and regulations for artificial intelligence?

RQ2: What are the differences in addressing the topic of deepfakes in the proposed or enacted laws, policies, and regulations for artificial intelligence?

RQ3: What, if any, laws, policies, and regulations have been proposed or enacted to specifically deal with the use of deepfake technology to address its effects on the natural disaster response AND/OR on the potential initiation or escalation of violent conflict?

RQ4: To what extent do these proposed or enacted laws, policies, and regulations for artificial intelligence provide useful tools or guidelines in specifically addressing how to deal with deepfakes in the two aforementioned cases: 1) the effects on the natural disaster response, and 2) the effects on the potential initiation or escalation of violent conflict?

Method

The study utilized qualitative thematic analysis of official documents collected in October and November 2024. Specifically, researchers focused on examining the treatment of deepfake-related issues within these documents: how the term "deepfake" is defined, what content is permissible within the legal framework, which activities are considered violations of the act, the available remedial measures, the scope of the regulatory or legal framework, the punitive measures outlined, and other rules and regulations involved in identifying deepfake-related irregularities and offenses.

This study examined 32 legal documents and articles, totalling over 1,200 pages, relevant to the G7 countries and the EU (see Table 1). We chose to investigate documents from G7 and EU countries because they started regulating deepfakes before other countries (Birrer & Just, 2024; Romero Moreno, 2024). The sample included 10 enacted legislation documents, eight proposed legislations, eight enacted/adopted and two proposed policies, and four enacted/in force and one proposed regulation acts and guidelines (see Table 1). We investigated the proposed and enacted deepfake laws and regulations to identify all relevant governance efforts that address potential harms associated with deepfakes in crisis scenarios. The final sample consisted of six US federal documents and five US state-level documents, four documents from EU, one from the Council of Europe, three from the United Nations, three from the UK, three from Germany, two documents from Italy and two from Canada, and one document from each France and Japan, as well as the policy framework of the G7 (see Table 2).

Table 1 - List of Analysed Documents

No	Jurisdiction	Document	Type of document	Status
1	Canada	Bill C-63 https://www.parl.ca/DocumentViewer/en/44-1/bill/C-63/first-reading	Legislation	Proposed
2	EU	EU Artificial Intelligence Act https://eur-lex.europa.eu/eli/reg/2024/1689/oj/eng	Regulation	In Force
3	France	SREN Act (Art. 226–8, French Criminal Code) https://www.legifrance.gouv.fr/jorf/id/JORFTEXT000049563368/	Legislation	Enacted
4	Germany	The Criminal Protection of Personal Rights Against Deepfakes https://www.bundesrat.de/SharedDocs/drucksachen/2024/0201-0300/222-24(B).pdf?__blob=publicationFile&v=1	Legislation	Proposed
5	Italy	Strategic Program on Artificial Intelligence 2022-2024 https://assets.innovazione.gov.it/1637777513-strategic-program-aiweb.pdf	Policy Document	Proposed
6	Japan	AI Guidelines for Business Version 1.0 https://www.meti.go.jp/shingikai/mono_info_service/ai_shakai_jisso/pdf/20240419_9.pdf	Policy Document	Enacted
7	USA	Identifying Outputs of Generative Adversarial Networks (IOGAN) Act https://www.congress.gov/bill/116th-congress/senate-bill/2904/text	Legislation	Enacted
8	USA	DEFIANCE Act of 2024 https://www.congress.gov/bill/118th-congress/senate-bill/3696/text	Legislation	Not Approved
9	USA	DEEPPAKES Accountability Act	Legislation	Proposed

		https://www.congress.gov/bill/118th-congress/house-bill/5586/text		
10	USA, Louisiana	ACT No. 457 https://legis.la.gov/legis/ViewDocument.aspx?d=1333325	Legislation	Enacted
11	USA, Texas	Section 255.004, Election Code https://statutes.capitol.texas.gov/docs/el/htm/el.255.htm	Legislation	Enacted
12	USA	Deepfake Report Act of 2019 https://www.congress.gov/bill/116th-congress/senate-bill/2065/text	Legislation	Not Approved
13	USA, Mississippi	Senate Bill 2577 https://billstatus.ls.state.ms.us/documents/2024/html/SB/2500-2599/SB2577IN.htm	Legislation	Enacted
14	USA	Protecting Consumers from Deceptive AI Act https://www.congress.gov/bill/118th-congress/house-bill/7766/text	Legislation	Proposed
15	UK	Criminal Justice Bill https://bills.parliament.uk/bills/3511	Legislation	Proposed
16	Canada	2023 Report: "The Evolution of Disinformation: A Deepfake Future" https://www.canada.ca/content/dam/isis-scrs/documents/publications/2023/The%20Evolution%20of%20Disinformation%20-%20Deepfake%20Report_EN_DIGITAL.pdf	Policy	Published
17	USA, California	Assembly Bill 2655 https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=20230240AB2655	Legislation	Enacted
18	USA, California	Senate Bill 942 https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=20230240SB942	Legislation	Enacted
19	EU	Digital Services Act (DSA) – Regulation (EU) 2022/2065 https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32022R2065&qid=1742834952559	Regulation	In Force
20	EU	General Data Protection Regulation (EU) 2016/679 https://gdpr-info.eu/	Regulation	In Force
21	EU	Code of Practice on Disinformation (2022 Revision) https://digital-strategy.ec.europa.eu/en/library/2022-strengthened-code-practice-disinformation	Regulation	Voluntary
22	Germany	Proposed StGB § 201b (Deepfake Misuse) https://www.bundesrat.de/SharedDocs/drucksachen/2024/0201-0300/222-24(B).pdf?__blob=publicationFile&v=1	Legislation	Proposed
23	Germany	State Media Authorities (Landesmedienanstalten) – Interstate Media Treaty Implementation https://www.die-medienanstalten.de/fileadmin/user_upload/Rechtsgrundlagen/Gesetze/Staatsvertraege/Interstate_Media_Treaty_en.pdf	Regulation	In force
24	Italy	Italian Data Privacy Code (Legislative Decree no. 196/2003) http://www.privacy.it/archivio/privacymcode-en.html	Legislation	Enacted
25	UK	Online Safety Act https://www.gov.uk/government/publications/online-safety-act-new-criminal-offences-circular/online-safety-act-new-criminal-offences-circular	Legislation	Enacted
26	UK	AI Regulation White Paper https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach/white-paper	Policy Document	Enacted
27	USA	National Defense Authorization Act (NDAA) – Deepfake Provisions https://www.wilmerhale.com/insights/client-alerts/20191223-first-federal-legislation-on-deepfakes-signed-into-law	Legislation	Enacted

28	United Nations	Governing AI for Humanity https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf	Policy Document	Endorsed
29	United Nations	Global Digital Compact https://www.un.org/global-digital-compact/sites/default/files/2024-09/Global_Digital_Compact_-_English_0.pdf	Policy	Adopted
30	United Nations	Convention against Cybercrime https://documents.un.org/doc/undoc/gen/n24/426/74/pdf/n2442674.pdf	Convention	Adopted
31	Council of Europe	Budapest Convention on Cybercrime https://rm.coe.int/1680081561	Convention	In Force
32	G7	Hiroshima AI Process	Policy Framework	In Force

Table 2 - Region/Country Summary

Region / Country	Number of Documents	Region / Country	Number of Documents
USA – Federal	6	Italy	2
USA – State	5	Canada	2
EU	5	France	1
United Nations	3	Japan	1
Germany	3	G7	1
UK	3		

To extract the data from this corpus of documents, a combination of keyword searches and an AI-based search was conducted. The keyword included words like “deepfake,” “deepfakes,” “deep fake,” “AI-generated,” and “AI-generated content.” Based on each of the research questions, we created prompts for ChatGPT to find quotations relevant to the research questions from all documents. After identifying the keyword language within each document, researchers manually verified all findings. All the relevant quotations were directly extracted from the legal documents by researchers, without relying on AI-generated responses. The extracted quotations or text from the regulatory documents were included in the dataset used for the thematic analysis, in which we open-coded data based on relevance to our research questions (Spiggle, 1994) and then created themes based on codes.

Two of the legal documents, the SREN Act of France and the Criminal Protection of Personal Rights Against Deepfakes of Germany, were unavailable in English. To analyse the documents, the authors resorted to Google’s translation service. This approach is supported by Balk et al. (2012), who found that Google’s translation from French and German is fairly accurate.

The thematic analysis was performed in two stages between December 2024 and May 2025. During the first stage of analysis, the first research team member independently analysed 15 out of 32 documents. During the second stage, the second team member independently conducted a thematic analysis of 22 out of 32 documents. In both stages, five documents were common: 1) EU’s AI Act (EU, in-force regulation); 2) France’s SREN Act (France, enacted legislation); 3) USA’s Identifying Outputs of Generative Adversarial

Networks Act (USA, enacted legislation; 4) Deepfakes Accountability Act (USA, proposed legislation); and 5) Defiance Act of 2024 (USA, not approved legislation). The researchers analysed each of the five documents separately, compared their notes, and discussed the preliminary thematic analysis.

Next, two researchers engaged in a triangulation process by examining together an additional 17 out of 32 legislative and policy documents. The researchers compared similar legislative and policy documents from different countries to check the consistency of the core analysis across these 17 documents. In a few cases of discrepancies, researchers worked through disagreements and resolved them upon agreeing on the themes and interpretation of findings. This within-method triangulation (Denzin, 2009) provided additional confirmability to the study (Ahmed, 2024).

Findings and Discussion

Overall, the findings demonstrated that, out of the G7 and the EU countries, only four—the EU, France, Japan, and the U.S.A.—have currently adopted regulations on the issue of deepfakes. The countries' early adoption of regulation may be linked to their technological advancements (Birrer & Just, 2024; Romero Moreno, 2024). However, the U.S.A. is yet to enact a federal law to address deepfake-related harms. Meanwhile, Japan's guidelines on AI indirectly mention deepfakes, but they are intended for AI businesses and do not include specific provisions to prevent deepfake-related risks or offences. Of the eight countries and regions, the EU AI Act was the most comprehensive, as it provided detailed guidelines on deepfake compliance and violations, consistent with previous studies (e.g., Birrer & Just, 2024). The French bill to Secure and Regulate Digital Space was also comprehensive. The regulatory measures in the other four countries—Canada, Germany, Italy, and the United Kingdom—were still at the proposal stage at the time of this analysis, either placed in legislatures for discussion or approved by the governments to be placed before the legislature.

The analysis below utilizes the Social Amplification of Risk Framework (Kasperson et al., 2022), focusing on whether the regulatory frameworks outlined specific measures in the stages where deepfakes as a risk can be amplified through the social system before they cause ripple effects. We present research question-based findings along with discussions on their implications in dealing with harmful deepfakes.

RQ1: Common Regulatory Themes

Across jurisdictions, five key common themes emerge: 1) emphasis on disclosure of the presence of AI-generated content; 2) platform responsibility; 3) application of privacy and consumer protection laws to deepfakes; 4) exceptions for free speech expression, satire, and artistic use; and 5) public interest and national security exceptions. In what follows,

we discuss each theme and provide concrete examples from the corpus. We numbered the documents for ease of citation and provided numbers instead of titles (e.g., doc. 2). The complete titles and the numbering legend are provided in Table 1.

First, nearly all jurisdictions emphasize labelling and disclosure requirements. The *EU's AI Act* (Article 52) mandates that users be informed when content is artificially generated or manipulated, unless it is evident by context (doc. 2, p. 34). Additionally, the EU law requires that deepfakes be “marked in a machine-readable format and detectable as artificially generated or manipulated” (doc. 2, p. 82). The latter requirement may help to ensure that AI systems are able to differentiate human-generated from AI-generated content. The *US DEEPFAKES Accountability Act* (proposed) and California's *Senate Bill 942* also require the disclosure of content produced by generative AI systems. Finally, Japan's AI guidelines for businesses require them to “develop and deploy reliable content authentication and provenance mechanisms, such as watermarking or other techniques to enable users to identify AI-generated content” (doc 6, p. 25).

Second, platform responsibility for removing certain kinds of harmful content is a common approach across several of these jurisdictions. The *Digital Services Act (EU)* and the *UK Online Safety Act* impose due diligence obligations on platforms, requiring detection and removal of harmful or deceptive content. Germany's *Interstate Media Treaty*, enforced by State Media Authorities (Landesmedienanstalten), assigns oversight responsibility to regional regulators to ensure compliance with media integrity standards. International frameworks reflect similar themes: the Hiroshima AI Process and the Global Digital Compact both call for transparency, content provenance, and ethical AI deployment. Canada's proposed Bill C-63 sets the standard of a platform operator determining that “there are reasonable grounds to believe that the content is content that sexually victimizes a child or revictimizes a survivor or intimate content communicated without consent” then the operator must make the content “inaccessible to all persons in Canada” (doc. 1, p. 25). In addition to concerns with certain kinds of content, largely related to victimization of minors, some policy approaches empower authorities to compel content removal rather than leaving decisions in the hands of firms. For example, a Mississippi state law empowers state courts to “order that any disseminated digitization be removed from any social media, electronic mail, electronic messaging, video-sharing services, or other physical or electronic method the digitization was disseminated through” (doc. 13, p. 4). In contrast, French law vests this power in state administrative authorities rather than in courts, but limits prevention of access to affected Internet addresses to three months (doc. 3, article 24, para. 5).

Third, many jurisdictions apply privacy and consumer protection laws to deepfakes. Italy's *Data Privacy Code* and the EU's *GDPR* regulate the unauthorized use of biometric data or image likenesses. Canada's 2023 report on “*The Evolution of Disinformation: A Deepfake Future*”, although not binding, recommends relying on existing defamation and fraud laws until specific deepfake legislation is developed. Because these kinds of harms are distant from our specific concerns with scenarios involving natural disasters and interstate conflict, we will not dwell on them here.

Fourth, several frameworks acknowledge exceptions for free speech, satire, and artistic use. Most explicitly, Germany's *proposed § 201b StGB* exempts "socially adequate uses," including content created "in pursuit of overriding legitimate interests," such as in the fields of art, science, journalism, teaching, or historical commentary (§ 201b(3)) (doc. 4). Similarly, the *EU AI Act* acknowledges that while deployers of AI systems used to generate or manipulate deepfakes must disclose the artificial origin of such content, this obligation should not impede rights to freedom of expression or artistic creation (doc. 2). When the content is evidently satirical, fictional, or artistic, the law only requires disclosure in a way that does not interfere with the display or enjoyment of the work. It also limits disclosure requirements for such artistic works, requiring "disclosure of the existence of such generated or manipulated content in an appropriate manner that does not hamper the display or enjoyment of the work" (doc. 2, p. 82). Balancing these kinds of important protections for speech and expression with efforts to protect public safety is a difficult and likely unresolvable tension.

Fifth, public interest and national security exceptions are present in some frameworks. *The EU AI Act* provides that transparency obligations do not apply if the deepfake is used for the detection, prevention, or prosecution of criminal offenses. The *U.S. NDAA* authorizes federal study and monitoring of foreign deepfake weaponization and implies permissive use in intelligence or defence contexts. Similarly, the proposed *DEEPFAKES Accountability Act* grants exceptions for "content 'produced by an officer or employee of the United States, or under the authority thereof, in furtherance of public safety or national security'" (doc. 9, p. 12). While research exceptions are sensible, the authorization of deepfakes for intelligence purposes is concerning, since use cases for intelligence purposes by some governments would likely explicitly include both exacerbation of natural disasters and potential efforts to initiate or escalate violent conflict.

RQ2: Divergences in Legal Design and Enforcement Mechanisms

The first major divergence relates to the scope of the regulatory frameworks regulating deepfakes and application of the framework. Some regulatory frameworks are comprehensive in nature while some are designed to address issue-specific deepfakes. For instance, the European Union, with instruments such as the *AI Act* and the *Digital Services Act*, utilizes a comprehensive and cross-sectoral regulatory structure that applies broadly to all AI deployers operating in or targeting the EU. The *EU AI Act* (doc. 2) is applicable for all kinds of AI service providers and any AI-generated contents. It addresses all the "deployers who use an AI system to generate or manipulate image, audio or video content" by making labelling for all kinds of "AI output" (doc. 2, p. 34). The act also incorporated a provision "to promote AI literacy tools, public awareness and understanding of the benefits, risks, safeguards, rights and obligations in relation to the use of AI systems" (doc. 2, p. 6).

Other regulations took a context-specific regulatory approach. For instance, the US is using a more fragmented and sector-specific set of laws. Specifically, federal initiatives, such as the NDAA, focus exclusively on national security and foreign manipulation, while states are introducing specific laws targeting election interference, non-consensual pornography, or consumer deception. The Protecting Consumers from Deceptive AI Act (doc. 14) only addresses the consumer deception aspect. France and Germany have introduced criminal law instruments prohibiting the unauthorized creation or dissemination of manipulated media but consider deepfakes primarily as attacks on personal dignity or authenticity and not as potential threats to democratic governance or security. Italy and Canada, which do not have specific legislation on the subject, include deepfake-related harms under general laws on privacy, consumer protection, or defamation. That approach creates uncertainty as to whether these laws actually apply to deepfakes, and to what extent, and it does not address the harms we are concerned with here.

The second major divergence is regulatory focus. Some documents focused on transparency and trust, while others focused on privacy protection and personal rights. In the EU, the focus is on systemic transparency and trust in artificial intelligence, with an emphasis on user information and platform risk mitigation as preventive measures. The *EU AI Act* (doc. 2) stressed ensuring “that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated” (doc. 2, p. 82). Similarly, Japan’s *AI Guidelines for Business Version 1.0* focuses on deployment of “reliable content authentication and provenance mechanisms” (doc. 6, p. 25).

In France, the *SREN law* emphasizes the preservation of personal autonomy and consent, especially when it comes to the theft of unauthorized audiovisual identities. The proposed Article 201b of the German Civil Code (StGB) focuses on protection against damage to reputation and deception that violates the subjective rights of individuals. Specifically, Germany’s legislation *The Criminal Protection of Personal Rights Against Deepfakes* (doc. 4) expresses great concern over personal rights protection. It says, “Over 90 percent of the deepfakes found on the Internet are said to relate to the area of pornography or nudity (‘deepnudes’)” (doc. 4, p. 6). In cases where personal rights are violated, the French SREN Act (Art. 226–8, French Criminal Code) (doc. 3) designates an entity to take measures, including blocking access to such harmful content. Canada’s proposed *Bill C-63* (doc. 1) also emphasizes personal rights and prevention of pornography. The *UK’s Online Safety Act* (doc. 25) is narrower and focuses on the psychological damage caused by sexually explicit or emotionally manipulative deepfakes.

Document analysis also revealed differences in enforcement structures. The EU relies on a decentralized model in which enforcement is led by national regulators, typically from the member states where a company is headquartered. The cross-border oversight is regulated through mechanisms like the European Digital Services Board. The *EU AI Act* (doc. 2) says, “The European Artificial Intelligence Board (the ‘Board’) should support the Commission, to promote AI literacy tools, public awareness and understanding...” (doc. 2, p. 6). Meanwhile, the UK centralizes enforcement in the hands of Ofcom, while

Germany and France pursue judicial enforcement through criminal sanctions. The *SREN Act* (doc. 3) empowered the administrative authorities in dealing with deepfakes. And Canada obligated the social media operators and AI operators to make the content inaccessible to users inside the country if required.

RQ3: Gaps about Natural Disasters and Violent Conflict Contexts

Our analysis reveals a notable lack of legal or policy measures that specifically target the use of deepfakes in the context of natural disasters or violent conflict escalation. None of the frameworks examined, except for the *DEEPFAKES Accountability Act*, contains relevant provisions. This is equally true of international frameworks, as neither the Hiroshima AI Process (doc. 32) nor the UN Cybercrime Convention (doc. 30) included provisions to address the possibility of deepfake-induced crisis at the time of natural disaster or violent conflict. Though the *DEEPFAKES Accountability Act* (doc. 9) incorporated the possibility of using deepfakes to incite violence or diplomatic conflict, the way it was incorporated lacks a framework to address such a challenge. It says, “whoever knowingly alters an advanced technological false personation record to remove or meaningfully obscure the disclosures required...with the intent to cause violence or physical harm, incite violent or diplomatic conflict, or interfere in an official proceeding, including an election, provided the advanced technological false personation record did in fact pose a credible threat of instigating or advancing such” (doc. 9, p. 6).

The emphasis across the regulatory frameworks remains on broad digital policy and AI safety principles, not rapid-response or content-specific interventions. As such, we conclude that there is a regulatory gap precisely in situations where the consequences of manipulated information could be the most destabilizing. This absence exposes a conceptual limit in the way harms arising from deepfakes are framed in current regulations and legislation. Most existing laws deal with deepfakes from the angle of personal injury (defamation, fraud, or invasion of privacy) or democratic integrity (electoral interference), not from an anticipated risk perception standpoint that may have a ripple effect over society at large.

Therefore, this study’s critical finding is that the existing and proposed policy frameworks, while useful for the general governance of AI-generated content, are not currently adequate to the challenge posed by deepfakes in crisis environments. Across the documents from various countries, the lack of specificity means that there is no clear authority to take preventative action. Additionally, there is no coordinated infrastructure for either domestic or cross-border response. For instance, Japan’s AI Guidelines for Business version 1.0 (doc. 6) acknowledged the potential social risk of GenAI-generated fake content, but lacks specification of coordinated remedial measure to prevent: “Generative AI has enabled everyone to forge fake information that seems to be true and fair, so recognize the increasing risk of destabilizing and confusing the society through

disinformation, misinformation, and biased information generated by AI...” (doc. 6, p. 14). In short, the current regulation cannot address the potential for risks to amplify through the social system and media ecosystem and cannot disrupt or tame the amplification process.

RQ 4: Guidelines to Mitigate Deepfakes in Relation to Natural Disaster Response and Conflict Escalation

The overall finding is that while current regulatory frameworks provide general tools to combat deepfakes, they lack the specificity to address the risks involved in natural disaster response or conflict escalation. Consequently, the core challenge is not the absence of deepfake regulation per se, but rather the mismatch between how existing laws and regulations are designed and what specific operational demands exist to deal with crisis scenarios arising from deepfakes.

Many laws, such as *EU AI Act* (doc. 2), *California’s SB 942* (doc. 18), and *UK Online Safety Act* (doc. 25), treat deepfakes as one threat among many within broader concerns like disinformation, platform liability, or AI transparency. So, when applied to emergency contexts, these tools can create ambiguity. This ambiguity raises doubts about whether existing standards on deception, data misuse, or psychological harm can be triggered quickly enough to prevent harm when it comes to situations like evacuations or ceasefires. By relying on laws built for other kinds of digital harms, enforcement bodies could face uncertainty about their jurisdiction and response time.

Moreover, deepfakes can spread at high speed and across various platforms before being detected. Detection tools may not be reliable or fast enough to act as effective safeguard mechanisms during real-time crises. This undermines the effectiveness of laws or regulations that depend on content identification as a prerequisite for platform action or legal intervention. Also, many frameworks delegate responsibility to platforms (e.g., *EU’s Digital Services Act*, *UK Online Safety Act*), but enforcement remains largely inconsistent. Some platforms are investing heavily in moderation infrastructure and transparency, while others might still lack the resources or are protected by legal boundaries. In addition, in the absence of clear standards for distinguishing harmful deepfakes from satire or dissent, platforms are caught in a bind. In some political contexts, they may be pushed to quickly remove content, which can lead to over-censorship or selective enforcement. In others, they may hesitate to remove harmful content even in a crisis, out of real or perceived risk that they may be accused of political bias. These risks are particularly high in politically sensitive contexts, where a misjudgement of an image or deepfake can either amplify misinformation or undermine self-expression. For instance, *AI Guidelines for Business Version 1.0* says, “Generative AI has enabled everyone to forge fake information that seems to be true and fair, so recognize the increasing risk of destabilizing and confusing society through

disinformation, misinformation, and biased information generated by AI...” (doc. 6, p. 14). This example captures the generalized definition with no crisis or risk angle. The following illustration points to consumer deception, not rapid harm prevention in public emergencies: “Deepfakes create consumer deception issues, where persons can create ‘deepfake’ images and videos to fool consumers...” (doc. 14, p. 3).

Next, the example from the *EU AI Act* (doc. 2) reflects a delegation to content deployers/platforms, requiring that “deployers of an AI system... shall disclose that the content has been artificially generated or manipulated” (doc. 2, p. 82). The Japanese AI guidelines outline expectations for platforms but lack consistent enforcement mechanisms. For example, they call for the introduction of “labelling and disclaimer labelling to help AI business users and non-business users know that they are interacting with the AI system” (doc. 6, p. 25). And, finally, the *Criminal Protection of Personal Rights Against Deepfakes* (doc. 4) emphasizes the heightened risk in political contexts.

In summary, while the existing and proposed country-specific laws and regulations reviewed offer useful principles, they were not designed to address the kinds of digital harms or risks arising from the use of deepfakes in natural disaster or violent conflict crises. Global initiatives also contained the same limitations. Despite encouraging responsible innovation and digital transparency, frameworks like the Hiroshima AI Process, Global Digital Compact, and Budapest Convention lack binding enforcement mechanisms. Overall, these initiatives failed to provide concrete strategies for crisis management or harm mitigation from deepfakes deployed to mislead the public in natural disasters or to initiate or escalate violent conflict.

SARF-Informed Risk Management Approach

Examining the results through SARF’s lens, we found that authorities view deepfake-induced risk from a content governance standpoint, not from the perspective of communication governance or risk mitigation. The commonality across the frameworks is to ensure content authenticity (e.g., labelling, disclosure), content removal (platform responsibility), and content creation prevention (criminal liability). From a SARF standpoint, this is insufficient. Pointing out this gap in the context of media research, Kaspersen et al. (2022) wrote that identifying or “clarifying the process by which media coverage [as a part of risk amplification] influences risk perceptions” (p. 1370). The documents under consideration identified some but not all of the risks that deepfakes may pose to society and lacked intervention mechanisms intended to disrupt or tame the amplification of risk through the social system.

Rather, the measures outlined were highly focused on content-level regulatory interventions. Specifically in the context of natural disasters, when deepfakes move fast and can shape the beliefs of people in distress, this content-focused regulatory approach may fail to address the crisis. Therefore, first, a comprehensive plan similar to the natural

disaster multi-faceted management plan (Rawluk et al., 2018) might be required. In the multi-faceted approach, different social actors have predefined roles and get activated in crises to intervene and identify deepfakes spread in the affected area. Deepening the roles these social actors play and accounting for the national-level measures, governments might be better prepared to deal with deepfake incidents during natural disasters. In any natural disaster situation, the objective of a disaster communication management plan is not to resist the source of the natural disaster itself, such as a tornado, cyclone, or heavy rainfall, but to resist misinformation that surrounds the disaster and reaches affected communities.

In the context of regulating deepfakes, measures can be aimed at eliminating the sources. But the elimination of the source is only one part of the deepfake-induced crisis. The second part, activation of relevant social actors, is still missing from current regulation.

According to SARF, the risk can be amplified based on the roles of societal actors and their behaviour. Feedback loops in the stages of risk amplification are critical for either neutralization or amplification. Hence, the focus should be on preparing a comprehensive and multi-faceted plan by incorporating social actors, including individual social members, community institutions, civil society groups, opinion leaders, journalists, and platform moderators into the plan to minimize the possible influence of deepfakes during natural disaster crises. Along with content-level regulatory measures, this framework should include community resilience through building awareness, AI literacy, and a nationwide community-specific deepfake crisis response team. Additionally, the framework would help disseminate messages through trusted community sources and build a sense of belonging.

This framework can be a defensive mechanism for deepfake-induced crises occurring during natural disasters and (with some additional measures) offers promise for managing deepfake campaigns intended to initiate or escalate violent interstate conflicts. In these latter cases, the activation of individuals and organizations to dampen risk amplification must include diplomats, military and security officials, as well as international organizations with conflict management mandates and expertise. The end goal is for society to resist deepfakes in crises, particularly today, when citizens are often engaged in digital spaces more intensely than their immediate, face-to-face communities.

Limitations

No study is without limitations. In this project, a keyword approach to identify relevant textual data for open coding and the use of ChatGPT to find initial quotations in documents could have limited the extraction of relevant data for the first phase of coding. However, the keyword approach and the AI tool allowed us to analyse the data, presenting a comprehensive overview of the current regulatory frameworks. Additionally,

regulatory mechanisms, policies, acts, and guidelines for dealing with deepfakes continue to evolve and expand rapidly. By the time this study is published, it may not be up to date on all G7, UN, and EU regulatory frameworks. At the same time, this research offers an important perspective on the status of deepfake regulations at the time it was conducted. Finally, recommendations offered based on the SARF-informed risk management approach remain normative and require additional testing in the context of natural disasters.

Conclusion and Future Studies

The contributions of this study are threefold. First, the study conceptualizes the harmful, ill-motivated, and unethical deepfakes as a risk rather than a mere content-level problem. Second, this study offers a systematic analysis of existing regulatory frameworks that identify the commonalities, divergences, and gaps in addressing deepfakes in the context of natural disasters or violent conflicts. Third, drawing on the conceptualization of deepfakes as a societal risk, this study provides a SARF-informed risk management approach (Kasperson & Kasperson, 1996) to mitigate deepfakes in times of crises.

In the documents analysed for this study, legal and regulatory responses to deepfakes revealed patterns that are both convergent and divergent. Most of the regulatory documents shared common objectives, such as transparency, user protection, and the containment of misleading synthetic content. However, the way in which these objectives are implemented varies considerably. We found no clear guidelines on how to manage the risks of deepfakes during natural disasters and the escalation of violent conflicts, which raises concerns about whether the existing frameworks are adequately equipped to deal with such scenarios. This finding supports the guidelines and recommendations from previous literature on the importance of managing deepfakes in crises, as malign actors might work to generate deepfakes to create chaos or incite violence (Hilberts et al., 2025; Topor, 2024).

Our research demonstrated that, in addition to national frameworks, several international organizations have proposed principles relevant to deepfake governance, including the Hiroshima AI Process, the Global Digital Compact, the UN Cybercrime Convention, and the Budapest Convention on Cybercrime. While these initiatives underscore the importance of transparency, safety, and digital accountability, they remain non-binding and therefore not enforceable, despite varying significantly in scope. As deepfake governance continues to grow worldwide, we hope that future studies will focus on expanding the SARF-informed deepfake-induced risk management framework that we introduced here. Specifically, future studies can develop a comprehensive and multi-faceted approach by inviting various social actors to recognize and minimize the presence of deepfakes in crises. Additionally, examination of the already existing legislation by the non-G7 and non-EU countries (e.g., specifically, Brazil, Colombia,

Kazakhstan, Nigeria, Saudi Arabia, South Korea, South Africa, and Viet Nam, among others) in the regions of Central and South-Pacific Asia, Africa, and Latin America will offer useful insights into the various approaches of dealing with deepfakes. Such future research has the potential to contribute to conversations about the global guidelines and regulations in the future to prevent the use of deepfakes to incite violence and compromise the natural disaster response worldwide.

Biographical notes

Katerina Tsetsura, Ph.D., is Gaylord Family Professor of Public Relations and Strategic Communication at the Gaylord College of Journalism and Mass Communication at the University of Oklahoma in the USA. Dr. Tsetsura is internationally known for her work in global public relations and media transparency. She is a co-author of *Transparency, public relations, and the mass media: Combating hidden influences in news coverage worldwide* (2017, Taylor & Francis) and co-editor of *Strategic Communications in Russia: Public Relations and Advertising* (2021, Taylor & Francis). Her research areas include studying societies in transition, global public relations development, media transparency, public diplomacy and government relations, and understanding and enhancing community resilience.

HM Murtuza, a doctoral student at the Gaylord College of Journalism and Mass Communication at the University of Oklahoma in the USA. After a decade-long career in journalism, he is now focused on studying how newsrooms and journalists are adopting Generative AI (GenAI) and what contextual factors shape technology adoption trajectories in non-Western journalism.

Mark Raymond, Ph.D., is the Wick Cary Associate Professor of International Relations and the Director of the Cyber Governance and Policy Center at the University of Oklahoma. He also serves as the Associate Director for International Security Policy with the Oklahoma Aerospace and Defense Innovation Institute. He is the Associate Editor of the journal *International Theory* and Co-Chair of the American Political Science Association's International Relations Theory Section. He is the author of *Social Practices of Rule-Making in World Politics* (New York: Oxford University Press, 2019). His research sits at the intersection of international relations theory and the role of digital technologies in world politics.

Typhaine Joffe recently completed her Master of Arts in International Studies at the University of Oklahoma, where she served as a Graduate Research Assistant. She also holds a Master's degree in International and European Studies from the University of Valencia and a Bachelor's degree in European Studies in English from the University of Clermont Auvergne. Her research explores the intersection of cybersecurity, international law, and diplomacy.

References

- Ahmed, S. K. (2024). The pillars of trustworthiness in qualitative research, *Journal of Medicine, Surgery, and Public Health*, 2, 100051.
<https://doi.org/10.1016/j.glmedi.2024.100051>
- Balk, E.M., Chung, M., Hadar, N., Patel, K., Yu, W.W., Trikalinos, T.A., & Chang, L. K. W. (2012). Accuracy of Data Extraction of Non-English Language Trials with Google Translate [Internet]. *Rockville (MD): Agency for Healthcare Research and Quality (US)*. Report No.: 12-EHC056-EF. PMID: 22624170.
- Barari, S., Lucas, C., & Munger, K. (2025). Political deepfakes are as credible as other fake media and (sometimes) real media. *Journal of Politics*, 87(2).
<https://doi.org/10.1086/732990>
- BBC (2019, June 13). *Deepfake videos could 'spark' violent social unrest*.
<https://www.bbc.com/news/technology-48621452>
- Birrer, A., & Just, N. (2024). What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape. *New Media & Society*, 27(12), 6819-6838. <https://doi.org/10.1177/14614448241253138>
- Business Software Alliance (2024, February 14). *BSA analysis: States intensify the work on AI regulation*. Press release. <https://www.bsa.org/news-events/news/bsa-analysis-states-intensify-work-on-ai-legislation>
- Cajueiro, D. O., & Rezende Celestino, V. R. (2026). A comprehensive review of Artificial Intelligence regulation: Weighing ethical principles and innovation. *Journal of Economy and Technology*, 4, 77-91. <https://doi.org/10.1016/j.ject.2025.07.001>
- Chapagain, D., Kshetri, N., & Aryal, B. (2024). Deepfake Disasters: A Comprehensive Review of Technology, Ethical Concerns, Countermeasures, and Societal Implications. In *Conference proceedings of the 2024 International Conference on Emerging Trends in Networks and Computer Communications (ETNCC)*, pp. 1-9. IEEE, 2024. <https://doi.org/10.1109/ETNCC63262.2024.10767452>
- Cochran, J. D., & Napshin, S. A. (2021). Deepfakes: Awareness, concerns, and platform accountability. *Cyberpsychology, Behavior, and Social Networking*, 24(3), 164-172. <https://doi.org/10.1089/cyber.2020.0100>
- Dasharathraj K. et al. (2025). Analyzing AI regulation through literature and current trends. *Journal of Open Innovation: Technology, Market, and Complexity*, 11(1), 100508. <https://doi.org/10.1016/j.joitmc.2025.100508>.
- Dauer, F. (2022, June 29). Law Enforcement in the Era of Deepfakes. *Police Chief Online*. <https://www.policechiefmagazine.org/law-enforcement-era-deepfakes/>
- Denzin, N. K. (2009). Apocalypse Now: Overcoming Resistances to Qualitative Inquiry. *International Review of Qualitative Research*, 2(3), 331-343.
<https://doi.org/10.1525/irqr.2009.2.3.331>
- EU (2025). *The EU Artificial Intelligence Act*. <https://artificialintelligenceact.eu/>

- Geng, Y. (2023). Comparing “deepfake” regulatory regimes in the United States, the European Union, and China. *Georgetown Law Technology Review*, 7, 167-178. <https://georgetownlawtechreview.org/wp-content/uploads/2023/01/Geng-Deepfakes.pdf>
- Hameleers, M. (2023). Disinformation as a context-bound phenomenon: toward a conceptual clarification integrating actors, intentions and techniques of creation and dissemination, *Communication Theory*, 33(1), 1–10. <https://doi.org/10.1093/ct/qtac021>
- Hilberts, S., Govers, M., Petelos, E., & Evers, S. (2025). The impact of misinformation on social media in the context of natural disasters: Narrative review. *JMIR Infodemiology*, 5, 70413. <https://doi.org/10.2196/70413>
- HybridCoE (2020). *Trends in the Contemporary Information Environment*. HybridCoE Trend Report 4. HybridCoE Publication. <https://www.hybridcoe.fi/wp-content/uploads/2020/07/Hybrid-CoE-Trend-Report-4.pdf>
- Kasperson, R. E., & Kasperson, J. X. (1996). The Social Amplification and Attenuation of Risk. *The ANNALS of the American Academy of Political and Social Science*, 545(1), 95-105. <https://doi.org/10.1177/0002716296545001010>
- Kasperson, R. E., Webler, T., Ram, B., & Sutton, J. (2022). The social amplification of risk framework: New perspectives. *Risk Analysis*, 42(7), 1367–1380. <https://doi.org/10.1111/risa.13926>
- Kerbage, R. (2025). An examination of AI edited photos and videos of actual footage during war times and natural disasters, and its impact on the perception of the public opinion. *Crossroads of Social Inquiry*, 1(2), 55-68. <https://doi.org/10.64851/csi.v1i2.31>
- Kugler, M. B., & Pace, C. (2021). Deepfake privacy: Attitudes and regulations. *Northwestern University Law Review*, 116(3), 611-680. https://heinonline.org/HOL/Page?handle=hein.journals/illlr116&div=21&g_sent=1&casa_token=&collection=journals
- Lawson, A. (2023, April 24). A look at global deepfake regulation approaches. Blog. *Responsible Artificial Intelligence Institute*. <https://www.responsible.ai/a-look-at-global-deepfake-regulation-approaches/>
- Lee, T. (2019). The global rise of “fake news” and the threat to democratic elections in the USA. *Public Administration and Policy: An Asia-Pacific Journal*, 22(1), 15-24. <https://www.doi.org/10.1108/PAP-04-2019-0008>
- Quirk, C. (2023). *The high stakes of deepfakes: The growing necessity of federal legislation to regulate this rapidly evolving technology*. Princeton University.
- Romero Moreno, F. (2024). Generative AI and deepfakes: A human rights approach to tackling harmful content. *International Review of Law, Computers & Technology*, 38(3), 297–326. <https://doi.org/10.1080/13600869.2024.2324540>
- Rawluk, A., Ford, R. M., & Williams, K. J. H. (2018). Value-based scenario planning: exploring multifaceted values in natural disaster planning and management. *Ecology and Society*, 23(4). <https://www.jstor.org/stable/26796851>

- Spiggle, S. (1994). Analysis and interpretation of qualitative data in consumer research. *Journal of Consumer Research*, 21(3), 491-503. <https://doi.org/10.1086/209413>
- de Rancourt-Raymond, A., & Smaili, N. (2023). The unethical use of deepfakes. *Journal of Financial Crime*, 30(4), 1066–1077. <https://doi.org/10.1108/JFC-04-2022-0090>
- Sample, I. (2020, January 13). What are deepfakes—and how can you spot them? *The Guardian*. <https://www.theguardian.com/technology/2020/jan/13/what-are-deepfakes-and-how-can-you-spot-them>
- Topor, L. (2024). Mis/disinformation and national resilience: Are countries immune to fake news? In *Cyber Sovereignty. Global Power Shift* (pp. 111-131). Springer, Cham. https://doi.org/10.1007/978-3-031-58199-1_5
- Van der Sloot, B., & Wagenveld, Y. (2022). Deepfakes: regulatory challenges for the synthetic society. *Computer Law & Security Review*, 46, 105716. <https://doi.org/10.1016/j.clsr.2022.105716>
- Vig, S. (2024). Regulating Deepfakes: An Indian perspective. *Journal of Strategic Security*, 17(3), 70-93. <https://doi.org/10.5038/1944-0472.17.3.2245>
- Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology innovation management review*, 9(11), 39–43. <https://timreview.ca/article/1282>

Risks of gender communication strategies in the Public Sector. The European Union perspective*

Gea Ducci^{**}

University of Urbino Carlo Bo

Camilla Folena^{*****}

University of Urbino Carlo Bo

Lucia D'Ambrosi^{***}

Sapienza University of Rome

Marica Spalletta^{*****}

Link Campus University of Rome

Paola De Rosa^{****}

Link Campus University of Rome

This article explores the risks of gender-sensitive communication in the public sector, with a focus on the European Union (EU) that adopted significant gender and inclusive community policies (e.g. the recent Gender Equality Strategy 2020-2025). The analysis is framed within a broader theoretical context, where public communication is understood as a strategic tool for strengthening democratic systems, fostering trust, and promoting social inclusion in an increasingly fragmented digital society. Theoretical perspectives on institutional public sector communication, inclusivity, intersectionality, and media polarization are considered to highlight the tensions between normative policy ambitions and the contested, often politicized, nature of gender public discourse in Europe. Against this background, the study aims to understand what kinds of European communication campaigns on gender issues are conducted, how they are disseminated across different platforms, and which risks emerge from the management of media and channels. Additionally, it seeks to examine how gender-related themes are framed and represented, and what risks may arise in terms of inclusivity and effectiveness. To achieve these objectives, the study considers four selected EU communication campaigns launched between 2020 and 2024. A qualitative approach is adopted, drawing on media content analysis and critical interpretation of campaign narratives. The findings, briefly outlined, point to structural, socio-cultural, and political risks that limit the transformative potential of EU gender-sensitive communication.

Keywords: Public communication, European diversity campaign, strategic communication risks, polarization, gender stereotypes.

* Article submitted on 20/09/2025. Article accepted on 15/12/2025.

** gea.ducci@uniurb.it

*** lucia.dambrosi@uniroma1.it

**** p.derosa@unilink.it

***** camilla.folena@uniurb.it

***** m.spalletta@unilink.it

Introduction

Within the scholarship on public communication undertaken by public sector organizations, the strategic communication approaches to inclusive and gender-related issues - such as the use of gender-sensitive language - have gained particular relevance in recent years. These issues have become especially salient in today's fragmented public debate and are marked by varying levels of controversy, reflecting a complex and ongoing cultural transformation within digital society.

This contribution aims to reflect on the challenges faced by public institutions in this regard, highlighting the main risks associated with strategic gender-sensitive public communication. The analysis focuses on the communication approaches adopted by the European Union over the past five years, within the framework of the Gender Equality Strategy 2020-2025, with specific reference to institutional communication practices (major communication campaigns).

Emerging sustainability, inclusion challenges, and risks for strategic public communication in the digital society.

Within a society characterized by turbulence, recurrent crises and emergencies, conflicts, and by the imperative to undertake profound and cross-cutting transitions towards sustainability - environmental, socio-economic, and digital - (Oliveira & Gonçalves, 2024), institutional public communication assumes a pivotal and strategic role.

Indeed, it can be regarded as an indispensable instrument enabling the public sector to steer ongoing processes rather than passively undergo them, while upholding and enhancing democratic systems and principles, and contributing to the construction of a relationship of trust between citizens and institutions (OECD 2021; Canel & Luoma-aho, 2019; Ducci & Lovari, 2021).

Institutional public communication (Lovari & Ducci, 2022) can, however, fulfill this role when it can confront emerging challenges by recognizing and navigating with awareness the current hybrid, convergent, and ever-evolving media ecosystem (Jenkins et al., 2013; Bentivegna & Boccia Artieri, 2021; van Dijck et al. 2018). This is a rather complex task, considering that we are faced with plural, dense, and fragmented public spheres, characterized by increasing informational disorder, the phenomenon of disinformation/misinformation, and a tendency toward opinion polarization driven by algorithmic logics (Boccia Artieri *et al.*, 2025).

In addition, recent studies highlight the rise of so-called 'incivility' in communicative practices which, although not a new phenomenon, now appears to be intensifying and taking on new characteristics in digital environments - emerging forms of violence expressed through words and images, aggressiveness, and denigration of others, among others - (Bentivegna & Rega, 2024).

In this context, among the many challenges posed by the transitions toward sustainability (Agenda 2030, etc.), public institutions - acting in step with the cultural movements and processes currently unfolding within society and enabled by digital technologies - are pursuing social inclusion and equal opportunity goals with greater intensity than in the past. They have done so through the adoption of policies supported by specific institutional public communication strategies (as will be illustrated in the following section with reference to the European Union). In particular, several studies emphasize that public communication is now widely recognized as a strategic tool for fostering an inclusive society, especially in the aftermath of the pandemic (Canel & Luoma-aho, 2019; Faccioli & D'Ambrosi, 2023; Spalletta, Ducci, D'Ambrosi & Folena, 2024; Oliveira & Gonçalves 2024).

However, addressing issues of inclusion and, more specifically, adopting gender-sensitive approaches in institutional communication entails that public administrations make carefully considered strategic choices, with full awareness of the potential risks involved.

Risks may stem partly from the characteristics of the contemporary public spheres mentioned above, and partly from factors related to organizational and communicative culture. These include organizational structure and available communication tools, the recognition of professional communicators (Grunig, 2008), prevailing communication models/approaches, one-way or two-way/relational communication planning, and the relationship with publics and stakeholders.

First and foremost, addressing emerging social issues is a duty for public administrations, since by their very nature, they are called upon to pursue the public interest. However, in the way these issues are approached - through communication style (language and tone of voice), the use of channels and communication tools, the tailoring of messages to different target audiences, and the fostering of engagement - institutions run the risk of failing to respect the diverse sensitivities present within society and, as a result, of generating rejection or alienation.

These 'precautionary measures' must, at the same time, be combined with the capacity to exercise agency, stimulating changes in citizens' attitudes and behaviors, or enhancing and supporting processes of change already underway within society. This does not mean that public administrations should never take a stand on certain issues, but rather that they should do so by explaining their reasons in a transparent and contextualized manner and, whenever possible, by linking the promotion of ideas, behaviors, or attitudes to tangible elements (such as the implementation of effective policies, the provision of accessible services, etc.) (Lee, Kotler & Colehour, 2023).

Today, in relation to gender-related issues, these risks are amplified by the 'information disorder' that characterizes public spheres (including the risk of disinformation/misinformation), by the multiplicity of actors and voices crowding digital environments (information overload), and by the fact that gender issues have unfortunately become 'controversial' and subject to politicization in recent years; as a result, they appear to be increasingly polarized and polarizing.

Recent gender and media studies significantly focus on how gender is represented in both traditional and digital media narratives, with particular attention to the construction of stereotypes (Ross & Padovani, 2017; Buonanno & Faccioli, 2020; Farci & Scarcelli, 2022).

Some studies highlight growing interest in the intersection of gender and diversity issues in politics and political communication (Campus, 2013; Saccà & Massidda, 2021; De Blasio & Selva, 2024). These dynamics are further exacerbated by the rise of digital polarization, where online spaces – initially seen as opportunities for democratic engagement – have also become arenas for hostility, incivility, and gender-based violence. There is clear evidence that women and gender nonconforming people are primary targets of hate speech and disinformation campaigns, which often follow specific patterns aimed at excluding them from public debate and cultural spaces (UNESCO, 2021).

Conflicts over gender issues have emerged within official and online channels, amplified by media coverage, resulting in greater polarization and incivility (Bentivegna & Boccia Artieri, 2021; Farci & Scarcelli, 2022). Gendered disinformation targets women and gender-diverse people with attacks based on stereotypes, including shaming them for entering male spaces, challenging inequality, or defying dominant norms (Di Meco, 2019; Martins, 2024).

Moreover, in an effort to counter potential biases (whether related to gender or to other aspects of inclusion from an intersectional perspective), within such a context, there is a risk of generating misunderstandings, reproducing existing stereotypes, or even reinforcing stigmatization.

The risk of reproducing gender bias is, moreover, increasingly under scrutiny in connection with the use of generative Artificial Intelligence technologies, which are being adopted with growing frequency by public sector communicators. These technologies tend to generate textual and visual content by drawing on materials already available on the web and, as a result, often replicate entrenched stereotypes in the representation of social roles and other gender-related characteristics.

Other risks may concern the ability, or lack thereof, to avoid fragmentation, to coordinate and ensure continuity in gender-sensitive institutional communication activities - especially at the inter-institutional level - and to establish, or fail to establish, 'communication alliances' with other institutions and non-institutional actors (such as non-profit organizations and profit-oriented companies) (Ibidem).

The evolution of a gender-sensitive and intersectional approach in the European Union public communication

The growing attention toward an intersectional approach in public sector organizations is fostering the adoption of inclusive communication practices, which support the integration of minorities and the incorporation of gender-sensitive perspectives into diversity management policies (Basaglia, Cuomo, & Simonella, 2022; Ravazzani, Mazzei, Butera, & Fisichella, 2023; Sebastião & Cotton, 2025; S. Kim, Buzzanell, Mazzei, & J.-N. Kim, 2025). Considering Grunig's (2008) theory of excellence, a proposed approach recognizes and values the inherent differences within organizations and the effect of the growing number of women in communication structures on excellent public relations. This approach improves organizations' relationships with their publics and plays a crucial role in ethical decision-

making processes. Furthermore, it creates a positive workplace culture by promoting diversity and inclusion as the main components of corporate identity in line with Agenda 2030 for Sustainable Development (Goal 5: Gender Equality and Goal 10: Reduced Inequalities).

At the international level regulations have increasingly set principles to guarantee accessible and inclusive communication (OECD, 2021, 2023). Specifically, European policies and measures on diversity, equity, and inclusion (DE&I) emphasize public communication as a strategic tool for promoting equality and well-being through balanced media representation and open, respectful discourse. The institution of the European Diversity Month, organised with the support of the 27 member states, reflects the commitment of EU institutions and stakeholders to promote diversity, equity, and inclusion, supporting the communication through strategic approaches and practical tools, such as social media toolkits. A significant cross-cutting objective is gender mainstreaming, which advocates the integration of the gender perspective into all community policies and actions.

Over the past decade, EU institutions have increasingly focused on gender-sensitive communication, embedding it within a gender mainstreaming and intersectional perspective (Crenshaw, 1991; McRobbie, 2009; Ghigi & Rottenberg, 2019). This evolution is reflected in key documents like the Gender Equality Strategy 2020-2025 and in the work of bodies such as the European Institute for Gender Equality (EIGE) to raise awareness about gender sensitive language and provide practical advice for policymakers (EIGE, 2019; European Parliament, 2019; European Commission, 2020). However, despite the attention devoted to these issues in European public policies and the availability of a range of resources for the promotion of Europe's linguistic and cultural diversity, such as the Council of Europe's *Inclusive Language Guidelines* (2024), gender-aware communication still tends to remain limited to specific actions within European institutions. Moreover, its implementation is often perceived as a risk in relation to issues of cultural tradition, which are sensitivities that warrant respectful consideration and debate. At the same time, the growing use of AI-powered tools in public communication can further reproduce or even amplify existing gender biases (European Institute for Gender Equality, *Words Matter* guide, 2025).

Audiovisual content has become central to public sector communication strategies in Europe, as institutions increasingly blend traditional methods with innovative, visually-driven storytelling to promote inclusivity and engagement and protect vulnerable groups in a converged environment (European Commission, 2013), particularly through visual storytelling strategies (Mergel & Bretschneider, 2013; Ducci, Lovari & D'Ambrosi, 2019; Luoma-aho & Badham, 2023). This strategic use of audiovisual content is particularly relevant in the context of gender equality, as the audiovisual sector plays a crucial role in shaping societal perceptions, ideas, attitudes, and behaviours, making it a key driver for fostering equality and inclusion (Recommendation on gender equality in the audiovisual sector).

From this perspective, public administrations are changing and innovating their communication strategies and languages (e.g., visual communication, storytelling techniques, etc.), by using appropriate communication channels, diversifying and adapting the messages in the digital ecosystem, even if new challenges arise concerning the

development of professionals' skills and abilities and the improvement of accessibility measures in public communication (OECD, 2021; Lovari & Ducci, 2022; D'Ambrosi, Ducci, Lovari & Folena, 2023; Luoma-aho & Badham, 2023). These communication practices started with the landing of public entities on visually-driven social media – such as Instagram, YouTube, or nowadays TikTok – and followed by the development of visual strategies, which seem to address the need to engage with different publics, particularly the younger segments, combining a strengthened digital identity of public sector communication with a long-term perspective (Luoma-aho & Canel, 2020). This shift is particularly evident in the domain of gender-related topics, where institutions face the dual challenge of developing communication strategies that reflect inclusivity and diversity, as highlighted in international debates (UNESCO, 2021; European Commission, 2020), while enhancing the skills and capacities of public sector professionals to address these issues effectively (OECD, 2021).

One of the key challenges in inclusive communication is overcoming the gender gap in public discourse, particularly in institutional storytelling. Too often, even well-intentioned communication strategies risk reproducing tokenistic or superficial approaches, rather than fostering profound structural change. Recognizing this, European institutions have increasingly integrated audiovisual media into their communication strategies, seeking to promote gender sensitive language through impactful and engaging content¹.

Within this theoretical background, the study advances an analytical lens that integrates gender sensitivity, intersectionality, and strategic risk communication to reconceptualise European inclusive communication within a risk-aware perspective attentive to the dynamics of the digital transition.

A study to explore the EU's approaches to gender communication

Grounded in this theoretical framework, a study investigates how, over the last five years (2020-2025), the European Union has communicated gender-related issues, focusing particularly on the growing reliance on audiovisual content, in their guise of "raising awareness tools" (EIGE, 2016) and the innovative use of high-engagement of social media platforms (e.g.: Meta, YouTube), as a key factor for change (Chadwick, 2006).

Research aims, methods, and data collection

The research aims to explore the potential risks that EU institutional communication campaigns face in navigating between normative policy strategies, particularly focused on gender issues, and the increasing politicization and ideologically driven reframing of gender topics in the public debate.

The research goal is to understand how EU institutional campaigns deal with the tension between normative policy strategies and the affective and ideologically-driven nature of gender discourse in the digital age.

The hypothesis could be summarised as follows: the particularly sensitive nature of gender-related issues – intersecting diverse socio-cultural contexts and heterogeneous political variables across the European Union – represents a significant risk to the effective translation of EU policy frameworks and guidelines into communicative practices that embrace the wide scenario of gender-related issues also from an intersectional perspective.

To test this hypothesis, two research questions were formulated:

- RQ1. How are the gender-related EU communication campaigns disseminated across different platforms and channels (media formats, audiovisual techniques, cross-platform strategies, target adaptation, etc.), and which potential risks may be detected in terms of the impact on connected publics?
- RQ2. How are gender-related themes framed (content, values, identity representation, discursive strategies, etc.), and which potential risks may arise?

From a methodological perspective, the research adopts a qualitative approach (Altheide & Schneider, 2013), carrying out a media content analysis (Macnamara, 2005; Tipaldo, 2014) on a selected sample of communication campaigns released by the European Union since 2020, namely: *CharacterHer – Empowering All Talents in the Film and Media Industries* (2021); *Keep Your Eyes Open* (2023); *#EndGenderStereotypes* (2023); *Spot the Violence Stop the Violence* (2024).

The decision to adopt a sample methodology was based on the need to analyse a limited textual corpus, while still representing – through both verbal/visual strategies and dissemination practices – the full range of campaigns released, as required in qualitative research.

The time frame was chosen based on two types of factors: on the one hand, a normative one, namely the launch of the *EU Gender Equality Strategy* and the *Creative Europe MEDIA Programme* in 2020; on the other hand, a socio-cultural factor, namely the growing public, political and institutional attention to gender-related issues in the last five years, particularly in the wake of the COVID-19 pandemic – which led to the implementation of specific EU policies to counter the effects of the “She-cession” (A.T. Kim, Erickson, Zhang, C. Kim, 2022) – as well as the impact of high-profile cases of gender-based violence and gender inequalities which have profoundly affected public debate.

To this purpose, a background analysis was conducted on the EU’s official websites and social media accounts to collect all campaigns addressing gender-related issues released in the mentioned timeframe (2020-2024). This mapping identified 18 campaigns, which collectively reflect the EU’s strong and intersectional focus on diversity and inclusion. The campaigns were highly heterogeneous in both themes and audiences, ranging from gender-based violence and minority rights to the representation of women in public life and broader civil-rights issues. From this wider set, four campaigns were selected for in-depth qualitative analysis because they mirrored the thematic diversity of EU gender-related communication while also providing the most complete and structured set of communication materials, in line with the goals of the Gender Equality Strategy 2020-2025. The selection criteria included

topical diversity (addressing different aspects of gender-related issues from an intersectional perspective), *target audience segmentation* (reaching varied demographic and social groups), *strategic variation* (using different formats, tones, and communicative goals), *dissemination practices* (centralized and decentralized communication models and varying levels of digital and audiovisual integration).

The following in-depth analysis of the selected case studies was carried out in two steps. The first one consisted of the analysis of the campaign’s description – if available – provided by the EU, considering its structural elements (e.g.: main goals, target groups, campaign structures and contents, dedicated websites and SoMe accounts, toolkits, dissemination activities, etc.).

The second step consisted in the development of an original codebook, specifically created for this research. The analytical grid was designed following Tipaldo’s (2014) qualitative content-analysis model and integrating typological categories emerging from previous studies (D’Ambrosi et al., 2024; Spalletta et al., 2024), alongside dimensions related to narrative structures, languages, tones (Gadotti & Bernocchi, 2010), primary goals (Kotler, Roberto & Lee, 2002), inclusivity markers, and portrayals of gender roles, with the aim of assessing whether and how audiovisual communication contributes to fostering equality (Table 1).

To ensure methodological transparency and rigor, all authors first coded the same campaign independently to test the analytical framework. Coding outcomes were then compared and discussed collectively, allowing the research team to refine category definitions, align interpretations, and resolve disagreements through extensive consultation to reach an agreement consensus. After that, the codebook was applied to the full dataset through a manual, iterative and inductive process, supported by ongoing discussion among coders.

<i>Dimension</i>	<i>Items</i>
Main Topic	Equality & Rights / GBV / Stereotypes / Empowerment / Education / Health
Objectives	Narrative + Declared campaign goals
Intersectionality	Yes / No
Diversity Categories	Gender identity / Sexual orientation / Ethnic origin / Age / Economic status / Religion / Disability
Portrayed Roles	Family roles / Professional roles / Other roles
Misrepresentation	Caregivers (women only) / Dominant men / Performativity ² / Tokenism / Trans/Non-binary exclusion / Hypersexualization / LGBTQ+ bias / Submissive women / Supporting roles
Tackling Stereotypes	Yes / No → Which stereotypes? (e.g. toxic masculinity, weak women...)
Tone/appeal	Emotional / Rational
Language	Sentimental-Poignant-Pathetic / Dramatic-Violent-Shocking Aggressive-Accusatory-Denunciatory / Reassuring-Gratifying-Positive Funny-Humorous-Ironic / Empowering-Paternalistic-Prescriptive Provocative-Irreverent-Transgressive / Informative-Descriptive-Documentary
Main Goal	Cognitive, Action, Behaviours, Value-based

Table 1- Codebook

To reach an intersubjective and shared interpretation of each campaign, the individual analyses carried out by different coders were tested through a joint visual analysis of the video content by the overall research group.

Main Research Findings

The main research findings will be presented by analyzing each communication campaign, following the chronological order of their release. Before introducing the in-depth analysis, the following Table 2 provides a synthetic comparative overview of how the four EU campaigns address gender-related dimensions, mapping convergences and divergences across key analytical categories, including thematic focus, intersectionality, role portrayals, and risks of stereotyped or performative gender constructions. Notably, while all campaigns include an intersectional perspective, this dimension remains somewhat limited: identity-related factors that are central to gendered vulnerability and marginalization - such as religion, sexual orientation, gender identity, or socioeconomic status - are mostly absent. Instead, intersectionality is expressed through broader social markers, such as age, ethnicity, or disability, resulting in a predominantly social rather than identity-based approach. This combined reading offers an essential entry point into the in-depth analysis discussed in the following section, highlighting both shared tendencies and critical gaps in the EU's gender-sensitive communication strategies.

<i>Dimension</i>	<i>CharactHer</i> (2021)	<i>End Gender</i> <i>Stereotypes</i> (2023)	<i>Let's Keep Your</i> <i>Eyes Open</i> (2023)	<i>Spot The Violence</i> (2024)
<i>Main Topic</i>	(1-3)	(2-3-4)	(1-2-3-4)	(1-2-3)
1) Equality & Rights	Supporting	Gender	Crimes against	Multidimensionality
2) Stereotypes	diversity and	stereotypes in	children;	of gender-based
3) Empowerment	inclusion	balancing family	xenophobic hate	violence
4) Education	enhancing women	and professional	and anti-LGBTIQ	
5) Health	professional	roles, focusing on	hate crimes; violent	
	empowerment in	role and social	crime and gender-	
	the film and media	expectations	based violence	
	industries			
<i>Intersectionality</i> (Yes / No)	Yes	Yes	Yes	Yes
1) Gender identity	(4-7)	(3)	(3-4)	(3-4)
2) Sexual orient.				
3) Ethnic origin				
4) Age				
5) Economic status				
6) Religion				
7) Disabilities				
<i>Portrayed Roles</i>	(2)	(1-2-3)	(3)	(2-3)
1) Family roles				

- 2) Professional roles
- 3) Other roles

<i>Stereotyped Representation</i> 1) Women as caregivers or secondary roles 2) Dominant men / Submissive women 3) Gender performativity 4) Tokenism 5) Trans/No-binary exclusion 6) Hypersexualization 7) LGBTQI+ biases	(3) Performative construction of gender, framing an empowerment-based meritocratic logic, stressing excellence and individual effort criteria (leader, bossy woman, hero...)	(3) Performative construction of gender, framing an empowerment-based meritocratic logic, stressing excellence and individual effort criteria (leader, bossy woman, hero...)	(4) Minimal or symbolic inclusion of marginalized groups, to signal diversity without granting them meaningful visibility, narrative relevance or decision-making power	(4) Minimal or symbolic inclusion of marginalized groups, to signal diversity without granting them meaningful visibility, narrative relevance or decision-making power
<i>Tackling Stereotypes</i> Yes / No and which ones	Yes Women in professional domains are traditionally considered masculine	Yes Women in: Secondary or supporting role or as sole caregivers / Men in: unconventional roles (subversion of conventional gender roles)	No	Both Yes: House-keeping / care roles are assigned to men No: Cooking / Care role assigned to women

Table 2 - Comparison of gender-related issues in the four selected campaigns

CharacterHer – Empowering All Talents in the Film and Media Industries (2021)

The *CharactHer* campaign, promoted by the European Commission within the *Media and Audiovisual Action Plan* (2020) in collaboration with the French-based action tank Collectif 50/50, represents the first initiative at the EU level explicitly focused on diversity and inclusion in the audiovisual sector.

Envisioned as part of the *EU Gender Equality Strategy 2020-2025*, the campaign seeks to challenge persistent gender stereotypes within film and media professions, traditionally viewed as male-dominated domains. The core message – “*There is no such thing as a ‘masculine’ job – women can do anything!*”³ – and the self-description provided highlight the campaign’s double aim, consisting of: dismantling the gender-related career stereotypes, by spotlighting underrepresented positions in the industry to promote diversity; overturning the ‘symbolic annihilation’ of women in the audiovisual sector (Gebner & Gross, 1976), giving them a direct voice, and a main stage, to raise awareness on the idea that “*Who tells the story matters*”.

The campaign's structure is articulated around a series of tools and formats intended to reach a broad public, while emphasizing women professionals and aspiring talents in the audiovisual sector. At its core is the main campaign video, supported by 12 video-interviews with women professionals from 9 EU countries, portraying personal and professional career paths, focusing on gender-related challenges and representations in the film and media industries⁴. These audiovisual contents are complemented by an educational toolkit, addressed to mid/high schools, film schools, and educational organizations across Europe, reflecting the campaign's ambition to serve as a guidance tool for younger generations.

CharactHer is conveyed through a dedicated website⁵ which serves as the campaign's central hub. It was launched by the European Commission via institutional news, mapping goals and activities online. On the one hand, the campaign does not rely on dedicated social media accounts or adopt a coordinated cross-platform and multi-level dissemination strategy, with the official website's social media icons redirecting to the Collectif 50/50 accounts. On the other hand, visibility is primarily achieved through participation in high-profile events, notably the Cannes Film Festival, where the campaign was officially launched in 2021.

From an analytical perspective, also on potential risks evaluation, by addressing the underrepresentation of women in the audiovisual sector, the campaign pursues a clear cognitive goal, aiming to inform and raise awareness of female empowerment in creative roles. The overall tone of voice of the campaign is rational and unfolds into an informative, descriptive, and documentary language. While it appeals to responsibility and reflection rather than emotion, adopting a restrained approach characterized by neutrality and low intrusiveness often risks being weak in persuasion and having limited impact (Gadotti & Bernocchi, 2010).

Consistent with the campaign narrative focus, the roles portrayed are exclusively professional and related to creative and technical aspects of the audiovisual sector. The analysis does not reveal any explicit issue of misrepresentation, with the only exception being an emphasis on gender performativity, which nonetheless seems functional to the campaign's goals.



Figure 1- *CharactHer's* campaign visuals

Regarding intersectional representation, the campaign addresses some diversity dimensions, notably age, disability⁶ and geographically spanning Eastern to Western Europe. A factor that highlights the EU's effort to construct a shared narrative and a broader sense of belonging, bringing together the perspectives of women from different member countries. However, the in-depth analysis clearly shows that *CharactHer* predominantly

represents white cisgender women, with the potential risk of neglecting key dimensions of diversity such as racial and ethnic backgrounds, gender identities beyond the binary, and other underrepresented social groups or experiences.

Keep Your Eyes Open (2023)

The *Keep Your Eyes Open* campaign, launched by the European Commission within the framework of the EU Strategy on Victims' Rights (2020-2025), was developed in response to the significant increase in crimes (e.g., domestic violence, child sexual abuse, cybercrime, xenophobic and LGBTIQ+ hate crimes) reported during the COVID-19 pandemic. The campaign's main goal is to raise awareness of victims' rights and empower individuals to recognize and support victims in their communities, encouraging an active stance embodied in the core message: "*I will keep my eyes open*".

The campaign is promoted in 10 focus countries (Belgium, Bulgaria, the Czech Republic, Croatia, Hungary, Estonia, Latvia, Poland, Romania, and Slovakia), and targets all citizens broadly, with a particular focus on young people aged 18-30. It addresses a wide spectrum of crimes – ranging from gender-based violence to child sexual abuse, cybercrime, and racist and xenophobic hate crimes – and aims to empower the target audiences to: increase awareness of victims' rights, support victims of violence, encourage crime reporting, and recognize signs of violence.

Keep Your Eyes Open is structured around a main campaign video and three thematic ones available in several languages, for a total of 45 audiovisual content items. The campaign's contents are accessible through a dedicated website⁷, which serves as a central hub available in 24 EU languages. The website includes an interactive digital tool enabling access to relevant resources and providing direct support: by selecting one of the EU's 27 countries and a type of crime, users can get tailored information about legal rights in the selected national context, as well as aggregated contacts of support organisations, NGOs, and government departments to ask for help. Moreover, the website offers a downloadable stakeholder kit⁸ and an open-access children's book⁹ targeted for 7-10-year-olds that introduces young readers, parents, and teachers to the experiences and rights of Ukrainian refugee children, highlighting awareness, empathy, and support for victims of war-related crimes.

Despite its more structured digital presence, with an institutional webpage and content organized around thematic areas, the campaign does not rely on dedicated social media accounts – resulting in the issue of limiting peer-to-peer engagement among younger audiences – and its dissemination strategy is shaped mainly through the DG EU Justice and Consumers' YouTube channel via a dedicated playlist¹⁰.

From an analytical perspective, the campaign combines a cognitive goal that aims at informing the target groups about rights and the support services available, with a behavioural and attitudinal one that encourages people to report crimes, offer active support to victims, and recognize various forms of violence. *Keep Your Eyes Open* predominantly

lays an emotional tone elaborated through a dramatic and poignant language that avoids pathetic or aggressive language. The campaign shows a good level of intersectionality, by paying particular attention to aspects such as ethnicity, age, and sexual orientation. Violence and crimes are also approached and presented as a composite and multifaceted phenomenon, including: war-related crimes, violence against a person, gender-based violence, xenophobia, LGBTQI-phobia, and violence against children (Figure 2). The campaign's language, which combines a dramatic tone with a sentimental and moving style, appears to effectively balance the strengths of these social communication registers (Gadotti & Bernocchi, 2010) while avoiding some of their adjacent stylistic pitfalls, such as accusatory or pathetic tones. Considering this potential limitations and risks may lie in a reduced incisiveness and memorability of the message the campaign seeks to convey. Additionally, *Keep Your Eyes Open*, in its attempt to develop a multifaceted and multidimensional perspective on violence, may risk not being fully accessible to its intended audiences, given differences in sensitivity, age, levels of awareness, and educational background.

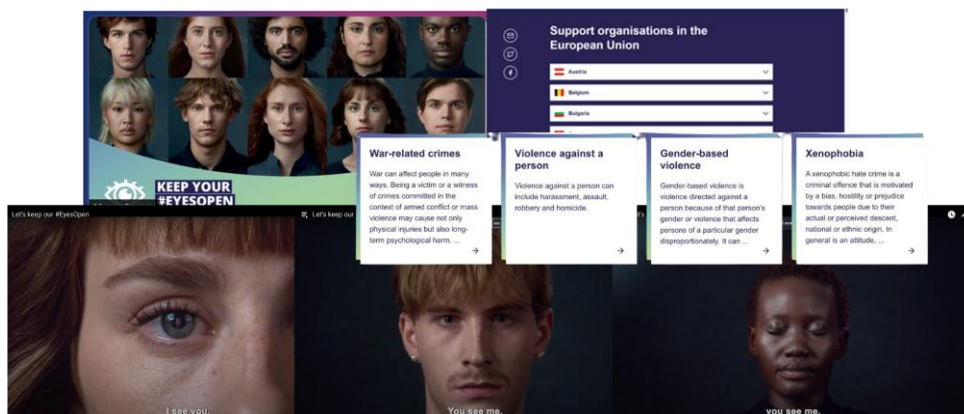


Figure 2 - *Keep Your Eyes Open's* campaign visuals

#EndGenderStereotypes (2023)

The *#EndGenderStereotypes* campaign, promoted by the European Commission and coordinated by the Directorate-General for Justice and Consumers, is structured as an institutional communication initiative to challenge the persistence of gender stereotypes in daily life. Conceived as part of the broader effort to promote gender equality across all areas of social and professional life, the campaign adopts a general awareness-raising approach, grounded in the idea – clearly stated in its official self-description – that gender stereotypes influence people's everyday decisions, behaviours, and aspirations, often unconsciously¹¹.

The campaign's core message is in the call to action "*It is time to #EndGenderStereotypes!*", which reinforces its main goal: to encourage reflection and promote individual and collective responsibility in recognizing and dismantling culturally

internalised gender norms. While the campaign positions itself as addressing a wide range of gender stereotypes, the actual structure of its communication strategy is articulated around three specific thematic areas: *career choices*, addressing early expectations around education and work; *work-life balance*, promoting shared responsibilities and care equity; *decision-making*, confronting gender gaps in leadership and representation.

Each theme is developed on a dedicated sub-page within the campaign's web platform¹², following a coherent communicative structure. Every section includes the main campaign video, visual graphics, explanatory texts, statistical references based on EU data sources (e.g., Eurobarometer, Eurostat, EIGE), and direct links to relevant EU policies and funded initiatives.

The campaign speaks directly to diverse target groups, including young people, families, employees, and employers, while paying particular attention to women and girls. The initiative's structure combines general messaging with thematic deep dives, tailored to audience-specific concerns (e.g., hiring bias, gender gaps in STEM, or shared caregiving).

Although the campaign uses hashtags and audiovisual content, it lacks a dedicated digital dissemination strategy. The absence of official social media channels specific to the campaign, combined with its integration into broader institutional profiles (Facebook, X, YouTube of the EU Justice and Consumers), reflects an approach more aligned with maintaining institutional coherence than with fostering campaign-specific visibility or engagement. This is particularly counter-intuitive given the campaign's branding through a hashtag, which typically implies a digitally driven and participatory communication strategy, which in this case is notably absent.

From an analytical perspective, also on potential risks evaluation, *#EndGenderStereotypes* presents itself as a value-based campaign that explicitly aims to challenge normative gender assumptions and promote a more inclusive and equitable society. Its communication strategy is designed to question common biases in a way that is both accessible and directive, relying on straightforward content but morally assertive messaging. The primary narrative strategy revolves around the reversal of gender roles, presented through short video scenes and graphic compositions that invite the audience to reflect on expectations commonly associated with gender. These scenes are typically accompanied by a single adjective followed by a question mark – such as “*Uncommon?*” or “*Surprising?*” – that prompt cognitive dissonance, exposing how ingrained and often unconscious such assumptions are. This approach effectively produces awareness, but its insistence on contrast and inversion lends the campaign a strongly didactic tone, which may limit its openness to diverse interpretations.

In terms of tone, the campaign adopts a predominantly rational register, favouring informative and empowering discourse. However, this is accompanied by specific paternalistic and prescriptive language: the audience is encouraged to reflect and implicitly guided toward “correct” conclusions. Rather than offering ambiguity or space for negotiation, the messaging appears to predefine the viewer's path, suggesting what they should question and how to respond. Regarding intersectionality, the campaign considers some elements of diversity – notably ethnicity and age. Nevertheless, the overall framing remains limited to a male-female binary, with no visible inclusion of non-binary identities or diverse

sexual orientations. As such, while the campaign addresses diversity to some extent, it stops short of fully embracing the complexity of contemporary gender experiences.



Figure 3 - End Gender Stereotypes campaign visuals

The gender performativity expressed in the campaign is presented as equal across male and female figures, with men and women portrayed in both traditional and non-traditional roles. However, the emphasis on balance and inversion reinforces a binary logic, which risks flattening more fluid or underrepresented gender positions. Furthermore, the campaign's protagonists are always placed in heroic, agentic positions (e.g., professional women at their best performance, caring men), reinforcing aspirational models that may reflect the ethos of a society grounded on excellence and performance.

The campaign's cognitive and behavioural goals aim to raise awareness and promote attitudinal change. The repetition of the same main video across all three thematic areas, combined with tailored visuals and explanatory content, reinforces a consistent message and narrative structure. Nevertheless, the lack of interactive dissemination strategies increases the campaign's risk of failing to adapt to diverse audiences and real-life contexts.

Spot the Violence Stop the Violence (2024)

The campaign *Spot the Violence / Stop the Violence*, launched by the European Commission in 2024, is inscribed into the UN initiative *UNiTE by 2030 to End Violence against Women*. Globally known through the hashtag #OrangeTheWorld, the UN initiative has a coordinated visual identity mainly characterized by the colour orange, symbolizing hope for a future free from violence.

Within this global framework, the EU campaign presents itself as a social media communication initiative coordinated by the Directorate-General for Justice and Consumers, in collaboration with the Directorate-General for International Partnerships, for *16 Days of Activism against Gender-Based Violence*. The campaign speaks to digital citizens on social media within the broader scope of connecting the European target to the global ones. For instance, it has been reshared by various official EU international accounts such as EU in Mexico, the UK, Liberia, South Africa, and by institutional social media related to EU Enlargement and Eastern Neighborhood policies.

Despite its symbolically powerful thematic relevance, from a communication strategy perspective, the campaign suffers from a lack of strategic coherence, visibility, and

dissemination. Although the campaign appears linked to the broader European External Action Service (EEAS) platform “*Break the Silence. Stop the Violence*”, there is no explicit institutional affiliation made visible in the campaign materials, aside from using a common hashtag. Moreover, unlike other analyzed campaigns, *Spot the Violence / Stop the Violence* lacks a clear self-description on the EU’s institutional websites and appears to be disseminated fragmentarily on digital platforms. On the one hand, it has no official website, dedicated landing page, or social media accounts, and content dissemination relies on sporadic posts with no clear storytelling structure or coordinated editorial planning, risking the dispersion of its impact due to a structurally incoherent multilevel communication initiative. On the other hand, it has a collection of visuals with key statistics and infographics that can be downloaded from the European Commission gender-based violence page, to be reshared via social media.

From an analytical perspective, the campaign’s core message, “*Spot the Violence / Stop the Violence*”, plays a central narrative and symbolic role. The verb “to spot” suggests an active process of seeing, noticing, detecting, and recognising violent actions and invites the audience to become more vigilant toward signs of violence that may go unnoticed. This approach implicitly shifts responsibility onto the viewer, reframing violence not only as a social problem, but as a cultural and perceptual one.

The campaign comprises short videos shared via Facebook and Instagram of the Justice and Consumers DG’s official accounts, developed through illustrated animations and real-life videos. Illustrated-animated content focus on two specific areas of life where violence occurs: public and private spaces. With a monocular moving around a flat, apartment by apartment, and on a bus, person by person, the video “spot the violence” among various scenes of everyday life: a domestic scene of violence or harassment on public transport. Real-life setting videos propose two other specific focuses spanning survivors’ interviews to a vox populi around Brussels’ tourist spot, where an interviewer tests people’s awareness regarding the International Hand Sign for Help.

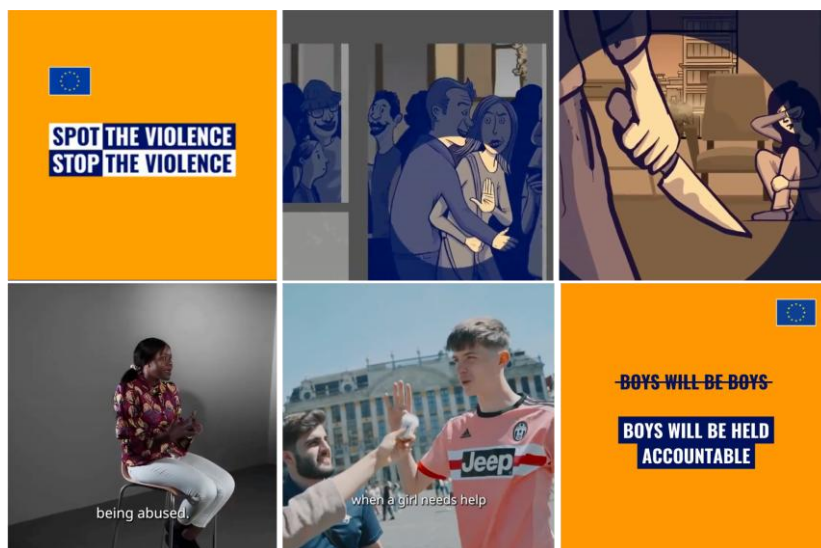


Figure 4 - *Spot the Violence Stop the Violence's* campaign visuals

In terms of tone, the campaign is rational, mainly using a documentary and informative language, to raise awareness on everyday situations where gender-based violence could occur, remaining unnoticed. The language also includes a shade of empowerment, providing both actionable strategies and cultural tools to increase public sensitivity to the diverse manifestations of violence. *Spot the Violence / Stop the Violence* invites reflection, encouraging viewers to question what they see and fail to see. The campaign has a clear cognitive goal, and its storytelling can effectively trigger cognitive dissonance. On a deeper level, it pursues a behavioural goal: from a knowledge check on daily-life gender-based violence, aiming to educate citizens to recognize all the shades of violence, and to train practical strategies, as the International Help Signal, to detect situations where a woman is reporting violence.

In terms of gender representation, the campaign simultaneously challenges and reproduces stereotypes. For instance, in one vignette, a man is vacuuming while a woman is reading on a sofa, while in another, a woman cooks and cares for children while the man reads the newspaper. Considering the evaluation of potential risks, the campaign maintains a strategically balanced, though ambivalent position seeking to reconcile divergent expectations while avoiding explicit alignment or fully committing to a clear ideological stance. It develops a reasonably good intersectional approach, depicting different ages, ethnicities, and religious backgrounds. Nevertheless, it is relevant to emphasise that gender-based violence is furthermore presented as an exclusively binary issue, with women as victims and men as perpetrators. Non-binary identities, LGBTQ+ issues, and less visible forms of structural violence are not addressed.

Discussion and Conclusion: open challenges and emerging risks

The analysis of the four selected campaigns shows an EU commitment to advancing gender equality and inclusive values, also from a strategic communication perspective, along with the strong effort in terms of policy-making. Nevertheless, the study reveals recurring challenges that could hinder the full communicative potential of these efforts. These difficulties arise both in the digital dissemination practices (RQ1) and in the framing criteria of gender-related themes (RQ2).

Weaving back into the theoretical background, answering RQ1, the detected risks for a gender-sensitive strategic public communication are structural-organizational, socio-cultural, and political in nature, unveiling broader tensions within the EU's institutional and communication landscape. Among the structural-organizational risks identified across the four EU campaigns under investigation in this study, the first major hurdle concerns fragmented and weakly coordinated digital dissemination. All campaigns lack dedicated social media channels, and *Spot the Violence* has no official website, nor any clear reference on the EU institutions' official websites. Several campaigns (*CharactHer*; *End Gender Stereotypes*; *Spot the Violence*) do not rely on coordinated cross-platform strategies or clear

editorial planning covering the entire communication process from beginning to end. Instead, they display a prevailing one-way communication approach, without implementing an accurate and structured relational communication strategy to engage campaign targets such as publics or stakeholders. EU gender communication campaigns thus depict a digital communication mosaic in which some initiatives have websites and structured content but lack social media strategies, while others are disseminated mainly on a single social media platform. Consequently, this patchworked and fragmented management of public communication on gender-related issues risks reducing efficacy and visibility, limiting peer-to-peer engagement – especially among younger audiences or groups with different educational backgrounds – and weakening the campaigns' adaptability to diverse socio-cultural contexts.

As a second point, answering RQ2, regarding the tones and languages, several stylistic constraints emerge. *CharactHer*, for instance, with its rational tone and informative, descriptive, and documentary style, risks emotionally disengaging the target audience. By merely stating its core message rather than persuading, the campaign favors messages that are more acceptable and neutral, therefore having limited impact, as highlighted in the literature on social communication (Gadotti & Bernocchi, 2010). Alongside these considerations, other campaigns – such as *End Gender Stereotypes* – rely on prescriptive or didactic registers that predetermine correct conclusions, leaving little space for audience negotiation. From a stylistic perspective, on the other hand, a campaign like *Keep Your Eyes Open* that goes to the opposite extreme – built on strong emotional appeal and dramatic tones aimed at eliciting compassion – may lack memorability or accessibility for certain target audiences, due to differences in age, sensitivity, or educational background.

Focusing on socio-cultural risks, the study highlights intersectionality gaps, partial representation, and a tendency toward ambivalence that could reinforce gender stereotypes. Despite inclusivity goals, the EU's communication campaigns largely depict white, cisgender women and binary gender roles. By excluding trans/non-binary identities, LGBTQ+ perspectives, and other ethnic minorities, they risk reproducing normative frames of gender equality, neglecting structural inequalities and silencing marginalized identities: failing to respect the diverse sensibilities of European contemporary societies.

Efforts to challenge gender stereotypes in the four campaigns often coexist with aspirational, performance-oriented portrayals that frame gender as a social issue, focusing on aspects such as work-life balance, career, and violence, rather than also addressing the identitarian dimension of gender equality. This tendency, together with the above-mentioned risk of flattening a wide range of complex gender experiences and inequalities (*End Gender Stereotypes*; *Spot the Violence*), could result in diluting the political message, making campaigns less disruptive toward dominant social norms, reproducing existing stereotypes, or even reinforcing stigmatization. Beyond these dynamics, further concerns arise in relation to new technologies and gender bias. This issue is increasingly under scrutiny in connection with the use of generative Artificial Intelligence, whose potential to replicate and amplify gender bias adds a new layer of complexity to the future of the public communication landscape.

Ultimately, both socio-cultural limitations identified – narrow intersectionality and partial representation on the one hand, and ambivalence with the risk of stereotype reinforcement on the other – can be interpreted as deeply intertwined with the EU's political sensitivities and ideological tensions. While the European Union promotes gender equality as a core value, the tension between normative commitments and policies, and heterogeneous political and ideological stances, exposes a gap between policy ambition and communication practice. Campaigns on gender equality are thus shaped within a highly ideologized political field, where the 27 Member States maintain diverging positions and sensitivities on gender issues, especially regarding identitarian versus social dimensions.

The need to balance institutional commitments with the risk of political backlash seems to result in cautious framing, diluted narratives, and the avoidance of intersectional 'contentious' topics, particularly those related to gender identities, sexual orientations, and migration backgrounds. Such dynamics ultimately seem to weaken the transformative potential of these campaigns, highlighting the contested position that the European Union is compelled to sustain in promoting gender equality within an ideologically and politically conflictual landscape.

Gender issues have become increasingly politicized and polarized in recent years, and this is the primary reason why further investigation is needed into how public communicators themselves perceive and navigate these constraints. Future research could therefore benefit from in-depth interviews with communication officers of the Directorate-General in charge of developing these campaigns, to better understand how political sensitivities, institutional mandates, and social expectations intersect in shaping the communicative framing of gender equality. This second phase of the project will also allow us to develop operational recommendations for institutional communicators, moving beyond symbolic inclusion and toward a genuinely transformative project for gender equality in EU public communication.

It is considered desirable to foster the growth of both theoretical and empirical reflections and studies, as well as to promote innovation and the strengthening of the training of professionals. The focus should be on how public communication can strategically play a role of agency within society, in terms of inclusion and respect for gender differences, in the changing media ecosystem.

Such an evolution should be based not so much on adapting to the algorithmic logic of digital platforms and, more generally, to the digital dynamics that characterize contemporary public spheres, but rather on the appropriation and in-depth understanding of the features of new digital environments (and their programming criteria). This would allow them to be employed in fostering a form of depoliticization and depolarization of gender-related issues, promoting exchange rather than conflict and separation between different cultures (in line with intercultural dialogue among the many European countries), thereby reducing the risk of disinformation and toxic dynamics (e.g., incivility), while adopting an ethical perspective (Bowen, 2024).

In summary, the EU's strategic gender-sensitive public communication must actively adopt and pursue ethical and democratic criteria to effectively exert its transformative *agency* within society. It is imperative that these criteria are grounded in intersectional framing to fully encompass all gender identities, thereby preventing partial representation

and mitigating the risk of reinforcing stereotypes or marginalizing complex identities. Crucially, the implementation of this approach must be contextualized and adapted to the specific socio-cultural and political characteristics of each EU Member State or region.

Furthermore, this core ethical commitment must be paired with the imperative to reduce the fragmentation in campaign management. This requires strengthening unified coordination and implementing two-way (bidirectional) communication strategies. Practically, this entails maintaining dedicated channels and coherent editorial planning, alongside developing structured listening mechanisms to facilitate active public engagement. From a stylistic perspective, it is necessary to achieve a balance that moves beyond overly rational or prescriptive tones, instead favouring emotional impact to enhance campaign engagement and memorability.

Finally, a critical appropriation of advanced technologies, particularly Artificial Intelligence, is fundamental. This is key to leveraging digital dynamics to depolarize gender issues, promote intercultural exchange, and contribute to the reduction of inequalities. Indeed, the primary aim of public communication is not to generate inequalities within the population, but to contribute to their reduction, in the general interest (Zémor 1995).

Biographical Notes

Gea Ducci is Associate Professor of Sociology of Culture and Communication processes and Deputy Head of the Department of Communication Sciences, Humanities and International Studies at the University of Urbino Carlo Bo (Italy). She serves as Director of Advanced Management Training for Healthcare Executives, a member of the PhD board in Humanistic Studies and of the Research Lab on Artificial Communication at the same university. She is a Board Member of the Italian Sociology Association for the 2023-2025 term. Her research focuses on public sector communication, social and gender issues, inclusion and sustainability public communication, health communication, digital media, and the use of AI in institutional communication.

Lucia D'Ambrosi is an Associate Professor of the Sociology of cultural and communication processes in the Department of Communication and Social Research (CoRiS) at Sapienza University in Rome. She teaches Public and Institutional Communication. Her research focuses on public communication in Italy and Europe, particularly with regard to gender issues and social inclusion. She is principal investigator of the project DeiCom Diversity, equity & inclusion in public communication. Empowering professionals for audience-centred approaches.

Paola De Rosa is an Assistant Professor in Sociology of Cultural and Communication Processes at Link Campus University of Rome, where she teaches Crisis Communication and Corporate & Institutional Communication. Her research focuses on crisis communication, the analysis of media representations and their effects on audiences, with a focus on gender issues and female empowerment. She's the co-coordinator of the GEMINI Observatory on Gender Equality and Diversity in Audiovisual Media.

Camilla Folena is a Post-Doctoral Research Fellow in Sociology of Cultural and Communication Processes at the Department of Communication, Humanities and

International Studies (DISCUI), University of Urbino Carlo Bo. She has taught Communication and Marketing, Sociology of Cultural Processes, and currently teaches Multimedial Communication while collaborating in Strategic Institutional Communication and Public and Social Communication courses. Her research focuses on journalism and gender studies, post-Western European perspectives, digital cultures, and political and institutional communication.

Marica Spalletta is an Associate Professor in Sociology of Cultural and Communication Processes at the Link Campus University of Rome, where she teaches Media Sociology and Media and Politics. She is also the scientific manager of LinkLAB (Social Research Centre). Her research focuses on the media representation of social issues and the audience perspective. Over the last 3 years, she has served as principal investigator of the European research-action project GEMINI-Gender Equality through Media Investigation and New training Insights.

References

- Altheide, D. L., & Schneider, C. J. (2013). *Qualitative Media Analysis*. Sage.
- Basaglia, A., Cuomo, M., & Simonella, M. (2022). *L'organizzazione inclusiva*. FrancoAngeli.
- Bentivegna, S., & Boccia Artieri, G. (2021). *Voci della democrazia: Informazione, partecipazione e nuovi media*. Laterza.
- Bentivegna, S., & Rega, R. (2024). *Un(Civil) democracy: Political incivility as a communication strategy*. Palgrave Macmillan, SpringerNature.
- Boccia Artieri, G., Bruns, A., Dehghan, E., & Iannelli, L. (2025). Fringe Democracy and the Platformization of the Public Sphere, *Comunicazione politica*, 26 (1), 3-21. <https://doi.org/10.3270/116607>
- Bowen, S. (2024). The Normative Imperative: Ethics, In E. Oliveira & G. Gonçalves (Eds.), *Ethics and society: Challenges in organizational and public communication*, 35-46. LabCom Books, Universidade da Beira Interior.
- Buonanno, M., & Faccioli, F. (2020). *Genere e media: Non solo immagini*. FrancoAngeli.
- Campus, D. (2013). *Women political leaders and the media*. Palgrave Macmillan.
- Canel, M. J., & Luoma-aho, V. (2019). *Public sector communication*. Wiley-Blackwell.
- Chadwick, A. (2006). *Internet politics: States, citizens, and new communication technologies*. Oxford University Press.
- Crenshaw, K. (1991). Mapping the margins: Intersectionality, identity politics, and violence against women of color. *Stanford Law Review*, 43(6), 1241–1299. <https://doi.org/10.2307/1229039>
- D'Ambrosi, L., Ducci, G., Lovari, A., & Folena, C. (2023). Questioni di genere: Il ruolo dei comunicatori pubblici tra innovazioni tecnologiche, strategie discorsive e resistenze al cambiamento. *Problemi dell'informazione*, 47(2), 225–250. <https://doi.org/11576/2726394>

- D'Ambrosi, L., Ducci, G., Folena, C., & Spalletta, M. (2024). Comunicazione istituzionale e prospettive di genere. Uno studio sul social posting delle Regioni italiane, fra regole e pratiche. *Mediascapes Journal*, 23(1), 3–26.
- De Blasio, E., & Selva, D. (2024). Politicisation of gender. In E. De Blasio & D. Selva (Eds.), *Gender and culture wars in Italy* (pp. 45–67). Palgrave Macmillan. https://doi.org/10.1007/978-3-031-60110-1_3
- Di Meco, L. (2019). *#ShePersisted: Women, politics & power in the new media world*. The Wilson Center.
- Ducci, G., Lovari, A., & D'Ambrosi, L. (2019). «Fra schermi e schermaglie»: Le sfide del visual nello storytelling istituzionale. *H-ermes. Journal of Communication*, 15, 313–352. <https://doi.org/10.1285/i22840753n15p313>
- Ducci, G., Lovari, A. (2021). The challenges of public sector communication in the face of the pandemic crisis. *Sociologia della comunicazione*, 61, 9-19. <https://doi.org/10.3280/SC2021-061002>
- EIGE – European Institute for Gender Equality (2016). *Gender equality training. Gender Mainstreaming Toolkit*. Available at: <https://eige.europa.eu/sites/default/files/documents/mh0716093enn.pdf>
- EIGE – European Institute for Gender Equality (2019). *Gender-sensitive communication*. Available at: <https://eige.europa.eu/publications/gender-sensitive-communication>
- Eurobarometer (2024). *Gender stereotypes*. Brussels: European Commission. Available at: <https://europa.eu/eurobarometer/surveys/detail/2974>
- European Commission (2013). *GREEN PAPER Preparing for a Fully Converged Audiovisual World: Growth, Creation and Values*. Available at: <https://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=COM:2013:0231:FIN:EN:pdf>
- European Commission (2020). *Gender equality strategy 2020–2025*. Luxembourg: Publications Office of the European Union. Available at: <https://op.europa.eu/en/publication-detail/-/publication/16d3f3f4-59b3-11ea-8b81-01aa75ed71a1>
- European Parliament (2019). *Inclusive communication guidelines*. Available at: <https://www.europarl.europa.eu/contracts-and-grants/files/grants/media-and-events/en-annex-9-inclusive-communication-guidelines-of-the-european-parliament.pdf>
- Oliveira, E., & Gonçalves, G. (Eds.). (2022). *Ethics and society: Challenges in organizational and public communication*. Routledge.
- Faccioli, F., & D'Ambrosi, L. (2023). Gender in COVID-19 campaigns: An analysis of the Italian government communications. *Journal of Italian Cinema & Media Studies*, 11(1), 123–145. https://doi.org/10.1386/jicms_00180_1
- Farci, M., & Scarcelli, C. M. (2022). *Media digitali, genere e sessualità*. Mondadori Università.
- Gadotti, G., & Bernocchi, R. (2010). *La pubblicità sociale*. Carocci.
- Ghigi, R., & Rottenberg, C. (2019). *Feminisms in neoliberal times: From neoconservative agendas to global protests*. Palgrave Macmillan.

- Gerbner, G., & Gross, L. (1976). Living With Television: The Violence Profile. *The Journal of communication*, 26, 173–99. <http://dx.doi.org/10.1111/j.1460-2466.1976.tb01397.x>
- Grunig, J. E. (2008). Excellence theory in public relations. In W. Donsbach (Ed.), *The international encyclopedia of communication*. Wiley-Blackwell. <https://doi.org/10.1002/9781405186407.wbiece047>
- Jenkins H., Ford S. and Green J. (2013). *Spreadable media*. New York University Press.
- Kim, A.T., Erickson, M., Zhang, Y., & Kim, C. (2022). Who is the ‘She’ in the Pandemic ‘She-Cession’? Variation in COVID-19 Labor Market Outcomes by Gender and Family Status. *Population Research and Policy Review*, 41(3): 1325–1358. <https://doi.org/10.1007/s11113-021-09693-0>
- Kim, S., Buzzanell, P. M., Mazzei, A., & Kim, J.-N. (Eds.). (2025). *The Routledge handbook of employee communication and organizational processes*. Routledge.
- Kotler, P., Roberto, N., & Lee, N. (2002). *Social marketing: Improving the quality of life*. SAGE.
- Lee N. R., Kotler P. & Colehour J. (2023). *Social Marketing: Behavior Change for Good*. SAGE.
- Lovari, A., & Ducci, G. (2022). *Comunicazione pubblica: Istituzioni, pratiche, piattaforme*. Mondadori Università.
- Luoma-aho, V., & Badham, R. (2023). *Handbook on digital corporate communication*. Edward Elgar Publishing.
- Luoma-aho, V., & Canel, M. J. (2020). *The Handbook of Public Sector Communication*, Wiley Blackwell.
- Macnamara, J. (2005). Media Content Analysis: Its Uses, Benefits, and Best Practice Methodology. *Asia Pacific Public Relations Journal*, 6(1), 1–34.
- Martins, P. (2024). *Placing “gender” in disinformation*. Association for Progressive Communications. Available at: <https://www.apc.org/sites/default/files/genderDisinformation.pdf>
- Mergel, I., & Bretschneider, S. I. (2013). A three-stage adoption process for social media use in government. *Public Administration Review*, 73(3), 390–400. <https://doi.org/10.1111/puar.12021>
- McRobbie, A. (2009). *The aftermath of feminism: Gender, culture and social change*. SAGE.
- OECD – Organisation for Economic Co-operation and Development (2021). *OECD report on public communication*. OECD Publishing.
- OECD – Organisation for Economic Co-operation and Development (2023). *Accessible and inclusive public communication*. Available at: https://www.oecd.org/en/publications/accessible-and-inclusive-public-communication_222b62d9-en.html
- Oliveira, E. & Gonçalves, G. (eds.) (2024), *Ethics and Society: Challenges in Organizational and Public Communication*, Covilhã: LabCom Books, Universidade da Beira Interior.
- Ravazzani, S., Mazzei, A., Butera, A., & Fisichella, C. (2023). *Diversity, Equity & Inclusion*. FrancoAngeli.

- Rocha Menocal, A. (2021). The role of identity in shaping prospects for more inclusive politics. *KIPA Public Policy Review*, 2(2), 9–35.
- Ross, K., & Padovani, C. (2017). *Gender equality and the media: A challenge for Europe*. Routledge.
- Saccà, F., & Massidda, L. (2021). *Culture politiche e leadership nell'era digitale*. FrancoAngeli.
- Sebastião, S. P., & Cotton, A.-M. (2025). *Global public goods communication: Mapping actors, policies, and narratives*. Springer.
- Spalletta, M., Ducci, G., D'Ambrosi, L., & Folena, C. (2024). Public sector communication and gender perspectives: Social media practices of Italian regions. In E. Oliveira & G. Gonçalves (Eds.), *Ethics and society: Challenges in organizational and public communication*, 191–233. LabCom Books, Universidade da Beira Interior.
- Tipaldo, G. (2014). *L'analisi del contenuto e i mass media*. il Mulino.
- UNESCO – United Nations Educational, Scientific and Cultural Organization (2021). *The chilling: Global trends in online violence against women journalists*. Paris: UNESCO Publishing. Available at: <https://unesdoc.unesco.org/ark:/48223/pf0000377223>
- Van Dijck, J., Poell, T., & De Waal, M. (2018). *Platform Society. Public Values in a Connected World*. Oxford University Press Inc.
- Zémor P. (1995), *La communication publique*. P.U.F., Que sais-je?

Notes

¹ Specifically, within the framework of the *Creative Europe MEDIA programme* (2021-27), the Commission has launched initiatives to promote greater gender balance and diversity in the media and audiovisual sector (Action Plan COM (2020) 784 final). These initiatives include funding opportunities for women filmmakers, producers, and screenwriters, as well as projects promoting inclusive storytelling and alternative gender representations.

² With the expression 'performativity', the authors refer to the performative construction of a precise stereotype through a narrative and visual emphasis on individual effort, merit, and hyper-excellence. These portrayals stage gendered roles as exceptional - e.g., the 'strong man in command', the 'self-made leader woman', the 'heroic career woman' - thus reproducing an empowerment-based, meritocratic logic rather than addressing structural gender inequalities.

³ «The *CharactHer* campaign aims at debunking these stereotypes: there is no such thing as a "masculine" job – women can do anything! The film and media industries offer a wide range of professional opportunities that require a variety of different competences. If you're ambitious, determined and passionate, there may be a job for you, no matter your gender or origin!». In <https://character.eu/>.

⁴ All interviews revolve around three key questions, which serve as the starting point for the reflections shared by the women professionals featured in the campaign: *What advice would you give to someone starting out in your field? Who were your heroes or models when you were younger? Name the film that has made the biggest impression on you (you can name several if you want)*.

⁵ In <https://character.eu/>.

⁶ One of the interviews features an autistic woman, and the selected profiles reflect different generations of audiovisual professionals: <https://character.eu/portraits/>

⁷ Keep your Eyes' Open official website: <https://victims-rights.campaign.europa.eu/en/country/netherlands>

⁸ The stakeholder kit is downloadable at: <https://victims-rights.campaign.europa.eu/en/resources>

⁹ The childer book is available at: [https://icfnextcloudshare.blob.core.windows.net/victims-rights/digital-book/digital book/DG JUST children-book EN.pdf](https://icfnextcloudshare.blob.core.windows.net/victims-rights/digital-book/digital%20book/DG_JUST_children-book_EN.pdf)

¹⁰ DG EU Justice and Consumer' official YouTube channel Keep Your Eyes Open' Playlist: <https://youtube.com/playlist?list=PLK3yPgVWaaEqK3uhKJUrIPefX-QZE5ZAk&si=MNvxu24xy2lgu6bo>

¹¹ The opening line of the website hosting the campaign provocatively asks: *"What gender instantly comes to mind when you think of a certain toy, a job, a sport, or even a colour?"*.

¹² There is no official standalone website for the campaign; instead, the initiative is hosted within the European Commission's institutional web infrastructure. The dedicated pages are managed by the Directorate-General for Justice and Consumers and available at:

https://end-gender-stereotypes.campaign.europa.eu/index_en

Cooperation rhetoric as risk-communication: An analysis of Biden Administration's China policy rhetoric through the lens of the Mixed-motive Model*

Juyan Zhang**

The University of Texas at San Antonio

This research applies the Mixed-motive Model of Public Relations to analyze the Biden Administration's rhetoric on China policy, focusing on crisis and risk communication. It explores how the administration rhetorically shifted from Trump's pure advocacy approach, known as "decoupling," to a nuanced mixed-motive strategy combining competition and cooperation. This shift redefines the US-China relationship, articulates the US vision, expresses underlying worldviews, and manages inherent tensions in policy rhetoric to reduce the risk of direct conflict. Using Burke's cluster-agon rhetorical criticism method, the study identifies key God terms, such as "rule-based international order," "shared vision of future," "diplomacy," "network of allies and partners," "Chinese people," "US strengths," and "comprehensive strategy/compete", and Devil terms, including "challenges," "conflict," and "China." The analysis reveals that Secretary Blinken's speech serves as strategic communication reflecting a blend of liberalism, realism, and constructivism, which justifies the use of mixed-motive strategies and mitigates tensions by aligning each strategy with a specific worldview. This complex rhetorical approach is shaped by the multifaceted nature of US-China relations and global dynamics. Importantly, the cooperative rhetoric functions as risk communication designed to prevent the US-China competition from escalating into armed conflict.

Keywords: Biden, China, mixed-motives, risk communication, strategic communication

* Article submitted on 20/09/2025. Article accepted on 15/12/2025.

** juyan.zhang@utsa.edu

Introduction

The power of discourse has become an important aspect of security analysis (Balzacq, 2005), and foreign policy is often legitimated through the mobilization of rhetoric (Goddard & Krebs, 2015). Rhetorical contestation shapes the boundaries of acceptable policy and the very possibilities of state action (Krebs & Jackson, 2007). U.S. foreign policy toward China is no exception. Recent research indicates that U.S. national discourse has increasingly constructed China as a potential enemy Other (Ooi & D’Arcangelis, 2017), and the mobilization of rhetorical resources from adjacent policy debates has become key to U.S. elites’ representation of China as an ideological threat (Winkler & Jerdén, 2023). Since Donald Trump became U.S. president in 2018, the U.S.-China relationship has entered a steady downward spiral marked by trade disputes, technological rivalry, geopolitical maneuvering, and human rights concerns. Bilateral relations between the two major powers have deteriorated to such an extent that both sides have begun discussing how to avoid an armed conflict (CSIS, 2024). In terms of rhetoric, Trump employed pure advocacy, exemplified by his call to “decouple” from China (Reuters, 2020). However, the subsequent Biden administration has adopted a different policy approach, and accordingly, the U.S. government’s rhetoric on China has shifted as well, moving from pure advocacy to a more nuanced and mixed representation that emphasizes both competition and cooperation (The Economist, 2020).

In this research, using rhetorical criticism and approaching the topic from the perspective of strategic communication, I examine the Biden administration’s rhetoric on its China policy. Specifically, I explore how the administration has rhetorically shifted from Trump’s approach of pure advocacy to redefine U.S.-China relations as a balance of competition and cooperation. I also analyze how the administration addresses the inherent tension within a policy mix that includes cooperation, contest, competition, confrontation, and the avoidance of conflict. This analysis aims to uncover the underlying motives behind the Biden administration’s China policy rhetoric: Is it merely a façade for a policy of pure advocacy, or does it reflect a genuine shift in U.S. strategic intent? The theoretical framework guiding this research is the Mixed-Motive Model of public relations, and the rhetorical criticism method employed is Burke’s (1969) cluster analysis. This study contributes to the strategic communication literature by testing one of its most influential theoretical models, the Excellence Theory’s Mixed-Motive Model, through rhetorical criticism.

Trump and Biden’s China rhetoric

Donald Trump adopted a policy and rhetoric of pure advocacy in U.S.-China relations. His China policy is widely regarded as a paradigmatic shift, centered on the embrace of “strategic competition” (Ashbee & Hurst, 2022), and marked a historic turnaround in U.S. policy toward China (Kubo, 2019). Trump imposed tariffs on Chinese imports, aiming to

curtail China's technological advancement. Geopolitically, he took a more confrontational stance, explicitly framing China as a strategic competitor. While some of Trump's measures sought to address legitimate concerns about China's economic and security practices, they often produced unintended consequences and strained bilateral relations (Black & Morrison, 2021). The trade war imposed economic costs on both countries, with limited long-term benefits for the United States. Tariffs disrupted global supply chains and increased costs for American consumers and businesses. Meanwhile, efforts at technological decoupling harmed U.S. companies reliant on Chinese technology and cooperation, while simultaneously accelerating China's push for technological self-sufficiency (Schuetz & Chan, 2019).

Biden's approach to China marks a departure from Trump's confrontational stance, emphasizing strategic competition within a multilateral framework (Shi, 2024). Whereas Trump employed fiery and combative rhetoric, Biden has adopted a more measured and diplomatic tone. His China policy was succinctly articulated by Secretary of State Antony Blinken as "competitive when it should be, collaborative when it can be, and adversarial when it must be" (Blinken, 2021). Unlike Trump's preference for unilateral actions, Biden has prioritized rallying traditional allies in Europe and Asia to form a coordinated and united front in response to China's assertive behavior (Johnson, 2022). Although the Biden administration continues to confront China on key issues, such as trade, human rights, and regional security, it does so within a framework of multilateralism and strategic alignment with allies. In terms of trade, while Biden has retained many of the tariffs imposed during the Trump era, his administration has emphasized more targeted and strategic measures aimed at addressing specific concerns, including intellectual property theft and unfair trade practices (Haas, 2024).

Biden's China policy as mixed strategy

When U.S. Secretary of State Antony Blinken met with Chinese President Xi Jinping in August 2024, he reiterated the strategic approach outlined at the beginning of the Biden administration: "President Biden is committed to responsibly managing this consequential relationship to ensure that competition does not veer into conflict or confrontation, and to work together where our interests align" (Hotham, 2024). The U.S. approach, which combines elements of both competition and cooperation, reflects what international relations scholars refer to as "mixed strategy games," where the interests of the actors are neither strictly opposed nor fully aligned (Aitken-Turff & Jackson, 2006; Blidaru, 2018). Similarly, the Mixed-Motive Model of public relations, the guiding theoretical framework of this research, posits that it is possible to be both cooperative and competitive within the same campaign (Grunig, 2001). In the following analysis, I examine the Biden administration's China policy during the first year of his presidency to assess whether the U.S. government has genuinely pursued a mixed strategy, or

whether the rhetoric of cooperation merely serves as a façade for a continuation of pure advocacy. The documentation of U.S. policy actions is sourced from *China Briefing*,¹ a publication that meticulously catalogs developments in U.S.-China relations during the Biden era.

On January 20, 2021, Joe Biden was sworn in as the 46th President of the United States. While signaling that he was in no rush to depart significantly from the Trump administration's China policies, his cabinet made it clear that Biden intended to adopt a multilateral approach to maximize U.S. leverage over China. On February 10, 2021, on the eve of the Lunar New Year, President Biden and Chinese President Xi Jinping held their first phone call. During the conversation, Biden emphasized U.S. concerns over China's economic practices, human rights issues, and Taiwan, while Xi focused on the need for mutual respect, cooperation, and dialogue. In the following months, the Biden administration began to clearly adopt a mixed strategy in both its China policy and accompanying rhetoric, balancing elements of competition and cooperation. On the one hand, the Biden administration inherited key advocacy strategies from the Trump era and, in many cases, expanded them. (See Table 1) On the other hand, it actively sought cooperation with China on climate, economic, and geopolitical issues. (See Table 2)

<i>Month</i>	<i>US sanction policy/actions toward China</i>
February	- Sanctioned 24 officials from Mainland China and Hong Kong.
March	- Took action against Chinese telecommunications firms over national security concerns. - Imposed sanctions (with allies) in response to alleged human rights abuses in Xinjiang.
April	- Blacklisted seven Chinese supercomputing entities.
June	- Expanded Trump-era ban on American investment in certain Chinese firms. - Banned imports of solar panel materials sourced from Xinjiang. - G7 leaders issued joint statement criticizing China over Xinjiang and Hong Kong. - NATO labeled China a global security challenge.
July	- Added 23 Chinese companies to the U.S. economic blacklist.
October	- Revoked China Telecom Americas' authorization to operate.
November	- Blacklisted several Chinese quantum computing firms.
December	- Finalized rules enabling delisting of Chinese companies from U.S. stock exchanges. - Announced diplomatic boycott of the Beijing Winter Olympics. - Banned all imports from Xinjiang. - Sanctioned four senior Chinese officials in Xinjiang. - Placed China's leading AI firm on investment blacklist. - U.S. Treasury blacklisted eight additional Chinese companies. - U.S. Commerce Department added 34 Chinese entities to its entity list. - Sanctioned five more Chinese officials based in Hong Kong.

Table 1 - Biden administration's sanction policies/actions toward China

<i>Month</i>	<i>US cooperative policy/actions toward China</i>
March	- Held first high-level U.S.-China meeting in Alaska. - Both sides issued statements identifying limited cooperation (e.g., climate change, geopolitical issues), while acknowledging disagreements (e.g., Hong Kong, Xinjiang, Tibet, Taiwan).

<i>Month</i>	<i>US cooperative policy/actions toward China</i>
April	- Jointly announced plans to cooperate on tackling the climate crisis.
May	- Held first U.S.-China trade talks since Biden took office.
	- Revoked Trump-era executive orders banning WeChat and TikTok downloads.
June	- Directed Department of Commerce to review apps linked to foreign adversaries (including China).
	- Secretary Blinken and Chinese counterpart Yang Jiechi held phone call discussing expanded cooperation opportunities.
July	- Deputy Secretary Wendy Sherman visited China; discussed U.S. concerns and potential areas of mutual interest.
August	- U.S. granted licenses for American firms to sell chips to Huawei.
October	- Treasury Secretary Yellen and Vice Premier Liu He held virtual meeting emphasizing global economic coordination and macroeconomic policy dialogue.
	- Issued U.S.-China joint declaration on climate action.
November	- Biden and Xi held first virtual summit; Biden reaffirmed "One China" policy while raising human rights concerns.
	- Agreed to relax visa restrictions on journalists from both countries.

Table 2.- Biden administration's cooperative policies/actions toward China

As demonstrated, the Biden administration has indeed adopted a mixed strategy toward China that combines elements of competition and cooperation, although the approach has notably leaned toward competition. This mixed strategy is succinctly captured in Secretary of State Antony Blinken's China policy speech at The George Washington University: "And to the people of China: we'll compete with confidence; we'll cooperate wherever we can; we'll contest where we must. We do not seek conflict" (Blinken, 2021). A close review of the U.S. policies and Blinken's rhetoric raises several important questions with both theoretical and practical implications: How is the U.S. mixed strategy, encompassing "compete," "cooperate," "contest," and "no conflict", rhetorically articulated by the Biden administration? How are the internal contradictions and tensions within such a strategy rhetorically negotiated? What kind of vision and worldview does this rhetoric project? Most critically, does the rhetoric possess ethos, namely credibility and moral appeal, that makes it persuasive to both China and U.S. allies? Answering these questions are important because the rhetoric, magnified through media, more than express policy. Rather, they help reconstruct collective imaginaries about "us," "them," and a common global order in an increasingly chaotic world (Silverstone, 1999).

This research contributes to the literature on strategic communication by applying rhetorical criticism to one of its most influential models: the Mixed-Motive Model of public relations. James Grunig argues that organizations and publics can each retain a strong sense of self-interest while being motivated to cooperate in limited ways to resolve conflict (Dozier et al., 1995). Grunig adopted the Mixed-Motive Model originally developed by Murphy (1991), asserting that it best represents the ideal of two-way symmetrical communication. However, empirical tests of the Mixed-Motive Model remain limited, and rhetorical examinations of this model, particularly in the context of international strategic communication, are even rarer. This research contends that rhetorical criticism offers a more suitable method for exploring the complexity and nuance embedded within the Mixed-Motive

Model than traditional quantitative approaches. By analyzing the Biden administration's China policy rhetoric, this study aims to show how rhetorical strategies can both reflect and shape the dynamics of mixed-motive public diplomacy.

The Mixed-Motive Model of Public Relations

Signitzer and Coombs (1992) linked public relations models to diplomacy in an international context. They note that the public relations models originally proposed for organizations can map onto different forms of diplomacy depending on communication goals. In this research, I examine the Biden administration's China policy and rhetoric through the lens of the Mixed-Motive Model of public relations. I argue that the Mixed-Motive Model effectively captures the nuance and complexity of the administration's rhetorical approach, particularly when analyzed through rhetorical criticism. According to Grunig, two-way symmetrical communication in situations of conflict represents the most effective and ethical form of communication (Grunig & White, 1992). However, what precisely constitutes two-way symmetrical communication has long been debated. To support his argument, Grunig drew upon the Mixed-Motive Model developed by Murphy (1991), which posits that both organizations and publics can maintain strong self-interest while also being motivated to cooperate, albeit in a limited way, to resolve conflict (Dozier et al., 1995). Grunig et al. (1992) contended that the Mixed-Motive Model best exemplifies the ideal of two-way symmetry, and that it offers a realistic foundation for symmetrical communication, without rendering the symmetrical worldview idealistic or unattainable. Grunig and White (1992) further argued that individuals within organizations often experience "divided loyalties," or mixed motives, that reflect both organizational and societal obligations. From a theoretical perspective, the Excellence Theory maintains that by critically evaluating the underlying worldviews of organizations, scholars and practitioners can determine "which worldview generates public relations programs that best resolve conflict in society, resolve national and international issues, [and] make organizations more effective" (Grunig & White, 1992, p. 37).

Murphy (1991) integrated game theory with the concept of symmetry to develop the Mixed-Motive Model of public relations. This model posits that an organization simultaneously pursues its own objectives while attempting to help others achieve theirs, demonstrating that cooperation and competition can coexist within the same campaign. Plowman et al. (1998) identified nine negotiation approaches within the public relations literature, which exist along a mixed-motive continuum ranging from asymmetric to symmetric. Approaches such as contention, avoidance, and principled negotiation lie closer to the one-way extreme, where the organization wins and the public loses. On the opposite end are accommodation, compromise, and mediated communication, where public interests are prioritized over organizational goals. Positioned in the middle, the win/win zone, are cooperation, unconditionally constructive tactics, and win-win or no-deal strategies. These approaches reflect the stances organizations may adopt in mixed-motive situations. Grunig

(2001) concurs that mixed motives aptly describe symmetry, suggesting that two-way communication can only be realized if its primary goal is to identify win/win solutions between an organization and its stakeholders (Plowman et al., 2001). To incorporate the Mixed-Motive Model into the two-way symmetrical framework, Grunig et al. (2002, p. 15) further define two-way symmetrical communication as a process that attempts to “balance the interests of the organization and its publics.” Symmetrical practitioners, therefore, hold mixed motives. They are loyal both to their employees and to the broader publics of their organizations. Grunig argues that asymmetrical tactics may sometimes be employed strategically to position the organization optimally within the win-win zone. However, because such tactics are constrained by a symmetrical worldview that values the integrity of long-term relationships, the two-way communication model remains fundamentally symmetrical. Grunig also introduced terms such as “collaborative advocacy” and “cooperative antagonism” to describe this balanced model, which “balances self-interests with the interests of others in a give-and-take process that can waver between advocacy and collaboration” (Grunig, 2001). Despite these theoretical developments, most empirical studies testing two-way symmetrical communication rely on quantitative methods, and few have examined the model from a rhetorical perspective. Tests of the Mixed-Motive Model are even rarer, with rhetorical criticism of the model being particularly scarce (Zhang, 2010).

Rhetoric and mixed-motives situation

Grunig argued that developments in rhetorical theory help us better understand and come to terms with mixed motives (Grunig & White, 1992). He noted that Booth’s four types of rhetoric closely parallel the four models of the Excellence Theory. Specifically, sub-rhetoric, which is employed to deceive or obscure issues, corresponds to the press agency model; mere rhetoric, used for sincere persuasion, aligns with the public information model; Rhetoric-B, the art of winning others through sound arguments, parallels the two-way asymmetrical model; and Rhetoric-A, which involves discovering and refining ends and purposes through critical dialogue, corresponds to the two-way symmetrical model (Grunig & White, 1992).

Simonson treats rhetoric as a socially embedded “space” shaped by institutions, media, and cultural rhetorics rather than merely as individual acts of persuasion (Simonson, 2014). The deployment of God-terms and Devil terms, therefore, should be analysed not only for their symbolic valence, but as instruments calibrated to operate within this structured rhetorical ecology. Burke (1966) argued that every work of language contains its own network of symbolic actions, and that its structure and function can be revealed through careful analysis. By examining which subjects cluster around others, we can identify what is significant within a text. Charting these clusters allows us to uncover the rhetor’s worldview (Burke, 1969). For Burke (1969), rhetoric is the use of language as a symbolic means of “inducing cooperation in beings that by nature respond to symbols” (p. 43). He posited that

inducing cooperation is essentially a process of negotiation for identification, whereby people come to identify with one another. At the same time, identification can produce alienation (Burke, 1951). To capture these dynamics, Burke introduced the concepts of “God terms”, words that promote identification and unity, and “Devil terms”, words that generate division and alienation. Both God and Devil terms serve as valuable tools for investigating an individual’s worldview (Burke, 1967).

This study employs Burke’s rhetorical analysis because Burke conceptualizes rhetoric as symbolic action through which political actors construct realities, frame motives, and manage tensions. Risk communication, particularly in high-stakes international contexts, is inherently rhetorical, relying on symbolic framing to reduce uncertainty and prevent escalation. Burke’s cluster-agon method reveals how “what goes with what” organizes key terms into God and Devil clusters, allowing us to see how the Biden Administration structures China as a challenge short of enemyhood, and diplomacy as a mechanism for managing competition. Burke proposed several critical approaches, including cluster analysis, to uncover a rhetor’s motivations and worldview. Cluster analysis involves four key steps: (1) identifying clusters to determine what elements are associated with each other; (2) agon analysis to examine what is opposed to what; (3) analyzing progressive form to trace the movement “from what, through what, to what”; and (4) analyzing transformations to understand what is changed into something or someone else (Koptak, 1990). Rueckert (1994) outlined guidelines for applying Burke’s cluster-agon analysis. The first step is to identify key terms or symbols within the text based on their frequency and intensity. At this stage, a “God term”, an expression around which all others are ranked as subordinate, and its opposite, the “Devil term,” typically emerge. The second step involves identifying and charting clusters surrounding these key terms, examining their combinations and equivalences. The third step, agon analysis, focuses on discerning patterns between God terms and Devil terms, determining the meanings of key terms, and uncovering underlying tensions and conflicts. It is during this phase that the rhetor’s worldview becomes apparent. Finally, the fourth step involves naming the rhetor’s orientation based on the patterns identified within the text. According to Foss (2008), the central question addressed here is: What motive likely underlies the production of this particular rhetoric in this specific way?

The primary text for this research is the speech entitled “The Administration’s Approach to the People’s Republic of China,” delivered by Antony J. Blinken on May 26, 2022, at The George Washington University. This speech provides the most comprehensive and detailed explanation of the Biden administration’s China policy. The cluster criticism analysis reveals several prominent God terms and Devil terms within the speech. Key God terms include “rule-based international order,” “shared vision of the future,” “diplomacy,” “network of allies and partners,” “Chinese people,” “U.S. strengths,” and “comprehensive strategy/compete.” In contrast, the primary Devil terms identified are “challenges,” “conflict,” and “China.” The following sections provide an elaboration and interpretation of these findings.

God Terms

Rule-based international order

The Biden administration's rhetoric centers on the existing "rule-based international order", which is a "system of laws, agreements, principles, and institutions" that "manage relations between states, to prevent conflict, to uphold the rights of all people." The system includes "the UN Charter and the Universal Declaration of Human Rights", which are "reflections of the world's shared aspirations." The system has made world "avoid another world war and armed conflict between nuclear powers," helped build "a global economy that lifted billions of people out of poverty," and "advanced human rights as never before." China's economic growth was made possible "by the stability and opportunity that the international order provides," and "no country on Earth has benefited more from that than China." The system "enshrined concepts like self-determination, sovereignty, the peaceful settlement of disputes" and has protected "all countries from being conquered or coerced." Ukraine "is fighting valiantly to defend its people and its independence." "Instead of weakening the international order," Putin has "brought countries together to defend it." The United States "must defend and reform" the rule-based international order. It "will defend and strengthen the international law, agreements, principles, and institutions that maintain peace and security," and make it possible for "all countries, including the United States and China, "to coexist and cooperate." It wants to "strengthen a system in which as many countries as possible can come together to cooperate effectively, resolve differences peacefully, write their own futures as sovereign equals."

Shared vision of future

"Future" and "shared vision of future" frequently appear in Blinken's speech. The United States wants "not just to sustain the international order", but to "modernize it", to make sure that it "represents the interests, the values, the hopes of all nations." To build "an open, secure, and prosperous future", countries in the world should "work in common cause to uphold the principles that make our shared progress possible" and "stand up for the right of every nation to write its own future." Blinken even states that "everything I've said today boils down to: advancing human progress, leaving to our children a world that's more peaceful, more prosperous, and more free." In this future world, energy is "cleaner, more secure, and more affordable," "trade and commerce support workers, raise incomes, create opportunity;" "countries are secure from coercion and aggression, and people, ideas, goods, and capital move freely."

Regarding to Asia and China, the United States hopes to forge "closer ties with the countries and people of Asia to try to enhance peace, prosperity, freedom, equality, sustainability. "But the United States "cannot rely on Beijing to change its trajectory." Instead, it will "compete with China to defend our interests and build our vision for the future." It will "shape

the strategic environment around Beijing to advance our vision for an open, inclusive international system.” To people of China, “we’ll compete with confidence; we’ll cooperate wherever we can; we’ll contest where we must. We do not see conflict.” In addition, the United States will align with its allies to “defend” and “reform” the rules-based order, and “lead a race to the top on tech, on climate, infrastructure, global health, and inclusive economic growth.” For Indo-Pacific region, the US vision is that it should be “one of a free and open Indo-Pacific where rules are developed transparently and applied fairly.”

Diplomacy

Blinken highlights the importance of diplomacy by noting that it is the “indispensable tool for shaping our shared future.” The United States has “put diplomacy back at the center of American foreign policy.” Through diplomacy, the United States is “not looking for conflict or a new Cold War,” but is “determined to avoid both.” It does not seek to “block China from its role as a major power,” nor to stop China “from growing their economy or advancing the interests of their people.” At times like these, “diplomacy is vital.” It is about “how we make clear our profound concerns, better understand each other’s perspective, and have no doubt about each other’s intentions.” The challenge posed by China “will test American diplomacy like nothing we’ve seen before.” Blinken said he has “never been more convinced about the power and the purpose of American diplomacy or sure about our capacity to meet the challenges of this decisive decade. “He is “determined to give the State Department and our diplomats the tools that they need to meet this challenge head on.”

Network of allies and partners

Aligning with allies and partners is a key US strategy. From day one, The US has worked to “re-energize America’s unmatched network of alliances and partnerships and to re-engage in international institutions.” Blinken highlighted the US efforts to build and strengthen alliances in Indo-Pacific region. The US “reaffirmed our vital security alliances with South Korea and Japan, “launched the Indo-Pacific Economic Framework for Prosperity, a first-of-its-kind initiative for the region,” took part in the leaders’ summit of the Quad countries, Australia, Japan, India, the United States.” The US is “reinvigorating our partnership with ASEAN,” “building bridges among our Indo-Pacific and European partners, including by inviting Asian allies to the NATO summit,” “enhancing peace and stability in the Indo-Pacific” through a new security partnership between Australia, the United Kingdom, and the United States (AUKUS).

In addition, the US has “deepened our alignment across the Atlantic” through the U.S.-EU Trade and Technology Council. The US also worked with the G20 and G7 on “fair trade” and meeting the “infrastructure needs in developing countries.” Both are meant to address China’s trade policy that “distorts” market. The US convened global summits on defeating COVID-19

and renewing global democracy, and rejoined the UN Human Rights Council and the WHO, re-energized NATO.

Chinese people

The Biden administration consciously made a distinction between the CCP and Chinese people. China has “formidable economic resources, including a highly capable workforce.” The US has attracted “millions of students from China, who have enriched our communities and forged lifelong bonds with Americans. Despite the pandemic, US issued “more than 100,000 visas to Chinese students in just four months”. “We’re thrilled that they’ve chosen to study in the United States.” “More than 80 percent of Chinese students who pursue science and technology PhDs in the United States have chosen to stay,” “They help drive innovation here at home, and that benefits all of us.” “Chinese Americans made invaluable contributions to our country; they’ve done so for generations. Mistreating someone of Chinese descent goes against everything we stand for as a country.” The US has “profound differences with the Chinese Communist Party and the Chinese Government. But those differences are between governments and systems – not between our people.” “The American people have great respect for the Chinese people. We respect their achievements, their history, their culture. We deeply value the ties of family and friendship that connect us. And we sincerely wish for our governments to work together on issues that matter to their lives and to the lives of Americans, and for that matter the lives of people around the world.” “We can stay vigilant about our national security without closing our doors.”

US strengths

The US is “endowed with many strengths.” These include “peaceful neighbors, a diverse and growing population, abundant resources, the world’s reserve currency, the most powerful military on Earth, and a thriving culture of innovation and entrepreneurship.” The American “open society” attracts flows of talent and investment.” It continues to “investing in our core sources of national strength” – starting with a modern industrial strategy, “making strategic investments in education and worker training”, investing in research, development, advanced manufacturing. “ The US has “long been a destination for talented, driven people from every part of the planet.” And of course, the “core source of national strength” of the US is “our democracy,” one of the most powerful assets in this contest with China.

Comprehensive strategy/compete

The Biden administration outlined a “comprehensive strategy” can be summed up in three words, “invest, align, compete.” The US “will invest in the foundations of our strength here at home – our competitiveness, our innovation, our democracy,” “align our efforts with our network of allies and partners, acting with common purpose and in common cause,” and “compete with China to defend our interests and build our vision for the future.” In this framework, the US will “compete with confidence; we’ll cooperate wherever we can; we’ll contest where we must. We do not see conflict.”

The US is “well-positioned to outcompete China in key areas.” It will unleash next wave of innovation and at the same time safeguard its technological competitiveness through a series of measures. The US will “push back on” China’s market-distorting policies and practices, and “boost supply chain security and resilience”. It will “stand together with others against economic coercion and intimidation.” It will “fight for American workers and industry with every tool we have.”

The US seeks to preserve peace through “integrated deterrence”: “bringing in allies and partners; working across the conventional, the nuclear, space, and informational domains; drawing on our reinforcing strengths in economics, in technology, and in diplomacy.” The US will “manage this relationship responsibly” to prevent unintended conflict from happening. US has “prioritized crisis communications and risk reduction measures with Beijing,” and remains “committed to intense diplomacy alongside intense competition. The US will “work to maintain economic and people-to-people ties connecting the United States and China.” “Competition need not lead to conflict. We do not seek it. We will work to avoid it. But we will defend our interests against any threat.” In fact, the US will seek to cooperate with China. “Even as we invest, align, and compete, we’ll work together with Beijing where our interests come together. We can’t let the disagreements that divide us stop us from moving forward on the priorities that demand that we work together, for the good of our people and for the good of the world.” The areas where the two countries can work together include climate, the COVID-19 pandemic, on which “our fates are linked,” nonproliferation and arms control, countering illegal and illicit narcotics, global food crisis, and global macroeconomic coordination between the United States and China.

Devil terms

Urgent challenges

The US believes that “the foundations of the international order are under serious and sustained challenge.” Russian “poses a clear and present threat.” In attacking Ukraine three months ago, he also attacked the principles of sovereignty. But the US will “remain focused on the most serious long-term challenge to the international order – and that’s posed by the People’s Republic of China.” China is “integral to the global economy and to our ability to solve challenges from climate to COVID.” The United States and China “have to deal with each other for the

foreseeable future.” In addition, “none of us can meet many global challenges alone.” From fight the COVID-19 pandemic and prepare for future global health emergencies, rebuild from economic shocks, from supply-chain disruptions to debt crises; “The common denominator across these efforts is the simple fact that none of us can meet these challenges alone. We have to face them together.” The US “take on this challenge with confidence.” “We face [the challenges] as a country and a planet”. China is number one. The United States is number two. Unless we all do much more, much faster, the financial and human cost will be catastrophic. Plus, competing on clean energy and climate policy can produce results that benefit everyone.

The progress that the United States and China make together – including through the working group established by the Glasgow Declaration – is vital to our success in avoiding the worst consequences of this crisis. I urge China to join us in accelerating the pace of these shared efforts.

Conflict

For the United States, the rule-based international order has helped the world avoid “another world war and armed conflict between nuclear powers.” With China, “we are not looking for conflict or a new Cold War. To the contrary, we’re determined to avoid both.” “Competition need not lead to conflict. We do not seek it. We will work to avoid it. But we will defend our interests against any threat.” The Biden administration “is shifting our military investments away from platforms that were designed for the conflicts of the 20th century toward asymmetric systems that are longer-range, harder to find, easier to move.” “We’re developing new concepts to guide how we conduct military operations.” “As President Biden likes to say, the only conflict worse than an intended one is an unintended one. We’ll manage this relationship responsibly to prevent that from happening. We’ve prioritized crisis communications and risk reduction measures with Beijing. And on this issue – and every other – we remain committed to intense diplomacy alongside intense competition.” “And to the people of China: we’ll compete with confidence; we’ll cooperate wherever we can; we’ll contest where we must. We do not see conflict.”

China

China as a devil term refers to the Beijing regime, sometime referred as “Beijing” and “Chinese Communist Party”. It is the “only country with both the intent to reshape the international order and, increasingly, the economic, diplomatic, military, and technological power to do it.” The US has “profound differences with the Chinese Communist Party and the Chinese Government.” China once “was isolated and struggling with widespread poverty and hunger.” Now it is “a global power with extraordinary reach, influence, and ambition.” It “seeks to

dominate the technologies and industries of the future. “It’s rapidly modernized its military and intends to become a top tier fighting force with global reach.” Beijing “is undermining” international rules. The “ruling Chinese Communist Party has become more repressive at home and more aggressive abroad.” “Beijing believes that its model is the better one; that a party-led centralized system is more efficient, less messy, ultimately superior to democracy.” “Beijing wants to put itself at the center of global innovation and manufacturing.” Blinken used many examples ranging from China’s legal system, state subsidies, and growing coercion to illustrate the “lack of reciprocity” in US-China relationship and pointed it is unacceptable and unsustainable.

Discussions and conclusions

Among the God terms identified, the “rule-based international system” and “shared vision of future” together depict a worldview that aligns primarily with the win-win zone in the Mixed-motive Model. This zone encompasses three strategies: “cooperation,” “unconditionally constructive,” and “win-win or no-deal.” Of these, “cooperation” is the most emphasized in Blinken’s speech. He highlights that the stability and opportunity provided by the international system made “China’s economic growth possible,” and that this same system enables the United States and China “to coexist and cooperate.”

The term “constructive” appears as well, though it is explicitly not “unconditional.” Blinken states, “we’ll engage constructively with China wherever we can, not as a favor to us or anyone else, and never in exchange for walking away from our principles.” This framing positions “constructive” engagement as conditional upon adherence to U.S. principles, which places it toward the advocacy end of the Mixed-motive Model’s continuum.

The God term “Chinese people” seems to invoke an “unconditionally constructive” stance rhetorically. Blinken acknowledges the valuable contributions of Chinese Americans and condemns mistreatment against them as contrary to American values. Yet, this stance is again qualified, as he asserts, “We can stay vigilant about our national security without closing our doors,” reflecting a balancing act between openness and security concerns.

Overall, the rhetorical clusters Blinken uses to elaborate on the “rule-based international order” and “shared vision of future” convey strong ethos, or source credibility. It is indisputable that China has significantly benefited from the liberal world system dominated by the United States. For example, China’s economy grew by 61% within five years after its entry into the WTO in 2001 (Li, Liu, & Zhou, 2023). This economic growth partly explains why China now positions itself as an advocate of globalization (Shrader, 2020). Moreover, Blinken’s rhetoric about defending foundational documents of the international order, including the UN Charter and the Universal Declaration of Human Rights, which enshrine principles such as self-determination, sovereignty, and the peaceful settlement of disputes, is rhetorically unassailable.

However, the credibility of this ethos faces challenges in the eyes of some international actors, particularly in the Arabic world, where U.S. policies regarding conflicts in Gaza and Ukraine

evoke perceptions of double standards and Western hypocrisy. Critics ask: if both Israel and Russia are occupiers, why is one shunned while the other is protected? (Abbas, 2024). Research supports these concerns, showing a high degree of voting alignment between the U.S. and Israel in the UN General Assembly and a liberal use of the U.S. veto power in the Security Council to support Israel, effectively creating a deadlock on the Question of Palestine at the United Nations (Sarsar, 2004). These tensions complicate the image of the U.S. as a fully impartial defender of the international order.

As such, the above rhetoric does not mean that the United States is a “do-gooder.” Rather, the US has its national interest in defending and reforming the international system. This is why Blinken articulated a series of advocacy strategies. The Devil terms “urgent challenges” and “China/CCP” demand asymmetric advocacy strategies. The whole range of US policies and “investments,” as is articulated by Blinken in elaborating the God terms “diplomacy,” “network of allies and partners,” and “comprehensive strategies,” center on advocacy. Among the three strategies on the advocacy extreme of the Mixed-motive Model, in which the organizations win and the publics lose, “contention” and “principled approaches” are stressed, whereas “avoidance” is absent. Blinken states, “we’ll compete with confidence; we’ll cooperate wherever we can; we’ll contest where we must. We do not see conflict.” The US will not “sacrifice” its “core values”; It “will defend” international principles, “work in common cause to uphold the principles,” and “never in exchange for walking away from our principles.”

The three strategies on the other extreme of the Mixed-motive Model, namely “accommodation,” “compromise,” and “mediated communication,” where the public wins and the organization loses, are all present in Blinken’s speech. “Accommodation” is reflected by the rhetoric of “reforming” the international system: “we must defend and reform the rules-based international order..... that should benefit all nations.” The statement that the US is not seeking “to transform China’s political system” to an extent represents a compromise to the US foreign policy that has promoted the value of democracy and liberty in the world. The extensive emphasis on “America’s unmatched network of alliances and partnerships, along with rejoining the UN Human Rights Council and the WHO, the World Health Organization” can be regarded as necessary instruments of “mediated communication.” “Compromise” is not clearly stated but in diplomacy it is very often practiced.

In terms of worldviews, Blinken’s speech shows a mixture of liberalism, realism, and constructivism, but ultimately rests in realism. His call of “advancing human progress, leaving to our children a world that’s more peaceful, more prosperous, and more free” paints a liberal world order. So is the rhetoric of defending and reforming the international institutions. At the same time, he notes the “growing convergence about the need to approach relations with Beijing with more realism.” He extensively elaborated on “investing” in US strengths and consolidating the network of alliances and partnerships, ensuring “that our military stays ahead.” The constructivist theory of international relations argues that the social world is constructed and not given; it emphasizes identities, ideas, values, and social norms, or a standard of appropriate behavior for actors (Wendt, 1999). Blinken’s elaboration on “rule-based international order,” “shared vision of future,” and the “urgent challenges” it faces due to China’s “misbehavior” all reflect a constructivist worldview. The statement that “We’ll continue to raise

these issues and call for change – not to stand against China, but to stand up for peace, security, and human dignity” highlights the power of persuasion and ideas. Such a mixture of worldviews helps justify the US use of multiple strategies identified by the Mixed-motive Model and reduces the potential tensions among them, because each of the strategies is valid within a particular worldview.

As can be seen, Blinken portrayed a mixed strategy approach to China and projects a worldview that mixes realism, liberalism, and constructivism. On the one hand, such a complicated vision and strategy mix are conditioned by the complexity of the bilateral relations and the world reality. As Blinken states, the US-China relationship is “one of the most complex and consequential relationships of any that we have in the world today,” and neither country can “meet many global challenges alone.” On the other hand, it appears that the cooperation rhetoric also serves as a safety valve and risk communication to prevent the US-China competition from turning into a real crisis such as a head-on armed conflict. This is clearly reflected in the rhetorical clusters elaborating on the Devil term “conflict.” Blinken states, “As President Biden likes to say, the only conflict worse than an intended one is an unintended one. We’ll manage this relationship responsibly to prevent that from happening. We’ve prioritized crisis communications and risk reduction measures with Beijing. And on this issue – and every other – we remain committed to intense diplomacy alongside intense competition.” This is also shown by US national security advisor Jake Sullivan’s emphasizing the importance of dialogue for preventing risks from turning into crisis. He said Washington was “willing to engage in dialogue” with China and others “to better understand risks and counter misperceptions” (South China Morning Post, 2024).

Lastly, in terms of theory building, a large body of literature testing the Two-way Symmetrical model through quantitative approaches has developed; however, research using rhetorical criticism to test the model is rare, and studies testing the Mixed-motive Model are even scarcer. This scarcity is likely because the latter model involves many variables and dimensions, making quantitative research difficult to design. The current research shows that rhetorical criticism can fully capture the sophistication and subtlety of the Mixed-motive Model and its various dimensions, potentially serving as a better method to test the model. Furthermore, research testing these models has largely focused on the organization. More research is needed to examine how other parties in the conflict perceive and rhetorically respond to the organization’s rhetoric. In the case of US-China relations, China does not appear convinced by the cooperation rhetoric of the United States. For example, when Chinese Foreign Minister Wang Yi met Antony Blinken in September 2024, he said to his US counterpart, “The U.S. cannot always approach China with ‘two faces,’ simultaneously trying to contain and suppress China while engaging in dialogue and cooperation as if nothing is wrong. Since the U.S. has repeatedly stated that it does not intend to conflict with China, it should fundamentally establish a rational understanding of China and build a proper way to coexist.” (Chinese Ministry of Foreign Affairs, 2024). The theoretical questions that arise from such rhetoric are: To what extent would the conflicting parties truly trust each other’s win-win rhetoric? What factors would make such rhetoric credible? How would the conflicting parties convince each other that their worldviews align around the win-win zone of the Mixed-motive Model, and that the cooperation rhetoric is not a façade

concealing advocacy motives or merely a form of risk communication to prevent conflict from spiraling out of control? If the latter is the case, should the Mixed-motive Model be regarded as a contingency model instead of a symmetrical model? More studies are needed to address these questions.

Biographical note

Dr. Juyan Zhang is Professor of Communication at the UTSA and Contributing Scholar for the USC Center on Public Diplomacy. He received his doctoral degree from the School of Journalism, University of Missouri-Columbia. His research focuses on strategic communication, public diplomacy, interfaith dialogue, religious tourism, and media studies. Zhang contributed two monographs for the USC Center on Public Diplomacy's faith diplomacy project, respectively on Buddhist diplomacy and the Sino-Vatican faith diplomacy.

References

- Abbas, F. (2024, October 22). Analysis: Memo from Moscow — How Russia courts the Arab world. *Semafor*. <https://news.yahoo.com/news/analysis-memo-moscow-russia-wins-133247673.html>
- Aitken-Turff, F., & Jackson, N. (2006). A mixed motive approach to lobbying: Applying game theory to analyse the impact of co-operation and conflict on perceived lobbying success. *Journal of Public Affairs*, 6(2), 84–101. <https://doi.org/10.1002/pa.216>
- Ashbee, E., & Hurst, S. (2022). The Trump administration and China: Policy continuity or transformation? In *The Trump Administration* (pp. 274–291). Routledge.
- Balzacq, T. (2005). The three faces of securitization: Political agency, audience and context. *European journal of international relations*, 11(2), 171-201. <https://doi.org/10.1177/1354066105052960>
- Baskin, O., Aronoff, C., & Lattimore, D. (1997). *Public relations: The profession and the practice* (4th ed., p. 43). Brown & Benchmark.
- Black, J., & Morrison, A. (2021, May–June). The strategic challenges of decoupling. *Harvard Business Review*. <https://hbr.org/2021/05/the-strategic-challenges-of-decoupling>
- Blidaru, S. (2018). Zero-sum games and mixed-motive games in the fight of the Russian Federation with the West for Ukraine. *Europolity: Continuity and Change in European Governance*, 12(2), 37–66. <https://doi.org/10.25019/europolity.2018.12.2.02>
- Blinken, A. (2021). A foreign policy for the American people. *U.S. Department of State*. <https://www.state.gov/a-foreign-policy-for-the-american-people/>
- Burke, K. (2023). *The philosophy of literary form*. University of California Press.
- Burke, K. (1969). *A rhetoric of motives* (p. 43). University of California Press.
- Burke, K. (1967). *The philosophy of literary form: Studies in symbolic action* (2nd ed., p. 315). (Original work published 1941). Louisiana State University Press.

- Burke, K. (1966). Poetics in particular, language in general. In *Language as symbolic action* (p. 42). University of California Press.
- Burke, K. (1951). Rhetoric—old and new. *Journal of General Education*, 5, 203.
- Chinese Ministry of Foreign Affairs. (2024, September 28). Wang Yi meets Antony Blinken. https://www.mfa.gov.cn/wjbzhd/202409/t20240928_11499554.shtml
- Center for Strategic and International Studies. (2024, January 3). *U.S.-China relations in 2024: Managing competition without conflict*. <https://www.csis.org/analysis/us-china-relations-2024-managing-competition-without-conflict>
- Dozier, D. M., Grunig, L. A., & Grunig, J. E. (1995). *Manager's guide to excellence in public relations and communication management*. Lawrence Erlbaum Associates.
- Foss, S. K. (2008). *Rhetorical criticism: Exploration and practice*. Waveland Press.
- Goddard, S. E., & Krebs, R. R. (2015). Rhetoric, legitimation, and grand strategy. *Security Studies*, 24(1), 5–36. <https://doi.org/10.1080/09636412.2014.1001198>
- Grunig, J. E. (2001). Two-way symmetrical public relations: Past, present, and future. In R. L. Heath (Ed.), *Handbook of public relations* (pp. 11–30). Sage.
- Grunig, J. E., & White, J. (1992). The effect of worldviews on public relations theory and practice. In J. E. Grunig (Ed.), *Excellence in public relations and communication management* (pp. 31–64). Lawrence Erlbaum Associates.
- Haas, R. N. (2024). How will Biden and Trump tackle trade with China? *Brookings*. <https://www.brookings.edu/articles/how-will-biden-and-trump-tackle-trade-with-china/>
- Hotham, O. (2024, August 29). Biden 'looking forward' to talks with Xi in coming weeks: Sullivan. *AFP*. <https://news.yahoo.com/news/biden-aide-sullivan-raises-taiwan-042915765.html>
- Johnson, I. (2022, May 27). Biden's grand China strategy: Eloquent but inadequate. *Council on Foreign Relations*. <https://www.cfr.org/in-brief/biden-china-blinken-speech-policy-grand-strategy>
- Koptak, P. (1990). *Judah in the biblical story of Joseph: Rhetoric and biography in the light of Kenneth Burke's theory of identification* (Doctoral dissertation, Northwestern University).
- Krebs, R. R., & Jackson, P. T. (2007). Twisting tongues and twisting arms: The power of political rhetoric. *European Journal of International Relations*, 13(1), 35–66. <https://doi.org/10.1177/1354066107074284>
- Kubo, F. (2019). Reading the Trump administration's China policy. *Asia-Pacific Review*, 26(1), 58–76. <https://doi.org/10.1080/13439006.2019.1633153>
- Li, J., Liu, Y., & Zhou, Y. (2023). Assessing the impact of WTO accession on China's economic growth: A synthetic control approach. In H. Gao, D. Raess, & K. Zeng (Eds.), *China and the WTO: A twenty-year assessment* (pp. 73–88). Cambridge University Press.
- Murphy, P. (1991). The limits of symmetry: A game theory approach to symmetric and asymmetric public relations. *Public Relations Research Annual*, 3, 115–131. https://doi.org/10.1207/s1532754xjpr0301-4_5

- Ooi, S.-M., & D'Arcangelis, G. (2017). Framing China: Discourses of othering in US news and political rhetoric. *Global Media and China*, 2(3–4), 269–283. <https://doi.org/10.1177/2059436418756096>
- Plowman, K. D., Briggs, W. G., & Huang, Y. (2001). Public relations and conflict resolution. In R. L. Heath (Ed.), *Handbook of public relations* (pp. 301–310). Sage Publications.
- Reuters. (2020, September 7). Trump again raises idea of decoupling economy from China. <https://www.reuters.com/article/idUSKBN25Z08T/>
- Rueckert, W. H. (1994). *Encounters with Kenneth Burke*. University of Illinois Press.
- Sarsar, S. (2004). The question of Palestine and United States behavior at the United Nations. *International Journal of Politics, Culture, and Society*, 17, 457–470. <https://doi.org/10.1023/B:IJPS.0000019613.01593.5e>
- Shrader, M. (2020, June 5). China is weaponizing globalization. *Foreign Policy*. <https://foreignpolicy.com/2020/06/05/china-globalization-weaponizing-trade-communist-party/>
- Schuetz, M., & Chan, E. (2019, May 25). Trump's throttling of Huawei could backfire on U.S. tech. *Bloomberg*. <https://www.bloomberg.com/news/articles/2019-05-25/trump-s-war-on-china-tech-could-backfire-on-u-s-companies>
- Shi, J. (2024, March 12). The China policy gap between Biden and Trump is bigger than you think. *The Diplomat*. <https://thediplomat.com/2024/03/the-china-policy-gap-between-biden-and-trump-is-bigger-than-you-think/>
- Signitzer, B. H., & Coombs, T. (1992). Public relations and public diplomacy: Conceptual covergences. *Public relations review*, 18(2), 137-147.
- Simonson, P. (2014). Rhetoric as a sociological problem. *Argumentation and Advocacy*, 50(4), 242-252. <https://doi.org/10.1080/00028533.2014.11821825>
- Silverstone, R. (1999). *Why study the media?* SAGE Publications.
- South China Morning Post. (2024, October 24). US issues AI national security memo to cut risks and avoid 'strategic surprise' by China. <https://finance.yahoo.com/news/us-unveil-ai-national-security-093000938.html>
- The Economist. (2020, November 19). Joe Biden's China policy will be a mix of Trump's and Obama's. <https://www.economist.com/china/2020/11/19/joe-bidens-china-policy-will-be-a-mix-of-trumps-and-obamas>
- Wendt, A. (1999). *Social theory of international politics* (pp. 1–4). Cambridge University Press.
- Winkler, S. C., & Jerdén, B. (2023). US foreign policy elites and the great rejuvenation of the ideological China threat: The role of rhetoric and the ideologization of geopolitical threats. *Journal of International Relations and Development*, 26(1), 159–184. <https://doi.org/10.1057/s41268-022-00288-6>
- Zhang, J. (2010). Exploring rhetoric of public diplomacy in the mixed-motive situation: Using the case of President Obama's 'nuclear-free world' speech in Prague. *Place Branding and Public Diplomacy*, 6(4), 287-299. <https://doi.org/10.1057/pb.2010.31>

Notes

¹ China Briefing. Available at< <https://www.china-briefing.com/news/us-china-relations-in-the-biden-era-a-timeline/>>

Crisis, Coercion and Conflict as the New Peace: A Critical Analysis of Western War Justifying Discourse^{*}

Greg Simons^{**}

Daffodil International University, Dhaka

George Orwell wrote and warned about the role of deceptive and misleading language in politics, a form of double speak to influence and persuade unwary audiences. War is an activity that requires the accumulation of political capital and legitimacy. By invoking a crisis situation, the political environment implies an extra-ordinary moment where standard laws and rules no longer apply. This is better done not by framing or narrating the 'desired' coming armed conflict in terms of unemotional national interests or the logical use of "just war", but through the emotional use of selective values and norms and through vague concepts like 'humanitarian war'. The Orwellian language of Western politicians and other actors promoting war in the name of peace during the era of the Global War on Terrorism (GWOT) and the Arab Spring from the year 2001 shall be subject to critical analysis, which reveals that the use of emotions and deceptive language are key components to the war justifying discourse. This literature review reveals that a rather systematic, symbolic and ritualistic use of war justifying discourse is the mechanism through which actors' attempt to engineer public perception and consent to an act that in all probability undermines the interests and security of that same public. This article looks at these aspects and factors as a sum of the individual parts, rather than identifying and analysing those individual pieces as the influence of the war justifying discourse comes from the totality through the cumulative effect of the interactions.

Keywords: Just war, George Orwell, discourse, GWOT, crisis (communication), West (Global North)

^{*} Article submitted on 24/09/2025. Article accepted on 15/12/2025.

^{**} gregmons@yahoo.com

Introduction

Throughout the ages in organised human society, war has been construed to be politics by a coercive means, intended to compel an opponent to do some action that they would not ordinarily do of their own free will (Clausewitz, 1982; Griffith, 1971). Machiavelli (2001) proposed the same line of thinking, but went even further as he believed the political elite faced no obstacles in deceiving foe or friend alike when it comes to pursuing the political aims of war, where all other considerations are subordinated to war's pursuit. It has been noted that the success of wars is being increasingly measured in political and not military terms, where there is a strong dependence on engineering domestic and international public opinion (Payne, 2005), where politics and mass media have formed a symbiotic relationship that witnesses mass media acting as an instrument of war as policy (DiMaggio, 2009). A series of dualistic, binary opposing projected discursive 'realities' are communicated to an audience to engineer their perception and consent by cognitively shifting them from a critical individual to an emotional herd in the politicised context of a 'crisis' to better ensure compliance with the war policy intent and also to overlook the otherwise obvious dilemmas and contradictions in the logic of the call to war.

Value-based approaches of modern Western society argue that organised warfare is inherently barbaric and should be left in the past. However, war also reflects the predominant political and cultural trends in the society from where it originates (Greene, 2007, p. xvii). There are paradoxes inherent in the political functioning of contemporary international relations, where seemingly evident self-contradictory propositions exist, where this paradoxical logic is best exemplified in the entire realm of armed conflict (Luttwak, 2001). These paradoxes are even hinted at in popular literature, such as the newspeak slogans of George Orwell's 1984. It has been argued that propaganda is much more than merely misleading rhetoric that has been communicated, rather it is an elaborate process that involves recontextualisation in order to ensure its effectiveness and success (Oddo, 2018).

There is a basic value dilemma that faces Western policymakers and mass media. On the one hand, wars are rightly condemned as being barbaric and they should be used only as a last resort requiring a high burden of proof. On the other hand, 'humanitarian wars' have been launched as a convenient short-term policy to bring about foreign policy 'successes' in selected countries. Therefore, warfare needs to be sold to the public, where public support before or during a military operation is of paramount importance (Western, 2005). There is also the need to not treat every war as being ethically and morally equal, but to divide between *good wars* (us) and *bad wars* (them) (Zollmann, 2017). The research question that is posed in this paper is the following: in Western warfare justifying discourse, what are the primary communicated values and outcomes and how are these motivated?

This paper begins with a literature review of academic literature on the topic of politics and the use and role of deception in language communicated to larger publics via mass communication. In the next section, the foundations for the 'moral' and 'ethical' reasoning of warfare is briefly described and explained. The following section will provide an overview of

the academic method selected and used in this paper. The fourth section consists of the main analysis of this paper and engages in a critical overview of the academic literature on how politicians and journalists construct and publicly communicate discourse on Western-led humanitarian warfare and regime change operations. Ultimately, the paper seeks to demonstrate that by taking a broader overview of the literature on the topic, the patterns reveal that through the cumulative effects of a range of varied discursive communication strategies an audience can have their consent engineered, rather than looking at more narrowly focused topics in text. The topic is not only of importance and relevance as audiences are being increasingly manipulated in what has been branded as the New Cold War, this study also intends to serve as a modest tool for helping to be able to decode and unmask those manipulations.

Politics and Deception in Language

An observation from 2010 noted a rather peculiar paradox in terms of the growing gap between Western rhetoric and the Western practice of warfare:

What perhaps remains a greater paradox is the divergence in the West between rhetoric and practice in attitudes to and the conduct of war, where one might have expected a rejection of certain forms of violence as modern states and societies have matured, we have in reality seen little change in the degree of conflict, even if its nature has modified (Kassimeras & Buckley, 2010, p. 2).

In describing the political climate of the 1920s, Walter Lippmann named indifference as the greatest challenge to 'democratic' politics and suggested means of countering it: "To conquer it men generally turn to their ancient comforter, self-deception: they complain about the stolid, inert masses and the apathy of the people" (Lippmann, 2005, 11). Lippmann's ideas have been subsequently expanded upon and developed by other researchers. According to Edelman (2001, pp. 3-4): "Virtually all political groups and individuals benefit at times from misleading and inaccurate assumptions and accordingly have an incentive to create and disseminate such beliefs. [...] A very high proportion of the beliefs that guide political conduct and political rhetoric accordingly are myths." Furthermore, with reference to social change language, Edelman (2001, p. 78) argues that "it helps determine beliefs about the past and present and what specific changes will mean for various groups in the future, and it shapes beliefs about which interest groups and public officials should be regarded as allies and which as threats or enemies." Language in its written and oral forms is present in the information realm in terms of its interpretations and representations of the physical realm in order to gain influence in the cognitive realm of audiences.

The above ideas seem rather diverse and scattered on the surface, but their intellectual logic and psychological effect are interconnected. Although the physical act of warfare is often associated and conceived within the theoretical framework of realism, the art of politics mixed with the art of language is needed to construct the idea of its supposed

necessity. This construction takes place in the information realm through the representation and interpretation to ignite and engineer emotions and ideas in the cognitive realm. The strong emotions of fear (of some risk) and/or altruism (to do something good) are effective primers and mobilisers for engineering consent by shaping audience opinion and perception of the elements of the physical realm. Rhetoric is key in manufacturing the feeling of comfort and/or self-deception among the audiences, which is necessary owing to the paradox and discrepancy between the intention of the political word and the outcome of the military action. Once the consent or lack of resistance to war as a policy is secured, the aftermath and outcome is of lesser importance as this is less commonly covered in 'good' Western wars as opposed to 'bad' enemy state wars (Simons, 2016; Zollmann, 2017).

In Orwell's dystopian novel 1984, language is used to keep the masses ignorant of their psychological manipulation and compliant with the wishes of the political system, where the public is unable to detect the contradictions present in society. This was combined with the use of catchphrases and slogans creating information and cognitive dominance (Hossain, 2017; Abdu & Khafaga, 2019, pp. 423-424). It is noted that linguistically speaking, strategies and tactics are used to persuade and manipulate the public in 1984:

[...] mind control is linguistically evidenced through a number of strategies, such as simplification, euphemism and morphologicalisation. All of which depend totally on the ideological use of lexis. These strategies have linguistically been manifested in different tactics, including slogans and caption (simplification); naming and/or renaming things and mollification (euphemism); and derivation, inflection, coinage, compounding and reduction (morphologicalisation) (Abdu & Khafaga, 2019, p. 428)-

The above quote alludes to the use of language as a means to prevent cognitive clarity of the politicised human realm in depiction and description. Orwell hated jargon and valued the clarity of simplicity, where his criticism gave rise to the entry of such concepts as 'big brother,' 'newspeak' and 'doublethink' (Slater, 1975, p. 460). Orwell pointed to the dualistic situation where language can either enrich or corrupt the intellectual thought process and the resulting cognitive evaluations. Johnson (2021, pp. 35-36) notes that language in 1984 is key to understanding that the freedom of thought, speech and action are a potential 'threat' to a political system, especially one where language is a mechanism of the construction of 'reality' and not simply its description.

Various tactics and mechanisms for manipulative political speech have been noted, especially with the use of projected or asserted values that reveal "metaphor, assertion-based patterns and thesis-antithesis in the service of axiological proximisation" (Sowinska, 2013, p. 792). For example, the rhetorical uses of values in the discourse of US President George W. Bush's presidential administration were "in fact re-definitions of particular axiological and value laden concepts, are an important tool in presidents' hands to establish who is 'us' and who is 'them', [and can be used to] garner support for particular policy proposals and legitimise political actions" (Sowinska, 2013, p. 806). Dunmire (2005, p. 481) notes, "an important ideological component of political discourse derives from its representation of the future and the rhetorical functions those representations serve in

implicating more immediate and discursive actions.” This is particularly valid in the nominalisation of threat functions as a means to construct a possible bleak future that requires drastic policy action, where the operationalised use of fear engineers public consent. Trotter (2015) observed that an audience develops a passively reactive and uncritically thinking herd mentality stance that subconsciously is guided by the communicator’s suggestion when emotionally primed and mobilised by fear.

The above cognitive conditions can be engineered and manipulated through the invocation of the concept of ‘crisis’ and its subsequent communication. By declaring a ‘crisis’, an extraordinary situation is implied and as such the laws and rules that govern political management can be justified as being insufficient to tackle the problem at hand. As its is extraordinary and this is the logic to attempt to legitimise risky and grey zone legal acts of aggression against another state or non-state actor in the dubious name of humanity and values as “freedom” and “democracy” (Chifu & Simons, 2023, pp. 219-245). The definition of a crisis is given as the simultaneous presence of three different environmental factors: 1) a threat to values (life, property, ideology, identity ... etc.), 2) presence of uncertainty (not known what shall happen and when), and 3) the issue of time constraint (problematic situation needs urgent resolution) (Boin et al., 2005, pp. 2-4). The perception of a crisis, regardless of whether an actual crisis exists, can simultaneously increase the operational choices of those seeking to create the context for war and decrease the defender’s operational choices in an environment rich with informational and cognitive deception and manipulation using language to define the situation and what is to be done politically.

The quality of deceptive and manipulative political language, using rhetoric, metaphors, simplifications, slogans, catch phrases, euphemisms and morphologicalisation are key to the (un)successful ability to persuade and influence an audience of the façade of an ethical and justifiable proposal for war. However, the proposal for war is anything but ethical and justifiable under ordinary situations, the pertinent point is for the audience under manipulation to believe that it is a ‘good’ and ‘necessary’ war. In the next section, the idea and practice of the process of moral reasoning for warfare shall be discussed, a key element in the step of the legitimization of warfare.

Moral Reasoning of Warfare

Coker (2004, pp. 19-20) noted that there are Western styles of warfare where, “Europeans are intent on normatising war; the Americans in instrumentalising it. [...] For in instrumentalising it more than ever the US seems to be intent on transforming not only its character but its nature.” There are different logical bases in use, which are rhetorically operationalised to attempt to justify and/or legitimize the resorting to the use of different forms of warfare as a foreign policy option (Walker, 2019). Two such paths are to base argumentations on legal/national interests or moral/ethical considerations. How these are applied is changing as the nature of warfare and the actors evolve and transform. Rochester

(2016) argues that there are an increasing number of emerging political and legal dilemmas as a result of changing warfare realities that are a consequence of the new realities of 21st century international relations and geopolitics. He observes that warfare is becoming increasingly messy in terms of the actors involved and their warfare practice.

Inherent dilemmas and contradictions can be detected in the war justifying political discourse of Western political leaders, especially within the frame of notions of security and insecurity within the context of the Global War on Terrorism (GWOT) or the so called Arab Spring. For example, the projected greatest threat to modern civilisation has been identified as the global jihad and the threat of Islamic terrorism. Yet, the United States and its allies supporting and encouraging proxy warfare and the use of state sponsored terrorism proving to be a significant threat to civilian populations (especially in the Middle East North Africa region) and undermining human security abroad as well as domestically (Hashmi, 2014). For example, the US supported Islamist groups in Afghanistan (including Osama bin Laden) to destabilise the Soviet-backed Afghan government and Soviet forces there, Islamist and terrorist proxy forces have also been supported in Libya and Syria by the US and its allies against the Gaddafi and Assad governments. Something not hidden by mainstream media, even if not 'advertised', but rather 'justified' as necessary in the name of Western values ('freedom', 'democracy' and so forth). Scahill (2013) notes the destructive and disruptive effects of the US covert militarised foreign policy operations that have resulted in perpetual war.

Walzer (2006) points out that the argument that links the practice of warfare with the evaluation of justice is in every sense a political and moral imperative. There are contradictory politicised conversations that condemn and seek to ban "unjust wars" and to promote and legitimise "just wars". The notion of "just war" or jihad among the Abrahamic faiths (Christianity, Islam and Judaism) involves competing and conflicting images of the gods with their demands for allegiance, victory and faithfulness in conflict and wars (Hoffmann, 2006). Legitimacy for engaging in a "just war" or jihad is found not only in the resonance of the message, but also the credibility of the messenger. The principles of "just war" and jihad form the basis of a moral judgement of the actions of the named enemy and to suggest a policy remedy in the name of the common good (Temes, 2003; Fisher & Wicker, 2010). This boils down to creating a sense of moral righteousness in politically legitimising the path to war, however, the execution of the physical dimensions of warfare are also subjected to moral evaluation and weighing and can be found wanting.

Although the nature of just wars is based on the premise of moral judgement, the process makes use of reasoned logic of the pros and cons of going to war (or not) based on the weighing of concrete national interests. Traditionally speaking, these interests have tended to be dominated by military-political considerations that led to intervention or non-intervention, which has gradually been broadened over time and circumstances in international relations (Forbes & Hoffman, 1993). A more emotionally based logic, which exhibits moral and ethical judgement as a basis for informing military intervention in international relations is the Responsibility to Protect (R2P) doctrine. The UN Secretary-General Ban Ki Moon argued that there was a need for countries to engage in regional and

global collaboration in order to prevent mass atrocity crimes, which was argued as a right to military intervention as a means to protect those civilians at risk. R2P was championed as a new norm in international relations and was adopted at the UN World Summit in 2005 (Kassim, 2014, pp. 8-13). However, the concept is somewhat vague and vulnerable to abuse, which some claimed was the case in R2P military intervention in Libya and why it was not able to be invoked in Syria as a consequence (Thakur, 2013; Brockmeier, Stuenkel & Tourinho, 2016). The basic questions surrounding R2P remained unanswered and could be subjected to selective interpretation and application for geopolitical expediency. Responsibility to protect whom, from what, and by which means?

Method

This paper involves a study of academic literature that interprets Western war justifying discourse, the highly politicised physical act of going to war requires the facades of justification and political legitimacy to enact. Even though not all academic literature is in complete agreement on the Western war justifying discourse, they do reveal some important conceptual and practical aspects on how those narratives are conceived and implemented by policymakers and practitioners. A discourse study is a valuable means of identifying the various patterns and trends present in each scenario or situation. This is an examination of how people use language to create meaning, influence others, and shape social realities. Narratives and discourse interact to make sense of political and social identities. "In short, identities are storied as we and what happens in the world is not simply discovered but made meaningful via the narrative form" (Forchtner, 2021, p. 305). McKenna noted that there is "the need for a strong social theory given the materialist and context-bound nature of discourse in daily activity. From this basis, the paper then characterizes the "new times" in which contemporary discourse occurs, and briefly surveys those issues typically analysed, namely political economy, race and gender, and critical literacy" (McKenna, 2004, p. 9). Of particular interest and relevance are the following contextual factors: informational power, cognitive influence, and political speech as key elements of Western war justifying discourse.

To capture the above-mentioned variables of interest for this paper, a literature review was conducted. Literature reviews can offer a broader and more comprehensive overview of the topic and subject matter at hand to provide clarity of the subject matter problems (Pati & Lorusso, 2018; Conz & Magnani, 2020). There are several different types of literature review – systematic, semi-systematic and integrative (Snyder, 2019, pp. 334-336). The integrative (critical) literature review has been selected for this paper. Its aim is to "assess, critique, and synthesize the literature on a research topic in a way that enables new theoretical frameworks and perspectives to emerge" (Snyder, 2019, p. 335). A literature review should be dialogic in nature and not simply a repetition of other peoples' writings and thoughts. Therefore, the literature selected needs to be relevant to the research task at hand where research is framed in the context of earlier works (Silverman, 2020, pp. 50-51). The sample of academic literature was collected through a digital search on Google Scholar

using the search terms “war justifying discourse”, which generated 17,400 results. This sample was further refined through a manual examination of the texts in the first 15 pages (10 results per page) of results to evaluate the texts relevance to the research question posed. A sample of some 60 academic texts were chosen on the above-mentioned qualitative criteria concerning their pertinence to analysing and explaining Western war justifying discourse from the year 2001.

Justifying Wars: Constructing and Representing ‘Good’ Wars

Information and communication are essential elements in Western politics and prosecution of warfare. Roberts (2017, p. 21) notes the West’s model for warfare consists of “deliberately designed information-centric forces, and information dependent people, believing that information itself is the key to victory.” Information is important on the battlefield, but it is also central to the politics of war and how warfare is communicated to public audiences. On the politicised issue of ‘national security’, it is noted that the emotive environment is easily subjected to manipulation by different stakeholders. Edelman (2011, p. 7) suggests, “because such anxieties are easily aroused and because they can be easily directed against any domestic or foreign group that is labelled a threat, worries about national security are constantly evoked.” The key outcome is that the threat is taken as a plausible argument or believed falsehoods by the audience rather than objective and factual evidence. Because then the suggested policy ‘remedy’ can be implemented with the façade of legitimacy. This is observable in the communication of politics on warfare. It should also be noted that discourse and narratives intended to justify and legitimise conflict involvement, evolve over time and circumstances, they are not static.

Rather than treating legitimacy as something decided only at the start of a war, the model views it as a shifting process shaped by events on the ground, political rhetoric, and public reaction. The wartime justification trajectory outlines four phases—initial justification, conflict dynamics, social reactions, and post-conflict evaluations—that evolve as conflicts progress (Alexandrescu, 2025, p. 1).

There is a continual process of negotiation and renegotiation, contestation and alignment, which demands new justifications for the façade of political legitimacy. Western (2005) notes the need to lobby and persuade different stakeholders (executive and legislative branches of government, mass media, expert community, domestic and international publics and so forth) to win hearts and minds to accept and support warfare as ‘necessary’ foreign policy. Not all wars are equal though. Mass media play a crucial role in preparing the cognitive realm of audiences in the build-up towards a military intervention. Politicians and mainstream media evoke the mantra of ‘humanitarian intervention’ and ‘R2P’ in their storyline when named enemy states (of the Western-centric order) are asserted to have committed human rights violations whereas the Western military interventions are asserted

to be means of preventing human rights violations (Zollmann, 2017). This is clearly seen in the communication of political leaders' speeches.

[...] particular constructions of freedom contributed to the overall representation of the invasion of Iraq and the Global War On Terrorism as an atemporal struggle between good and evil forces, and also enabled the Bush administration to broaden the axiological spectrum to seek allies in 'freedom's cause' and thereby solicit legitimisation for US military involvement and democracy promotion in the Middle East on a global scale (Sowinska, 2013, p. 806).

The executive branches of Western governments have an established rhetorical approach in their war justifying discourse to public audiences in the post-Cold War world of 'wars of choice'. News media and journalists tend to cover the official war justifying discourse in an uncritical manner to mass audiences, where the press is assessed as having failed in its duty to hold power to account (Thussu & Freedman, 2003; Bennett, Lawrence & Livingston, 2007). A number of discourse strategies intended to legitimize (to make acceptable as the 'right' thing to do) militarized foreign policy have been noted: 1) creating and using emotions (fear in particular), 2) use of a hypothetical future, 3) use of rationality, 4) voices of expertise, and 5) projecting altruism (Reyes, 2011, p. 781). These discursive strategies are consciously and deliberately embedded into political speech, which is in turn relayed by news media to mass audiences as an instrument of war through the engineering of public consent by obtaining their approval and support.

Loseke (2009, p. 517) notes that "rhetoric might achieve effectiveness to the extent that it deploys only the most widely held and deeply embraced symbolic codes and their accompanying emotion codes, avoids codes whose meanings and evaluations are widely contested, and is one-dimensional, binary, and dramatic, and therefore discourages disagreements." Emotions are a powerful and effective tool of cognitive manipulation of audiences. According to Reyes (2011, p. 785): "The appeal to emotions allows social actors to skew the opinion of their interlocutors or audience regarding a specific matter." Therefore, emotions can be skewed in the audience towards accepting and responding to a negative representation of a 'them' or a positive representation of an 'us' to prime and mobilise audiences against one actor and against another. There are different types of emotions that can be exploited from the negative spectrum of fear or hatred or revenge to the positive spectrum of patriotism or pride. Fear is used in the information realm to affect and shape the cognitive realm of audiences to engineer and manufacture their consent to a suggested course of action. The 11 September 2001 terror attacks on the US mainland are an iconic moment in the collective memory of the US, where powerful emotions such as fear, anger and desire for revenge were instrumentalised to rally public support and legitimacy for what became known as the GWOT (Reyes, 2011). Fear of further attacks and an unseen enemy, anger at the human losses and attack on US soil, and the desire to engage in revenge to overcome the fear and anger were utilised in war justifying discourses. As Dunmire (2005, 507) argues: "Bush's vision offers us an Orwellian future in which the specific embodiment

of the threat may change, but the threat is always there, occasionally diminished but never fully defeated.”

Hypothetical future provides another angle for Western war justifying discourse. As noted by Reyes (2011, p. 786), “to pose a threat in the future that requires our imminent action in the present. Political actors exert their power addressing the future by employing specific linguistic choices and structures, such as conditional sentences [...]” A false logic can be used to nudge audience perception and opinion on a projected threat. For instance, it may be wrongly argued that failure to act or do something in one part of the world can have disastrous consequences in another part of the world and therefore military action is required (Chang & Mehan, 2008; Reyes, 2011, p. 794). This was used with the hypothetical arguments that Saddam Hussein was gathering weapons of mass destruction (WMDs) in Iraq and had links with terrorists, which were used as a pretext of a pre-emptive strike on Iraq in 2003 and the subsequent occupation of the country. Another deceptive discourse logic used to raise anxiety among the audience was the threat of leaving the ‘job’ of fighting terrorism incomplete in Afghanistan, which was also subsequently followed up as an excuse for the continued occupation of the country or fighting ‘terrorism’ in Syria to prevent its spread elsewhere (especially that of the Islamic State). It is possible to amalgamate different discourse strategies, such as emotions with a hypothetical future. As Loseke (2009, p. 518) suggests: “[...] the promised happy ending of Bush’s story, the triumph of virtue over morality, idealizes American values. Audience members outside the United States may not be amused with this story glorifying American values and anointing Americans as saviours of civilization.” Therefore, the war justifying discourse is based upon constructing the premise of a ‘promise’ of a ‘good’ future, which is projected on the basis that a ‘bad’ future shall prevail because of the risks and threats if nothing is done. There is also the use of projecting a bad future if “something” is not done to support another side in a proxy war as the Ukraine War. There is an attempt to justify and legitimise political action by using the false logic that the war will come to ‘our’ borders if something is not done, and to narrate and use discourse that makes the ‘good’ side (Ukraine) as a mirror to Western society and the ‘bad’ side (Russia) a binary opposite of us (Alyaya, 2023). This is done in a similar vein to the *Why We Fight* series of Frank Capra, such as the *Battle for Russia Part One*, which attempted to mirror the Soviet society to the US after decades of demonisation (please see <https://www.youtube.com/watch?v=Y6mkjqSxOlo&t=13s>).

This use of binary opposing dichotomies of discourse in political speech has been observed and confirmed by other researchers. Even though Donald Trump began the delivery of lethal military aid to Ukraine, he associates the war as “Biden’s war”, and Biden’s speeches followed various political and humanitarian imperatives that were framed in a crisis context. “The data analysis reveals that the discursive strategies used in speech are authorisation, moral evaluation, rationalisation, and mythopoesis” (Aluya, 2025, p. 117). The opposing logics of the discourse and narrative construction of the Ukraine War is an unbridgeable one at this stage, where one may be mistaken for thinking these sets of justifications and legitimacy were not even about the same conflict (Krickovic & Sakwa, 2022). There is a perfectly logical and operationally oriented reason for this situation, the

information flows on a crisis event as the Ukraine War are used to define reality, which in turn is used by the sides to try and restrict their opponent's operational choices while simultaneously enlarging their own operational choices. It is possible to see the shifts, backwards and forwards, on President Trump's discourse on the Ukraine War. When the war seems to be a winning proposition he supports escalating the conflict. When the war seems to be lost on the battlefield then it is "Biden's war" and he threatens to withdraw direct support in the proxy war and hand it over to the Europeans. In this way, evading being seen as being associated with losing a war (as Biden was in 2021 with Afghanistan) and to potentially make money from selling weapons to the Europeans to provide to Ukraine, minimising political risk and maximising economic opportunity. This may well explain Trump's obsession with Venezuelan regime change on the flimsy and not credible excuse of President Maduro being a drug cartel head and Venezuela a narco-state, while also acknowledging the largest oil reserves are to be found in Venezuela. A fact that has been mentioned by Trump in both of his presidencies.

Rational logic is established "when political actors present their legitimization as a process where decisions have been made after a heeded, evaluated and thoughtful procedure" (Reyes, 2011, p. 786). At times, false logics can be used, especially those that convey a sense of inevitable threat, such as the false claims by the Bush administration about Iraq's leader Saddam Hussein having weapons of mass destruction and links with terrorists (Chang & Mehan, 2008, p. 453), but they do need to be plausible and resonate with the audience. A rationally constructed discourse is articulated by the speaker and carried without further interpretation by mass media, which is used to justify a course of action that is likely to involve a certain cost or sacrifice by the audience being addressed. For example, the rational logic that "we" did not ask for a fight but have been forced by circumstances may be employed. Otherwise, any arguments used that invoke notions of 'consulting' with allies/partners or the 'international community's' opinion on a specific military issue or event imply conscious intentionality for war, rather than having an 'unwilling' government's policy hand 'forced'. In other instances, a decision is "not made lightly." Therefore, the constructed discourse is constituted on the basis of hard decisions, consultations with stakeholders, a general lack of choices available, or being 'forced' into a course of action (Chang & Mehan, 2008; Oddo, 2011, p. 388; Reyes, 2011, pp. 797-799). Another tested rational logic strategy of war justifying discourse by an actor is to divide and construct a cognitive distance in the audience between the constructed "us" and "them" as a means of projecting moral evaluation as a source of political legitimacy for war as policy (Oddo, 2011). This was experienced in President George W. Bush's infamous binary delineation of *you are either with us or with the terrorists* (them) after the 9/11 attacks.

Rational logic was also used in the humanitarian discourse that accompanied some of the uprisings during the so called "Arab Spring." Malito (2019, p. 104) identified three rhetorical mechanisms used in the war against Libya: the first refers to "regime change" which "was defined as a universal interest through the Manichean representation of Gaddafi opposed by a unified Libya (universalisation)." A second point is that "contradictions in the resort to violence have been marginalised and alternatives to militarisation have been

ignored, such in the case of the African Union's roadmap (simplification)." The third point made is that "the media and scholars have perpetuated dominant narratives portraying Gaddafi as a 'mad dog' of the Middle East (reiteration)." By dominating the information realm, it is hoped to depoliticise, delegitimise and marginalise alternative arguments or proposals to ensure the cognitive realm only imagines one realistic policy path, the road to war. This was clearly visualised in the 2015 parliamentary debates in Britain on the use of military force in Syria: "Crucially, pacifism's subjugation means that political discourse, and thus political action, is critically limited. And thus, war, and the practices and institutions that support it, seem normal and inevitable" (Reeves O'Toole, 2020, p. 20). The Syrian War (as have a multitude of other wars) proved that the role and value of the "dynamics of (de)legitimation are key in armed conflict as they are linked to actors' power and authority at local, regional and international levels. [...] Political discourses can become powerful discursive tools and have real consequences on conflict dynamics. They can affect the legitimisation of violent actions and in (re)producing actors' legitimacy" (Martini, 2020, p. 737). The importance of an effective and publicly visible framing and counter-framing in shaping actors' opportunities and threats in the Syrian conflict is also highlighted by Schoon and Duxbury (2019). Although, rational and emotional logics are not strictly separated in politicised discourse on the topic of warfare, they can become entangled and mutually reinforcing in the persuasiveness of the argument. To illustrate the mixing of styles of appeal and justification, the Ukraine War provides numerous examples. Where justification and opposition to a conflict go hand in hand, the rhetorical opposition to a conflict can serve as a board to deeper engagement in the conflict further ahead in time when it is convenient. This is found when juxtaposing Russian and versus NATO discourse on the Ukraine War.

First, both sides describe the events in a way that combines the factual and moral reading of them that favour them. Second, continuity is rhetorically established to justify the actions of both factions as rooted in their pre-conflict status rather than being seen as reacting to the war. Finally, both sides use threats and exhortations for others to act in accordance with their wishes and demands. These three strategies showcase how both sides rhetorically manage their accountability, moral rightness and, at the same time, work up the moral guilt of their opponents. Similarity between the literature on expressions and denials of prejudice are found here, in the case of discursively sanitizing military action (Demasi, 2023, p. 257).

Ethos and the role of expert voices in the field are instrumentalised to back political proposals with statements based upon their expertise. The persuasive power is derived from the potential of a credible expert to bring context and strength to the official argument to legitimize the proposed actions (Reyes, 2011, p. 786). In this category of discursive persuasion, 'experts' may consist of political leaders (political authority), military leaders (military authority), and subject knowledge experts (knowledge authority). There is an attempt to justify and legitimize the ideological positioning of warfare by the projected 'authority' of the speaker(s) to have the course of action supported or at least not resisted (Reyes, 2011, pp. 797-799). A great deal of contrived technical and political expertise was used to successfully establish the path for war in Iraq, which was based on the presumed

knowledge, expertise and public profile of key influencers for the US-led invasion of Iraq in 2003 (Chang & Mehan, 2008). The role of former General and Secretary of State Colin Powell in presenting fake WMD evidence at the United Nations, for example, where the messenger rather than the message was intended to be the more persuasive element.

Projecting altruism occurs when “public speakers, in particular, and social actors, in general, make sure their proposals do not appear driven only by personal interests. They present themselves, for example, as serving their voters, and therefore they legitimize proposals as a common good that will improve the conditions of a particular community” (Reyes, 2011, pp. 786-787). For example, altruism may be utilised by way of invoking the notion of a “just war” or R2P. Hence there is an attempt to project the act of warfare as being beneficial to other individuals or groups, especially if they are presented as being innocent and/or powerless against powerful perpetrators of ‘unspeakable’ violence. This category of discursive argumentation can include such ‘noble’ acts as bringing ‘democracy’ and ‘freedom’ to a disadvantaged group, preventing war crimes or crimes against humanity, as well as bringing peace and stability to a region or country (Reyes, 2011, pp. 801-803). In other words, some forms of tangible or intangible benefits are alleged for disadvantaged groups of a third party through the use of military force against another party (often a government or politically organised group). The war justifying discourse of the GWOT relied on the spirit of a rational logic and approach aligned with the notion of a “just war”, which has differed from the Arab Spring interventions that applied an emotional logic and approach more in tune with R2P. Haqpana and Tsouroufli (2023, p. 1) observed that “international development initiatives in the Global-South are often shaped by Western and Eurocentric constructions of gender, female agency and empowerment embedded in popular narratives of universal rights and white mainstream feminist ideals.” Although, such discourses and narratives appear after a target country has been regime changed, therefore demanding a justification and legitimisation for a continued occupation, such as the Afghanistan War (2001-2021). The age old “tried and true” imperial excuse of bringing ‘civilisation’ and ‘democracy’ to a country.

There is clear power hierarchies and cultural connotations apparent in the war justifying discourse used in the branded conflicts of the Arab Spring, such as Libya in 2011, and the failed attempt to use it in Syria later, where the right to intervene has evolved to the responsibility to intervene. “This idea not only has heavy moral connotations but also imperial ones. It builds upon John Stuart Mill’s exception to the rule of non-intervention for non-civilised states” (Xypolia, 2022, p. 17). Thus, the projected notion of an altruistic civilising mission of the West, assisting ‘barbarians’ (oppressed beings and worthy of assistance) that seek to progress their civilizational development, and removing ‘savages’ (tyrannical and illegitimate) that hold them back, is part of this discourse. Additionally, the accusation of the use of sexualised violence in war has been used to fit political agendas and to legitimise military intervention, which is sensationalised and highlighted above other brutality encountered in warfare. Three discursive legitimisation strategies leading to the acceptance of political policy in the form leading up to military action are: 1) “[C]onstructing sexualised violence as a weapon of war, 2) condemning concrete architects of the

machinery (typically one of the conflicting parties), and 3) proclaiming the urgent need to protect the victims through all necessary means” (Kolmasova & Krulisova, 2019, p. 147). This provides the façade and pretence that any military action for geopolitical gains is obscured through the discourse of altruism by the pursuit of ‘humanitarian’ objectives based upon cultural and value-laden reasoning and appeals.

Discourse and narrative-based justification and legitimacy appeals can and do evolve with circumstances and events in the operational and political environments. For example, the 20 year long conflict in Afghanistan. The strategic narrative began as a defensive imperative and necessity born from the insecurity emanating from the 9/11 attacks. “For nearly two decades, U.S. foreign policy was locked in the iron cage of a robust liberal narrative, centred around anti-terrorism.” This evolved in to “From 2018 to 2021, the severity/frequency of terrorist attacks declined significantly, the anti-terrorism narrative weakened, audience costs of inaction declined, and costs of action rose” (Walldorf, Jr, 2022, p. 136). This came at a convenient geopolitical moment when geopolitical crises in Europe and the Indo-Pacific were emerging that required additional military resources. A shifting of military assets to meet what were considered more serious and pressing security challenges.

Analysis and Discussion

Politicised symbolic language is used to engineer and construct deceptive language, which tends to make use of a number of the discursive strategies identified by Reyes (2011): emotions, hypothetical future, project altruism, ‘expert’ testimony, and logic (including false logic). In this respect, those creating and communicating the war justifying discourse seek to engineer a sense of deceived fear or comfort, a means of priming and mobilising apathetic audiences in democracies. Both Lippmann (2005) and Edelman (2001) note the cognitive challenge in overcoming that apathy. Even though warfare is barbaric, it is also being normalised, which potentially makes it more socially and politically acceptable to Western audiences. The moral reasoning seeks to create a perception of urgent threat to a “foreign people” that can ultimately influence “home audiences” to accept a firm and urgent response such as the use of military force. This is intended to create a sense of comfort in audiences, by helping a ‘like-minded’ foreign people and averting the sense of fear through removing a hypothetical future threat. Hence, morality and ethics are artificially introduced into an activity that lacks ethical or moral pursuits.

The hypothetical future discursive strategy is noted by Roberts (2017), where the belief in information itself is key to victory. There is a great deal of agreement among authors (such as Western, 2005; Thussu & Freedman, 2003; Bennett, Lawrence & Livingstone, 2007) who suggest that in Western war justifying discourse, the role of the information war is to support and enable the physical (kinetic) war through engineering consent through projecting the sense of moral and political legitimacy for war. Those wars tend to be communicated, not as wars of choice, but as a ‘moral’ duty forced upon them (altruism) to protect human life

and values. This is representative of Orwell's newspeak as an instrumental aspect in the efforts to prime and mobilise audiences to emotionally accept engaging in a course of risky policy action (warfare). Though the risky action is communicated as being necessary ethically and for the sake of human security (including at 'home').

There is an observable evolution in war justifying discourse and narratives underway, which is in turn a reflection on the evolving political, economic, social and military circumstances of the West. The days of asymmetric warfare, where a far superior Western military goes to war against a far weaker military opponent from the days of the Global War On Terrorism and the Endless Wars are declining. The West is in relative decline, increasing political and economic pressure on the prioritisation of resources and where public sentiment is hardening against more war. The final part of the evolution puzzle is relating to the military opponents, who are better equipped militarily and politically motivated, such as Russia (2022), Iran (2025) and China. This makes war a much riskier and harder to justify proposition. Hence the decline of humanitarian wars and the rise of geopolitical wars of interests and resources, rather than values and human rights. Even if the familiar priming mechanism of the us versus them has a familiar ring.

Conclusion

The research question posed in the opening of this paper asks, what are the primary communicated values and outcomes in Western warfare justifying discourse, and how are these discourses motivated? The five discursive strategies identified by Reyes (2011, p. 781) are absolutely relevant and connected to the Western approach to employing war justifying discourse. This has formed the basis for informing the final section of this paper in delineating the discursive strategies in different Western-led wars of the 21st century. These did not prove to be neatly delineated and operating as separate or individual discursive strategies but can be observed as operating in a complementary manner to each other to boost the persuasive capability and capacity of mainstream mass media outlets.

Edelman noted (2001, p. 10) "the willingness of the general public to vest power in a small group of people by accepting their right to rule and by obeying their laws and orders, even when these are contrary to the interests of the great majority." This is particularly evident in the Western world's war justifying discourse, especially in the sphere of 'humanitarian' wars or regime change. Here, Reyes's (2011) discursive strategies meet with Orwell's warnings on the use of deceptive language, such as double speak or newspeak. There is also an attempt to cognitively project the sense of being forced into wars and 'making mistakes', when the wars being justified by discourse are intentional and deliberate. Therefore, Western war justifying discourse is rooted in the universalisation, simplification and reiteration of interests, culture, emotions, morality and information space dominance.

Contemporary mass mediated political persuasion is made problematic and difficult for the communicator, which is due in no small part to national and international audiences

being heterogeneous in the understanding and interpretation (hence conclusions and reactions) to the emotional and symbolic codes used in the discourse. Therefore, the influence and effects of cultural production on emotional meaning is a variable that needs to be taken into account and understood when an actor uses war justifying discourse. Cultural codes are a communicational mechanism that are used to guide and prime particular ways of thinking, and emotional codes are used as a means of generating particular ways of feeling among audiences.

From the literature review it can be deduced that information operations and influence activity is intended to play a supporting role to enacting selective military operations by powerful brokers rather than the intended façade of an honest broker. Interestingly, the current Western war justifying discourse concerning the Russian Ukrainian War has evolved away from the familiar patterns and templates described above. Even before the war began the emphasis was on protecting Ukraine's national sovereignty and right to choose its own path, which fits with the strategies of hypothetical future, rationality, expert voices and altruism. However, as the war has progressed, there have been an increasing number of public statements from high-ranking Western officials that rely on emotions, hypothetical future (negative) and introducing self-interest. For example, referring to the conflict as a proxy war against Russia and using the war to weaken and degrade Russia's power and influence (Simons, 2022). The justification for this course of indirect military conflict is the necessity to contain Russia. Although, having stated this point, it should and must be noted that this is a very tentative observation as the Ukraine conflict is still ongoing and transforming.

There are limitations to this paper that need to be made transparent and acknowledged. First and foremost, the sample is relatively small owing to the manual nature of the process, which has implications and namely that the result is indicative in nature. Secondly, where the sample was drawn from has unseen constraints and especially the effects of algorithms on generating the sample and available works, including the ranking of the total sample. If another means and source was used for collecting the sample, a different result could be delivered. A third point that needs to be mentioned is that the sample are largely drawn from Western-based or oriented academic journals, which adds another potential bias, ideological and/or perceptual.

The power of discourse has become an important aspect of security analysis (Balzacq, 2005), and foreign policy is often legitimated through the mobilization of rhetoric (Goddard & Krebs, 2015). Rhetorical contestation shapes the boundaries of acceptable policy and the very possibilities of state action (Krebs & Jackson, 2007). U.S. foreign policy toward China is no exception. Recent research indicates that U.S. national discourse has increasingly constructed China as a potential enemy Other (Ooi & D'Arcangelis, 2017), and the mobilization of rhetorical resources from adjacent policy debates has become key to U.S. elites' representation of China as an ideological threat (Winkler & Jerdén, 2023). Since Donald Trump became U.S. president in 2018, the U.S.-China relationship has entered a steady downward spiral marked by trade disputes, technological rivalry, geopolitical maneuvering, and human rights concerns. Bilateral relations between the two major powers

have deteriorated to such an extent that both sides have begun discussing how to avoid an armed conflict (CSIS, 2024). In terms of rhetoric, Trump employed pure advocacy, exemplified by his call to “decouple” from China (Reuters, 2020). However, the subsequent Biden administration has adopted a different policy approach, and accordingly, the U.S. government's rhetoric on China has shifted as well, moving from pure advocacy to a more nuanced and mixed representation that emphasizes both competition and cooperation (The Economist, 2020).

In this research, using rhetorical criticism and approaching the topic from the perspective of strategic communication, I examine the Biden administration's rhetoric on its China policy. Specifically, I explore how the administration has rhetorically shifted from Trump's approach of pure advocacy to redefine U.S.-China relations as a balance of competition and cooperation. I also analyze how the administration addresses the inherent tension within a policy mix that includes cooperation, contest, competition, confrontation, and the avoidance of conflict. This analysis aims to uncover the underlying motives behind the Biden administration's China policy rhetoric: Is it merely a façade for a policy of pure advocacy, or does it reflect a genuine shift in U.S. strategic intent? The theoretical framework guiding this research is the Mixed-Motive Model of public relations, and the rhetorical criticism method employed is Burke's (1969) cluster analysis. This study contributes to the strategic communication literature by testing one of its most influential theoretical models, the Excellence Theory's Mixed-Motive Model, through rhetorical criticism.

Biographical note

Professor Greg Simons has a PhD from the University of Canterbury in New Zealand and has an Associate Professor title awarded in Political Science from Uppsala University in Sweden. He is a Professor at the Department of Journalism, Media and Communication at Daffodil International University in Dhaka, Bangladesh. His research interests include the changing political dynamics and relationships, mass media, public diplomacy, political marketing, crisis management communications, media and armed conflict, and the Russian Orthodox Church. He also researches the relationships and connections between information, politics and armed conflict more broadly, such as the GWOT and Arab Spring.

References

- Abdu, K. S. T., & Khafaga, A. F. (2019). A critical discourse analysis of mind control strategies in George Orwell's *Nineteen Eighty-Four*. *International Journal of English Linguistics*, 9(6), 421–430. <https://doi.org/10.17645/pag.10044>
- Alexandrescu, M. (2025). The wartime justification trajectory: A dynamic approach to justifying wars in the 21st century. *Politics and Governance*, 13.
- Aluya, I. (2025). Language in Conflict Situation: Investigating Discursive Strategies in President Joe Biden's Speech on the Russian-Ukraine War. *Language in India*, 25(2).

- Alyahya, A. F. (2023). Critical discourse analysis of newspaper articles declaring the outbreak of war in Ukraine: The *Washington Post* vs. *The Moscow Times*. *International Journal of Society, Culture & Language*, 11(2), 47–59.
- Baskin, O., Aronoff, C., & Lattimore, D. (1997). *Public relations: The profession and the practice* (4th ed.). Brown & Benchmark.
- Bennett, W. L., Lawrence, R. G., & Livingstone, S. (2007). *When the press fails: Political power and the news media from Iraq to Katrina*. University of Chicago Press.
- Blidaru, S. (2018). Zero-sum games and mixed-motive games in the fight of the Russian Federation with the West for Ukraine. *Europolicy: Continuity and Change in European Governance*, 12(2), 37–66. <https://doi.org/10.25019/europolicy.2018.12.2.02>
- Boin, A., t'Hart, P., Stern, E., & Sundelius, B. (2005). *The politics of crisis management: Public leadership under pressure*. Cambridge University Press.
- Brockmeier, S., Stuenkel, O., & Tourinho, M. (2016). The impact of the Libya intervention debates on norms of protection. *Global Strategy*, 30(1), 113–133. <https://doi.org/10.1080/13600826.2015.1094029>
- Chang, G. C. (2008). Why we must attack Iraq: Bush's reasoning practices and argumentation system. *Discourse & Society*, 19(4), 453–482. <https://doi.org/10.1177/0957926508089939>
- Chifu, I., & Simons, G. (2023). *Rethinking warfare in the 21st century: The influence and effects of the politics, information and communication mix*. Cambridge University Press.
- von Clausewitz, C. (1982). *On war*. Penguin Classics.
- Coker, C. (2004). *Is there a Western way of warfare?* (IFS Info No. 1/2004). Norwegian Institute for Defence Studies.
- Conz, E., & Magnami, G. (2020). A dynamic perspective on the resilience of firms: A systematic literature review and a framework for future research. *European Management Journal*, 38, 400–412. <https://doi.org/10.1016/j.emj.2019.12.004>
- Demasi, M. A. (2023). Accountability in the Russo-Ukrainian war: Vladimir Putin versus NATO. *Peace and Conflict: Journal of Peace Psychology*, 29(3), 257–265. <https://doi.org/10.1037/pac0000653>
- DiMaggio, A. R. (2009). *Mass media, mass propaganda: Examining American news in the "War on Terror"*. Lexington Books.
- Dunmire, P. L. (2005). Pre-empting the future: Rhetoric and ideology of the future in political discourse. *Discourse & Society*, 16(4), 481–513. <https://doi.org/10.1177/0957926505053052>
- Edelman, M. (2001). *The politics of misinformation*. Cambridge University Press.
- Fisher, D., & Wicker, B. (Eds.). (2010). *Just war on terror? A Christian and Muslim response*. Ashgate.
- Forbes, I., & Hoffman, M. (Eds.). (1993). *Political theory, international relations and the ethics of intervention*. Palgrave Macmillan.
- Forchtner, B. (2021). Introducing 'narrative in critical discourse studies'. *Critical Discourse Studies*, 18(3), 304–313. <https://doi.org/10.1080/17405904.2020.1802765>

- Greene, R. (2007). *The 33 strategies of war*. Penguin Books.
- Haqpana, S., & Tsouroufli, M. (2023). 'Powerless, poor and needy?': Reproducing colonial discourses of gender and Muslim women through educational interventions by I-NGOs in Afghanistan. *Women's Studies International Forum*, 98, 102714. <https://doi.org/10.1016/j.wsif.2023.102714>
- Hashmi, T. (2014). *Global jihad and America: The hundred-year war beyond Iraq and Afghanistan*. SAGE.
- Hoffmann, R. J. (Ed.). (2006). *The just war and jihad: Violence in Judaism, Christianity, & Islam*. Prometheus Books.
- Hossain, M. (2017). Language as the device for psychological manipulation in George Orwell's Nineteen Eighty-Four: a psycholinguistic analysis. *European Journal of English Language and Linguistics Research*, 5(8), 25-31.
- Johnson, P. (2021). The Role of Implicated Meaning in George Orwell's Nineteen Eighty-Four: Language, Thought & Power. *The English Languages: History, Diaspora, Culture*, 7, 22-39.
- Kassim, Y. R. (2014). *The geopolitics of intervention: Asia and the responsibility to protect*. Springer.
- Kassimeris, G., & Buckley, J. (Eds.). (2010). *The Ashgate research companion to modern warfare*. Ashgate.
- Kolmasova, K., & Krulisova, K. (2019). Legitimizing military action through "rape-as-a-weapon" discourse in Libya: Critical feminist analysis. *Politics & Gender*, 15, 130–150. <https://doi.org/10.1017/S1743923X18000326>
- Krickovic, A., & Sakwa, R. (2022). War in Ukraine: The clash of norms and ontologies. *Journal of Military and Strategic Studies*, 22(2).
- Lippmann, W. (2005). *A preface to politics*. Prometheus Books.
- Loseke, D. R. (2009). Examining emotion as discourse: Emotion codes and presidential speeches justifying war. *The Sociological Quarterly*, 50(3), 497–524. <https://doi.org/10.1111/j.1533-8525.2009.01150.x>
- Luttwak, E. N. (2001). *Strategy: The logic of war and peace*. Belknap Press of Harvard University Press.
- McKenna, B. (2004). Critical discourse studies: Where to from here? *Critical Discourse Studies*, 1(1), 9–39. <https://doi.org/10.1080/17405900410001674498>
- Machiavelli, N. (2001). *The art of war*. Da Capo Press.
- Malito, D. V. (2019). Morality as a catalyst for violence: Responsibility to protect and regime change in Libya. *Politikon: South African Journal of Political Studies*, 46(1), 104–121. <https://doi.org/10.1080/02589346.2019.1572296>
- Martini, A. (2020). The Syrian wars of words: International and local instrumentalizations of the war on terror. *Third World Quarterly*, 41(4), 725–743. <https://doi.org/10.1080/01436597.2019.1699784>
- Oddo, J. (2011). War legitimization discourse: Representing 'us' and 'them' in four presidential addresses. *Discourse & Society*, 22(3), 287–314. <https://doi.org/10.1177/0957926510395442>

- Oddo, J. (2018). *The discourse of propaganda: Case studies from the Persian Gulf and the War on Terror*. Penn State University Press.
- Pati, D., & Lorusso, L. N. (2018). How to write a systematic review of the literature. *Health Environments Research & Design Journal*, 11(1), 15–30.
- Payne, K. (2005). The media as an instrument of war. *The US Army War College Quarterly: Parameters*, 35(1), 10.
- Reeves-O'Toole, O. (2020). The subjugation of pacifism in UK parliamentary discourse: Analysing the 2015 debate on bombing Syria. *Global Society*, 34(1), 4–20. <https://doi.org/10.1080/13600826.2019.1668362>
- Reyes, A. (2011). Strategies of legitimisation in political discourse: From words to actions. *Discourse & Society*, 22(6), 781–807. <https://doi.org/10.1177/0957926511419927>
- Roberts, P. (2017). Designing conceptual failure in warfare: The misguided path of the West. *The RUSI Journal*, 162(1), 14–23. <https://doi.org/10.1080/03071847.2017.1301512>
- Rochester, J. M. (2016). *The new warfare: Rethinking rules for an unruly world*. Routledge.
- Scahill, J. (2013). *Dirty wars: The world is a battlefield*. Serpent's Tail.
- Schoon, E. W., & Duxbury, S. W. (2019). Robust discourse and the politics of legitimacy: Framing international intervention in the Syrian civil war, 2011–2016. *Sociological Science*, 6, 635–660. <https://doi.org/10.15195/v6.a24>
- Silverman, D. (2020). *Interpreting qualitative data*. SAGE.
- Simons, G. (2016). 'Good' battles and 'bad' battles: A comparative analysis of Western media coverage of the battles of Mosul and Aleppo. *Tractus Aevorum*, 3(2), 114–138.
- Simons, G. (2022, May 12). Sweden and Finland NATO membership and the militarisation of the Baltic Sea. *Politics Today*. <https://politicstoday.org/sweden-finland-nato-membership-baltic-sea-militarization/>
- Slater, I. (1975). Orwell, Marcuse and the language of politics. *Political Studies*, 23(4), 459–474.
- Snyder, H. (2019). Literature review as a research methodology: An overview and guidelines. *Journal of Business Research*, 104, 333–339. <https://doi.org/10.1016/j.jbusres.2019.07.039>
- Sowinska, A. (2013). A critical discourse approach to the analysis of values in political discourse: The example of freedom in President Bush's State of the Union addresses (2001–2008). *Discourse & Society*, 24(6), 792–809. <https://doi.org/10.1177/0957926513486214>
- Temes, P. S. (2003). *The just war: An American reflection on the morality of war in our time*. Ivan R. Dee.
- Thakur, R. (2013). R2P after Libya and Syria: Engaging emerging powers. *The Washington Quarterly*, 36(2), 61–76. <https://doi.org/10.1080/0163660x.2013.791082>
- Thussu, D. K., & Freedman, D. (Eds.). (2003). *War and the media*. SAGE.
- Trotter, W. (2015). *Instincts of the herd in peace and war*. Filiquarian Publishing.
- Walker, S. (2019). *American foreign policy and forced regime change since World War II: Forcing freedom*. Palgrave Macmillan.

- Walldorf, C. W., Jr. (2022). Narratives and war: Explaining the length and end of US military operations in Afghanistan. *International Security*, 47(1), 93–138.
- Walzer, M. (2006). *Just and unjust wars: A moral argument with historical illustrations*. Basic Books.
- Western, J. (2005). *Selling intervention and war: The presidency, the media, and the American public*. Johns Hopkins University Press.
- Xypolia, L. (2022). From white man's burden to responsible saviour: Justifying humanitarian intervention in Libya. *Middle East Critique*, 31(1), 1–19.
<https://doi.org/10.1080/19436149.2022.2030981>
- Zollmann, F. (2017). *Media, propaganda and the politics of intervention*. Peter Lang.

Why risky and amoral ‘transparency’ fails as a lone ethical concept*

Shannon A. Bowen**

University of South Carolina; Global Strategic Communication Consortium

Transparency is a concept frequently arising in ethics discussions but poorly conceptualized and, in some cases, even harmful. Searching outside the term's current general popularity in media ethics and in public relations, a literature review revealed that many scholars from various disciplines have misgivings and concerns about transparency. Other professions are used to discuss the problems that arise when relying on transparency for ethical guidance. Consequentialist, virtue, and deontological ethics all object to the use of transparency as an ethical construct. Many problems emerge, based on a Kantian reading; for example, transparency does not ensure good intention; and, transparency does not equal veracity, as transparency can also be used for unethical purposes. Immanuel Kant did not discuss transparency as an *a priori* (reasoned from cause to effect) moral concept but awarded *prima facie* (on its face) moral worth to *good intention* (the highest form of good), dignity and respect, and duty – in a context of rational moral autonomy. Transparency is not on its own considered a moral construct but rather its use should be accorded to intention. Transparency can be used for either ethical or unethical *intention* - therefore the term transparency is rendered amoral (without moral status on its own) and should not be used as a lone ethical construct but must be combined with stronger concepts for more rigorous moral analyses. Transparency as a professional standard is also explored yet proved problematic across other professions from health care to AI. Ethical analyses, as added to the issues management process, is advised and discussed. The lexicon of strategic communication management should replace transparency with more meaningful terms rooted in moral philosophy and conceptually grounded for not only clearer understanding but also for enhancing organizational responsibility, ethical principle use, and ethical outcomes. Good intention, veracity, full disclosure, visibility, candor, clarity, completeness could aid precision in our ethical analyses. As a lone unqualified good, pure intention should drive veracity, disclosure, dialogue, and so on. These terms are offered in conclusion as a normative and practical alternatives (or amendments) to risky and amoral transparency. Greater acumen in our lexicon can enhance the professionalism of the field and strengthen its analyses by aligning it with the discipline and analytic of moral philosophy, using the well-examined traditions of ethics therein, and stand the field in organizational and ethical leadership with other professions.

Keywords: Veracity, disclosure, intention, respect, issues management

* Article submitted on 26/09/2025. Article accepted on 22/12/2025.

** sbowen@sc.edu

Purpose

Transparency is a concept frequently arising as an ethical standard in discussions of strategic communication academics and professionals. Transparency is discussed in relation to ethics in numerous contexts. But it is assumed that we know what transparency consists of and it is without question or unequivocally an ethical answer to complex questions.

Yet, transparency is often bandied about without further study, elucidation, or elaboration, as in, for example, “This case needs transparency.” Therefore, a rigorous philosophical examination of transparency based on moral philosophy as well as professional and academic use is warranted. This study offers an issues management framework for understanding risk reduction as the forum for examining ethical constructs and analyses such as transparency; as compared with ethical concepts from moral philosophy, specifically Kantian deontology, such as responsibility, duty, dignity and respect, intention, or moral autonomy.

Immanuel Kant was a modern philosopher (1724-1804) who originated the principle-based, non-consequentialist paradigm known as deontology (literally the study of duty). As a moral philosophy, deontology used the approach of virtue ethics and sought to make it universal, or applied equally to every rational being by virtue of reason (as opposed to aristocracy, education, or affluence). In this way, Kantian philosophy uses radical equality to state that all rational beings can use moral discernment to understand their duty to the moral law, to create a universal position that can be applied equally to all people, maintains dignity and respect, and is based on good will (as opposed to self-interested outcomes). These precepts are what Kant called the categorical imperative, or three moral tests he designed to eradicate bias and ensure equality and pure intention.

Kant accorded the highest moral worth to good intention, a concept that cannot be assailed (even in excess) but results in a universal good (Sullivan, 1989). This universal approach is sometimes known as the law of non-contradiction, as Kant’s first principle of reason, meaning a logical application of the rule would not contradict itself (Guyer, 2019). Kant (1785/1964) conceived several concepts as involving innate moral worth, goodness or ethicality associated with the term and its use but accorded the *highest* moral worth to goodwill or pure moral intention as the only *incorruptible* good that is unqualified in nature (Sullivan, 1989). As an unqualified good, pure intention should drive veracity, disclosure, dialogue, and so on (DeGeorge, 2010). These terms are offered in conclusion as an alternative or amendment to risky and amoral transparency.

Risk, Issues, and Issues Management

No study of risk is complete without a thorough understanding of issues management and how it is applied in a rapidly changing environment. The rapidly changing landscape of

strategic communication includes numerous ethical challenges. Applications of strategic communication are diverse, ranging from leadership or organizational communication to marketing and consumer outreach, to governmental regulation and public sector communication, to artificial intelligence use and regulation, to international trade and finance, to management itself (Bowen & Erzikova, 2025).

These strategic communication areas (among numerous other specialties within strategic communication), all hold a broad range of issues that normally contain numerous ethical components. Issues can be societal trends, shifts in demands or a competitive environment, or even a changing international environment. These issues are problems, regulations, or public policy changes comprised of varying levels of risk, to be managed – known as strategic issues management (Heath, 2025).¹ [Author note: Risk management is a financial tool not to be confused with the issues management of problem solving]. Issues management was conceived of as a strategy for managing public policy issues in response to increasing activist pressure, increased regulatory scrutiny, and environmental turbulence. Issues management was conceived of in the 1970s but it is, by its very nature, futuristic. This futurist perspective has always been a part of issues management: one of the canonical works by Chase (1984) is named *Origins of the Future*.

Futurism is demonstrated in one classic definition of issues management offered by Coates, Coates, Jarrat, and Heinz (1986):

Issues management is the organized activity of identifying emerging trends, concerns, or issues likely to affect an organization in the next few years and developing a wider and more positive range of organizational responses toward that future. Business and industry, in adopting issues management, seek to formulate creative alternatives to constraints, regulations, or confrontation (p. ix).

These scholars observed something key: An increased risk level necessitates research and a strategic response of organizations, considering the future and creative problem solving. Issues management is futuristic in that it allows time to research and create strategy options in a proactive or interactive sense. This future-minded approach to strategy lessens reaction time, offers management numerous alternatives, enhances autonomy, and allows an organization to increase efficiency.

Organizations not using issues management futuristically lose efficiency and the opportunity to create proactive strategy and are often caught in a reactive or accommodative loop responding to changes (Heath & Palenchar, 2009), rather than helping to create change proactively (Bowen et al., 2024) and build interactive strategy (Buchholz et al., 1994). Public affairs and policy creation is an inextricable part of issues management strategy (Heath, 1997; Lerbinger, 2006) which allows organizations to change and react to their environment with strategic intelligence (Bowen & Heath, 2020) rather than a vague and unrealistic goal of mutual benefit which is often impossible in a heavily regulated industry.

Missing from the classic Coates, Coates, Jarrat, and Heinz (1986) definition of issues management offered above is the concept of identifying and studying ethical dilemmas. Although not stated explicitly, ethical problems would constitute issues, and most issues

management problems also contain an element of ethical responsibility. These scholars, no doubt, viewed ethics as one part of the issues management equation, though the hindsight of observing numerous scandals would conclude that ethics should be an explicit part of issues management to avoid oversight and worse: corruption. Infamous cases of unethical behavior have seen organizations lay claim to transparency as a defense. For example, though large amounts of disclosure, problematic data can be hidden or obscured. One such example can be seen in Pacific Gas and Electric's cover-up of hexavalent chromium as a simple "rust inhibitor" yet its presence in the water supply led to birth defects and cancer. Avoiding using transparency to obscure negative information may be more problematic than perceived if negative information is not shared with communication professionals (Berger, 2005; Berger & Reber, 2006).

There are many reasons to attend to avoiding ethical complications and using ethical issues management to build relationships based on trust and respect. Ethical issues management can add to positive organizational outcomes, such as overall efficiency among a workforce (Karnes, 2009) and commitment among consumers (Uzunoğlu et al., 2017), attributed to organizations who do identify and address ethical concerns proactively.

Adding Ethics to Issues Management

Ethical analyses were first suggested as an explicit concern in the issues management process by Bowen (2000) who suggested a duty-based framework to guide issue decision-making based highly on Kantian deontological ethics. This Kantian framework of ethical issues management (Bowen, 2005) included a section asking questions to ensure moral autonomy and objective rationality, or freedom from bias, and an examination of organizational policies. The Kantian framework included the three forms of Kant's categorical imperative (a 3-part universal moral test of duty, dignity and respect, and good intention) to examine the dilemma with stakeholders, publics, the self, organization and governing elements, as well as society as a whole. Applying deontological ethics to issues management was tested and confirmed as the predominant useful paradigm in two global organizations (Bowen, 2002).

The issues management process was summarized in six steps by scholars (Buchholz et al., 1994) seeking to make the process useful and meaningful in management practice. Ethics were added explicitly as an analysis component in step three by Bowen, Rawlins, and Martin (Bowen et al., 2019) and again expanded upon in more detail by (Bowen et al., 2024). These scholars (Bowen et al., 2024) suggested that ethical analyses play an explicit role in at least three of the steps of the issues management process:

1. Identify issues and trends. (incl. ethical responsibilities)
2. Evaluate issue impact, *ethical values*, and set priorities.
3. Conduct research and analyses, *including ethical analyses*.
4. Develop *ethical* strategy and strategic options.

5. Implement strategy.

6. Evaluate strategy. (p. 288, emphases original)

These scholars did not mention transparency, or offer the specific ethical tests to be used as a step in this model, but discussed moral autonomy in the publication as a transition to a codified ethical perspective, with specific means of analysis to follow. Bowen, Lovari, and Timmermans (Bowen et al., 2024) focused upon institutionalizing ethics as a routine component of the issues management process for managers and scholars in this field. Explicitly including ethics in several steps of the issues management process, it is not an afterthought but an overt part of the research and analyses of any problem.

Linking issues management to ethics also necessitates a macro-perspective on the role of that function not only within organizations, but also within society. Heath argued that society exists for the collective management of risk, necessitating issues management (Heath, 2001), or the enacting of strategic issues in the public square along with appropriate policymaking by organizations (Heath & Nelson, 1986). Scholars have also noted that civil society enactment (Taylor, 2010) and a normative role for ethics in the civil discourse of problem resolution (Bowen, 2010) strengthen a commitment to ethical analyses not only for an organization to do well, but also for it to do good.

As a way for strategic communicators to manage and reduce risk as well as maintain responsibility, issues management offers a framework that can inculcate research, ethical decision-making, and strategy. Egregious ethical scandals such as those at Volkswagen or Wells Fargo, among numerous others, offer ample rationale for including ethics as a codified part of the issues management process. As strategic communication grows in complexity and importance, as part of the public policy and public affairs process on numerous issues from human rights to artificial intelligence, the gravity of ethical analysis weighs heavy on the shoulders of communication professionals. An uncertain future, rife with risk and conflict, ranging vastly from natural disasters to warfare, require acumen in ethical analyses (Bowen, 2025; Bowen & Erzikova, 2025).

But how does transparency relate to issues management analysis? Is the concept “transparency” strong enough to warrant its frequent recommendation and use as an ethical standard? Imprecise standards, varied meaning, and differing usage render the term transparency vague and suspect as an ethical basis for strategic communication.

A closer examination of the concept of transparency should elucidate whether transparency is warranted to include in numerous ethics models and discussions offered by strategic communication scholars and as a basis for the analyses required in strategic issues management. These questions haunt ethics in the strategic communication discipline and serious study of the often-used term transparency is needed. Thus, the research questions driving this study are:

RQ. 1. Does transparency succeed or invite risk when used as a basis for ethics?

RQ. 2. Do other professions successfully employ transparency as an ethical construct?

RQ. 3. Are there superior ethical standards than transparency?

Parameters of this Review

Many disciplines have much to say on ethics and moral decision making: Moral philosophy, behavioral ethics, cognitive psychology, organizational psychology, theology, developmental psychology, management, and business ethics, to name a few. Of these, moral philosophy is the oldest and most direct form of reasoning related to ethics, based in philosophy or the love of knowledge. Within moral philosophy, metaethics (meaning), normative ethics (principle), and applied ethics are all different pursuits within the discipline, with practical pursuits and the professions dominating applied ethics (e.g., medical ethics, legal ethics, and so on). Thoroughly reviewing the literature of numerous fields is an impossibility in the space of one article, so this piece focuses on the intersection of issues management and strategic communication with transparency.

Due to the common use of the term transparency but its relatively underexplored parameters, a two-step exploratory literature review of numerous fields was used. In the first phase, a search of the literature for “transparency” in Academic Search Complete, Web of Science, and Scopus library databases was conducted (search terms included: transparent, transparency). Articles with transparency in the title or keywords were identified, downloaded, and searched for relevance. Those with a significant *focus* on transparency ($n = 56$) were included in the review for this study while those using transparency as a descriptor or tangential term were excluded from this study. Although this process is a qualitative one, it was selective in searching for articles containing a definition of transparency, a discussion of or justification for the term, an analysis or a critique of the use of transparency in a philosophical or professional context, or an explication of when the concept was relevant or not useful. In phase two of this research, a number of books (about 50) and articles (about 30) in moral philosophy were consulted, searching for the term transparency or related concepts such as obfuscation, veracity, truth, and untruth, visible, opaque, and opacity as expanded search terms to the lack of the term transparency in philosophical literature and indices. Those discussions were used in building the arguments and subsections of this paper. A small number of strategic communication, public relations, and media ethics sources were also used to compare and contrast different perspectives on and uses of transparency, to create a dialectic with expected usage versus philosophical usage.

As part of this interdisciplinary approach, other professions offer comparative examples in their literature to illustrate the application of transparency in conceptual and real world situations. Creation of case examples from news stories and legal filings was used as illustration. To expand communication literature, the very nature of this study is interdisciplinary, though it is based in moral philosophy. In conclusion, philosophical conceptions from virtue and deontological traditions were brought to bear on the research questions, with resulting practical recommendations.

Defining Transparency

Transparency is defined by Heald (2006) as the principle of enabling the public to gain information about the operations and structures of a given organization. This form of transparency has become a goal in and of itself in recent times, or one that is encouraged as enabling the market to self-correct, through consumers with knowledge and access to transparent information. In this manner, economic transparency is presumed to foster the “invisible hand” that will allow self-correction in *laissez faire* capitalism without a need for government regulation. In media ethics and public relations, the term is rarely defined but used colloquially and assumed to mean visibility.

Media ethics scholar Plaisance (2007) discussed transparency as an ethical recommendation and with a positive regard for the term rather than seeing it as one of amoral status depending on usage. Plaisance (2007) read transparency alone as what he termed a ‘gift’ of veracity from Kant, though one can argue that veracity and transparency are not synonymous. The difference may be one of what discipline is involved. Transparency and veracity are not equivalent in moral philosophy. Veracity is indeed a moral/philosophical principle, from which springs truthfulness; transparency stands amoral (without moral status in and of itself) as its ethical stance is determined by usage.

Kant himself distinguished between truthfulness, lying, social convention, discernment, privacy, deceit, and untruthfulness (Guyer, 2019; Price, 2003). The term veracity is embraced in virtue ethics (Aristotle 1910/322BCE; Singer, 1994), where one does not find transparency, but focus on the principle of veracity as means of defining truthfulness or honesty. Many communication and media ethics books are silent on the topic of transparency (Patterson & Wilkins, 2002; Parsons, 2004; Knowlton & Parsons, 1994; Christians et al., 2001). Other communication scholars (Oliveira & Goncalves, 2024; Pearson, 1989) regard transparency as one ethical consideration among many that can be used to support responsible organizational commitments in society.

Also positively explained, political philosopher Etzioni (2010) pointed out that the *Wall Street Journal* has couched transparency as the antidote to burdensome government regulation. Even in this strong conception of transparency conceived in economic theory, there is a problem with the conception in that its use is assumed to be good or ethical. The risk with this assumption is that transparency cannot assure a level playing field of fairness or that it is used with only good intention – unethical transparency exists and is used on a routine basis, as explained herein. The problem of not having fuller accurate information access also impedes using transparency as a stand-alone ethical construct. As Kronewald argued, “The concept of transparency is contingent upon the availability of clear and unambiguous information about the originators of communication or messages” (p. 49).

Arguing in favor of transparency, Plaisance (2007) wrote, “Even if transparency is not always a sufficient condition for more ethical behavior, its absence is a prerequisite for deception” (p. 193). But, this is not the case in the professional world of public relations and strategic communication, whose practitioners state in interviews (Bowen, 2008) that

transparency is used as a form of deception. So, not only is its presence no assurance of ethical behavior, it is often used as an excuse, obfuscation, or “window dressing similar to greenwashing without intent to have true environmental impact.

Further, transparency is often mandated by government rather than enforced due to sociological or communitarian social norms and controls. Making the operations of an organization visible (Edgett, 2002) or discoverable is an admirable starting position, but does nothing to ensure that the inner workings of that organization, including priorities, management, or disclosures, are ethical. In fact, transparency is more likely unethical than ethical because it places the burden of surveillance, analysis and discovery on the observer – those outside the organization, rather than inside it and in authority over its decisions. This displacement of responsibility can lead to a moral abrogation of organizational ethics, such as in a prior study’s communication officer’s cynical statement, “We knew no one would find it, but we put it out there so they couldn’t say we hid it” (Bowen et al., 2016). This type of unethical case of ‘transparency’ constitutes what Fleming (2025) called, “abuses of knowledge in the public realm” (p. 187).

It is for these reasons that in moral philosophy, transparency is not routinely discussed as an ethical construct, as the term itself is imprecise and its usage based on other ethical constructs such as good or bad intent. Analytical philosophy reveals far too many problems with the concept of transparency for it to pass the bar of being included in an ethical analysis, as it is *prima facie* amoral. To be amoral means that transparency’s ethical status is determined by the user rather than by the concept as one holding innate goodness. Transparency is similar to a tool such as a stone cutter’s saw. This saw can create great works of art when used with good intention, or be used to fell a structure on unsuspecting occupants when used with unethical intention. The intention drives the tool and hence defines its moral status.

In one case example, in a critique of the pharmaceutical industry’s drug marketing and clinical trial database practices, Falit (2005) wondered if increased transparency was actually worth anything, given the rollout problems explained by four pharmaceutical associations, or if it were simply “contrived public appeasement” (p. 391) illustrating that the intention behind transparency is the important variable. This example alone, if not even for the support of thousands of others, illustrates why transparency holds no innate moral worth of its own as a standard for ethical decision making or analyses.

Transparency in Extra-Disciplinary Literature

Searching outside the term’s current general popularity in media ethics and in public relations, a literature review revealed that many scholars from various disciplines have misgivings and concerns about transparency. The potential for artificial intelligence (AI) to reshape entire industries and nation-states is well-discussed in the literature as are calls for algorithmic transparency. Therefore, this area provides an especially rich discussion of

transparency as well as problematics in its application. Klenk (2023) explained that algorithmic transparency is a common goal but can also be manipulative. Klenk (2023) argued that manipulative transparency may have numerous motives other than revealing reasoning to users such as regulatory compliance or impression management necessitating an urgent need for a critical perspective on transparency.

Philosophers of technology (Wang, 2022, 2023; Zerilli et al., 2019), likewise argued that humanity needs to worry about manipulative transparency because the social and power related implications of algorithmic transparency offer it motive for deception toward observers. Current public policy standards for requiring transparency were assessed (Felzmann et al., 2019) to be not rigorous enough to generate actual transparency of the algorithm better understood as relational transparency or explainability.

Licon (2024) concluded, “transparency can be morally and politically corrupting.”

In the realm of politics and economics, transparency is said to have the potential for corrupting decisions because it applies reputational concerns onto purely moral judgments. Hypocrisy, half-truths, and deception are the results (Licon, 2024). Although Licon (2024) cites evidence showing the deontological judgments are preferred above the consequential and reputational, the author argued that inauthentic moral and political judgments are often undertaken for show - counter to personal moral convictions, increasing a political cost for popularity, concluding that “transparency corrupts” (p. 41).

Paul (2015) also described using transparency in conjunction with intention is not a workable ethical construct. She explained that “the Transparency approach to explaining knowledge of our intentions fails” (p. 1529) because it can introduce schematic bias. Paul (2015) argued that we know of our intentions by theorizing about ourselves but these decisions are often first-person privileged. Counter, Boyle, drawing on Sartre’s non-positional self consciousness, posited transparency does not exist in the mind because knowledge of “the world constitutively involves a kind of implicit self awareness” (p. 1014).

Noting its ambiguity, while confusing the self-evident good of a concept with the intention supporting its use, Metzen (2025) argued that there are three types of transparency: “mere transparency, publicity, and accountability” (p. 1) that could redeem the vague concept. Noting that philosophers have broadly discussed the danger of the transparency concept, Metzen (2025) tried unconvincingly to support it with other concepts such as regulation, disclosure, and trust – nonetheless belying the weakness of the original concept noting numerous ways that it “can backfire” (p. 10).

According to Mitchell (2025), trust is not congruent with transparency because transparency requires the truster to constantly check and verify information, implying that a lack of trust exists. If trust is a result of transparency, what would be the motivation to continue to use and monitor transparency for either party? The resulting tautology once again illustrates the conceptual frailty of transparency as an ethical concept.

Although the concept of transparency was not apparent in the issues management literature, it does appear more commonly in the general lexicon of strategic communication and public relations. It is a particularly prevalent part of the codes of ethics of some professional associations, yet is still ill defined and used with great assumption as to its

meaning, use, and role as an ethical standard. A study (Bowen, 2016) regarding the use and precision of ethical terms in the field found transparency problematic due to public relations professionals using transparency to abrogate their moral responsibility of disclosure. For example, one participant of the prior study (Bowen, 2016) said, “as long as it was put out there and we have been transparent then it’s up to them.” Another public relations participant in an ethics study (Bowen et al., 2006) reported, “We are transparent with so much information they will never find the thing we are worried about. No one would.”

Other critiques included that transparency inspired an “audit culture” rather than one of critical moral judgement. Bowen (2016) recommended replacing the term transparency with honest disclosure, contextual disclosure, or veracity. In answering RQ 1, the conclusion, based upon the literature discussed above, is that transparency poses a risk for organizations or professionals attempting to employ it as an ethical construct and it fails to be satisfactory as a standard for ethics. The amoral status of the transparency concept renders it useless in an ethical framework, even when supported by more powerful concepts, as it would introduce an amoral element counter to ethics and undermine ethical decision making.

For issues management and strategic communication, transparency poses great risk through its amoral status and is not acceptable for use as a term related to ethics or ethical analyses. Strategic communication has long argued for its status as a profession and for a seat within the dominant coalition, so it is bound to act as professions act with regard to high and knowable ethical standards (Berger & Reber, 2006; Wright & Turk, 2007). However, because of the term transparency’s prevalence among the professions, such as engineering, medicine, or finance, one wonders if transparency is something useful in those limited contexts?

Exemption for Professional Services?

In professional services such as accounting, finance, engineering, law, or medical care, transparency may indeed prove valuable. The potential for collusion, corruption, and usuary (e.g., data exploitation or similar) are rife when those in positions of power hold access to information and data that others (in less exclusive positions, i.e., clients) do not. Exposure, disclosure, and clarity of standards and processes, as well as protections, are needed when handling sensitive and private information (Bowen, 2024). A case study analysis in this area may prove helpful.

A recent case in the United States elucidates the need for disclosure and ethical standards within professional service communities. In this case, the University of Alabama – Birmingham was recently found to be using (or “stealing”) the body parts of penitentiary inmates without permission or authorization, much to the shock and horror of inmates’ families, who filed subsequent lawsuits alleging there was no informed consent (Hrynkiw, 2024). How was this procedure allowed to happen as part of a university teaching hospital

with an institutional review board and medical ethics panel (Wood & Gauntt, 2024)? What were the standards of “transparency” and full disclosure regarding body parts? Why had no one on the ethics panel or institutional review board determined that inmates themselves, or family members, must agree to organ donation -- rather than using secretive organ harvesting? The exploitation, loss of rights, will, and body determination in this case offer a staggering glimpse into what happens when a system is in charge of “transparency” as opposed to valuing the moral determination of individual autonomy.

The hospital defended itself by using transparency -- mentioning a policy that body parts are not returned after an autopsy, concluding that everyone should know that was standard procedure (Hrynkiw, 2024). How subjects should know that information was not specified and the hospital declined to comment on the case (Wood & Gauntt, 2024). In this case, a legalistic approach had overridden all ethical concerns but was couched in terms of transparency as if families were familiar with fine print in voluminous forms related to the potential future use of inmates’ body parts.

Another shocking factor of the University of Alabama – Birmingham’s case is that though these organ harvesting operations after autopsy without permission happened at the state’s largest hospital, the news media has been relatively silent on the case. Silence may be viewed as complicity at worst (Bok, 1983), or viewed at best as a complete failure of the fourth estate “watchdog” function of the news media (Theus, 1993). The news value alone of organ harvesting should have placed the issue high on the media agenda and it should have been framed as a case of exploitation of rights of the weak, disadvantaged, or underserved.

The case example of the University of Alabama – Birmingham’s missing body parts also shows us why transparency is amoral and fails as a guiding ethical construct. Although legal cases and arguments are still underway, it is possible, that state officials had some legal right to body parts after an autopsy of those in custody. However, facts and chain of ownership are unclear and disputed, and this custody would conflict with the ideas of human rights and dignity strongly represented in Kantian ethics. This is what the hospital alleged but is unclear; it could be something that families even acknowledged - at some point – unknowingly, due to the arcane and voluminous amount of legal forms, documentation, disclosures and other information provided to them. If this hypothetical proves true, then technically the state had transparency; but, in actuality, it violated every standard of normative ethics, including Kantian dignity and respect, as the humanity form of the categorical imperative.

In summary, transparency can be effectively used to obscure important facts and information by its sheer volume or intent. This use of transparency to obscure is common, yet patently unethical in this case because it violates the ethical tradition of deontological respect and dignity (Place et al., 2021), as well as good moral intention. It also violates the responsibility of duty to *legitimately* disclose pertinent information rather than conceal it while disclosing it along with volumes of other difficult and often less than relevant information. Without discussion, one cannot be certain that inmates or relatives understood the “practice”

of organ harvesting, especially with differing language, hearing, and reading abilities present in prison populations (Whitehead, 1985).

This case may seem exceptional, but public relations professionals have previously acknowledged using transparency to “hide” unethical actions or problematic facts (Bowen, 2008). Legitimate disclosure is not transparency, but full and contextual understanding and discussion that offers a feedback loop of symmetrical communication including opportunities to ask questions, study the problem, listen, and engage in dialogue to ensure full understanding (Place & Flamme, 2024). Transparency is thus not a moral concept but a legalistic one that fails the test of providing veracity or ethical valence.

Professional services often require an element of confidentiality that can, unfortunately, also lead to concealment, secrecy, exploitation of this access, and wrongdoing. No doubt, financial incentives for ethical misdeeds or working ‘outside of the rules’ are indeed compelling. So strong is the need for “daylight” in this area that Justice Brandeis of the US Supreme Court questioned if transparency acted as a “disinfectant.” Although the term transparency is weak, one can argue for contextual/full disclosure, informed consent, non-ambiguity, dignity and respect of involved persons, veracity, and symmetrical communications as cornerstones of professions (Wright & Turk, 2007). These ethical securities would be particularly important in professions that use secrecy.

In the healthcare context, researchers (Alizadeh et al., 2025) argued that opacity or a lack of transparency was found to decrease access to information, reduce service quality, increase corruption, erode trust, and create ineffective decision making. Another study in that sector (Mooghali et al., 2024) found that a lack of transparency can decrease physician confidence in and use of even effective devices, with this lack of understanding hampering patient outcomes. Pharmaceutical communication researchers (Lopert & Rosenbaum, 2007) focused instead on the term fairness, but argued that transparency in drug pricing was desirable yet perhaps impossible to implement internationally. It is clear that transparency is central in medical and professional ethics -- yet it is compromised by its amoral status. Because transparency can be used for good or for harm, it is amoral, making it not an ethical concept but one that is determined by the intention of its user. Moral intention directs transparency to moral purposes or as illustrated in the above case, unethical purposes veiled behind a wall of transparency with the burden of discovery on the user, not the decision maker.

Despite the term’s prevalence in professions, transparency (as amoral) should not be used alone as it invites a great deal of risk. Disclosure, discussion, and contextual disclosure should supplant the use of transparency, and if it must be used for common parlance the term should be combined with more rigorous Kantian ethical constructs. A stronger and more thoroughly conceptualized ethical analysis is supported by concepts that link to the *a priori* good, relying upon those that have moral status, such as honesty and servant leadership, as well as other ethical constructs that foster patient care (patient rights, full consent, and dignity, for example).

Therefore, RQ 2 concludes that transparency is not a rigorous or fully conceptualized ethical concept that can support professional ethical standards and practice.

Philosophical Analysis & Alternative Concepts

Philosophical perspectives most often use an outcome-based approach (consequentialist reasoning, such as utilitarianism) or a principle-based framework (virtue and deontology). Though the differing approaches have their own offshoots, strengths, and weaknesses, a general guideline is that the ethical can be found in either creating a greater good or in living up to the highest moral principles or virtues. Even a simplistic conception of each school of thought can offer critique of the transparency concept when used as an ethical guide.

In a consequence-based framework, transparency is potentially corrupting, particularly for heavily regulated or public traded organizations (Mill, 1861/1957). Transparency is undertaken from compulsion or for public view rather than from moral regard, in this environment of consequence. Then transparency incentivizes inauthentic behavior for public consumption rather than reasoned judgment based upon moral ends, leading to a loss of autonomy and the potential for corruption. Licon (2024) reasoned:

Transparency incentivizes people to foster distinct (inauthentic) moral identities—one personal and one public—or to make moral judgments based on reputational reasons that sometimes diverge from the moral facts. In either case, transparency can be morally and politically corrupting. (p. 40)

Transparency undertaken for the public good in a utilitarian sense does not work as a motivating factor for creating the greatest good in an ethical sense, but in a perceptual and reputation sense that quickly becomes self-serving. The utilitarian calculus of the maximum good created with the minimum amount of harms (Elliott, 2007) then can no longer apply to the ethical situation because transparency has compelled outcomes or outsized implications over that which may be more ethical.

Philosophers (Licon, 2024; Metzen, 2025) agree with O'Neill (2002) who argued that transparency can encourage decision makers to be less honest, engaging in half-truths or compromised ideals to appear favorable, rendering transparent information not trustworthy. This problem is disqualifying of using transparency in a consequentialist paradigm, because the ethical solution output is tainted by incorrect or biased information input. A logical fallacy ensues, meaning transparency undermines the utilitarian calculus. Numerous problems with applying transparency a consequentialist or utilitarian approach are reviewed below, as well; however, incorrect input renders the greater good outcome calculation null before it begins. Virtue ethicists see no benefit of transparency as an amoral construct because it does nothing to practice the core virtues of building an ethical character. In fact, placing the burden of truth and discovery on the recipient of information is patently unethical to virtue ethicists.

Deontological philosophers who, akin their foundations in virtue ethics, base moral worth on principle rather than consequence (Scruton, 1982), are perhaps uniformly opposed to

the concept of transparency. Because transparency itself holds no principle to be evaluated, it is rarely discussed as even a remote part of ethics. This disdain for the concept is not in response to its overuse in the professions, but because it has no bearing on true moral analyses, which have rigorous conceptual frameworks posited, dissected, criticized, and reconstructed over thousands of years.

Kant asked for decision makers to judge independently, using their will to rationality and moral autonomy as guides rather than prudential self interest (Kant, 1963/1930). Although deontologists acknowledge that the self is part of this conception (Kant, 1997; Newton, 2019), analytic, research, and reason should drive moral decisions rather than bias that often results when using a transparency idea or determining what facts to offer. Transparency asks for comprises in moral autonomy. Deontologists hold that the internal locus of control of the autonomous decision maker, categorically obligated by duty to the moral law to do the same act other rational decision makers would undertake, is a superior ethical framework that seeks to use moral principle (Sullivan, 1989).

Control is a crucial variable of moral autonomy in this argument, though deontology insists that duty, respect, and good intention also alight with moral autonomy to uphold a moral principle that would be upheld in a perfect kingdom of ends (Kant, 1783/1977). Deontology offers these terms as analytically superior to the amoral term transparency due to their ability to further a kingdom of ends and harmonize with a prior good that is the ultimate ideal of normative ethics (Kant, 1948).

RQ 3: Superior alternatives to transparency exist through the traditions of moral philosophy and are summarized above and synthesized below for ease of implementation.

In conclusion, strategic communicators create risk when using an amoral concept to attempt to provide ethical direction. This ambiguity leaves the ethical framework open to interpretation, situational perspectives, and bias, as well as the lack of a foundation in moral intention that is needed to drive ethics.

Conclusions and Recommendations

In conclusion, returning to an ethical issues management framework allows organizations to minimize risk introduced by a failure to include ethical analyses and management decision-making and organizational activities. In order to include ethics in the issues management process the adapted framework reviewed above is recommended. More specific inclusions include philosophical frameworks such as the principle of veracity from virtue ethics and a deontological decision-making framework based on Kant's three forms of the categorical imperative (1. duty to universal, rational moral principle; 2. dignity and respect, and 3. good intention). These moral frameworks offer well researched and rigorous means of including ethics in organizational decision-making. Yet, the ease and popularity of recommending transparency can be problematic by placing the burden of responsibility on receivers of information (stakeholders, publics, regulatory agencies, and so on) as well as by using disclosure as a substitute for truth.

Returning to the research questions listed above, studied through literature analysis from various fields outside of communication and compared with our own, and is illustrated by small case examples, reveals additional concerns. Transparency as an ethical construct has failed to meet a minimum standard in each of the Research Questions (RQs) posed in this research. To summarize, these RQs, with new answers in *italic*, are:

RQ. 1. Does transparency succeed or invite risk when used as a basis for ethics? *Transparency fails as a loan ethical concept because of its immoral status and invites risk.*

RQ. 2. Do other professions successfully employ transparency as an ethical construct? *No, transparency obscures or reduces moral intention in the professions, as seen from small cases in artificial intelligence, health, or energy.*

RQ. 3. Are there superior ethical standards than transparency? *Yes, numerous well conceptualized alternatives are offered in moral philosophy, as synthesized in the list below, most prominently guided by good intention or the pure moral will and the principle of veracity.*

In fact, analyses in case examples showed that the inclusion of an amoral term such as transparency can invite risk through numerous intentions and biases, instead of reducing that risk through ethical issues management. Transparency can be used as the defrayment of responsibility for an obfuscation and can be harmful to ethical decision making.

When asked if transparency can still be included as part of an ethical model, framework, or approach in strategic communication, the philosophically astute say no. Transparency is not based on good intention alone but is amoral, defined by how it is used, and thus should not be included in an ethical framework. Why would one weaken an ethical framework within an amoral concept that adds nothing other than risk? It is likely due to the colloquial usage and common understanding of transparency that lends itself to easy discussion among management. However, the lack of precision in the concept of transparency and how it is applied leave numerous definitional questions open, inviting risk rather than using ethical issues management to reduce uncertainty and risk.

In fact, inclusion of transparency would weaken the rest of the concepts by acting as an amoral non-contributor that introduced a level of risk and uncertainty inappropriate when conducting an *ethical* analysis. Transparency should be wholly replaced with ethical terms involving a sense of goodness and *a priori* moral worth, such as good intention, veracity, full contextual disclosure, honesty, dialogue, questioning, and understanding. Such concepts are geared toward moral worth - though not entirely incorruptible with the exception of the good moral will (Kant, 1785/1964). As the only incorruptible concept, good intention or good will should, of course, be a cornerstone of ethical analyses. The inclusion of good will is the primary concept which places the analysis firmly in the school of Kantian deontology, providing an unassailable cornerstone then assisted by the other two forms of the categorical imperative (those being duty, and humanity's dignity and respect), and grounding in rational moral autonomy.

Issues management, including rigorous ethical analyses, based on the superior and thoroughly conceptualized terms from moral philosophy will help organizations be more responsible and effective, and prevent numerous problems, disasters, or scandals. Using concepts based in moral philosophy also foster the ethical leadership of the communication

function in business ethics (Versteegt et al., 2025), in managing ethical issues with moral duty, excellent analytical prowess, and noteworthy philosophical mastery.

Synthesizing the shortcomings of transparency and offering stronger alternatives for implementation in practice

The numerous shortcomings of transparency as an ethical principle arise from its status as an amoral concept. It is simply without moral status and can be used for good or evil, productivity or corruption, and is neither ethical nor unethical a priori.

Transparency is a risky concept to discuss with ethics because it can be neither good nor bad but it is defined and evidenced by its usage. And, who knows? Intention for usage is vague and mutable, at best. To summarize the discussion above, transparency fails as an ethical construct due to numerous concerns, as listed below. Following this list are recommendations for implementing a normative, deontological ethical analysis that is rigorous and based on sound moral philosophy. The following list should be a helpful, practical guide for implementing the arguments presented in the above article.

Risks, Pitfalls, and Drawbacks of Transparency

- amoral status (intention is ambiguous)
- riskiness
- allows concealment
- obfuscates
- places burden of truth/discovery on external observers (locus of control is not with the ethical decision maker as it should be)
- denies moral agency and individual control
- denies autonomy in decision-making
- incapable of conveying moral intention
- can be used to scapegoat, victimize, gaslight, or blame
- can be regarded as “enough”
- assumes information access for all
- places the burden of data analyses on stakeholders
- may offer short-term unfair advantage in the marketplace
- can be used for insincere appeasement
- encourages half-truths and corruption
- Schrödinger’s cat problem - one does not know the intent behind transparency until it is “too late” and consequences have resulted
- data dumps can actually suppress laissez faire capitalism
- can meet the regulatory standard but not an ethical one

Kantian Alternatives to Transparency based on Moral Philosophy

- good intention or the pure moral will (the only concept that is considered good without qualification) (This is the kingdom of ends form of Kant's categorical imperative)
- veracity
- dignity and respect (The humanity for of Kant's categorical imperative)
- honesty
- universal decision holds equally for others (The duty form of Kant's categorical imperative)
- reversibility
- full disclosure
- contextual disclosure (Bowen, 2016)
- moral autonomy (locus of control is with the ethical decision maker)
- ethical issues management
- dialogue and symmetrical understanding
- virtuous character and integrity
- question asking is encouraged (understanding is goal rather than compliance)
- data collection to verify 'assumptions'
- rigorous testing
- openness eschewing secrecy
- duty and responsibility
- practiced virtuous character in problem solving and issues management
- acts harmonizing with a possible kingdom of ends where only good will remains

The above practical reconceptualization can be used in implementing ethical analysis as part of decision making in any type of organization, with any size or mission. Routinely using issues management to conduct ethical analyses using these terms will help professionals and their organizations make more reliably ethical decisions. It will also serve to enhance the credibility of the strategic communication discipline as one in line with the intellectually rigorous discipline of moral philosophy, and foster an understanding of the role of normative ethics in building more responsible and trustworthy organizations as positive contributors to society.

Biographical Note

Shannon A. Bowen, Ph.D., (University of Maryland) is a Professor at the University of South Carolina where she founded (along with Dr. Bruce Berger and Dr. Elina Erzikova) the *Global Strategic Communication Consortium* to study ethics and futurism in strategic communication, resulting in the *Handbook of Innovations in Strategic Communication: AI, Futurism and Directions* (2025). Her research areas are applied ethics, artificial intelligence (AI), internal communication, and issues management. Bowen won numerous awards and top papers, including the Robert L. Heath Award from ICA and the Jackson, Jackson and

Wagner Behavioral Science Research Prize, and served as principal investigator on grants from the National Science Foundation, the International Association of Business Communicators, Arthur W. Page Center for Public Integrity, and others. She has served as a journal co-editor of *Ethical Space: The International Journal of Communication Ethics* and as regular ethics columnist in *#PRethics Community* on LinkedIn and *PRWeek*, as well as offered ethics training in numerous professional settings. She has several books, a research dictionary, encyclopedia entries, and more than one hundred articles, many catalogued under Orcid: 0000-0001-7675-5002

Acknowledgement

This research was supported by the Global Strategic Communication Consortium (GSCC), University of South Carolina.

References

- Alizadeh, M., Azizi, N., Mahdavi, S., & Baghlani, F. (2025). Unveiling the shadows: obstacles, consequences, and challenges of information opacity in healthcare systems. *Philosophy, Ethics, and Humanities in Medicine*, 20(1), 6. <https://doi.org/10.1186/s13010-025-00170-6>
- Aristotle. (1910). *Nicomachean ethics* (J. E. C. Welldon, Trans.). Clarendon Press. (Original work published ca. 340 BCE)
- Berger, B. K. (2005). Power over, power with, and power to relations: Critical reflections on public relations, the dominant coalition, and activism. *Journal of Public Relations Research*, 17(1), 5–28. https://doi.org/10.1207/s1532754xjpr1701_3
- Berger, B. K., & Reber, B. H. (2006). *Gaining influence in public relations: The role of resistance in practice*. Lawrence Erlbaum Associates.
- Bok, S. (1983). *Secrets: On the ethics of concealment and revelation*. Vintage Books.
- Bowen, S. A. (2000). *A theory of ethical issues management: Contributions of Kantian deontology to public relations' ethics and decision making* [Doctoral dissertation, University of Maryland].
- Bowen, S. A. (2002). Elite executives in issues management: The role of ethical paradigms in decision making. *Journal of Public Affairs*, 2(4), 270–283. <https://doi.org/10.1002/pa.119>
- Bowen, S. A. (2005). A practical model for ethical decision making in issues management and public relations. *Journal of Public Relations Research*, 17(3), 191–216. https://doi.org/10.1207/s1532754xjpr1703_1
- Bowen, S. A. (2008). A state of neglect: Public relations as 'corporate conscience' or ethics counsel. *Journal of Public Relations Research*, 20(3), 271–296.
- Bowen, S. A. (2010). The nature of good in public relations: What should be its normative ethic? In R. L. Heath (Ed.), *The Sage handbook of public relations* (pp. 569–583). Sage.

- Bowen, S. A. (2016). Clarifying ethics terms in public relations from A to V, authenticity to virtue. *Public Relations Review*, 42(4), 564–572. <https://doi.org/10.1016/j.pubrev.2016.03.012>
- Bowen, S. A. (2024). If it can be done, it will be done: AI ethical standards and a dual role for public relations. *Public Relations Review*, 50(5), 102513. <https://doi.org/10.1016/j.pubrev.2024.102513>
- Bowen, S. A. (2025). Artificial intelligence, warfare, and the surveillance/information state. In F. Roumate & A. La Rosa (Eds.), *Artificial intelligence, media and international security: The weaponization of AI use in media and international security* (pp. 1–16). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-95757-4_1
- Bowen, S. A., & Erzikova, E. V. (Eds.). (2025). *Handbook of Innovations in Strategic Communication: Artificial Intelligence, Futurism and Directions*. Edward Elgar.
- Bowen, S. A., & Heath, R. L. (2020). Intelligences in strategic issues management: Challenging the mutually beneficial relationship paradigm. *Partecipazione e Conflitto*, 13(2), 1002–1021. <https://doi.org/10.1285/i20356609v13i2p1002>
- Bowen, S. A., Heath, R. L., Lee, J., Painter, G., Agraz, F. J., McKie, D., & Toledano, M. (2006). *The business of truth: A guide to ethical communication*. International Association of Business Communicators.
- Bowen, S. A., Lovari, A., & Timmermans, A. (2024). Issues management: A new model for a key function in public affairs. In A. Timmermans (Ed.), *Research Handbook on Public Affairs* (pp. 282–297). Edward Elgar. <https://doi.org/10.4337/9781803920283.00031>
- Bowen, S. A., Rawlins, B. L., & Martin, T. M. (2019). *An overview of the public relations function* (2nd ed.). Business Expert Press.
- Boyle, M. (2019). Transparency and reflection. *Canadian Journal of Philosophy*, 49(7), 1012–1039. <https://doi.org/10.1080/00455091.2019.1565621>
- Buchholz, R. A., Evans, W. D., & Wagley, R. A. (1994). *Management responses to public issues: Concepts and cases in strategy formulation* (3rd ed.). Prentice Hall.
- Chase, W. H. (1984). *Issue management: Origins of the future*. Issue Action Publications.
- Christians, C. G., Fackler, M., Rotzoll, K. B., & McKee, K. B. (2001). *Media ethics: Cases and moral reasoning* (6th ed.). Longman.
- Coates, J. F., Coates, V. T., Jarratt, J., & Heinz, L. (1986). *Issues management: How you can plan, organize and manage for the future*. Lomond.
- De George, R. T. (2010). *Business ethics* (7th ed.). Prentice Hall.
- Edgett, R. (2002). Toward an ethical framework for advocacy in public relations. *Journal of Public Relations Research*, 14(1), 1–26. https://doi.org/10.1207/S1532754XJPRR1401_1
- Elliott, D. (2007). Getting Mill right. *Journal of Mass Media Ethics*, 22(2&3), 100–112.
- Etzioni, A. (2010). *Is transparency the best disinfectant?* *Journal of Political Philosophy*, 18(4), 389–404. <https://doi.org/10.1111/j.1467-9760.2010.00366.x>

- Falit, B. (2005). Pharma's commitment to maintaining a clinical trial registry: Increased transparency or contrived public appeasement? *The Journal of Law, Medicine & Ethics*, 33(Summer, 391-396).
- Felzmann, H., Villaronga, E. F., Lutz, C., & Tamò-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Society*, 6(1), 2053951719860542. <https://doi.org/10.1177/2053951719860542>
- Fleming, C. (2025). On doing your own research: Conspiracy theory, distributed epistemic labour and the politics of trust. *Subjectivity*, 32, 170–191. <https://doi.org/10.1057/s41286-025-00219-z>
- Guyer, P. (2019). *Kant*. Cambridge University Press.
- Heald, D. (2006). Transparency as an instrumental value. In C. Hood & D. Heald (Eds.), *Transparency: The key to better governance?* (pp. 25–43). Oxford University Press.
- Heath, R. L. (1997). *Strategic issues management: Organizations and public policy challenges*. Sage.
- Heath, R. L. (2001). A rhetorical enactment rationale for public relations: The good organization communicating well. In R. L. Heath (Ed.), *Handbook of public relations* (pp. 31–50). Sage.
- Heath, R. L. (2025). The ethical challenges of transparency and the mandate for ethical standards in public relations. In D. M. Topić & H. A. Z. Wæraas (Eds.), *Public relations and social theory: Key thinkers and concepts* (pp. 13–26). Routledge.
- Heath, R. L., & Nelson, R. A. (1986). *Issues management: Corporate public policymaking in an information society*. Sage.
- Heath, R. L., & Palenchar, M. J. (2009). *Strategic issues management: Organizations and public policy challenges* (2nd ed.). Sage.
- Hryniw, I. (2024, April 18). UAB stealing dead Alabama prison inmates' organs after autopsies, families claim in lawsuit. *AL.com*. <https://www.al.com/news/2024/04/uab-stealing-dead-alabama-prison-inmates-organs-after-autopsies-families-claim-in-lawsuit.html>
- Kant, I. (1783/1977). *Prolegomena to any future metaphysics*. (T. E. Humphrey, Trans.). Hackett Publishing.
- Kant, I. (1785/1964). *The groundwork of the metaphysic of morals* (H. J. Paton, Trans.). Harper Torchbooks.
- Kant, I. (1930/1963). *Lectures on ethics* (L. Infield, Trans.). Harper & Row.
- Kant, I. (1948). *The groundwork of the metaphysic of morals* (H. J. Paton, Trans.). Harper Torchbooks. (Original work published 1785)
- Kant, I. (1997). *Critique of pure reason* (P. Guyer & A. W. Wood, Trans.). Cambridge University Press.
- Karnes, R. E. (2009). A Change in Business Ethics: The Impact on Employer-Employee Relations. *Journal of Business Ethics*, 87(2), 189–197. <http://www.jstor.org/stable/40294915>

- Klenk, M. (2023). Algorithmic Transparency and Manipulation. *Philosophy & Technology*, 36(4), 79. <https://doi.org/10.1007/s13347-023-00678-9>
- Knowlton, S. R., & Parsons, P. R. (Eds.). (1994). *The journalist's moral compass: Basic principles*. Praeger.
- Kronewald, E. (2024). Adapting to the challenges of communication in the digital age: ethical principles for PR practice in Germany. In E. Oliveira & G. Goncalves (Eds.), *Ethics and society: Challenges in organizational & public communication* (pp. 47–66). LabCom Books.
- Lerbinger, O. (2006). *Corporate public affairs: Interacting with interest groups, media, and government*. Lawrence Erlbaum Associates.
- Licon, J. A. (2024). Transparency as morally and politically corrupting. *Asian Journal of Philosophy*, 3(2), 41. <https://doi.org/10.1007/s44204-024-00169-4>
- Lopert, R., & Rosenbaum, S. (2007). What is fair? Choice, fairness, and transparency in access to prescription medicines in the United States and Australia. *The Journal of Law, Medicine & Ethics*, 35(4), 643–656. <https://doi.org/10.1111/j.1748-720X.2007.00187.x>
- Metzen, H. (2025). *Trust in Scientific Expertise and the Varying Demands of Value Transparency*. Canadian Journal of Philosophy, 1–16. <https://doi.org/10.1017/can.2025.10004>
- Mill, J. S. (1861/1957). *Utilitarianism*. The Liberal Arts Press.
- Mitchell, T. (2025). Trust and Transparency in Artificial Intelligence. *Philosophy & Technology*, 38(3), 87. <https://doi.org/10.1007/s13347-025-00916-2>
- Mooghali, M., Stroud, A. M., Yoo, D. W., Barry, B. A., Grimshaw, A. A., Ross, J. S., Zhu, X., & Miller, J. E. (2024). Trustworthy and ethical AI-enabled cardiovascular care: a rapid review. *BMC Medical Informatics and Decision Making*, 24(1), 247. <https://doi.org/10.1186/s12911-024-02653-6>
- Newton, A. M. (2019). Kant and the transparency of the mind. *Canadian Journal of Philosophy*, 49(7), 890–915. <https://doi.org/10.1080/00455091.2019.1565613>
- Oliveira, E., & Goncalves, G. (Eds.). (2024). *Ethics and society: Challenges in organizational & public communication*. LabCom Books.
- O'Neill, O. (2002). *A question of trust*. Cambridge University Press.
- Parsons, P. J. (2004). *Ethics in public relations: A guide to best practice*. Kogan Page.
- Patterson, P., & Wilkins, L. (2002). *Media ethics: issues & cases*. McGraw-Hill.
- Paul, S. K. (2015). The transparency of intention. *Philosophical Studies*, 172(6), 1529–1548. <https://doi.org/10.1007/s11098-014-0363-3>
- Pearson, R. (1989). Albert J. Sullivan's theory of public relations ethics. *Public Relations Review*, 15(2), 52–62.
- Place, K. R., & Flamme, C. (2024). E-activism as an issues management strategy: New directions for strategic issues management in public relations. *Public Relations Review*, 50(2), 102474. <https://doi.org/10.1016/j.pubrev.2023.102474>

- Place, K. R., Edwards, L., & Bowen, S. A. (2021). Dignity and respect or homocommodification? Applying moral philosophy to LGBTQ public relations. *Public Relations Review*, 47(4), 102085. <https://doi.org/10.1016/j.pubrev.2021.102085>
- Plaisance, P. L. (2007). Transparency: An Assessment of the Kantian Roots of a Key Element in Media Ethics Practice. *Journal of Mass Media Ethics*, 22(2-3), 187–207. <https://doi.org/10.1080/08900520701315855>
- Price, R. (2003). From a review of the principal questions in morals. In S. Darwall (Ed.), *Deontology* (pp. 34–54). Blackwell.
- Scruton, R. (1982). *Kant*. Oxford University Press.
- Singer, P. (Ed.). (1994). *Ethics*. Oxford University Press.
- Sullivan, R. J. (1989). *Immanuel Kant's moral theory*. Cambridge University Press.
- Taylor, M. (2010). Public relations in a troubled world. In R. L. Heath (Ed.), *The SAGE handbook of public relations* (pp. 5–15). Sage.
- Theus, K. (1993). Organizations and the media: Structures of miscommunication. *Management Communication Quarterly*, 7(1), 67–94.
- Uzunoğlu, E., Türkel, S., & Kip, S. M. (2017). Engagement and dialogue in digital public relations: Studying online stakeholder engagement and dialogue. *Public Relations Review*, 43(5), 1025–1034. <https://doi.org/10.1016/j.pubrev.2017.09.006>
- Versteegt, J., van Dijke, M., & Kaptein, M. (2025). *The status of normative ethical decision-making models: The development and application of an analytical framework*. Journal of Business Ethics. Advance online publication. <https://doi.org/10.1007/s10551-025-06094-7>
- Wang, H. (2022). Transparency as Manipulation? Uncovering the Disciplinary Power of Algorithmic Transparency. *Philosophy & Technology*, 35(3), 69. <https://doi.org/10.1007/s13347-022-00564-w>
- Wang, H. (2023). *Why Should We Care About the Manipulative Power of Algorithmic Transparency?* *Philosophy & Technology*, 36(1), 9. <https://doi.org/10.1007/s13347-023-00610-1>
- Whitehead, T. (1985). *Methods for doing ethnography: The natural culture learning process*. meeting of A Training of Trainers Workshop, U.S. Peace Corps, Bukavu, Zaire.
- Wood, S., & Gauntt, J. (2024, April 18). 5 families file lawsuits claiming UAB took organs from ADOC inmates “without consent”. *WBRC*. <https://www.wbrc.com/2024/04/18/5-families-file-lawsuits-claiming-uab-took-organs-adoc-inmates-without-consent/>
- Wright, D. K., & Turk, J. V. (2007). Public relations knowledge and professionalism: Challenges to educators and practitioners. In E. L. Toth (Ed.), *The future of excellence in public relations and communication management: Challenges for the next generation* (pp. 571–588). Lawrence Erlbaum.
- Zerilli, J., Knott, A., Maclaurin, J., & Gavaghan, C. (2019). *Transparency in Algorithmic and Human Decision-Making: Is There a Double Standard?* *Philosophy & Technology*, 32(4), 661–683. <https://doi.org/10.1007/s13347-018-0330-6>

Notes

¹ Risk management is a financial tool not to be confused with the issues management of risks and problem solving.

Le strade della prevenzione.

Comunicare i vaccini e gli screening oncologici nelle regioni italiane*

Laura Solito**

Letizia Materassi***

Ester Macri****

Erika Greco*****

Università degli studi di Firenze

Il presente lavoro si propone di analizzare i cambiamenti avvenuti nella comunicazione della prevenzione sanitaria in Italia negli ultimi 20 anni, letti attraverso una ricerca empirica su due specifici ambiti: gli screening oncologici e i programmi vaccinali. Background: il contributo si iscrive nel panorama più ampio degli studi sulla comunicazione pubblica della salute e sull'evoluzione della comunicazione sociale, sia in termini di produttività, sia nell'impiego di specifici linguaggi e registri comunicativi. In particolare, partiamo dal presupposto che le disuguaglianze presenti tutt'oggi nel campo della salute siano legate anche a fattori culturali di cui la comunicazione fa parte. Per tale motivo si rende necessaria una lettura diacronica e contenutistica di come gli screening oncologici e i programmi vaccinali sono stati comunicati nel nostro Paese. Metodo: attraverso l'esplorazione dei siti web istituzionali delle Regioni italiane e dei loro canali social, si sono raccolti i materiali informativi e comunicativi prodotti e pubblicati dal 2003 al 2023. Il corpus, composto da 434 unità di analisi, è stato investigato attraverso l'analisi descrittiva del contenuto e l'individuazione di variabili quali-quantitative. Risultati: la ricerca offre un panorama regionale frammentato e disomogeneo. La differenziazione dei territori dal punto di vista della produttività di materiali informativi e comunicativi è solo uno dei fattori significativi che va messo in relazione alla continuità/discontinuità delle azioni intraprese nel tempo, all'articolazione dei temi e dei contenuti prodotti, alla maggior/minore articolazione dei linguaggi, dei registri e delle funzioni narrative. Implicazioni: ciò che emerge è una "geografia" complessa, che non restituisce solo una disomogeneità nelle strategie di comunicazione, ma rischia di essere causa di ulteriori disuguaglianze nella salute.

Keywords: comunicazione, salute, screening, vaccini, regioni

* Article submitted on 27/03/2025. Article accepted on 20/12/2025.

** laura.solito@unifi.it

*** ester.macri@unifi.it

**** letizia.materassi@unifi.it

***** erika.greco@uniroma1.it

Introduzione

“La salute è un bene sociale la cui tutela non dovrebbe dipendere unicamente dal servizio sanitario, ma dall’impegno di tutti i settori che possono avere un impatto sulla distribuzione dei determinanti del benessere fisico e mentale dei cittadini e della comunità.”

Dichiarazione di Roma dei Ministri europei della Sanità
per la strategia Salute in tutte le politiche, del 18/12/2007

L’equità nella salute rappresenta tutt’oggi un obiettivo planetario da raggiungere, che necessita di un approccio integrato e multisettoriale, un sinergismo capace di affrontare le profonde disuguaglianze sanitarie che permangono nella popolazione e che restano difficili da scardinare (Marmot, 2007; Terraneo, 2018).

Queste disuguaglianze derivano non solo da fattori materiali e quindi legati a differenti condizioni economiche e di vita della cittadinanza, ma anche da una dimensione culturale e immateriale, che influisce sull’accesso e sull’utilizzo dell’offerta sanitaria (Costa et al., 2015). Il cosiddetto “*divide culturale*” che ne deriva (Kreuter & McClure, 2004) è a sua volta collegato a due principali ordini di fattori: da un lato, alle determinanti sociali e culturali (World Health Organization, 2007), ossia a quei fattori, come la religione, le credenze popolari, i livelli di istruzione, i valori, le comunità o le subculture di appartenenza, le esperienze individuali pregresse e le loro reciproche influenze, che possono produrre e riprodurre forme di disuguaglianza nella salute della popolazione (Yadav et al., 2022); dall’altro lato, il ruolo rivestito dalla comunicazione e dall’informazione sui temi della salute, che non deriva soltanto da una disponibilità o indisponibilità di contenuti informativi adeguati, ma anche dalla capacità di questi ultimi di ascoltare, rispondere e accogliere le eterogeneità presenti nella popolazione, facilitandone l’accesso, il consumo e la comprensione (Dutta, 2016; Parrott, 2009). Infatti, è da più parti sottolineato come l’elaborazione di strategie comunicative efficaci debba partire da un ascolto più attento alle articolazioni dei pubblici e dalla scelta di diete mediali maggiormente sfaccettate per superare le barriere che ancora limitano l’equità nella salute (Kreuter & McClure, 2004; Lovari, 2017; Collicelli, 2018).

Quando si parla di “comunicazione della salute” dobbiamo considerare che in letteratura si rintracciano almeno due distinti ambiti (Lovari e Ducci, 2022) o strategie (Cucco, Pagani e Pasquali, 2005): 1. la comunicazione sanitaria, nella quale rientrano le attività di informazione sui servizi sanitari e di assistenza ai cittadini, nonché sulle modalità di accesso agli stessi; 2. la comunicazione per la promozione della salute, con le attività che, da un lato, mirano a promuovere comportamenti corretti e stili di vita salutari per un completo benessere psico-fisico del cittadino, dall’altro vogliono educare e incrementare i livelli di literacy sanitaria nella popolazione. Rientrano in questo secondo ambito le attività di comunicazione sociale per la prevenzione di malattie, che, come vedremo, sono oggetto del presente contributo.

Entro questo secondo versante, numerosi studi hanno esplorato empiricamente il nesso tra fattori culturali e comunicazione, evidenziando come spesso le disuguaglianze nella salute nascano o si accentuino a seguito di una comunicazione inadeguata e non attenta alle specificità culturali e tecnologiche dei cittadini (Viswanath & Kreuter, 2007).

Ad esempio, è nota in letteratura la correlazione tra “*divide culturale*” e “*divide digitale*” (Lupton, 2016), che oltre a riprodurre le disuguaglianze comunicative caratterizzanti alcune fasce della popolazione, possono far emergere timori circa la perdita di controllo dei propri dati sensibili o di imbattersi in fenomeni di disordine informativo (Lovari e Ducci 2022).

Altrettanto forte, in taluni contesti, è il nesso tra convinzioni religiose e bassa alfabetizzazione sanitaria, che può costituire un ostacolo importante alla diffusione di informazioni mediche vitali, come quelle relative all'adesione agli screening oncologici (Hoover et al., 2019); oppure, è stato esplorato l'atteggiamento verso i programmi vaccinali e il ruolo-chiave giocato in alcune società dai leader religiosi (Melki et al., 2021) come mediatori di fiducia verso i servizi sanitari e come fonti informative ritenute autorevoli dalla popolazione che, se non opportunamente coinvolti, possono rendere le strategie fallimentari. Ma potremmo anche riferirci alla stigmatizzazione sociale associata a quelle patologie dell'apparato riproduttore e collegate a questioni di sessualità e moralità che non sempre sono tenute in considerazione dalle campagne di comunicazione, incidendo sulla partecipazione ai percorsi di prevenzione oncologica (Omondi et al., 2014), come nel caso del vaccino per l'HPV (Colòn-Lopez et al., 2017). In ciascuno di questi casi, l'efficacia comunicativa è strettamente correlata alla considerazione di quelle credenze e di quei valori che possono scoraggiare o, al contrario, avvicinare i cittadini ai programmi di salute, problematizzando l'esistenza di barriere psicosociali e istituzionali di cui la comunicazione deve tener conto per il raggiungimento di una maggiore equità (Mosavel & Ports, 2015; Ragas et al., 2014; Albright et al., 2014; Dutta, 2021). Eppure, nonostante tali consapevolezza, proprio la recente pandemia da Covid-19 ha riportato al centro del dibattito il rischio che le strategie di comunicazione adottate da governi e istituzioni sanitarie siano foriere di disuguaglianze sociali, andando ad interessare in modo particolare le fasce della popolazione maggiormente vulnerabili, le minoranze etniche e linguistiche, i cittadini con livelli più bassi di *health literacy* o coloro che devono prendere decisioni sulla salute dei propri familiari - es. persone disabili, anziani, donne in gravidanza - o assistiti, come nel caso dei *caregiver* (Kalocsanyiova et al., 2023; Team & Manderson, 2020; Boccia Artieri e Farci, 2021).

Oltre ad essere investigato in numerosi Paesi al mondo (Collyer & Smith, 2020), il *divide culturale* si rintraccia anche sul territorio italiano, dove sono state messe in luce variabili socio-antropologiche come la paura del cambiamento, la responsabilità individuale e gli atteggiamenti verso la prevenzione e la cura; questi fattori, a loro volta, vanno ad intersecarsi con caratteristiche quali l'età, il genere, il livello di studio, le convinzioni religiose, l'alfabetizzazione digitale, etc., rendendo evidente la natura intersezionale e complessa del fenomeno (Macchione e Passoni, 2024; Battistelli e Galantino, 2020).

A questi aspetti culturali, fanno eco fattori di natura organizzativa e istituzionale che da tempo sono oggetto di osservazione da parte degli studiosi, come ad esempio: un'attenzione discontinua alle attività di comunicazione da parte delle istituzioni sanitarie (Lovari, 2017), la mancanza di coordinamento tra livelli organizzativi diversi e la carenza di professionalità dedicate (Ducci, 2021), la scarsa integrazione delle politiche tra loro, dovuta a una impostazione segmentata delle strategie di promozione della salute e del benessere

e la permanenza di fattori di differenza e disparità entro i contesti regionali italiani (Gola et al., 2018). D'altronde, la regionalizzazione del Sistema sanitario italiano può portare non solo a disomogeneità strutturali, ma anche alla elaborazione di strategie dettate da volontà politiche diverse che indirizzano le scelte e fissano le priorità anche nelle attività di comunicazione (Mazzoleni, 2023).

Un tentativo di superamento di alcuni fattori di discontinuità territoriale risiede nell'accordo Stato-Regioni, sancito l'8 agosto 2001 tra il Governo, le Regioni e le Province Autonome di Trento e Bolzano, che affronta diversi aspetti legati al finanziamento e all'organizzazione del Servizio Sanitario Nazionale (SSN). L'obiettivo dell'accordo era stabilire un quadro finanziario stabile e sostenibile, mediante impegni specifici per contenere i costi e ottimizzare la gestione dei servizi, garantendo equità territoriale e qualità dei servizi sanitari a livello nazionale. Eppure, le disparità continuano a sussistere, anche nell'ambito della spesa sanitaria privata sostenuta dai cittadini. La quota di famiglie che effettuano spese sanitarie di "tasca propria" risulta molto più elevata nel Sud, dove raggiunge il 79,9%, rispetto al Centro (76,4%) e al Nord-Ovest (75,1%) nel 2015. Inoltre, l'impatto economico di queste spese è particolarmente marcato nel Mezzogiorno, dove il 2,8% delle famiglie si impoverisce a causa delle spese sanitarie (Crea, 2015). Questa percentuale è significativamente superiore rispetto all'1,1% delle famiglie del Centro, allo 0,4% del Nord-Ovest e allo 0,6% del Nord-Est (Collicelli, 2018). Dai dati Istat (2019), emerge anche l'aumento dei dati percentuali dell'emigrazione ospedaliera verso un'altra regione per ricoveri ordinari acuti sul totale dei ricoverati: 4.4% Nord, 6.5% Centro, 9.5% Mezzogiorno. Inoltre, le regioni del Nord Italia registrano una speranza di vita più alta rispetto a quelle del Sud. Ad esempio, la Provincia Autonoma di Trento ha una speranza di vita di 84,2 anni, contro gli 80,9 anni della Campania (Giuffrida et al, 2024).

Si evincono differenze regionali importanti anche rispetto al tema della vaccinazione. Esistono marcate disuguaglianze tra Nord e Sud Italia nella copertura vaccinale, sia per le vaccinazioni pediatriche che per quelle degli adulti e anziani. Le regioni del Centro-Nord, come ad esempio l'Emilia-Romagna e il Veneto, tendono ad avere tassi di copertura vaccinale più elevati rispetto a quelle del Sud come la Campania e la Calabria, spesso dovute a differenze organizzative e infrastrutturali (Armeni et al, 2024). In Puglia, le vaccinazioni per gli over 65 sono erogate principalmente dalla medicina generale. Tuttavia, nonostante le iniziative, i livelli di copertura rimangono sotto gli obiettivi nazionali. La Regione ha recentemente testato strategie innovative, come la "chiamata attiva" tramite SMS, per migliorare la sensibilizzazione e l'accesso ai vaccini (es. pneumococcico e herpes-zoster), ma di cui ancora non conosciamo l'esito. Il Lazio ha implementato campagne specifiche per gruppi di rischio, registrando incrementi significativi nella vaccinazione di persone con diabete (+21%), broncopneumopatia cronica ostruttiva (+30%) e insufficienza renale cronica (+47%). La Sicilia si distingue per aver aggiornato il suo calendario di immunizzazione con l'introduzione della vaccinazione contro il RSV (virus respiratorio sinciziale) sia per i neonati sia per gli over 60 con patologie croniche. Questo dimostra un approccio proattivo per fronteggiare le malattie infettive nelle fasce più vulnerabili. La Liguria, invece, si è concentrata sul miglioramento delle strategie già consolidate, come

l'offerta del vaccino contro l'herpes zoster e lo pneumococco per gli anziani, evidenziando una lunga tradizione di prevenzione (Giuffrida et al, 2024). Per alcune vaccinazioni pediatriche (es. esavalente e MPR), molte regioni raggiungono o superano l'obiettivo del 95%, mentre altre, specialmente al Sud, rimangono sotto questa soglia. Toscana, Emilia-Romagna e Lombardia si distinguono per modelli organizzativi efficaci, inclusa una forte integrazione tra servizi territoriali e medici di medicina generale. Campania e Calabria presentano invece difficoltà legate alla carenza di personale e infrastrutture, che influiscono negativamente sui tassi di copertura (Armeni et al, 2024). La pandemia da COVID-19 ha poi influenzato negativamente la somministrazione di altre vaccinazioni, soprattutto nelle regioni con sistemi sanitari già sotto pressione (Ibidem, 2024; Lovari et al. 2020). Un panorama, dunque, articolato, da cui partiamo, cercando di problematizzare il ruolo che la comunicazione della prevenzione ha assunto nel nostro Paese e come questo sia cambiato nel tempo.

Screening oncologici e vaccini: vent'anni di comunicazione

Metodologia della ricerca

La ricerca presentata si inserisce all'interno di un progetto competitivo di Ateneo finanziato nell'ambito dei fondi PNRR "ECoHeTE - Effective Communication for Healthcare: Theory and Evidence". Il progetto "EcoHeTE" mira a valutare, attraverso un approccio multidisciplinare, l'efficacia delle campagne di comunicazione sulla salute, in particolare sui temi della prevenzione vaccinale e oncologica. Nell'ambito del progetto generale, l'obiettivo del nostro lavoro ha riguardato una ricognizione dei materiali di comunicazione realizzati in Italia negli ultimi vent'anni sulle vaccinazioni e sugli screening oncologici, analizzandone l'andamento nel tempo, le trasformazioni relative ai contenuti, alle modalità narrative, ai linguaggi e ai registri comunicativi. Il tema dell'analisi delle disuguaglianze territoriali nella comunicazione per la prevenzione, oggetto specifico di questo contributo, non era stato definito a priori nel disegno di ricerca originale ma è emerso in maniera significativa in fase di analisi dei dati, tanto da suggerire al gruppo di ricerca di dedicarvi un approfondimento. Il corpus analizzato è stato costruito attraverso un processo di raccolta manuale e sistematica condotto da una ricercatrice con il continuo confronto con l'intero team di ricerca, formato da 4 ricercatrici, tra gennaio e dicembre 2023. La selezione è avvenuta accedendo ai siti web istituzionali delle Regioni e Province Autonome italiane, del Ministero della Salute, dell'Istituto Superiore di Sanità e dell'Organizzazione Mondiale della Sanità. Oltre ai siti istituzionali, la ricerca è stata estesa alle pagine ufficiali social (Facebook, Instagram e Twitter/X) e ai canali Youtube degli enti presi in considerazione. Il materiale è stato acquisito nella sua forma originaria (PDF, immagini, video, post digitali) e archiviato in un database appositamente creato. Sono stati inclusi nel dataset solo i materiali afferenti a due

macrotemi: vaccinazioni (es. anti-HPV, anti-Covid, antinfluenzale, ecc.) e screening oncologici (mammografico, cervicale, colon-retto).¹

Non sono stati considerati contenuti divulgativi generici, notizie giornalistiche o pubblicazioni scientifiche. Inoltre, sebbene nella comunicazione sociale un insieme coordinato di azioni comunicative relative ad uno stesso contenuto e mosse da un medesimo obiettivo strategico si possa definire “campagna di comunicazione” (Puggelli e Sobrero, 2010; Gadotti e Bernocchi, 2010), in ambito sanitario si rintraccia una certa polisemia di tale espressione che può dar vita a qualche ambiguità. In quest’ultimo campo, infatti, per “campagna” si può anche intendere la disponibilità e la promozione per un arco di tempo limitato di un servizio o di un bene - come, ad esempio, un vaccino - da diffondere entro la popolazione, senza necessariamente richiamare il ruolo della comunicazione. Dunque, nel contributo che segue ci riferiremo, più genericamente, alla selezione e analisi di un insieme di materiali comunicativi e informativi che, in tale accezione e per motivi di brevità, definiremo “campagna”. Ogni unità di analisi è stata associata, in una matrice dati creata con Excel, a una materiale identificabile attraverso anno, soggetto promotore e target specifico, anche in assenza di un naming ufficiale della campagna al quale il materiale afferisce. Nella stessa matrice si è proceduto a classificare ciascun materiale raccolto. La classificazione è avvenuta secondo un set di variabili definito a priori, ispirato alla letteratura sulla comunicazione sociale (Puggelli e Sobrero, 2010; La Rocca, 2018). Tali variabili includono: anno di pubblicazione, tema (vaccino/screening), specifica patologia, Regione o ente promotore, altro soggetto coinvolto nella promozione, formato del materiale (poster, video, locandina, social card, brochure, podcast, spot radio), presenza del logo e dell’hashtag, titolo, canale di diffusione ma anche informazioni più qualitative come lo stile narrativo (se formale/informale), il linguaggio utilizzato, il registro, la funzione narrativa, i destinatari, il setting e i protagonisti. La codifica è stata condotta dalla stessa ricercatrice con verifica incrociata da parte del team. Il metodo impiegato ha privilegiato la disponibilità pubblica online dei materiali. Sono stati selezionati anche tutti i materiali che il Sistema Sanitario chiama “campagna” (es. “giornata mondiale contro il cancro”, “giornata nazionale salute donna”, “giornata mondiale tumore al seno”) con cui intende promuovere i servizi di prevenzione. Tuttavia, è possibile che alcuni materiali, pur realizzati, non siano stati archiviati online o risultino difficilmente reperibili. Il disegno metodologico adottato si configura quindi come un’analisi quali-quantitativa descrittiva dei materiali raccolti. Non è stato implementato, in questa fase, un protocollo di analisi più tradizionalmente qualitativa dei contenuti (es. semiotica, *frame analysis*, *health literacy*), che potrebbe rappresentare uno degli sviluppi previsti nelle successive fasi del progetto.

I dati di ricerca

La maggior parte dei materiali comunicativi su vaccinazione e screening oncologici reperiti online proviene dalle Regioni (434 su 719 complessivi), seguite da enti nazionali e

internazionali come l'OMS (152), l'Istituto Superiore di Sanità (77) e il Ministero della Salute (56). Tra le Regioni più attive, in termini di numero assoluto di materiali prodotti si rintraccia la Toscana (60), l'Emilia-Romagna (55), Trentino-Alto Adige (49), Sicilia (43). Al contrario, Abruzzo, Friuli-Venezia Giulia, Basilicata e Calabria risultano tra le meno attive nella produzione di materiali.

È però necessario sottolineare che l'assenza o la scarsità di materiali rilevati online non implica necessariamente una totale mancanza di attività comunicativa: i materiali potrebbero essere stati veicolati attraverso altri canali, non digitali o oggi non più accessibili.

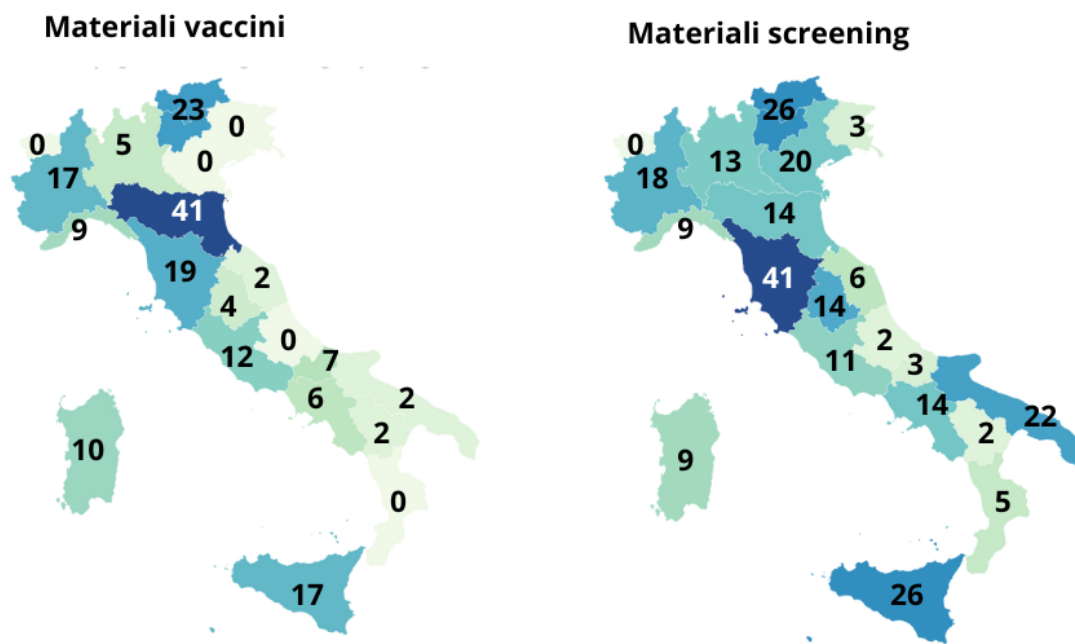


Grafico 1 - Macrotema per Regione.

Dall'analisi dei dati emerge una netta differenziazione tra regioni in termini di produzione di materiali per la comunicazione delle due tematiche. Alcune regioni, infatti, si contraddistinguono per un'elevata produzione di materiali, altre sembrano essere decisamente meno attive. In particolare, alcune regioni presentano una scarsa attenzione soprattutto in merito agli screening oncologici. Questo sembra essere in linea con i dati raccolti dall'Osservatorio nazionale sulla salute nelle regioni italiane (2022): "in Italia i programmi di screening mammografico e coloretale si stanno estendendo su tutto il territorio nazionale, ma con tempistiche diverse tra Centro-Nord e Meridione, dove la copertura è ancora sub-ottimale. Sebbene si stia assistendo ad un'espansione dell'offerta degli screening al Sud e nelle Isole, esistono ancora delle differenze significative rispetto al Centro-Nord. Tutto questo ha sicuramente contribuito a rendere il quadro epidemiologico dei tumori non omogeneo a livello nazionale" (Demu et al., 2022). La gestione

dell'emergenza sanitaria, inoltre, ha provocato ritardi dal punto di vista organizzativo dell'offerta dei programmi di screening organizzati da parte delle ASL e minore adesione da parte della popolazione, tant'è che nel 2020 la copertura degli screening oncologici si è ridotta, come sottolinea la stessa Organizzazione Mondiale della Sanità (WHO, 2020). Tuttavia, alcune realtà territoriali non erano in grado di assicurare l'offerta di screening già da prima dell'emergenza pandemica (Rapporto Osservasalute, 2022).

È possibile suddividere le regioni italiane in quattro gruppi distinti in base al numero di materiali comunicativi prodotti su vaccini e screening oncologici. Ogni gruppo rappresenta regioni con simili livelli di attenzione nei confronti di questi temi, permettendo di identificare pattern comuni tra di esse.

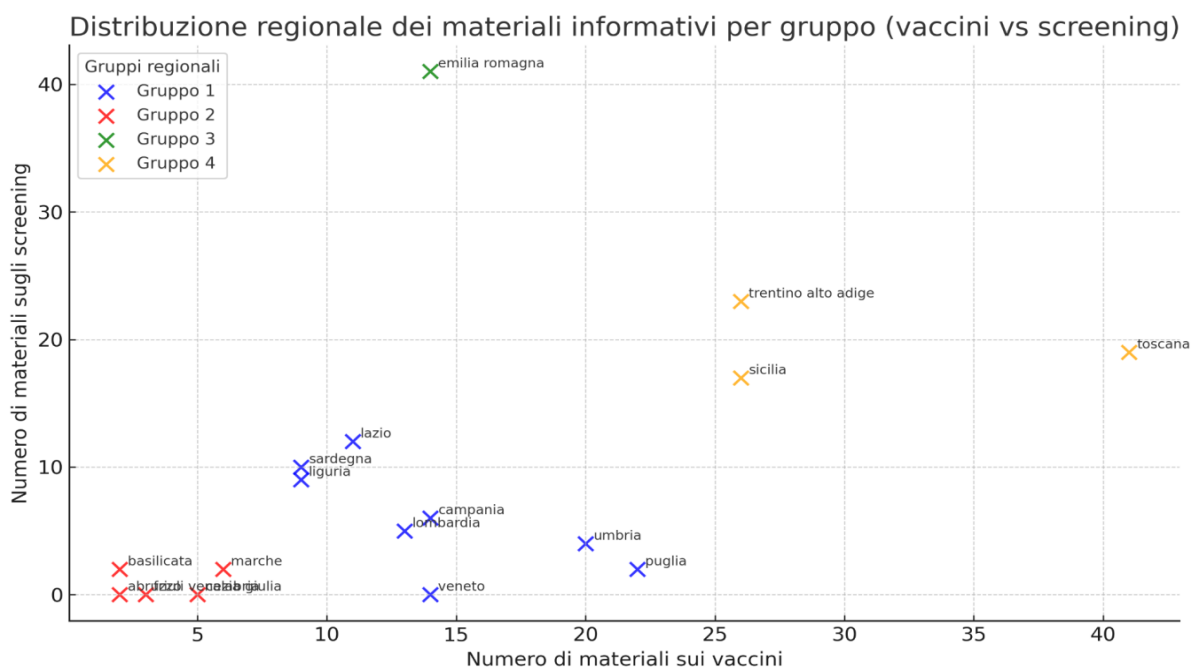


Grafico 2 - Materiali sui vaccini e materiali sugli screening a confronto

Il primo gruppo include Lazio, Sardegna, Liguria, Campania, Lombardia, Umbria, Puglia, e Veneto, regioni che si distinguono per un numero medio-alto di materiali sui vaccini e uno sforzo moderato per quanto riguarda invece gli screening oncologici. Le regioni appartenenti a questo gruppo si dimostrano generalmente attive su entrambi i fronti, ma senza eccellere in maniera particolare in nessuno dei due temi. Il secondo gruppo include Abruzzo, Basilicata, Calabria, Marche, Friuli-Venezia Giulia, cioè le regioni con un basso numero di materiali prodotti su entrambi i temi. Queste regioni mostrano una bassa attenzione sia per i vaccini che per gli screening oncologici, risultando meno attive nella comunicazione rispetto alle altre. Nel terzo gruppo si trova solo l'Emilia-Romagna, un caso a parte, che ha prodotto un numero particolarmente elevato di materiali, soprattutto sugli screening

oncologici. Sembra esserci quindi in questa regione un'alta attenzione agli screening, spesso accompagnata da un impegno considerevole anche sui vaccini. Infine, il quarto gruppo include Toscana, Trentino-Alto Adige e Sicilia, che si distinguono per un elevato livello di produzione di materiali sia sui vaccini che sugli screening oncologici. Queste regioni sono le più bilanciate e attive, con un forte impegno su entrambi i fronti della comunicazione sanitaria presa in esame.

Comunicare gli screening oncologici

Lo screening oncologico per il quale sono stati reperiti più materiali è quello del tumore alla mammella (91 materiali), seguito da quello per il tumore della cervice uterina (81 materiali) e del colon retto (79 materiali). È interessante notare che l'Emilia-Romagna, regione molto attiva nella promozione degli screening, si rivela molto produttiva anche nei materiali sullo screening del tumore del colon retto, con ben 27 materiali su questo tema. Non sono stati rilevati molti materiali relativi a screening che riguardino altri tumori rispetto ai tre principali screening, ossia quelli offerti dal Sistema Sanitario Nazionale: mammella, cervice uterina e colon retto.

<i>Regione</i>	<i>Screening mammella</i>	<i>Screening cervice uterina</i>	<i>Screening colon retto</i>	<i>Altro</i>	<i>Totale</i>	
Abruzzo	0	0	0	0	0	0
Basilicata	2	1	1	0	0	4
Calabria	0	0	0	0	0	0
Campania	4	4	4	0	0	12
Emilia-Romagna	19	12	27	0	0	58
Friuli-Venezia Giulia	0	0	0	1	1	1
Lazio	8	5	5	1	1	19
Liguria	7	4	1	2	2	14
Lombardia	5	5	5	0	0	15
Marche	1	2	1	0	0	4
Molise	6	4	5	0	0	15
Piemonte	3	11	4	0	0	18
Puglia	0	1	0	1	1	2

Sardegna	6	6	6	0	18
Sicilia	15	7	7	4	33
Toscana	6	6	9	2	23
Trentino-Alto Adige	9	10	3	0	22
Umbria	0	3	1	0	4
Veneto	0	0	0	0	0
Totale generale	91	81	79	11	262

Tabella 1 - Ambiti tematici degli screening oncologici per Regione².

Comunicare i vaccini

La disaggregazione tematica dei materiali comunicativi dedicati alle vaccinazioni³ restituisce un quadro articolato, che rivela significative differenze regionali non solo in termini quantitativi ma anche rispetto alle scelte di focalizzazione. I materiali relativi ai vaccini contro il Covid-19 e contro il papilloma virus rappresentano le aree comunicative più presidiate a livello nazionale, riflettendo da un lato l'urgenza dettata dalla crisi pandemica e, dall'altro, una crescente attenzione verso la prevenzione dei tumori correlati all'HPV. In particolare, regioni come la Toscana, la Sicilia, la Puglia e il Piemonte si distinguono per un impegno comunicativo rilevante in quest'ultimo ambito. Anche la comunicazione sul vaccino antinfluenzale mostra una buona presenza, seppure non uniforme, con regioni come la Lombardia e il Trentino-Alto Adige che spiccano per numero di materiali, mentre altre aree, in particolare del Sud, appaiono meno attive o meno visibili online. Più limitata risulta la comunicazione dedicata ai vaccini pediatrici, fatta eccezione per alcune regioni del Centro-Nord come il Veneto e la Toscana. Ancora più esigua è la produzione di materiali rivolti a target specifici, come le donne in gravidanza o le persone con patologie croniche, con un impegno comunicativo evidente solo in pochi casi, tra cui Sicilia e Piemonte. Alcune regioni, come la Toscana o il Trentino-Alto Adige, propongono una comunicazione ampia e articolata su più fronti vaccinali; altre, come la Calabria, l'Abruzzo o la Basilicata, mostrano invece un impegno comunicativo più debole o frammentato, almeno per quanto rilevabile online. Questo insieme di elementi suggerisce l'opportunità di leggere i dati non soltanto attraverso la lente quantitativa, ma anche considerando la varietà e la coerenza tematica dei materiali, nonché la loro accessibilità e visibilità nei contesti digitali, aspetti che influenzano fortemente la capacità della comunicazione pubblica di farsi strumento efficace di promozione della salute.

<i>Regione</i>	<i>Vaccinazione popolazione specifica (gravidanza, diabete)</i>	<i>Vaccinazioni pediatriche</i>	<i>Vaccino antinfluenzale</i>	<i>Vaccino hpv</i>	<i>Vaccino covid-19</i>	<i>Totale</i>
Abruzzo	0	0	2	0	0	2
Basilicata	0	0		2	0	2
Calabria	0	0		0	5	5
Campania	1	4	6	1	1	13
Emilia- Romagna	0	9	2	3	0	14
Friuli-Venezia Giulia	0	1	1	0	0	2
Lazio	1	1	3	4	1	10
Liguria	0	2	3	2	3	10
Lombardia	0	12	12	0	1	25
Marche	1	0		0	5	6
Molise	0	1	1	0	1	3
Piemonte	3	0	4	7	4	18
Puglia	0	3	7	7	12	29
Sardegna	0	1	2	1	8	12
Sicilia	6	2	4	8	7	27
Toscana	0	5	7	21	7	40
Trentino-Alto Adige	0	1	8	4	16	29
Umbria	0	0	6	1	13	20
Veneto	0	7	7	5	2	21
Totale	12	87	114	66	102	381

Tabella 2 - Materiali relativi alle vaccinazioni per Regione

È interessante notare l'andamento delle frequenze dei materiali relativi alle vaccinazioni pediatriche nel tempo, con un solo materiale nel 2004-10, 10 nel 2011-2015, 40 tra il 2016-2020 e 36 materiali tra il 2021 e il 2023. In questo senso la crescita di attenzione sul tema può anche legarsi all'introduzione del decreto Lorenzin⁴ nel 2017.

I dati ufficiali sugli screening

I programmi di screening oncologico in Italia rientrano tra i Livelli Essenziali di Assistenza (Lea), come previsto dal DPCM del 12 gennaio 2017, e il loro andamento è costantemente monitorato grazie alle analisi condotte dall'Osservatorio Nazionale Screening (Ons) e dal sistema di sorveglianza Passi⁵. I tre screening principali (mammografico, cervicale, coloretale) seguono le linee guida aggiornate del Consiglio europeo del dicembre 2022, che puntano a offrire tali esami ad almeno il 90% dei cittadini idonei entro il 2025. Nel 2019, prima della crisi pandemica, gli indicatori di prevenzione per la popolazione adulta in Italia erano in miglioramento. Il 70,9% delle 50-69enni risultava aver fatto un controllo mammografico nei due anni precedenti (65,9% la media Ue27), sebbene con la persistenza di un forte divario regionale: al Nord quote elevate simili a quelle della Danimarca, al Sud e nelle Isole valori bassi, poco sopra quelli dei paesi dell'Europa dell'Est", (Istat, 2019). In questo senso sono evidenti alcune differenze a livello regionale. Pur consapevoli che non sussista una relazione lineare, né un nesso causale tra le azioni di comunicazione intraprese dalle istituzioni e i livelli di adesione agli screening da parte dei cittadini, non si può altresì ignorare che i materiali prodotti nell'ambito della comunicazione sociale si pongano degli obiettivi di cambiamento da raggiungere, cognitivo o comportamentale che sia (Puggelli e Sobrero, 2010). Per cui, mettendo in relazione i dati dei materiali comunicativi prodotti dalle diverse regioni con quelli dell'adesione agli screening oncologici nel 2023⁶, emerge un interessante quadro dell'intensità della produzione di materiali a livello regionale e l'adesione della popolazione agli screening. Questa correlazione non deve tuttavia essere letta come corrispondenza univoca tra numero di materiali informativi prodotti e adesione ai programmi di screening, quanto piuttosto come lo studio di due fenomeni che possono influenzarsi reciprocamente. La lettura congiunta dei dati, pertanto, va intesa soprattutto come uno strumento esplorativo utile a porre interrogativi, a "problematizzare" il ruolo della comunicazione e a orientare approfondimenti futuri.

Per leggere il dato in maniera immediata, si è deciso di mettere in relazione il numero totale di materiali comunicativi sui tre principali screening (mammella, cervice uterina e colon retto) per ciascuna regione con il tasso medio regionale di adesione agli screening per i medesimi tumori. Il grafico seguente mostra i gruppi regionali che è possibile individuare attraverso la lettura congiunta di questi dati.

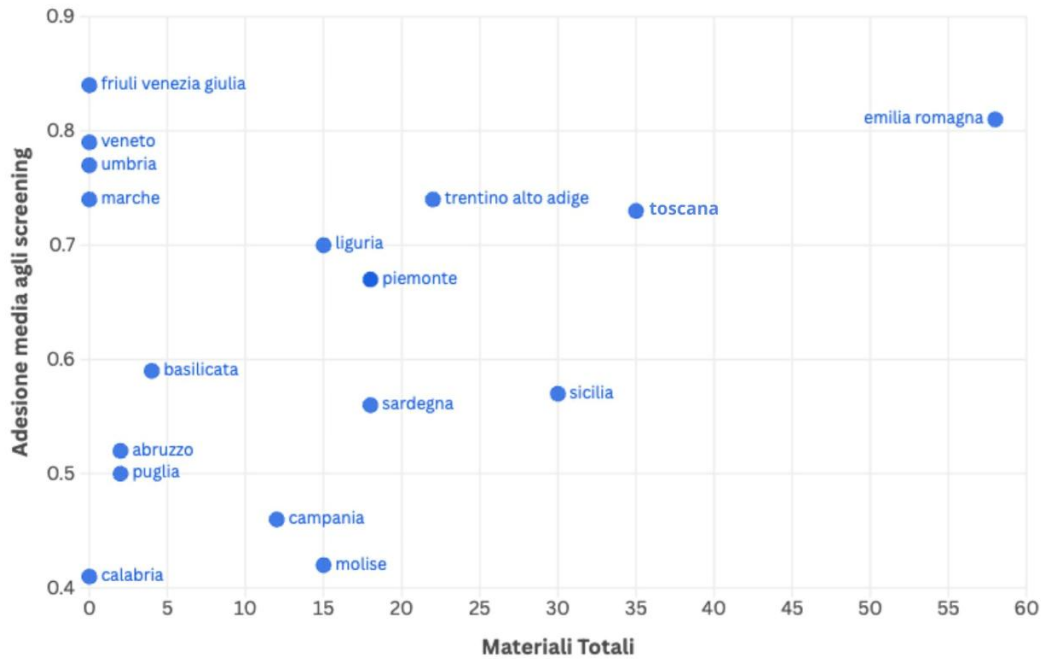


Grafico 3 - Distribuzione congiunta di materiali comunicativi e livelli di adesione agli screening oncologici.

Da questa analisi emerge una possibile articolazione delle regioni italiane in tre gruppi principali, cui si aggiungono due casi che meritano una riflessione a parte. Il primo gruppo comprende le regioni caratterizzate sia da una bassa produzione di materiali comunicativi sia da livelli contenuti di adesione agli screening oncologici: si tratta di Calabria, Puglia, Abruzzo, Campania, Molise e Sardegna. Un secondo gruppo è composto da regioni che, pur producendo un numero relativamente ridotto di materiali informativi, registrano comunque buoni tassi di adesione: ne fanno parte Friuli-Venezia Giulia, Veneto, Umbria, Marche e Liguria. Un terzo gruppo raccoglie regioni che presentano un equilibrio tra quantità di materiali informativi prodotti e livelli di adesione: è il caso di Toscana, Trentino-Alto Adige, Lazio e Piemonte. Sicilia ed Emilia-Romagna costituiscono due casi peculiari. La Sicilia si distingue per un'alta produzione di materiali, che tuttavia non corrisponde ad un pieno riscontro nei livelli di partecipazione agli screening, sebbene comunque questi risultino superiori a quelli di altre regioni del Mezzogiorno. L'Emilia-Romagna, al contrario, coniuga un'elevata intensità comunicativa a tassi di adesione tra i più alti del Paese, delineandosi come un'esperienza particolarmente virtuosa.

Analisi longitudinale dei materiali raccolti

L'analisi longitudinale condotta sui materiali informativi prodotti dalle Regioni italiane nel corso degli ultimi vent'anni restituisce un quadro articolato e disomogeneo, segnato da momenti di intensificazione comunicativa e da fasi di relativa inattività. Il dato complessivo

mostra un incremento netto a partire dal 2013, con un'accelerazione significativa nel triennio 2021-2023. Questo andamento non è però interpretabile come un semplice fenomeno quantitativo: al contrario, la variabilità osservata riflette l'intreccio di fattori politici, normativi e organizzativi che hanno inciso nel tempo sulle priorità di comunicazione pubblica in ambito sanitario. In particolare, il contesto emergenziale della pandemia da Covid-19 ha segnato un punto di svolta, generando in molte Regioni una spinta comunicativa inedita, soprattutto rispetto alla promozione vaccinale e, in parte, agli screening oncologici.

A livello regionale, emergono configurazioni differenti che vale la pena esplicitare. Si possono sintetizzare:

- Regioni con traiettoria di crescita regolare già a partire dal periodo 2013–2015, segno di un investimento costante nella promozione della prevenzione, ben prima della spinta emergenziale prodotta dal Covid. Emilia-Romagna e Toscana si confermano tra le realtà più attive e coerenti sul piano della comunicazione istituzionale. La Toscana, in particolare, raggiunge il numero più elevato di materiali prodotti a livello nazionale, a testimonianza di una strategia di comunicazione strutturata e continuativa nel tempo.
- Regioni con numeri rilevanti di materiali comunicativi ma concentrando gran parte della produzione nel periodo più recente (Trentino-Alto Adige; ma è una dinamica che si riscontra anche in regioni come Sicilia e Puglia, il cui impegno comunicativo si intensifica visibilmente solo a partire dal 2021).
- Regioni che mostrano una presenza documentale più frammentaria e tardiva. L'Abruzzo, ad esempio, non presenta materiali rintracciabili prima del 2021, con appena due prodotti informativi nel periodo più recente. Anche la Calabria si attiva tardi, e il Friuli-Venezia Giulia produce esclusivamente dopo il 2020. Il Molise, sebbene presenti un picco relativo tra il 2016 e il 2020, riduce significativamente la sua produzione negli anni successivi. Il caso del Piemonte è peculiare: registra un'elevata produzione nella prima fase (2013–2015), seguita da una marcata contrazione, suggerendo forse un disinvestimento o un cambiamento nella strategia di comunicazione.

Queste differenze non sono interpretabili solo in chiave organizzativa, ma riflettono anche modelli culturali e amministrativi differenti nella gestione della comunicazione pubblica. Alcune regioni sembrano aver integrato la comunicazione sanitaria in un disegno politico più ampio, legato alla promozione del benessere e all'empowerment dei cittadini; altre appaiono più reattive, orientate a comunicare in funzione di sollecitazioni esterne o emergenze. Vi è infine chi, pur essendo attivo, potrebbe adottare forme comunicative meno visibili o strutturate, non sempre rintracciabili attraverso i canali istituzionali ufficiali. È tuttavia fondamentale precisare che l'assenza di materiali in determinati anni o territori non implica automaticamente l'assenza di attività comunicativa. L'indagine si è basata sull'analisi sistematica dei siti istituzionali regionali e delle sezioni dedicate alla prevenzione vaccinale e oncologica; di conseguenza, è possibile che alcuni contenuti siano stati diffusi attraverso canali diversi (come materiali cartacei, conferenze stampa, eventi locali o media regionali), pubblicati su piattaforme non più accessibili, o successivamente rimossi dagli

archivi digitali. La mappatura restituita va dunque interpretata come una rappresentazione parziale ma significativa, che privilegia la tracciabilità e la comparabilità delle fonti disponibili.

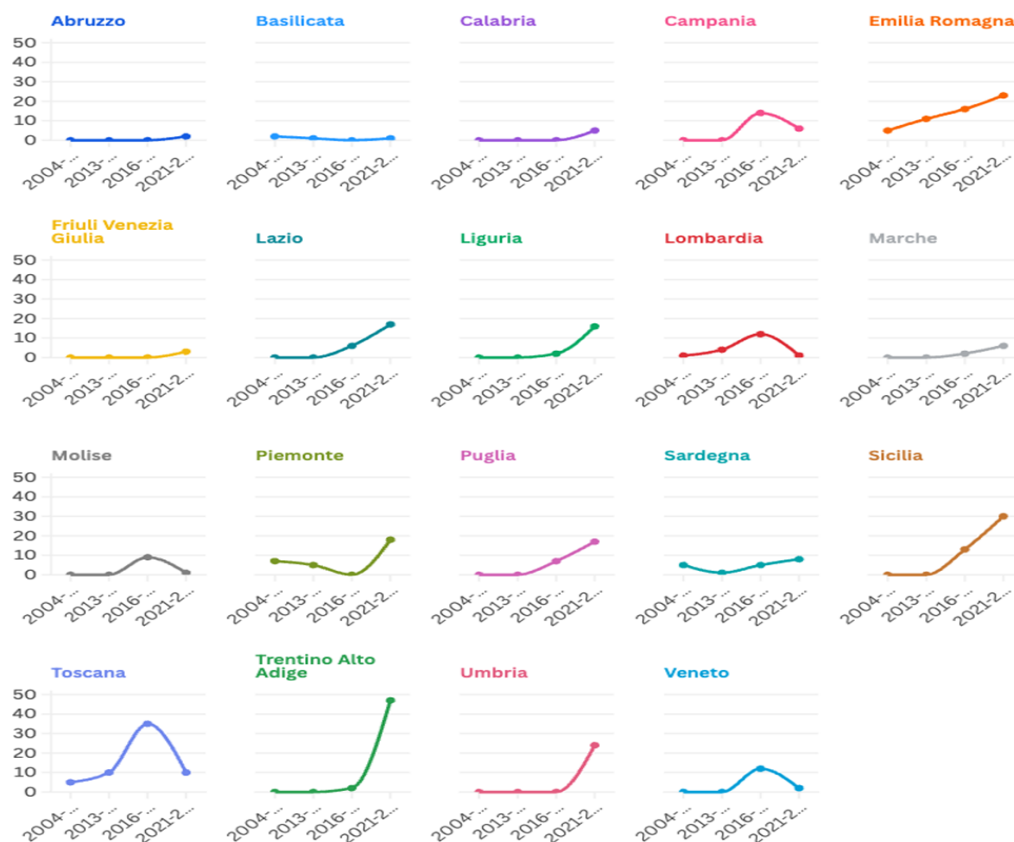


Grafico 4 - Andamento della produzione di materiali nelle Regioni Italiane.

Analisi qualitativa dei materiali raccolti

Oltre all'analisi di tipo quantitativo, si è proceduto con la lettura del contenuto dei materiali raccolti ed una loro riclassificazione in chiave più qualitativa. Per facilitare l'interpretazione dei risultati emersi, presentiamo alcuni dati suddivisi per macroaree geografiche, al fine di offrire una visione d'insieme più compatta e coerente, riducendo la frammentazione derivante dal confronto regione per regione. Tale impostazione consente di cogliere con maggiore immediatezza le tendenze comuni e le differenze più rilevanti tra le diverse parti del Paese, valorizzando gli elementi strutturali e culturali che caratterizzano ciascun contesto territoriale. Le tre macroaree individuate corrispondono a una classificazione consolidata nella letteratura sociosanitaria e comprendono: il Nord (Valle d'Aosta, Piemonte, Lombardia, Province Autonome di Trento e Bolzano, Friuli-Venezia Giulia, Veneto, Liguria ed Emilia-Romagna), il Centro (Toscana, Marche, Lazio e Umbria) e il Sud e Isole (Abruzzo, Campania, Puglia, Molise, Basilicata, Calabria, Sicilia e Sardegna). Le analisi che seguono

adottano dunque questa chiave di lettura per restituire un quadro più organico e comparabile delle strategie comunicative regionali.

I destinatari dei materiali

L'analisi della distribuzione dei destinatari dei materiali comunicativi evidenzia una tendenza complessiva delle regioni italiane a rivolgersi in via prioritaria alla popolazione generale, che rappresenta mediamente oltre la metà dei soggetti target in tutte le macroaree geografiche. Questo dato, pur attestando l'intento di raggiungere un pubblico ampio e indifferenziato, evidenzia anche una certa cautela nell'adottare strategie comunicative pienamente segmentate e orientate ai bisogni specifici dei diversi gruppi. Le donne costituiscono il secondo target per frequenza, con una presenza particolarmente rilevante nelle Regioni del Centro Italia (33,3%), a conferma della centralità delle azioni relative alla prevenzione oncologica femminile, in particolare per quanto riguarda lo screening cervicale e mammografico⁷. Altri segmenti di popolazione appaiono meno sistematicamente coinvolti: adolescenti e giovani, destinatari cruciali per le campagne vaccinali HPV, rappresentano una quota modesta, pur con un picco nel Centro (7,7%), mentre i genitori, interlocutori strategici per la prevenzione in età pediatrica, risultano target in meno del 10% dei materiali prodotti, con una prevalenza nel Nord Italia. Ancora più residuale è la presenza di contenuti rivolti ad anziani e soggetti fragili, pubblici prioritari per la vaccinazione antinfluenzale e Covid-19, che trovano spazio solo marginale, in particolare nel Sud e nelle Isole. L'attenzione rivolta agli operatori sanitari è molto contenuta, fatta eccezione per alcune iniziative localizzate nel Mezzogiorno, evidenziando una carenza di strategie mirate a rafforzare la comunicazione professionale come leva per l'alleanza terapeutica e la promozione della prevenzione. In generale, la lettura dei dati evidenzia una struttura comunicativa ancora largamente generalista, che sembra privilegiare la diffusione di messaggi universalistici piuttosto che una segmentazione calibrata dei contenuti, capace di rispondere in modo differenziato ai bisogni informativi dei diversi gruppi di popolazione.

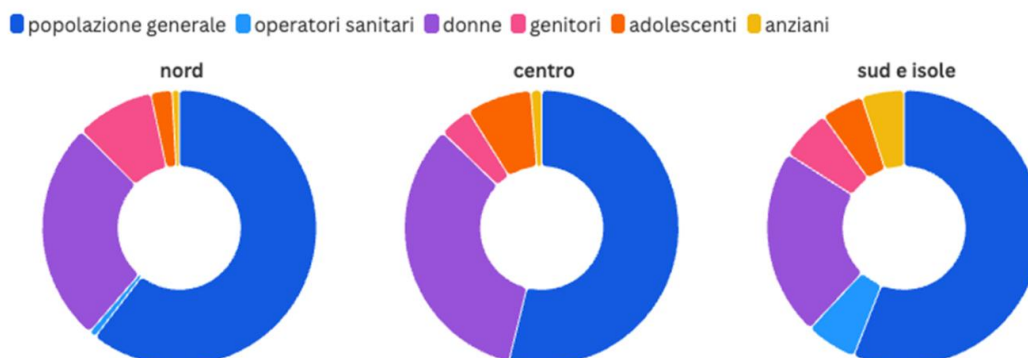


Grafico 5 - Destinatari dei materiali raccolti per aree geografiche.

L'analisi delle funzioni narrative rintracciabili nei materiali comunicativi raccolti consente di superare l'approccio puramente quantitativo e di accedere a una dimensione qualitativa capace di restituire una lettura più profonda delle strategie discorsive adottate dalle istituzioni sanitarie regionali. Il team di ricerca ha individuato cinque principali funzioni narrative a partire dalla letteratura esistente (Cucco et al., 2005; Solito e Materassi, 2021): suggerire, prescrivere, spiegare, coinvolgere e rendicontare, ciascuna delle quali risponde a una specifica modalità di costruzione del messaggio e di relazione con il destinatario. I materiali che assolvono alla funzione del "suggerire" si limitano a proporre comportamenti o scelte sanitarie desiderabili, senza tuttavia far ricorso a toni imperativi o vincolanti; la funzione "spiegare" implica invece uno sforzo esplicativo, volto a rendere comprensibili le ragioni alla base delle indicazioni sanitarie, facendo appello a dati, evidenze o processi; la funzione "coinvolgere" mira a suscitare un coinvolgimento emotivo o personale dell'utente, attraverso storie, testimonianze o linguaggi narrativi; infine, la funzione "rendicontare", benché marginale, si riferisce a quei materiali che documentano risultati, attività svolte o traguardi raggiunti, con una finalità di trasparenza o di legittimazione dell'azione pubblica. Gli unici materiali che hanno esplicitamente adottato una funzione "prescrittiva", riconducibile a toni normativi, obblighi o divieti, si rintracciano nell'immediato periodo post-pandemico (dal 2021 al 2023), seppur con una contenuta significatività.

<i>Funzione narrativa</i>	<i>Nord</i>	<i>Centro</i>	<i>Sud e isole</i>	<i>Totale</i>
Suggerire	44,44%	78,72%	69,42%	62,29%
Prescrivere	0,00%	0,00%	0,00%	0,00%
Spiegare	53,33%	19,15%	17,36%	31,71%
Coinvolgere	1,48%	2,13%	13,22%	5,71%
Rendicontare	0,74%	0,00%	0,00%	0,29%
Totale generale	100,00%	100,00%	100,00%	100,00%

Tabella 3 - Funzione narrativa per area geografica

L'analisi del linguaggio adottato nei materiali comunicativi istituzionali consente di approfondire ulteriormente le scelte discorsive che caratterizzano i materiali su vaccini e screening, restituendo una mappa delle posture comunicative privilegiate dalle Regioni italiane. I materiali sono stati classificati in base a quattro categorie linguistiche prevalenti e presenti in letteratura (Gadotti e Bernocchi, 2010): informativo, rassicurante/emotivo/empatico, fear arousing appeal e paternalistico. Ciò che si osserva è una predominanza marcata del registro rassicurante, emotivo ed empatico, presente nel 76,92% dei materiali complessivi, con un picco nel biennio 2021-22. Questo tipo di linguaggio, che mira a costruire prossimità affettiva e fiducia tra istituzioni e cittadini, appare trasversalmente diffuso su tutto il territorio nazionale, con punte leggermente più elevate nel Sud e nelle Isole (79,55%) e nel Nord (76,79%), e una lieve flessione nel Centro Italia

(73,33%). Il linguaggio informativo, che privilegia l'oggettività, la trasparenza dei dati e la chiarezza espositiva, si attesta invece su valori più contenuti (22,31% del totale), con una presenza più marcata nel Centro Italia (26,67%) e nel Nord (23,21%), e una minore incidenza nel Sud (18,18%). Marginale è la presenza di materiali che fanno uso di fear arousing appeal (0,77%), ossia di un linguaggio che fa leva sulla paura per sollecitare comportamenti protettivi. Questi compaiono esclusivamente nei materiali del Sud (2,27%) e a partire dal 2021.

<i>Linguaggio</i>	<i>Nord</i>	<i>Centro</i>	<i>Sud</i>	<i>Totale</i>
Paternalistico	0	0	0	0
Informativo	23,21%	26,67%	18,18%	22,31%
Rassicurante/emotivo/ empatico	76,79%	73,33%	79,55%	76,92%
Fear arousing appeal	0	0	2,27%	0,77%
Totale	100,00%	100,00%	100,00%	100,00%

Tabella 5 - Tipo di linguaggio per area geografica

L'analisi dei registri narrativi adottati nei materiali comunicativi raccolti offre uno spunto rilevante per comprendere le intenzioni strategiche sottese alla produzione istituzionale in ambito sanitario. I principali registri individuati, già noti nelle ricerche sulla comunicazione sociale (Cucco et al., 2005) sono: sensibilizzare, educare e informare. Ciascuno di essi esprime un diverso modo di porsi nei confronti del destinatario e di orientare la comunicazione. Il registro della sensibilizzazione, largamente prevalente nel corpus analizzato (64,39% a livello nazionale), si fonda su un linguaggio volto a stimolare la coscienza civica, l'attenzione al bene comune e il coinvolgimento emotivo, senza necessariamente attivare processi di approfondimento cognitivo o di alfabetizzazione sanitaria. Questo approccio è dominante in tutte le macroaree del Paese, con un picco nel Sud e Isole, dove raggiunge l'80,31%, mentre nel Centro e nel Nord si attesta rispettivamente al 68,70% e al 64,06%. Molto più contenuta è invece la presenza del registro educativo (11,40% complessivo), che si propone di accrescere la consapevolezza del cittadino. Anche in questo caso si osserva una distribuzione disomogenea: il Sud mostra una presenza maggiore (5,51%) rispetto al Centro (2,61%) e al Nord (2,60%), indicando una lieve tendenza ad affiancare alla sollecitazione emotiva un tentativo di strutturare messaggi dal contenuto formativo. Ancora più residuale è l'uso del registro informativo (5,84%), fondato su un'esposizione neutra e descrittiva dei dati o delle misure sanitarie.

<i>Registro narrativo</i>	<i>Nord</i>	<i>Centro</i>	<i>Sud e isole</i>	<i>Totale</i>
Sensibilizzare	64,06%	68,70%	80,31%	64,39%
Educare	2,60%	2,61%	5,51%	11,40%
Informare	3,65%	10,43%	9,45%	5,84%
Totale generale	100,00%	100,00%	100,00%	100,00%

Tabella 6 - Registro narrativo per area geografica

Al di là delle specifiche differenze dei linguaggi emerse tra le diverse Regioni, quello che va sottolineato è la loro evoluzione negli anni. I materiali restituiscono una ricchezza e un'articolazione delle forme assunte dalla comunicazione pubblica della salute oggi, soprattutto dopo la svolta digitale, e come intorno ad essa si siano create aspettative crescenti che si allontanano sempre più dalla sola esigenza di “far sapere” e informare. Oggi alla comunicazione pubblica è chiesto di rilanciare la partecipazione civica, di accrescere l'engagement del cittadino, di costruire e gestire relazioni. Obiettivi sempre più necessari - anche conseguenza di stili di vita e di consumo informativo più veloci, complessi e asistematici da parte dei cittadini.

Nell' articolazione dei temi affrontati e delle modalità narrative utilizzate, il racconto della prevenzione si snoda attraverso tre macro-dimensioni –tra loro strettamente correlate e rintracciate, peraltro, in precedenti ricerche (Solito et al. 2021) - che travalicano e intersecano temi, modalità narrative e linguaggi, che possiamo definire: *di servizio*, *conoscitiva* e *identitaria/comunitaria*. Tutto questo sembra visibile nei materiali raccolti, seppur con differenze che raccontano e confermano, anche su questi aspetti, una diversa “geografia” della comunicazione pubblica sulla salute: in termini di sensibilità, evoluzione, organizzazione, competenze.

Riflessioni conclusive e limiti dello studio condotto

L'analisi condotta evidenzia un quadro complesso e articolato delle disuguaglianze territoriali e sociali nel sistema sanitario italiano. Queste disparità, seppur radicate da decenni, continuano a influenzare la capacità del Servizio Sanitario Nazionale di garantire il principio costituzionale di uguaglianza nell'accesso alle cure e alla prevenzione. Le differenze si manifestano sia nei tempi e nelle modalità di accesso ai servizi sia nella disponibilità e distribuzione delle risorse ma anche nel sistema informativo e comunicativo sulla salute.

In termini quantitativi, ossia di produzione di materiali informativi e della loro distribuzione nel range temporale considerato, alcune regioni si contraddistinguono per un'elevata produttività e presenza continuativa, mentre altre sono caratterizzate da un'azione scarsa o discontinua; ma le differenze sono anche qualitative, relativamente all'articolazione dei temi

trattati, i linguaggi e i registri impiegati, i frame narrativi adottati. Ne è emersa una diversa “geografia” della comunicazione della salute nelle regioni italiane che a sua volta si interseca con le diverse scelte strategiche adottate dalle istituzioni sovranazionali, nazionali e locali.

I limiti, le “carenze” a cui abbiamo fatto riferimento come la frammentarietà e l’eterogeneità, la discontinuità, la scarsa integrazione e coordinamento delle azioni comunicative, capacità (o meno) di intercettare pubblici diversi e bisogni diversificati per fasce di pubblico riflettono (e rispondono) sicuramente a fattori esterni (Ministero, Enti sovranazionali), a necessità (come nel caso del Covid), indirizzi politici nazionali e locali (come fattori abilitanti o freni a certe policy e allo stesso investimento nella comunicazione della salute), a sensibilità, obiettivi, priorità e investimenti diversi (come emerge chiaramente nell’analisi longitudinale dei materiali); ma riflettono anche modelli culturali e amministrativi differenti nella gestione della comunicazione pubblica e nell’investimento fatto sulla leva della comunicazione della salute, sulle strutture comunicative, sulle competenze professionali. Tali differenze pongono pertanto interrogativi rilevanti anche sul tema dell’equità comunicativa, sull’efficacia e sulla continuità delle strategie di prevenzione adottate nei diversi contesti territoriali.

Insomma, un intreccio tra sfide nuove e irrisolti problemi storici, che non trovano soluzione nel solo attivismo comunicativo o negli entusiasmi creativi, che l’ambiente digitale asseconda e abilita, bensì necessita di investimenti organizzativi, culturali, professionali, per l’adeguamento delle strutture e delle competenze coinvolte.

Questo studio presenta alcuni limiti che meritano di essere esplicitati. In primo luogo, la raccolta dei materiali è stata condotta esclusivamente attraverso fonti istituzionali online, le cui modalità di archiviazione e aggiornamento non sempre garantiscono una completa omogeneità tra le regioni, soprattutto andando indietro nel tempo. In secondo luogo, l’approccio adottato, basato sulla riclassificazione dei materiali e l’analisi qualitativa e quantitativa dei contenuti, privilegia la dimensione testuale e simbolica, ma non consente di esplorare in modo sistematico le mediazioni organizzative, politiche e culturali che ne influenzano produzione e ricezione. In prospettiva, sarebbe pertanto utile affiancare a questa analisi documentale ulteriori strumenti metodologici, quali interviste, focus group o rilevazioni empiriche, per approfondire il ruolo degli attori istituzionali e la percezione degli stessi comunicatori.

Nota biografica

Laura Solito è Professore Ordinario di Sociologia dei processi culturali e comunicativi all’università di Firenze, dove insegna Sociologia della comunicazione e Sociologia della comunicazione pubblica e istituzionale. Tra le più recenti pubblicazioni: *Linguaggi visuali per la comunicazione pubblica digitale. Prospettive oltre l’emergenza*, in “Hermes” 2023; *The Utilisation of News in the Fourth Industrial Revolution*. in AA.VV. *Journalism, Digital Media and the Fourth Industrial Revolution*, Palgrave Macmillan, 2024.

Letizia Materassi è Professoressa Associata in Sociologia dei Processi Culturali e Comunicativi all'Università di Firenze. Insegna Sociologia della comunicazione, a livello triennale, e Comunicazione, Organizzazioni e Responsabilità Sociale alla magistrale di Strategie della Comunicazione Pubblica e Politica nella Scuola di Scienze Politiche. Sul tema della comunicazione della salute ha recentemente pubblicato: *Accorciare le distanze, costruire fiducia. La comunicazione pubblica alla prova di una pandemia 2.0* (2024), in "Pandemie. Globalizzazione, società e politica tra crisi e catastrofe", a cura di Cea R., Il Mulino, Bologna; con Laura Solito, *Tra interferenze e integrazioni. Istituzioni e attori politici su Facebook durante l'emergenza pandemica* (2022), in "Comunicazione Politica", 3, pp. 331-354.

Ester Macrì è docente a contratto presso l'Università di Firenze, tutor didattico di Metodologia della ricerca politica e sociale presso l'Università di Bologna e presidente dell'impresa sociale Retesviluppo. Tra le più recenti pubblicazioni: con Silvia Bacci, Bruno Bertaccini e Anna Pettini, *Measuring sustainability consciousness in Italy* (2024) in "Quality & Quantity", pp. 1-28.

Erika Greco è dottoranda in Peace Studies presso Sapienza Università di Roma. Da marzo 2023 a febbraio 2025 è stata assegnista di ricerca presso l'Università di Firenze e formatrice della Piattaforma LoveStorm. Collabora con la rivista La città manifesta del laboratorio politico Per Un'altra Città Firenze. Tra le più recenti pubblicazioni: Greco, E., & Scotto, G. (2024). "Love Storm". *Una proposta di formazione per la cittadinanza digitale attiva*. EPALE JOURNAL.

Bibliografia

- Albright, K., Richardson, T., Kempe, K. L., & Wallace, K. (2014). Toward a trustworthy voice: Increasing the effectiveness of automated outreach calls to promote colorectal cancer screening among African Americans. *The Permanente Journal*, 18(2), 33. <https://doi.org/10.7812/TPP/13-139>
- Anessi Pessina, E. (2022). Avanzo/disavanzo sanitario pubblico pro capite. In *Rapporto Osservasalute 2022* (pp. 385–388). Prex S.p.A.
- Armeni, P., Cavallaro, L., Costa, F., Rognoni, C., & Tarricone, R. (2024). La copertura vaccinale nelle regioni italiane: Un'analisi dell'eterogeneità. In *Rapporto OASI 2024* (pp. 521–552). Egea.
- Battistelli, F., & Galantino, M. G. (2020). *Sociologia e politica del coronavirus: Tra opinioni e paure*. FrancoAngeli.
- Bernocchi, R., Contri, A., & Rea, A. (2018). *Comunicazione sociale e media digitali*. Carocci.
- Bertolazzi, A., Lombi, L., Lovari, A., Ducci, G., & D'Ambrosi, L. (2023). I know that I know: Online health information seeking, self-care and the overconfidence effect. *Sociological Forum*, 38(3), 793–812. <https://doi.org/10.1111/socf.12925>
- Boccia Artieri, G., & Farci, M. (Eds.). (2021). *Shockdown: Media, cultura, comunicazione e ricerca nella pandemia*. Meltemi.
- Calabrò, G. E., Rizzo, C., Domnich, A., De Waure, C., Rumi, F., Bonanni, P., Boccia, S., Bechini, A., Panatto, D., Amicizia, D., Amodio, E., Costantino, C., Bert, F., Lo Moro, G., Di Pietro, M. L., Giuffrida, S., Giordano, V., Conversano, M., Russo, C., ...

- Ricciardi, W. (2024). Health technology assessment del vaccino ricombinante adiuvato contro il virus respiratorio sinciziale (Arexvy®). *Journal of Preventive Medicine and Hygiene*, 65(2 Suppl 1), E1–E159. <https://doi.org/10.15167/2421-4248/jpmh2024.65.2s1>
- Collicelli, C. (2018). Salute e sanità nel Mezzogiorno: Tendenze e criticità. *Rivista Giuridica del Mezzogiorno*, 32(3), 703–724.
- Collyer, T. A., & Smith, K. E. (2020). An atlas of health inequalities and health disparities research: “How is this all getting done in silos, and why?”. *Social Science & Medicine*, 264, 113330. <https://doi.org/10.1016/j.socscimed.2020.113330>
- Colón-López, V., Fernández-Espada, N., Vélez, C., Gonzalez, V. J., Diaz-Toro, E. C., Calo, W. A., Savas, L. S., Pattatucci, A., & Fernández, M. E. (2017). Communication about sex and HPV among Puerto Rican mothers and daughters. *Ethnicity & Health*, 22(4), 348–360. <https://doi.org/10.1080/13557858.2016.1246938>
- Corradi, L. (2001). Quando il corpo è ambiente. In *Complessità eco-sistemica e sviluppo eco-sostenibile* (pp. 173–179). Edizioni Messaggero.
- Costa, G., Bassi, M., Gensini, G. F., Marra, M., Nicelli, A. L., & Zengarini, N. (Eds.). (2015). *L'equità nella salute in Italia. Secondo rapporto sulle disuguaglianze sociali in sanità: Secondo rapporto sulle disuguaglianze sociali in sanità*. FrancoAngeli.
- Cucco, E., Pagani, R., & Pasquali, M. (Eds.). (2005). *Primo rapporto sulla comunicazione sociale in Italia*. RAI ERI.
- D'Angela, D., Polistena, B., & Spandonaro, F. (2024). *Opportunità di tutela della salute: Le performance regionali* (12th ed.). Locorotondo Editore.
- De Marvao, A., Alexander, D., Bucciarelli-Ducci, C., & Price, S. (2021). Heart disease in women: A narrative review. *Anaesthesia*, 76, 118–130. <https://doi.org/10.1111/anae.15376>
- De Nard, F., Deandrea, S., Bastiampillai, A. J., Cavazzana, L., Carnevali, D., Principi, N., Luconi, E., Schulz, P. J., Castaldi, S., & Silvestri, A. R. (2021). The development and testing of balanced communication material for a population-based breast cancer screening program. *Frontiers in Communication*, 6, 719555. <https://doi.org/10.3389/fcomm.2021.719555>
- Demu, E., Rossi, S., & De Angelis, R. (2022). Malattie oncologiche. *Rapporto Osservasalute*.
- Ducci, G. (2021). Alcuni nodi della comunicazione pubblica digitale di fronte alla pandemia: Le regioni italiane su Facebook durante il lockdown. *Mediascapes Journal*, 18, 141–160.
- Dutta, M. J. (2016). Cultural context, structural determinants and global health inequities: The role of communication. *Frontiers in Communication*, 1(5), 1–9. <https://doi.org/10.3389/fcomm.2016.00005>
- Dutta, M. J. (2021). Communication inequality, structural inequality, and COVID-19. In M. Lewis, E. Govender, & K. Holland (Eds.), *Communicating COVID-19*. Palgrave Macmillan.
- Eiriksdottir, V. K., Baldursdottir, B., Fridriksson, J. O., & Valdimarsdottir, H. B. (2022). How much information do Icelandic men receive on pros and cons of prostate-specific antigen testing prior to undergoing testing? *American Journal of Men's Health*, 16(3), 15579883221097805. <https://doi.org/10.1177/15579883221097805>
- Gadotti, G., & Bernocchi, R. (2010). *La pubblicità sociale: Maneggiare con cura*. Carocci.
- Giordani, C., & Spandonaro, F. (2015). Rapporto CREA. *Salute e Territorio*, 36(205), 613–616.

- Giuffrida, S., Giordano, V., Conversano, M., Russo, C., Spadea, A., Amicizia, D., Grammatico, F., Ansaldi, F., Costantino, C., Amodio, E., & Vitale, F. (2024). Vaccinazione nell'adulto/anziano. Regioni a confronto: L'esperienza delle Regioni Calabria, Campania, Puglia, Lazio, Liguria, Sicilia. *Journal of Preventive Medicine and Hygiene*.
- Gola, E., Meloni, F., & Porcu, R. (2018). *Comunicare la salute: Metodi e buone pratiche per le amministrazioni pubbliche*. Carocci.
- Hoover, D. S., Pappadis, M. R., Houston, A. J., Krishnan, S., Weller, S. C., Giordano, S. H., Bevers, T. B., Goodwin, J. S., & Volk, R. J. (2019). Preferences for communicating about breast cancer screening among racially/ethnically diverse older women. *Health Communication*, 34(7), 702–706. <https://doi.org/10.1080/10410236.2018.1431026>
- Istat. (2019). *Rapporto "Prevenzione e fattori di rischio per la salute in Italia e in Europa"*. <https://www.istat.it/wp-content/uploads/2021/12/REPORT-PREVENZIONE-2019.pdf>
- Kalocsanyiova, E., Essex, R., & Fortune, V. (2023). Inequalities in Covid-19 messaging: A systematic scoping review. *Health Communication*, 38(12), 2549–2558. <https://doi.org/10.1080/10410236.2022.2088022>
- Kazepov, I. (2009). *La dimensione territoriale delle politiche sociali in Italia*. Carocci.
- Kreuter, M. W., & McClure, S. M. (2004). The role of culture in health communication. *Annual Review of Public Health*, 25, 439–455. <https://doi.org/10.1146/annurev.publhealth.25.101802.123000>
- La Rocca, G. (2018). *Nuove forme di comunicazione sociale: Antifragilità, communication voice e studio di caso*. Carocci.
- Lazzara, F., Conti, F., Giuffrida, E., Eandi, C. M., Drago, F., Platania, C. B. M., & Bucolo, C. (2024). Integrating network pharmacology: The next-generation approach in ocular drug discovery. *Current Opinion in Pharmacology*, 74, 102425. <https://doi.org/10.1016/j.coph.2023.10242>
- Lovari, A. (2017). *Social media e comunicazione della salute: Profili istituzionali e pratiche digitali* (Vol. 14). Angelo Guerini e Associati.
- Lovari, A., & Ducci, G. (2022). *Comunicazione pubblica*. Mondadori Università.
- Lovari, A., Martino, V., & Righetti, N. (2020). Blurred shots: Investigating the information crisis around vaccination in Italy. *American Behavioral Scientist*, 65(2), 351–370. <https://doi.org/10.1177/0002764220910245>
- Macchione, N., & Passoni, D. (2024). *Pene, maschile plurale*. Fabbri Editori.
- Marmot, M. (2007). Achieving health equity: From root causes to fair outcomes. *The Lancet*, 370(9593), 1153–1163. [https://doi.org/10.1016/S0140-6736\(07\)61385-3](https://doi.org/10.1016/S0140-6736(07)61385-3)
- Mazzoleni, M. (2023). *Comunicare le politiche: Scenari e strumenti*. Mondadori.
- Melki, J., Tamim, H., Hadid, D., Makki, M., El Amine, J., & Hitti, E. (2021). Mitigating infodemics: The relationship between news exposure and trust and belief in COVID-19 fake news and social media spreading. *PLOS ONE*, 16(6). <https://doi.org/10.1371/journal.pone.0252830>
- Mosavel, M., & Ports, K. A. (2015). Upward communication about cancer screening: Adolescent daughter to mother. *Journal of Health Communication*, 20(6), 680–686. <https://doi.org/10.1080/10810730.2015.1012245>
- Omondi Aduda, D. S., & Mkhize, N. (2014). Ethical issues evolving from patients' perspectives on compulsory screening for syphilis and voluntary screening for cervical cancer in Kenya. *BMC Medical Ethics*, 15, 1–12. <https://doi.org/10.1186/1472-6939-15-27>
- Parrott, R. (2009). *Talking about health: Why communication matters*. Wiley-Blackwell.

- Puggelli, F. R., & Sobrero, R. (2010). *La comunicazione sociale*. Carocci.
- Ragas, D. M., Nonzee, N. J., Tom, L. S., Phisuthikul, A. M., Luu, T. H., Dong, X., & Simon, M. A. (2014). What women want: Patient recommendations for improving access to breast and cervical cancer screening and follow-up. *Women's Health Issues*, 24(5), 511–518. <https://doi.org/10.1016/j.whi.2014.06.011>
- Rosenthal, E. L., Buffington, S. D. C., & Cole, G. (2018). From the small screen to breast cancer screening: Examining the effects of a television storyline on awareness of genetic risk factors. *Journal of Communication in Healthcare*, 11(2), 140–150. <https://doi.org/10.1080/17538068.2018.1438766>
- Solito, L., & Materassi, L. (2021). If a picture is not worth a thousand words: Digital infographics use during the Covid-19 pandemic crisis. *Sociologia della Comunicazione*, 61. <https://doi.org/10.3280/SC2021-061005>
- Solito, L., Materassi, L., Greco, E., & Macrì, E. (2025). Le disuguaglianze di genere nelle campagne di comunicazione sugli screening oncologici e i vaccini degli ultimi 20 anni. *Cambio. Rivista sulle Trasformazioni Sociali*, 14(28), 159–172. <https://doi.org/10.36253/cambio-16798>
- Spandonaro, F., & D'Angela, D. (Eds.). (2015). *Una misura di performance dei SSR* (2nd ed.). Sidera Edizioni.
- Team, V., & Manderson, L. (2020). How COVID-19 reveals structures of vulnerability. *Medical Anthropology*, 39, 671–674. <https://doi.org/10.1080/01459740.2020.1830281>
- Terraneo, M. (2018). *La salute negata: Le sfide dell'equità in prospettiva sociologica*. FrancoAngeli.
- Timelli, L., Marino, M. G., Menzano, M. T., Vaia, F., & Galeone, D. (2024). I programmi di screening oncologico organizzati in Italia: Differenze tra macroaree da un'indagine su Regioni e Province Autonome (2023). *Bollettino Epidemiologico Nazionale*, 5(1), 15–22. https://doi.org/10.53225/ben_083
- Viswanath, K., & Kreuter, M. W. (2007). Health disparities, communication inequalities and eHealth. *American Journal of Preventive Medicine*, 32(5S), 131–133. <https://doi.org/10.1016/j.amepre.2007.02.012>
- World Health Organization. (n.d.). *World Health Organization*. Retrieved December 14, 2024, from <https://www.who.int/>
- World Health Organization. (2020). *WHO report on cancer: Setting priorities, investing wisely and providing care for all*. World Health Organization. https://www.who.int/health-topics/cancer#tab=tab_1
- Yadav, A., Dobe, M., Paul, B., & Taklikar, C. (2022). A cross-sectional study on assessment of perceived threat to cervical cancer using health belief model among women in a slum area of Kolkata. *Journal of Education and Health Promotion*, 11. https://doi.org/10.4103/jehp.jehp_392_21

Note

¹ L'obiettivo specifico del progetto ECoHeTE, da cui scaturisce l'operazione di raccolta dati, è quello di fornire raccomandazioni politiche efficaci e a basso costo; per questo sono state selezionate due aree di salute legate alla prevenzione.

² Il totale dei materiali divisi per Regione e tipo di screening è più alto del totale dei materiali sugli screening perché alcuni materiali riguardano contemporaneamente diversi tipi di screening. Spesso gli screening ginecologici sono presentati insieme in un unico materiale, in alcuni casi tutti e 3 gli screening principali sono oggetto di un unico materiale.

³ Anche in questo caso il totale dei materiali divisi per Regione e tipo di vaccino è più alto del totale dei materiali sui vaccini perché alcuni materiali riguardano contemporaneamente diversi tipi di vaccinazione, ad esempio Covid-19 e vaccino antinfluenzale o vaccino antinfluenzale e vaccinazioni pediatriche.

⁴ Il Decreto Lorenzin (D.L. 73/2017, convertito in L. 119/2017) ha introdotto in Italia l'obbligo di dieci vaccinazioni per i minori da 0 a 16 anni. L'iscrizione ai nidi e scuole dell'infanzia è subordinata all'adempimento vaccinale, mentre nella scuola dell'obbligo i non vaccinati non vengono esclusi ma i genitori sono sanzionati. Le vaccinazioni sono gratuite e verificate dalle ASL con l'obiettivo di aumentare la copertura vaccinale per proteggere la salute pubblica.

⁵ Il sistema di sorveglianza "Passi" dell'Istituto Superiore di Sanità si declina attraverso indagini campionarie con informazioni dalla popolazione italiana adulta, 18-69 anni, sugli stili di vita e fattori di rischio comportamentali connessi all'insorgenza delle malattie croniche non trasmissibili e sul grado di conoscenza e adesione ai programmi di intervento che il Paese sta realizzando per la loro prevenzione. I temi indagati sono il fumo, l'inattività fisica, l'eccesso ponderale, il consumo di alcol, la dieta povera di frutta e verdura, il controllo del rischio cardiovascolare, l'adesione agli screening oncologici e l'adozione di misure sicurezza per prevenzione degli incidenti stradali, o in ambienti di vita di lavoro, la copertura vaccinale antinfluenzale e lo stato di benessere fisico e psicologico, e ancora alcuni aspetti inerenti la qualità della vita connessa alla salute.

⁶ <https://www.osservatorionazionale screening.it/content/i-dati-dello-screening>

⁷ Per un approfondimento si rimanda a: Solito, L., Materassi, L., Greco, E., & Macrì, E. (2025). Le disuguaglianze di genere nelle campagne di comunicazione sugli screening oncologici e i vaccini degli ultimi 20 anni. *Cambio. Rivista sulle Trasformazioni Sociali*.

Ageing Eastwood Between Films and Media*

Roy Menarini**
Università di Bologna

This article aims at analysing the last years of Clint Eastwood's artistic production and, at the same time, the physical ageing processes of the film star, as witnessed by images and photos that circulated on the Internet between his 93rd and 94th birthdays. Thanks to this case study, we can identify more clearly some aspects hidden by media rumors: the ambivalent relationship between on-screen and off-screen; the differences that are now imposed between the third and fourth age in storytelling and media representation; the negotiation of Eastwood's masculinity and his role as a male star; the clear difference between the artistic longevity of film directors depending on whether they are men or women.

Keywords: Celebrity Culture, Star, Ageing, Masculinity, Clint Eastwood, Testament Film

* Article submitted on 06/05/2025. Article accepted on 15/12/2025. This article is part of the research activities of the PRIN 2022 PNRR project “CELEBR-ETÀ – Ageing celebrity in contemporary media and sport contexts: a model of social and cultural well-being” (project code P2022XS937, CUP D53D23019930001), funded by the European Union – Next Generation EU.

Article submitted on 25/03/2025. Article accepted on 15/06/2025.

** roy.menarini@unibo.it

The longevity of filmmakers is a well-cultivated myth in discourses relating to cinephilia and the admiration for directors. From the most striking example (Manoel de Oliveira died in 2015 at the age of 107, after having shot his last film in 2012 at the age of 104, with an active career spanning 80 years) to the most popular ones (the over-80s who have made feature films since 2020, from Francis Ford Coppola to Ridley Scott, from Marco Bellocchio to Frederick Wiseman), even despite the lack of specific statistics, there is much empirical evidence of this long-lasting presence in the film industry. Actually, it is more likely a tangible example of the positive results brought about by the continuation of one's professional activity, by the passion linked to artistic production, by the proven well-being that constant planning generates even in years of physical frailty – in addition, of course, to the increase in average life expectancy in the Western world.

The testament paradox

Nowadays, the most admired actor in this sphere is certainly Clint Eastwood, who not only made *Juror #2* (2024) when he was 94, but has also made 5 other films since he was 85: *Sully* (2016), *The 15:17 to Paris* (2018), *The Mule* (2018), *Richard Jewell* (2019), *Cry Macho* (2021). In two cases – *The Mule* and *Cry Macho* – Eastwood also took the liberty of appearing in front of the camera as the leading actor, openly denouncing his age and coming to terms – through his characters – with physical difficulties (difficulty in walking on the set, stiffness of the actor's body, profound transformation of his somatic features).

So far, no paradoxes are involved. The theoretical interest of this matter comes from another, previous film, made when he was 78, *Gran Torino* (2008). With hindsight, we can consider that title as one of many in a filmography so dense and precise that ten other productions would follow. However, the meaning of the film was different when it came out, or at least the interpretations that were drawn from it. In particular, the main thesis that it was a so-called “testament film” became popular.

Testament film is a critical category that is used when an author decides to make a film that appears as a summa of their career, either because it touches on most of the themes that characterised it, or because it reflects on the passage of time, or because it openly cites their previous titles. This testament film also evokes a sense of anxiety and a loss of cinematic values, as well as the feeling of a disappearing world (Keller, 2020). Some examples of testament films are *Porto da minha infância* (2001) by Manoel de Oliveira, *Deconstructing Harry* (1997) by Woody Allen, *Il sol dell'avvenire* (2023) by Nanni Moretti, just to mention a few directors of different nationalities who are well known on the international market. As to the first film, de Oliveira, who was 93 years old at the time, recalls his childhood home in Oporto and reflects melancholically on the passage of time, on memories, on the generations that fade away and on the homes that somehow distinguished them:

Understanding the past is crucial to creating the present. In this way, the addition to the spectral image of the childhood home of a speech by the autodiegetic narrator (the filmmaker), extracting from the rubble of

memory their own experience of that place, produces the oneiric sense that the birthplace represents. (Barini Camargo, 2020, p. 96, tr. Valentina Cristiani)¹.

Similarly, *Deconstructing Harry* tells the parable of an intellectual with a personal life that is contradictory at the very least, travelling to his old university that wants to honour him as an *alumnus*, whilst dealing with memories, fantasies and reveries that generate small episodes and stories (some of which are taken from Allen's early literary works). The coexistence of quotes from Ingmar Bergman, references to his own filmography, inside jokes and meditations on his personal memory, led to talk of a testament film (Bocchi, 2010, pp. 123-141).

Nanni Moretti's case is even more obvious. In addition to the frequent self-references he makes, the last sequence is conceived as a march made up of many of the characters (and therefore actors) from his previous films, who all find themselves together, led by the same actor/director, and walking towards an indistinct destination (again a kind of *summa* of other endings of Moretti's films, in particular *Palombella rossa*) (Di Paolo, 2023, pp. 173-177).

The same paradox applies to all the aforementioned examples: the authors who publicly made their "artistic testament" have then – happily and successfully – continued their career with many other productions, downgrading what appeared to be mournful and definitive to "ongoing evaluations". Upon closer inspection, we can therefore confirm that this cinematic longevity is a result of the biological longevity achievements. The authors' ageing process was so well known to them that they wanted to dedicate – ahead of their time (as we only understood *afterwards*, however) – a summary work on their career and poetics.

Clint Eastwood: a case study

But let's go back to Eastwood and his *Gran Torino*. The film tells the story of the redemption of Walt Kowalski, a grumpy old man full of prejudice, recently widowed and with a terrible relationship with his sons. Isolated and aloof, Walt is frustrated by the multicultural transformation of his neighbourhood and hates his neighbours, in particular some Hmong families, with whom Walt has a conflictual relationship, fuelled by his obvious racism. The only thing he seems to care about is the car that gives the name to the feature film, a 1972 Ford that he looks after as if it were a living being. But the car is nearly stolen by Thao, a young neighbour, who is forced by a local gang to try and steal the Gran Torino as a rite of passage. Walt figures out the situation, slowly comes to understand his neighbours' rites and customs, faces his inner demons (the Korean War, not surprisingly) and ends up protecting Thao and his sister from the criminals threatening them. With the usual distrust of the authorities and the rules of ordinary justice that is typical of Eastwood's films, the protagonist Walt avoids involving the police and acts on his own. Old and frail, he can't challenge the criminals on a physical level, so he devises a clever but self-sacrificing plan: he provokes the gang and gets himself killed, thus ensuring that the criminals are arrested by the police.

This is the plot, which alone seems to deal with the long debate on Clint Eastwood's conservative ideology – actually recently questioned by critical reception (Ciotta, Silvestri, 2023) – to which are added narrative aspects that emphasise the transformation of the typical character of the actor/director (generally a tough man, uncompromising, contemptuous of danger, intolerant of injustice, sensitive to private and personal justice, bound to the traditional values of the American nation) (Engel, 2007). All recognisable features that can be found in Walt, with the difference that he is too old to be able to assert his personality by physical force and who is going through a process of historical and civil “re-education” – although still based on direct knowledge of others, as real people and not social or ethnic categories.

The key sequence, the one in which Walt openly confronts the gang, was considered highly symbolic. The man, ready for the ultimate sacrifice, pretends to draw a gun (perhaps the Magnum 44 made mythical by the *Dirty Harry* saga) but in the end he only uses his index finger and thumb to mime the presence of the weapon. Once shot down by enemy fire, Walt falls and his body lies down like a cross, in the Christological sense. A complete reversal: just as the characters created by Eastwood were known for their dexterity in the use of guns and for the indifference with which they killed the “bad guys”, Walt gets himself killed unarmed because he knows that the only way to defend his young Korean friends is to sacrifice himself.

According to Giulia Carluccio:

the Eastwood/Kowalski character moves all the time like a zombie, as do the others, especially and significantly the antagonists [...] are caricatures, images, spectres of a theatre that sometimes has the lateral features of the stage, as in the final sequence, of Walt's sacrificial death, witnessed by neighbours from their windows, through gestures and actions that are reduced to simulacra, to ghosts of gestures and actions, and have nothing real or realistic about them. (Carluccio, 2009, pp. 18-19, tr. Valentina Cristiani)².

And again: “The return of Eastwood himself as he plays and re-plays his character, bringing his own real aged, vulnerable (and irresistible) physicality into play, encourages the spectator to identify with him” (p. 21, tr. Valentina Cristiani)³. And even French film critics emphasise the self-reflexive nature of the text and Eastwood's ability “to articulate its classical power [...] with the creation of an extraordinarily complex mental space” (Malausa, 2009, p. 20, tr. Valentina Cristiani)⁴.

These critical-interpretative sources are elaborated a few days or a few months after the film's release, and precisely because they are expressed within an ongoing process of film ageing of which they cannot predict the outcome or possible consequences (not only regarding the health and longevity of the director but also in terms of a possible retirement from the scene, at least as an actor), they are even more useful to us than others.

Eastwood's Performance

To broaden the discussion on Eastwood, however, we should go beyond the symbolic meanings of the plot, the intertextual play with characters from the past, or the frankness with which he portrays the vulnerability of the “great old man” (the great old men being, according to gerontology, people over 85 years of age, although recently the categories have been easily delayed, so that old age now begins at 75 rather than 65, the fourth age at 85, and the great old over 90). The representation also passes through directing the actor (and therefore himself as a figure who works in front of and behind the camera) and the acting performance, with body, posture and gestures.

Between 2008, the year *Gran Torino* was released, and *Juror #2*, 16 years passed in which Eastwood repeatedly returned to the passage of time and the age of the protagonists. He does this mainly in the feature films in which he appears as an actor, as well as a director, such as *The Mule* and *Cry Macho*, respectively shot at the age of 88 and 91. In the former, he plays a very old flower grower, forced by the decline of his business to accept an offer to transport drugs for some traffickers – using his old and unsuspecting pick-up – in exchange for the money he desperately needs. In the latter, he plays a broken-down former rodeo champion who, despite being full of the aches and pains of old age, goes looking for the thirteen-year-old son of a colleague who has got mixed up in shady dealings in Mexico. In both films, not only is Eastwood's physical ageing shown and filmed on screen with no frills or euphemisms, but the stories themselves are marked by the walking and physiological difficulties of both characters (Canadè, 2021).

A brief analysis of two sequences will be useful. In *The Mule*, at one point the elderly protagonist comes across a group of motorcyclists who have parked their bikes and are trying to figure out what is wrong with one of them, which has broken down. Leo is attracted by the gathering and the noise of the engines, and immediately engages them in conversation. The camera observes him in full shot as he walks slowly towards the bikers, hunched over and holding a container in his hand. His baseball cap, short-sleeved shirt and slow gait immediately identify him as an elderly time-waster, eager for social interaction (the film actually tells the story of how this elderly man unexpectedly becomes a drug courier involved in daring adventures, with an obvious reversal of expectations). Leo calls the person bent over “son” but soon realises that they are all women. At that point, he calls them “gals”, a dismissive and paternalistic term. The motorcyclists rebuke him for the term and introduce themselves as “Dykes on Bikes”. As we know, for a long time, “dyke” was considered a derogatory term for lesbian women who looked masculine because they did not conform to gender stereotypes. Its reappropriation (as with queer and other insulting terms) is the result of an ironic but powerful claim to the political value linked to identity, to the point of overturning the symbolic violence that surrounded the word. Leo walks away, surprised and amused, looking like someone who feels cut off from modernity, but he does not give up – stepping away with a smile – and helps them by giving advice on the technical problem with the motorbike, which he seems to know well.

The sequence is very sophisticated. We have an elderly white man, apparently traditionalist and conservative (who actually takes on the role of outlaw in the rest of the story), who is mistaken for an elderly male stereotype by the lesbian motorcyclists. At the same time, however, Leo perpetuates the image he has been assimilated into through the most classic mansplaining about engines, by using a vocabulary that is unsuitable for the times. Although the communicative imbalance seems to confirm these symbolic social structures, the humanity of both Leo and the women (who are, all things considered, well-disposed and not aggressive towards him) establishes fertile ground for an albeit rapid confrontation.

In *Cry Macho*, on the other hand, filmed even more recently and with even more evident ravages of age, Eastwood operates in the opposite way. His character is, even in his clothing, close to the icon through which we have come to know him. The cowboy hat, jeans, boots and Texan shirt accompany Mike – a battered and weary former rodeo star – for most of the film. In this way, by bringing the star back within the Tex/Mex border and Western iconography, the contrast between the physical and cynical strength of the classic Clint and his equally contemptuous but vulnerable version is emphasised. Eastwood, as director, decides to maintain the stylistic codes of the past, framing himself in medium shots, close-ups or sometimes extreme close-ups (where his physical decay is most apparent). On the other hand, however, it is once again the story that claims a symbolic space for the old hero, tasked with a rescue operation that is anything but simple – helping a young boy in a geographical area dominated by gangs and crime – and offering the protagonist even a fistfight and a romantic seduction. *Cry Macho* thematises Mike's de-escalation from machismo, right from the title. The road movie between the elderly saviour and the young man rescued from his destiny in petty crime is also, allegorically, a journey around masculinity and its most harmful expressions. In this sense, the journey that began with *Gran Torino* reaches a complete revision of the concept in *Cry Macho*, almost a retrospective commentary on the actor's filmography, without this leading to repentance or renunciation, but rather to a lucid examination of what ageing brings, in terms of artistic wisdom and strategies for representing the performing body.

Ageing Eastwood in the media

But if in this way we can reconstruct the artistic choices of the American author, and therefore the filmic methods through which he personally manages his cultural capital and his appearance (especially taking into account the fact that he directs himself and therefore is not obliged to negotiate with other subjects behind the camera), we must now explore some episodes in which Eastwood's star persona was not managed directly by the director or the film medium.

In particular, it is worth focusing on the issue of Eastwood's ageing outside the set. With some preliminary remarks.

As a film star, Eastwood has capitalised on a conservative, binary, virile masculinity linked to patriarchal schemes (Cornell, 2009), also emphasised in various interviews and in his political commitment to the Republican Party. In a famous interview given to *Esquire* in 2016, Eastwood declared that he wasn't too close to Trump's ideas but certainly less distant to those than to the Democrats, who he felt were too politically correct. And, to demonstrate how politically incorrect he was, he explained the state of US culture at that time:

Secretly everybody's getting tired of political correctness, kissing up. That's the kiss-ass generation we're in right now. We're really in a pussy generation. Everybody's walking on eggshells. We see people accusing people of being racist and all kinds of stuff. When I grew up, those things weren't called racist (Michael Heaney, 2016).

In fact, the films played and directed by an older Eastwood wouldn't be that different from the others:

He continues to play violent vigilantes reinstituting order, even when his characters begin to creak and crave a quieter life. Enormous effort is made to shore up each male star's increasingly fragile sense of virility at the expense of others, usually women dismissively treated as potential or rejected sexual objects and racialized men easily killed off. While I am far from arguing that white patriarchy is losing its grasp on hegemony, these films reveal a social concern about that possibility and worry that the aging of society will clinch its demise (Chivers, 2011, p. 102).

Compared to this pattern of reception of the Eastwood figure, images of the ultra-nonagenarian Eastwood taken from his family, by the paparazzi or in other public situations show a much greater fragility than those in the films, even if they too deal with the issue of ageing. In fact, if the titles we mentioned above can be considered strategic negotiations between the actor and the coherence of his typical character, or between the symbolic dimension of the aged Eastwood star figure (Cohen-Shalev, 2009) and the story of the effects that time has on his image, the feeling is different when looking at shots where apparently there is no symbolic or narrative balance.



Figs. 1 e 2 - Clint Eastwood in public events in 2024, published on LinkedIn

The photos taken in April 2024, initially posted on LinkedIn by a bystander, show a public event in Carmel, California (the city where Eastwood lives and where he was mayor), where the 90-year-old anthropologist Jane Goodall made an appearance for the Sunset Cultural Center. To put it bluntly, in this case Eastwood seems to be dressed in a casual and careless way, with thick and wild hair and beard, with a mix of untidiness and neglect that reminds us of the guests of retirement homes.

Perhaps to rebalance the “stolen” image, Eastwood's daughter (Alison, also an actress) then posted a series of photos of her father celebrating his 94th birthday that May, on social media.



Figs. 3 4 - Clint Eastwood in photos taken by the daughter Alison in 2024 on Instagram

In particular, in the photo on the left, Alison Eastwood writes on Instagram: “I'm not sure what we were laughing about, but I always laugh with this guy #dad May” (2024).

While the photo on the left shows Eastwood wearing the same casual clothes, the one on the furthest right shows him wearing a jacket and tie for the celebration, presumably in a restaurant or some other public place.

Another type of photo in which Eastwood has appeared recently (and which he has authorised for distribution and sharing) is linked to more recognisable and decidedly more popular coolness effects, as they involve the meeting of old glories (one more famous, while the other more niche, but both iconographically striking). Here are the two pictures of Eastwood published by the legendary bass player Leland Sklar (long-time collaborator of James Taylor and Crosby, Stills, Nash and Young).



Figs. 5 e 6 - Pictures of Clint Eastwood in private life

In the one on the right, the star's ironic and unsettling attitude stands out, as if to show (it is 7 July 2024) a completely different vitality from that glimpsed on previous occasions, where the overly domestic traits of the great icon had somewhat disappointed fans. The thousands of likes under the photo of Eastwood giving the finger on Sklar's Facebook page, and the enthusiasm aroused by the actor's arrogant gesture, show that his closeness to the world of rock and his message of “maverick” irreverence towards the system revitalise the star's image and restore him to the role of proud individualist that he has always been known for.

Moreover, there is no evidence to confirm the difference between “stolen” and authorised photos, given the fragmented and confusing flow of images on the Internet. Everything could be reconsidered in light of Eastwood's conscious and extremely lucid self-narrative, which – although he is clearly not very familiar with digital marketing or self-promotion strategies – is nonetheless likely interested in confusing his audience's expectations, consciously playing with his fandom, and creating further connections with fictional films, always in the context of ageing and its various stages.

We are therefore faced with a continuous swing, where the on-screen and off-screen discourse continues seamlessly, depending on the context. After all, the instability of the star figure and his physical suffering are aspects that bounce back and forth between the screen and the news, in a complex terrain of symbolic exchange that involves stardom and masculinity (Scandola, 2017, pp. 7-20).

Open issues

As Deborah Jermyn pointed out, these public or mediatised ageing processes cannot hide gender disparity. Clint Eastwood is a man who, through his white masculinity and adherence to the conservative values of a white, traditionalist America, has developed a cultural and symbolic approach that has often transcended such values with intelligence, yet without ever eliminating or renouncing them. The processes of female ageing are way different: “Ours is

a culture where the representation of ageing and 'older' female celebrities is by turns seemingly hopeful and newly affirming one moment and destructive and retrograde the next" (Jermyn, 2012, p. 22).

In Eastwood's case, according to Philippa Gates, there isn't too much contradiction between on-screen and off-screen: "The star's aged face and body have been presented as a spectacle for the audience to behold; in doing so, his body's physical vulnerability is offered to, and read by, audiences as an emotional one" (P. Gates, 2012, pp. 169).

If anything, given that the process of ageing up negotiated with his fans and audience goes back to the nineties and *The Unforgiven* (1992), we should perhaps distinguish two stages of old age for the star: the third and the fourth age. If *Gran Torino* assumes the role of a testament film for a longer period that began in the previous decade – where we collectively face a new phase of the representation of the hero – we can define it no longer as a conclusive work that only a good set of chromosomes has been able to extend to the present day, but rather as an explicit turning point from the third to the fourth age (Dolan, 2018). The fourth age is the result of a technical society: medical and technological progress, social inclusion, cosmetic development, processes dedicated to the psycho-physical well-being, widespread awareness of nutrition and nutraceuticals (Higgs, Gilleard, 2015). These processes began in the 1970s and have been apparently unstoppable for advanced Western societies up to now.

Obviously, in the case of male celebrities, this long passage is different from that of female celebrities (Bruzzi, 2013; De Rosa, Mandelli & Re, 2021). Let's imagine if the same neglect (observed in some of Eastwood's shots) had been noticed in a female star of the same age. It is likely that the comments would not have simply swung between affectionate support and embarrassed silence, but they would have criticised the appearance of the subject photographed – perhaps even arguing for the very need not to publish the picture. And this brings us back to where we started, namely the legendary longevity of cinema professionals, both in terms of age and artistic career. In this case too, the gender inequality in cinema hits women hard. The planning skills that are so beneficial for male directors can very rarely be extended to female directors, as very few have been able to establish themselves in the years when Eastwood was making a name for himself and therefore continue their career after their 80s or 90s. It is no coincidence that, in debates on the subject, only Agnès Varda is ever mentioned. She died in 2019 aged 91 and was the director of four feature films between 2008 and 2019.

In conclusion, it is only by starting from the features of his masculinity and without ignoring a widespread privilege that the male film star (and his forms of celebrity translated into digital media) has been able to enjoy, that it is possible to understand Clint Eastwood's public relationship with old age and the way he has chosen to include it in the themes and storytelling of his films.

Biographical note

Roy Menarini is Full Professor of Pemm/01-B at the University of Bologna, where he teaches Cinema and Cultural Industry and various other courses. He is senior editor of the academic journal *Cinergie – Il cinema e le altre arti* (and directs the series of the same name for Mimesis) and director of the academic journal *Zone Moda Journal*, both of which are ANVUR A-level subjects in some areas. He is Principal Investigator of the *PRIN 2022 PNRR Celebr-age* and Unit Manager of the *PRIN 2022 Reframing Film Festivals*. He has written numerous books on contemporary cinema, film criticism, cinephilia, film analysis and the poetics of authors. He collaborates with various institutions and is on the scientific committee of many magazines and series. He is a consultant for the Fondazione Cineteca di Bologna and coordinator of media education courses in the context of the ministerial project “Cinema and Images for the School”.

References

- Abruzzese, A. (1994). *Elogio del tempo nuovo: Perché Berlusconi ha vinto*. Costa & Nolan.
- Anania, F. (2003). *Immagini di storia: La televisione racconta il Novecento*. RAI Libri.
- Appadurai, A. (2013). *The future as cultural fact: Essays on the global condition*. Verso. (Italian translation published 2014 as *Il futuro come fatto culturale: Saggi sulla condizione globale*, Raffaello Cortina)
- Barini Camargo, F. (2020). Visit or memories and confessions: Manoel de Oliveira's film-testament. *Revista do CESP*, 40(64), 95–110. <https://doi.org/10.17851/2359-0076.40.64.95-110>
- Bocchi, P. M. (2010). *Woody Allen: Quarant'anni di cinema*. Le Mani.
- Bruzzi, S. (2013). *Men's cinema: Masculinity and mise-en-scène in Hollywood*. Edinburgh University Press.
- Canadè, A. (2021). *Prima che il gallo canti: Cry Macho di Clint Eastwood*. Retrieved December 31, 2025, from <https://www.fatamorganaweb.it/cry-macho-di-clint-eastwood/>
- Carluccio, G. (Ed.). (2009). *Clint Eastwood*. Marsilio.
- Chivers, S. (2011). *The silvering screen: Old age and disability in cinema*. University of Toronto Press.
- Ciotta, M., & Silvestri, R. (2023). *Spettri di Clint: L'America del mito nell'opera di Eastwood*. Baldini+Castoldi.
- Cohen-Shalev, A. (2009). *Visions of aging: Images of the elderly in film*. Sussex Academic Press.
- Cornell, D. (2009). *Clint Eastwood and issues of American masculinity*. Fordham University Press.
- De Rosa, P., Mandelli, E., & Re, V. (Eds.). (2021). *Aging girls: Identità femminile, sessualità e invecchiamento nella cultura mediale italiana*. Meltemi.
- Di Paolo, P. (2023). Questo è fare un film! *Bianco e Nero*, 606, 173–177.
- Dolan, J. (2018). *Contemporary cinema and “old age”: Gender and the silvering of stardom*. Palgrave Macmillan.

- Engel, L. (2007). *Clint Eastwood: Actor and director*. University of Utah Press.
- Gates, P. (2012). A good vintage or damaged goods? Clint Eastwood and aging in Hollywood film. In L. Engel (Ed.), *New essays on Clint Eastwood* (pp. 168–189). University of Utah Press.
- Gilleard, C., & Higgs, P. (2015). *Rethinking old age: Theorising the fourth age*. Bloomsbury Academic.
- Heaney, M. (2016). Clint and Scott Eastwood: No holds barred in their first interview together. *Esquire*. Retrieved December 31, 2025, from <https://www.esquire.com/entertainment/a46893/double-trouble-clint-and-scott-eastwood>
- Jermyn, D. (2012). Get a life, ladies. Your old one is not coming back: Ageing, ageism and the lifespan of female celebrity. *Celebrity Studies*, 3(1), 22–25. <https://doi.org/10.1080/19392397.2012.644708>
- Keller, S. (2020). *Anxious cinephilia: Pleasure and peril at the movies*. Columbia University Press.
- Malausa, V. (2009). L'épreuve de force. *Cahiers du Cinéma*, 642, 19–22.
- Scandola, A. (2017). *Hollywood men: Immagine, mascolinità e performance nel cinema americano contemporaneo*. Kaplan.

Notes

¹ “A compreensão do passado é crucial para a criação do presente. Desse modo, o acréscimo à imagem espectral da casa da infância de um discurso do narrador autodiegético (o cineasta), extraindo dos escombros da memória a própria experiência vivida naquele lugar, produz o sentido onírico que a morada natal representa” (Barini Camargo, 2020, p. 96).

² “il personaggio di Eastwood/Kowalski si muove per tutto il tempo come uno *zombie*, così come anche gli altri, soprattutto e significativamente gli antagonisti [...] sono caricature, immagini, spettri di un teatro che talvolta ha in effetti le caratteristiche laterali del palcoscenico, come nella sequenza finale, della morte sacrificio di Walt, vista dai vicini dalle finestre, attraverso gesti e azioni che sono ridotti a simulacri, a fantasmi di gesti e azioni, e non hanno nulla di reale o di realistico” (Carluccio, 2009, pp. 18-19).

³ “Il ritorno dello stesso Eastwood a interpretare e reinterpretare il suo personaggio, mettendo in gioco la propria e reale fisicità invecchiata, vulnerabile (e irresistibile), spinge l'identificazione dello spettatore” (p. 21).

⁴ “articuler sa puissance classique [...] à la mise en place d'un espace mental d'une extraordinaire complexité” (Malausa, 2009, p. 20).

Mediascapes journal 26/2025

Copy, Paste, Create. Authorship, Creative Labor, and Intellectual Property after Generative AI*

Federico Pilati**
Università di Bologna

Maria Tartari***
CNR - Istituto di Scienze del Patrimonio Culturale

Artificial Intelligence is reshaping cultural and creative industries by challenging long standing paradigms of creativity. Tracing a trajectory from traditional models to the current digital landscape, this article argues that generative AI represents more than a technological innovation—it is also a renewed manifestation of participatory and decentralized creative practices. By examining ongoing changes and challenges through the interconnected lenses of authorship, labor, and intellectual property, our inquiry reveals how AI simultaneously democratizes and disrupts traditional notions of cultural production, suggesting a transformative shift with profound implications for both artists and publics.

Keywords: Artificial Intelligence, Computational Creativity, Authorship, Intellectual Property

* Article submitted on 24/12/2024. Article accepted on 28/08/2025.

** federico.pilati2@unibo.it

*** maria.tartari@ispc.cnr.it

Introduction

The trajectory of cultural expression and creation has been characterized by significant paradigmatic shifts, each intricately linked to technological and societal transformations. From the collective oral traditions of preliterate societies to the individualistic ethos of the industrial age, and now to the emerging digital paradigm influenced by Artificial Intelligence (AI), the conceptualization and praxis of creativity continue to evolve, challenging established notions of authorship, originality, and cultural value (Boden, 2004).

Digital technologies have facilitated unprecedented ease in remixing and recontextualizing existing content (Jenkins, 2006). Lessig (2008) in 'Remix' argues that this read/write culture represents a return to a more participatory form of cultural creation, challenging the read-only culture that dominated much of the 20th century. Similarly the very idea of open, collaborative and accessible use of coding in projects like Wiki and GitHub have facilitated new forms of collective authorship (Bruns, 2008). Moreover, the ubiquity of smartphones with high-quality image sensors and intuitive editing applications has further democratized content creation. To this extent authors like Manovich (2018) argue that these technologies have given rise to new aesthetic forms and practices, blurring the demarcation between amateur and professional content creation.

These observations serve as a point of departure for our examination of AI's role in creative processes and its potential to innovate cultural production. Indeed, AI can be seen as a technological response to a social demand for collective creativity that has existed for decades. This is reminiscent of the early 20th century, when technologies like the phonograph, radio, and cinema responded to the need for easily reproducible and accessible sounds and images, as scrutinized by Benjamin (1935) in 'The Work of Art in the Age of Mechanical Reproduction'. However, the scale and nature of AI's impact is unprecedented. Unlike previous technological innovations that primarily affected the reproduction and distribution of cultural works, AI has the potential to be directly involved in the creative process itself. This raises profound questions about the nature of creativity and the role of human agency in cultural production. As asked by Esposito (2022), can Artificial Intelligence be genuinely creative, or are so-called generative technologies merely tools amplifying human creativity?

Moreover, the collective feedback loops embedded in AI-assisted creativity challenges our individualistic notions of genius and artistic expression that have dominated Western thought since the

Renaissance (Simonton 2000). AI's integration into cultural creation represents both a rupture and a continuity with historical shifts in creativity. While it democratizes access and enables new modes of expression, it also forces society to confront enduring questions about authorship, value in cultural labour and intellectual property.

By examining the emerging modes of AI-assisted creativity, this article aims to focus on how AI is reconfiguring our understanding of authorship and creativity, leading us towards new forms of cultural creation that blur the boundaries between individual and collective,

human and machine agency. To do so we will analyse AI's integration into cultural production based on three interconnected dimensions. Firstly we will highlight the uses of AI as a creative tool. AI tools have indeed the potential to democratize creativity by lowering the technical barriers to entry in various creative fields while the use of AI in cultural production challenges traditional notions of authorship. Secondly, we will focus on creative labor markets, as traditional organizational models of cultural production are being reshaped by AI. Finally, we will study how the collective and AI-assisted nature of emerging forms of creativity poses challenges to the existing intellectual property legal framework.

Are AI technologies reconfiguring the creative ecosystem?

Despite the significant challenges, the integration of AI into cultural production signals a transformative moment, not merely a technological evolution but a deep reshaping of social, cultural, and creative paradigms. Generative AI operates as a synthesizer of collective human knowledge, trained on vast datasets that distill the creative contributions of diverse, often unacknowledged, individuals. This shift suggests that cultural production is increasingly moving away from hierarchical, top-down structures toward more distributed, participatory, and networked forms of creation. In this context, authorship is no longer a static attribution but becomes an accessible, fluid, multi-layered phenomenon involving human actors, algorithms, and the datasets that underpin AI systems.

This perspective resonates with posthumanist theory, which challenges anthropocentric models of agency. Haraway (1991) famously proposed the “cyborg” as a hybrid entity that destabilizes boundaries between human and machine, while Braidotti (2019) highlights how posthuman subjectivity emerges from relational entanglements with technological and ecological others. Likewise, software studies scholars such as Fuller (2008) and Chun (2011) remind us that algorithms and code are not neutral tools but cultural agents embodying particular assumptions, values, and histories. Placing AI creativity within this lineage reveals its situatedness: generative systems are relational artifacts, co-produced by human labor, cultural archives, and computational infrastructures.

However, this democratization must be critically examined through the lens of its unintended consequences. While AI can enhance individual creativity, it may paradoxically reduce collective diversity in content creation (Doshi & Hauser, 2024). The algorithmic training of AI systems often amplifies dominant cultural patterns, potentially marginalizing alternative narratives and perspectives. Addressing these biases is crucial for ensuring that AI-facilitated creativity does not erode the richness of global cultural diversity.

Generative AI techniques, such as text-to-image models, also play an increasingly prominent role in fostering collective creative exploration. These technologies can facilitate collaborative speculation and community engagement by enabling diverse groups to envision and discuss future scenarios. Indeed, AI holds the potential to redefine how collective intelligence operates, bridging gaps and creating synergies between human and

machine cognition. Verhulst (2018) argues that AI can enable collective intelligence to scale, while collective intelligence in turn, humanizes AI by embedding it in systems of shared meaning and social interaction. Leveraging AI for collective intelligence could transform how we address complex global challenges, from climate change to urban planning, by augmenting individual cognition and enhancing the functioning of collaborative systems (Evans et al., 2021).

In this context, AI acts not as a substitute for human creativity but as a tool that amplifies the capacity for collective problem-solving and innovation. The unexpected divergences between human-imagined and AI-generated outputs often spark new insights and lateral thinking. These moments of cognitive dissonance serve as creative catalysts, encouraging participants to refine and expand their ideas, making generative AI an invaluable tool in participatory design and speculative innovation. For example, community workshops might use AI-generated visualizations to reimagine urban spaces, allowing stakeholders to engage in dynamic and iterative processes of co-creation. These augmentations offer exciting possibilities but also demand a nuanced understanding of human agency in processes mediated by algorithmic systems (Esposito 2022).

In their chapter *Artificial Intelligence and Collective Intelligence* from the *Handbook of Collective Intelligence*, Weld, Lin, and Bragg (2015) present a compelling case for the integration of AI in crowdsourcing, emphasizing its potential to unlock new possibilities for collective intelligence. By exploring the growing intersection between artificial intelligence (AI) and crowdsourcing, highlighting how these two domains can complement and enhance each other. They identify two key connections: the use of AI to improve crowdsourcing processes and the reliance on crowdsourcing to support AI development, particularly through the creation of labeled datasets for supervised learning algorithms. While their primary focus is on the former, the authors recognize the reciprocal influence of these fields and their potential to reshape collective problem-solving.

Crowdsourcing, by design, is an online participatory practice that brings together large and distributed groups of workers and creatives to achieve common goals. This approach leverages the diversity, availability, and collective intelligence of the wisdom of the crowd (Surowiecki, 2005), enabling innovative applications across various domains. Bansal (2024) argues that AI offers powerful solutions to these challenges, making it an essential tool for advancing crowdsourcing. By automating task allocation, improving quality control, and optimizing workflows, AI has the potential to make crowdsourcing platforms far more efficient, cost-effective, and accessible to novice users.

For example, AI algorithms can dynamically match tasks to the most suitable workers based on their skills, interests, or past performance, thereby increasing the likelihood of high-quality outputs. Similarly, AI can model the accuracy of individual contributors, aggregate responses to resolve conflicts or errors, and identify outliers to ensure reliability in results. These capabilities not only address the logistical difficulties of crowdsourcing but also enhance its overall effectiveness. The authors envision a future where AI fundamentally transforms crowdsourcing, enabling it to operate at unprecedented levels of efficiency and scalability.

AI-enhanced collective intelligence can surpass the capabilities of either humans or AI alone by leveraging their complementary strengths within multilayered network frameworks (Cui & Yasseri, 2024). As Cui and Yasseri claim, the pressing challenges faced by contemporary society often surpass the capacity of human efforts, whether undertaken individually or collaboratively.

As AI continues to develop, its role within human collectives is poised to evolve, transitioning from a mere assistive tool to a more integrated and participatory agent. By combining their distinct yet

complementary strengths, humans and AI have the potential to achieve levels of collective intelligence that neither could attain independently. However, these human-AI systems are marked by inherent complexity, driven by intricate interactions and interdependencies; humans bring diversity that spans surface-level traits, such as demographics, to deep-level characteristics like values and expertise.

Similarly, AI agents vary in their functionalities and degrees of anthropomorphism, influencing how they interact with human collaborators. This multilayer perspective highlights the capacity of human-AI collaboration to address artistic and cultural challenges that neither could tackle independently. They claim that blending the computational power of AI with the nuanced judgment and emotional resonance of human creativity, this synergy introduces novel possibilities for artistic expression and cultural production.

Another fundamental aspect that suggests that AI technologies may reconfigure the creative ecosystem, is that human-AI co-creation enables artists to tap into an unprecedented pool of collective memory and data, fundamentally supporting collaborative strategies. AI systems, trained on vast datasets comprising historical and contemporary works, serve as repositories of collective cultural memory. These systems can uncover connections and patterns that might escape human perception, inspiring artists to push creative boundaries. For instance, artists employing generative AI tools can explore unfamiliar artistic styles, juxtapose disparate cultural references, or simulate entirely new forms of artistic media. Such tools democratize access to creative resources, opening up pathways for individuals and communities traditionally excluded from the art world.

Ultimately, the relationship between AI and collective art creation reflects broader societal shifts toward hybrid human-machine systems. These systems not only redefine the creative process but also offer a lens through which to examine our evolving relationship with technology, culture, and each other. In this context, the integration of AI into the arts serves as both a mirror and a catalyst for reimagining the possibilities of collective human creativity.

Artificial Intelligence as a Creative Tool

In the very last few years, AI is increasingly being utilized as a tool to assist human creators in various domains. As discussed by Brynjolfsson and McAfee (2014) in 'The Second Machine Age', AI could augment human creativity, allowing more individuals to express

themselves artistically and contribute to cultural production. In music, for example, AI systems like AIVA can compose original pieces in various styles. In their exploration of the relationship between artificial intelligence (AI) and artistic creation, Mazzone and Elgammal (2019) highlight the transformative impact of GANs on the field of algorithmic art, as exemplified by the AI-generated portrait Edmond de Belamy sold at Christie's in 2018 (Elgammal, 2018). Parallely, emerging applied projects are exploring how AI can facilitate collaborative creation. For example, Magenta, a research project from Google, is developing machine learning models that can collaborate with human musicians in real-time improvisation. AI systems are also being utilized to aggregate and analyze collective human knowledge and creativity. Projects like MIT's Collective Intelligence Design Lab are exploring how AI can be used to enhance group decision-making and creative problem-solving (Malone, 2018).

Taking this new and still unsettled context, AI tools have the potential to democratize creativity also by lowering the technical barriers to entry in various creative fields. Manovich (2018) in 'AI Aesthetics' explores how AI is transforming the nature of creative work, arguing that it represents a new stage in the automation of visual creation that began with photography in the 19th century. In this term, the proliferation of AI in cultural production is likely to redefine what we consider valuable

artistic skills. As routine technical tasks become automated, there may be an increased emphasis on uniquely human capabilities such as emotional intelligence, cultural understanding, and the ability to create meaningful narratives. Traditionally, digital art required artists to meticulously code the rules and parameters that defined the desired aesthetic outcome.

This approach meant that the creative process was heavily reliant on the artist's direct control over the computational system, with outcomes that adhered strictly to predefined rules. However, the advent of GANs has introduced a new paradigm in which the artist's role shifts from coding specific aesthetic rules to setting up machine learning algorithms capable of learning these aesthetics independently (Manovich 2024). By analyzing extensive datasets of images, these algorithms internalize patterns and styles, enabling them to generate new images that align with the aesthetics they have learned. This shift broadens the possibilities for artistic creation and redefines the interaction between the artist and the algorithm. Rather than directly producing the artwork, the artist now focuses on curating the dataset, configuring the algorithm, and refining the results in both pre- and post-production phases (Grilli and Pedota 2024).

Importantly, failure cases of AI art attempts highlight a critical distinction in the creative process: the absence of explicit artistic intention in the unexpected results. While the artist guides the overall process and curates the output, the specific deformations or anomalies generated by the machine are not deliberate choices. Instead, they emerge as byproducts of the algorithm's attempt to imitate learned patterns. Despite this lack of intentionality, these outcomes can still resonate with viewers, challenging traditional notions of artistic agency and authorship. Colton and Wiggins (2012) in their work on computational creativity

anticipated that this shift towards machine serendipity might lead to a greater focus on the conceptual aspects of art, with technical execution becoming less central to artistic value.

A posthumanist lens clarifies that intentionality in human–AI creative systems is distributed across socio-technical assemblages rather than located solely in either the human or the model. Haraway's (1991) cyborg and Braidotti's (2019) posthuman subject foreground hybrid agencies that destabilize human/machine dichotomies and explains how happy accidents emerge from material–discursive entanglements of prompts, training data, model architectures, interfaces, hardware, and institutional settings. From this standpoint, algorithmic serendipity is not the model intending, but an emergent property of the assemblage—co-authored by designers, dataset curators, prompt engineers, and end-users. In practice, authorship becomes a choreography of iterative selections and curatorial judgments whereby purposes are negotiated and revised across human and computational agents (relational intentionality). Algorithms are cultural artifacts embedding histories and norms, thus serendipity is patterned by infrastructural choices (e.g., datasets, sampling, loss functions) rather than neutral randomness. This reframing strengthens the claim that human–AI creativity is fundamentally relational and contextual, with agency emerging from the ongoing alignment between human goals and machine affordances.

AI's capacity to analyze and synthesize vast amounts of cultural data could lead to both increased diversity and potential homogenization of cultural output. In this sense AI systems, particularly those based on machine learning, can be conceptualized as capitalizing on collective creativity. These systems are trained on vast datasets of human-created content, essentially distilling and recombining the collective creative output of humanity. Boden (2004) in 'The Creative Mind' discusses how computers enable new forms of combinatorial creativity on a scale previously unimaginable. This raises intriguing questions about the nature of creativity itself and the role of human agency in AI-assisted creation. On one hand, AI could help surface and combine diverse cultural influences in novel ways. On the other, if AI systems are primarily trained on dominant cultural outputs, they could reinforce existing cultural biases and hegemonies.

Despite these promising developments, the integration of AI into art creation raises critical questions. The boundaries between human and machine contributions often blur, complicating traditional notions of creative ownership. The notion of democratization of creative practices requires critical scrutiny. As Terranova (2000) argues in her work on free labor, digital participation often conceals forms of invisible exploitation, while platforms channel the creative energies of users toward market-oriented logics. Similarly, van Dijck, Poell, and de Waal (2018) emphasize that platform-based participation is rarely neutral, being shaped by infrastructures of visibility, ranking, and monetization. From this perspective, democratization through AI must be seen as ambivalent: it lowers barriers to entry, but simultaneously reinforces new asymmetries of control over cultural production. Furthermore, the use of AI in art challenges the role of artistic practice as a space for scientific inquiry. While traditionally viewed as a medium for self-expression and cultural critique, art increasingly serves as a testing ground for AI technologies, blurring the boundaries between creative and scientific endeavors. The intersection of these domains

raises ethical considerations about the purpose of art and the responsibilities of artists, technologists, and audiences in shaping its future.

Addressing these challenges requires coordinated action across multiple domains. Governments, private markets, and community-driven initiatives must collaborate to ensure the safe and equitable development of AI in the arts. Regulatory measures can establish guidelines for ethical AI use, intellectual property rights, and equitable compensation for creators whose work contributes to AI training datasets (de Neufville & Baum, 2021). Meanwhile, private markets can incentivize innovation by supporting human-AI co-creation projects that prioritize inclusivity and social impact. Community self-organizing efforts, such as open-source AI art platforms or decentralized funding models, can empower grassroots artistic movements and ensure that the benefits of AI are distributed more equitably.

Creative Labor Market

Empirical research offers nuanced insights into the resilience of creative occupations in the face of automation. Frey and Osborne (2017) highlight that creative professions are relatively resistant to full automation, as many tasks within these roles require a combination of originality, emotional intelligence, and contextual awareness—qualities that AI has not yet fully mastered. However, the resilience of creative occupations varies by domain. Fields like graphic design, content generation, and algorithmic art have seen rapid integration of AI technologies, with tools that can produce visually appealing designs or text-based content autonomously. In contrast, roles that involve high levels of interpersonal interaction, such as theater direction or curatorial work, remain less susceptible to automation due to their reliance on human judgment and cultural interpretation.

The integration of AI into creative processes is likely to have further significant impacts on creative labor markets, as it transforms the roles and skills required of creative professionals (Grilli and Pedota 2024). On the one hand, AI systems can automate routine creative tasks and empower creators to produce and distribute high-quality content at a fraction of the cost, fostering a more democratized approach to creativity. This automation creates efficiency but also raises concerns about the displacement of jobs that rely on these tasks (Frey and Osborne 2017).

The rapid advancement of AI technologies has driven significant transformations in areas traditionally reliant on human creativity, such as music, visual arts, literature, and scientific discovery. These technologies are not only augmenting human creative processes but, in some cases, are beginning to replace them, bringing the question of whether AI can truly be considered creative closer to the realm of reality. Creative professionals who can adapt to these changes by integrating AI into their practice—using it as a tool to enhance, rather than replace, their creative processes—are likely to find new avenues for growth (Epstein et al. 2023).

In the music industry, for instance, AI-based systems have ushered in a new era of innovation following the disruptions caused by digitization, streaming, and free downloading. Today, autonomous AI systems are capable of composing, arranging, and even performing music (Carnovalini and Rodà 2020). Technologies such as holographic representations of musical icons and humanoid robots delivering live performances are reshaping the boundaries of artistic expression. Many AI-driven musical composition tools produce original, copyright-free works, enabling users to integrate their own compositions with AI-generated pieces. This dynamic opens new avenues for co-creation, offering composers and audiences fresh ways to engage with music while maintaining human agency as a central element of creative expression.

In these and other ways, AI technologies have redefined the processes and outputs of creativity, enabling new forms of artistic and cultural expression that were previously unimaginable. These tools lower barriers to entry in creative fields by providing accessible platforms for non-specialists to produce professional-quality work (Manovich 2018). For instance, generative AI tools like OpenAI's DALL·E and AIVA enable users without formal training to create compelling visual art and music compositions; Adaptive AI systems personalize creative outputs based on user preferences, fostering more intimate and tailored artistic experiences. AI is not merely a substitute for human creativity but a collaborator. Hybrid workflows combine human intuition and contextual understanding with AI's computational power, enabling the production of sophisticated works that neither could achieve alone. For example, in fashion design, AI can analyze trends and suggest patterns, while designers bring the vision and cultural sensibility to refine and contextualize the outputs (Manovich 2024). These shifts require businesses and professionals to adopt more agile and collaborative approaches, leveraging AI not only as a tool for production but also as a means to engage audiences in more interactive and participatory ways.

In this sense, humanities professionals, trained in disciplines such as language, communication, and fine arts, provide a depth of contextual awareness that is vital for designing effective and meaningful AI prompts. The burgeoning field of AI prompt engineering, which involves crafting inputs to guide AI systems' outputs effectively, is enriched by the contributions of humanities scholars and creative professionals. While the field is often seen through the lens of technical expertise, the nuanced insights brought by these disciplines are indispensable for advancing the ethical, cultural, and creative dimensions of AI development.

These experts possess the ability to interpret and convey complex ideas with subtlety and precision, ensuring that AI-generated content resonates with diverse audiences. For instance, a novelist or playwright might design prompts that encourage AI systems to simulate complex character dialogues or generate compelling narratives. Similarly, visual artists could contribute to prompts that inspire AI-driven creative outputs, such as designing illustrations. This collaborative infusion of creativity not only enhances the functionality of AI systems but also expands their potential to inspire and engage users. By introducing elements of surprise, playfulness, and aesthetic appeal, creative professionals ensure that AI interactions resonate on a deeper, more human level.

The integration of humanities perspectives into AI prompt engineering is particularly crucial for addressing ethical challenges and mitigating biases. Scholars from fields such as philosophy, anthropology, and critical theory are adept at interrogating the ethical dimensions of technology, providing a framework for identifying and addressing potential risks. AI systems, trained on vast datasets, often inherit the biases embedded in those datasets. Without careful consideration, prompts can unintentionally reinforce stereotypes, marginalize certain groups, or propagate harmful narratives. Humanities scholars, with their critical lens, can analyze the underlying assumptions and values encoded in prompt design, advocating for strategies that prioritize fairness and inclusivity (Floridi 2024). By embedding ethical considerations into the very fabric of prompt engineering, they help guide AI development toward outcomes that respect human dignity and diversity.

Copyrights and Intellectual Property

When an AI system generates a piece of music or art, questions arise about who the author is - the AI system, its creators, the people whose work it was trained on, or some combination of these? Gunkel (2017) in 'Robot Rights' explores these questions, arguing that our current legal and ethical frameworks may be inadequate to deal with the creative output of AI systems. The ease of copying and modifying digital content has already posed significant challenges to traditional copyright frameworks. Boyle (2008) in 'The Public Domain' argues that current copyright laws, designed for an analogue age, are ill-suited to the realities of digital cultural production and may stifle creativity and innovation. The emergence of Creative Commons licenses, as discussed by Aufderheide and Jaszi (2018) in 'Reclaiming Fair Use', represents an attempt to create more flexible copyright arrangements that better suit the realities of digital creation and distribution.

For instance, paraphrasing, often seen as a way to reframe existing content, occupies a legally contentious space when it comes to copyright law. In many jurisdictions, including Italy, copyright infringement extends beyond exact replication. The Italian Supreme Court (Cassazione) has ruled that even paraphrasing a text with superficial changes—such as replacing a few words or restructuring sentences—can constitute copyright infringement if the core structure and substance of the original work remain intact. This interpretation complicates the use of AI tools like ChatGPT, which often summarize or paraphrase content with only minor alterations. Copyright law has traditionally been reserved for works that exhibit originality and are created by humans. In most jurisdictions, the absence of human authorship renders AI-generated works ineligible for copyright protection (Guadamuz 2021). These works are often considered part of the public domain, available for use without restriction. However, the situation becomes more nuanced when humans play a significant role in shaping the AI-generated output. For instance, if a person uses AI to produce a draft and then refines or adds substantial creative elements, the resulting work could potentially qualify for copyright protection.

In the European Union, the Copyright Directive introduces a text and data mining exception, which allows large-scale data extraction and reproduction for non-commercial purposes unless explicitly restricted by rights holders (Rosati, 2019). Compounding these challenges is the ethical dilemma posed by the training of AI models. These systems rely on vast datasets, often compiled from publicly available online content with several authors and co-authors, much of which is copyrighted. Reflecting on the integration of AI in creative processes, we can identify and expand upon several key considerations, particularly regarding the transformation of copyright systems, equitable remuneration for creators, and the implications of new production and distribution models.

Lemley and Casey (2019) in 'Fair Learning' argue for the need to rethink copyright law in light of machine learning, suggesting that we may need new frameworks to balance innovation incentives with the public interest in AI-generated works.

Konrad Gliściński, in his 'Reclaim the state: public interest in copyright and Modern Monetary Theory' (2022) highlights the opportunity to reimagine cultural production as a shared, equitable endeavor, leveraging AI not as a tool of consolidation for platform capitalism but as a democratizing force within a restructured system of public support for creativity.

The traditional copyright system, built around exclusive rights, has historically aligned with the development of capitalism, prioritizing market-driven incentives for creators. However, the system is not inherently tied to this structure and can be reimaged. In the context of AI and cultural production, AI-generated works, which challenge notions of authorship and originality, expose the limitations of a copyright model reliant on exclusivity. Platform capitalism, characterized by centralized control of content production and distribution, exacerbates inequalities in creative industries. The author claims that AI tools are deeply embedded in this system, often owned and operated by large corporations that benefit disproportionately: Companies that control AI training datasets and algorithms profit from the aggregated contributions of countless creators without redistributing value equitably. This mirrors the broader critique of platform capitalism, where intermediaries capture most of the revenue. AI complicates the already fraught question of compensating creators. For instance, if an AI model trained on copyrighted works, which means having a long chain of authors and contaminations, with several generates new outputs, who should benefit? We invite to embrace the vision of Gliściński, whose proposal for a publicly funded model aligns with the challenges posed by AI by shifting the focus from exclusivity to accessibility, ensuring equitable remuneration for multiple creators in a system where AI complicates traditional notions of authorship, and encouraging the development of tools and systems that prioritize public benefit over private profit.

Conclusion

The integration of AI in cultural production represents not just a technological shift, but a profound social and cultural transformation. It challenges our understanding of authorship and creativity, potentially leading us towards new forms of collective cultural creation. This shift brings with it both possibilities and significant challenges. The democratization of creative tools and the potential for AI to augment human creativity could lead to an explosion of cultural production and new forms of artistic expression. At the same time, it raises complex practical, ethical and legal questions about the nature of creativity, the value of human artistic labor, and the ownership of AI-generated works.

As we stand at the potential threshold of this momentum, it's clear that the impact of AI on cultural production will be far-reaching and multifaceted. The need to preserve human agency, cultural diversity, and fairness in creative processes underscores the importance of developing frameworks that balance technological innovation with equitable opportunities for creators. The fusion of humanities, creative disciplines, and technical expertise is crucial to navigating this transformative moment.

While technologists provide the foundation for AI systems, humanities scholars bring contextual understanding, ethical insight, and cultural sensitivity, and creative professionals contribute imagination and innovation. At the same time, it is crucial to acknowledge the risks inherent in the current configuration of generative AI. As Zuboff (2019) demonstrates, the concentration of data and computational power within platform capitalism tends to reinforce asymmetries of control and surveillance, rather than dismantle them. Moreover, as Couldry and Mejias (2019) argue, the very extraction of human creativity and cultural expression as “raw data” constitutes a new form of colonial appropriation, which complicates optimistic narratives of democratization. Thus, the cultural future shaped by AI will depend not only on technical affordances but on the political capacity to resist exploitative infrastructures and imagine alternative, more equitable socio-technical arrangements. Together, these perspectives shape a multidisciplinary approach that ensures AI systems not only enhance productivity but also reflect the complexity and diversity of human culture.

Moving forward, it is essential to embrace this collaborative vision while addressing systemic issues in copyright, authorship, and labor markets. Publicly funded and decentralized models of cultural production could provide a more inclusive and equitable alternative, prioritizing access and fair compensation over proprietary control. Such systems would amplify marginalized voices and foster cultural innovation, reflecting the richness of diverse perspectives.

Ultimately, the future of creativity in the age of AI will be determined by the choices we make as societies—choices that define how we integrate, regulate, and adapt AI within our creative practices. By leveraging AI as a partner rather than a competitor, and by centering ethical and inclusive principles, we can ensure that this transformative technology serves as a tool for enriching the cultural landscape, preserving human creativity as its guiding force.

Biographical note

Federico Pilati is a researcher at the University of Bologna and teaches *Generative Artificial Intelligence for Social Research* at the University of Milan–Bicocca.

Maria Tartari is a researcher at the CNR Institute of Cultural Heritage Sciences and teaches *Cultural Economics* at IULM University in Milan. She works as a cultural planning consultant and actively collaborates with EIT Culture & Creativity.

References

- Aufderheide, P., & Jaszi, P. (2018). *Reclaiming fair use: How to put balance back in copyright*. University of Chicago Press.
- Bansal, G. (2024). Wisdom of the experts, not the wisdom of the crowds: The power of case-based research in the age of AI. *Journal of Information Technology Case and Application Research*, 26(1), 90–91. <https://doi.org/10.1080/15228053.2024.2322860>
- Benjamin, W. (1935). The work of art in the age of mechanical reproduction. In *Illuminations*. Schocken Books.
- Boden, M. A. (2004). *The creative mind: Myths and mechanisms*. Routledge.
- Boyle, J. (2008). *The public domain: Enclosing the commons of the mind*. Yale University Press.
- Braidotti, R. (2019). *Posthuman knowledge*. Polity.
- Bruns, A. (2008). *Blogs, Wikipedia, Second Life, and beyond: From production to produsage*. Peter Lang.
- Brynjolfsson, E., & McAfee, A. (2014). *The second machine age: Work, progress, and prosperity in a time of brilliant technologies*. W. W. Norton & Company.
- Carnovalini, F., & Rodà, A. (2020). Computational creativity and music generation systems: An introduction to the state of the art. *Frontiers in Artificial Intelligence*, 3, 14. <https://doi.org/10.3389/frai.2020.00014>
- Chun, W. H. K. (2011). *Programmed visions: Software and memory*. MIT Press.
- Colton, S., & Wiggins, G. A. (2012). Computational creativity: The final frontier? In *ECAI 2012*. IOS Press.
- Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.
- Cui, H., & Yasserli, T. (2024). AI-enhanced collective intelligence. *Patterns*, 5(11). <https://doi.org/10.1016/j.patter.2024.101074>
- De Neufville, R., & Baum, S. D. (2021). Collective action on artificial intelligence: A primer and review. *Technology in Society*, 66, 101649. <https://doi.org/10.1016/j.techsoc.2021.101649>

- Doshi, A. R., & Hauser, O. P. (2024). Generative AI enhances individual creativity but reduces the collective diversity of novel content. *Science Advances*, 10(28), eadn5290. <https://doi.org/10.1126/sciadv.adn5290>
- Elgammal, A. (2018). AI is blurring the definition of artist. *American Scientist*, 107(1), 18–21. <https://doi.org/10.1511/2019.107.1.18>
- Esposito, E. (2022). *Artificial communication: How algorithms produce social intelligence*. MIT Press.
- Epstein, Z., Hertzmann, A., Investigators of Human Creativity, Akten, M., Farid, H., Fjeld, J., ... & Smith, A. (2023). Art and the science of generative AI. *Science*, 380(6650), 1110-1111. <https://doi.org/10.48550/arXiv.2306.04141>
- Evans, J., Hinds, R., Jung, M. F., & Pescetelli, N. (2021). Leveraging artificial intelligence for collective intelligence. In *Academy of Management Proceedings* (Vol. 2021, No. 1, p. 15722). <https://doi.org/10.5465/AMBPP.2021.15722symposium>
- Floridi, L. (2024). On the future of content in the age of artificial intelligence: Some implications and directions. *Philosophy & Technology*, 37(3), 112. <https://doi.org/10.1007/s13347-024-00806-z>
- Frey, C. B., & Osborne, M. A. (2017). The future of employment: How susceptible are jobs to computerisation? *Technological Forecasting and Social Change*, 114, 254–280. <https://doi.org/10.1016/j.techfore.2016.08.019>
- Fuller, M. (2008). *Software studies: A lexicon*. MIT Press.
- Gliściński, K. (2022). Reclaim the state: Public interest in copyright and modern monetary theory. *Internet&Sociedade*, 3(2). <http://dx.doi.org/10.2139/ssrn.5063272>
- Grilli, L., & Pedota, M. (2024). Creativity and artificial intelligence: A multilevel perspective. *Creativity and Innovation Management*, 33(2), 234–247. <https://doi.org/10.1111/caim.12580>
- Guadamuz, A. (2021). Do androids dream of electric copyright? Comparative analysis of originality in artificial intelligence generated works. In J.-A. Lee, R. Hilty, & K.-C. Liu (Eds.), *Artificial intelligence and intellectual property* (pp. 147–176). Oxford University Press.
- Gunkel, D. J. (2017). *Robot rights*. MIT Press.
- Haraway, D. (1991). *Simians, cyborgs, and women: The reinvention of nature*. Routledge.
- Jenkins, H. (2006). *Convergence culture: Where old and new media collide*. NYU Press.
- Lee, H. K. (2022). Rethinking creativity: Creative industries, AI and everyday creativity. *Media, Culture & Society*, 44(3), 601–612. <https://doi.org/10.1177/01634437221077009>
- Lemley, M. A., & Casey, B. (2019). Fair learning. *NYU Law Review*, 94(6). <http://dx.doi.org/10.2139/ssrn.3528447>
- Lessig, L. (2008). *Remix: Making art and commerce thrive in the hybrid economy*. Penguin.
- Malone, T. W. (2018). *Superminds: The surprising power of people and computers thinking together*. Little, Brown and Company.
- Manovich, L. (2018). *AI aesthetics*. Strelka Press.

- Manovich, L. (2024). *Diffusions in architecture: Artificial intelligence and image generators*. Polity.
- Mazzone, M., & Elgammal, A. (2019). Art, creativity, and the potential of artificial intelligence. *Arts*, 8(1), 26. <https://doi.org/10.3390/arts8010026>
- Rosati, E. (2019). Copyright as an obstacle or an enabler? A European perspective on text and data mining and its role in the development of AI creativity. *Asia Pacific Law Review*, 27(2), 198–217. <https://ssrn.com/abstract=3452376>
- Simonton, D. K. (2000). Creativity: Cognitive, personal, developmental, and social aspects. *American Psychologist*, 55(1), 151–158. <https://doi.org/10.1037/0003-066x.55.1.151>
- Surowiecki, J. (2005). *The wisdom of crowds*. Vintage.
- Terranova, T. (2000). Free labor: Producing culture for the digital economy. *Social Text*, 18(2), 33–58. https://doi.org/10.1215/01642472-18-2_63-33
- Van Dijck, J., Poell, T., & de Waal, M. (2018). *The platform society: Public values in a connective world*. Oxford University Press.
- Verhulst, S. G. (2018). Where and when AI and CI meet: Exploring the intersection of artificial and collective intelligence towards the goal of innovating how we govern. *AI & Society*, 33, 293–297. <https://doi.org/10.1007/s00146-018-0830-z>
- Weld, D. S., Lin, C. H., & Bragg, J. (2015). Artificial intelligence and collective intelligence. In *Handbook of collective intelligence* (pp. 89–114).
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.

Liveness generativa. L'incertezza come dispositivo spettacolare nel live coding e nelle AI-driven performance*

Dellapasqua Alex**

Università degli Studi di Urbino Carlo Bo

This article investigates the intersection of algorithmic mediation and live performance, focusing on how generative systems transform the experience of uncertainty and unpredictability from something to be minimized into a creative resource. The contribution situates itself within the broader debate on the risk society, examining how algorithmic systems reduce present uncertainty through probabilistic calculation. However, as recent scholarship on liveness emphasizes, live artistic practices often revalue unpredictability as a central aesthetic and communicative component. Drawing on theories of artificial communication the study explores how deep learning algorithms generate contingency within live performance contexts. Through qualitative analysis of semi-structured interviews and three case studies – two performances integrating generative AI (*Spazio Latente*, *Radio Pentothal*) and one live coding performance (*L'Embodiment dell'Effimero*) – the article examines how artists strategically engage with algorithmic systems to foreground risk and contingency. The interviews with performers, programmers, and creative technologists further reveal how unpredictability is deliberately designed and cultivated as an integral part of the communicative process in live performance. The article contributes to performance and media studies by exploring how liveness generated by leverages uncertainty to shape interactions between performers, computational systems, and audiences.

Keywords: Liveness; Artificial Communication; Algorithmic Mediation; Artificial Intelligence; Uncertainty

* Article submitted on 02/10/2025. Article accepted on 22/12/2025.

** a.dellapasqua1@campus.uniurb.it

Introduzione

Vi è una situazione instabile che può essere sfruttata (...) per fini poetici, singolari. Voglio dire che la stessa situazione, l'incertezza, può essere angosciante, ma può anche essere esaltante, a condizione di farne un gioco e di fare quindi del principio di incertezza una regola, la regola del gioco. (Baudrillard, 2000)

La crescente presenza di sistemi algoritmici generativi nell'attuale panorama mediale mostra un rapporto ambivalente con la gestione dell'incertezza. Da un lato, l'incertezza prodotta dalla sovrabbondanza di opzioni (Luhmann, 1991; Beck, 1992), tende a essere mitigata attraverso il calcolo probabilistico: nel quotidiano, gli algoritmi agiscono per ridurre l'indeterminatezza del "qui e ora", orientando verso la scelta ritenuta più ottimale in una data situazione. Dall'altro, nelle arti performative dal vivo gli algoritmi possono diventare agenti che consentono di abitare la contingenza del presente e trasformarla in esperienza condivisa. Difatti, nell'attuale centralità della *liveness* nelle pratiche artistiche e culturali (Gemini e Brilli, 2023), l'incertezza assume un valore comunicativo desiderabile e specificamente ricercato. È in questa tensione – tra algoritmi impiegati come tecnologie di riduzione del rischio nella vita quotidiana e algoritmi riattivati come generatori di imprevedibilità nei contesti performativi – che si colloca il focus di questo contributo. Nonostante gli agenti algoritmici e le forme dal vivo siano al centro di numerosi dibattiti, le loro intersezioni restano ancora poco esplorate in prospettiva sociologica, in particolare per quanto riguarda il ruolo che tali sistemi assumono nella produzione e nella valorizzazione dell'incertezza nel tempo dell'evento dal vivo.

Da un punto di vista teorico, la presente ricerca si inserisce nel quadro degli studi sulla *mediatizzazione* (Krotz, 2007; Boccia Artieri, 2015; Hepp & Couldry, 2017), con l'obiettivo di indagare come le logiche algoritmiche plasmino e riorientino le modalità dell'agire individuale e collettivo. Più nello specifico, il contributo esplora gli impatti sociali della mediazione algoritmica in relazione agli studi sulla *liveness*, assumendo le arti performative dal vivo come un campo privilegiato per osservare come la contingenza prodotta dalla comunicazione artificiale (Esposito, 2017, 2022) venga valorizzata e resa significativa nel contesto artistico.

L'analisi si concentra su un confronto tra due diverse forme performative di interazione live con gli algoritmi: *AI-driven performance* e pratiche di live coding. Con le prime s'intendono quelle performance artistiche e teatrali basate su una comunicazione dal vivo con un'IA che viene addestrata e codificata precedentemente all'evento, mentre le seconde sono performance in cui è il processo stesso di scrittura e programmazione dell'algoritmo a venire eseguito in tempo reale. L'analisi comparata di queste due tipologie di performance consente di comprendere come l'impiego di modelli algoritmici generativi intervenga sulle qualità della *liveness* legate all'incertezza e all'imprevedibilità e, al contempo, di osservare in che modo la contingenza che ne emerge si differenzi da quella prodotta dall'agente umano o da altri artefatti tecnici.

Dal calcolo dell'incertezza alla liveness generativa

La tensione richiamata in apertura, tra la funzione ordinaria degli algoritmi come tecnologie di gestione e riduzione dell'incertezza e il loro uso straordinario, nelle arti performative, come generatori di contingenza, può essere compresa più a fondo se collocata all'interno del più ampio dibattito sulla società del rischio (Beck, 1992).

Al fine di delineare con maggiore precisione il quadro teorico, si rendono necessarie alcune precisazioni terminologiche. In filosofia e nella logica modale, il termine *contingente* designa ciò che non è né necessario né impossibile: qualcosa che può esistere, ma che potrebbe anche non esistere o essere altrimenti (Hughes & Cresswell, 1968). La contingenza rinvia dunque a un orizzonte di possibilità aperte e alternative. Da ciò discende una prima conseguenza fondamentale: la contingenza implica incertezza, poiché la presenza di più opzioni possibili rende azioni e decisioni intrinsecamente non determinate e sempre suscettibili di essere diverse (Esposito, 2022). L'incertezza, a sua volta, comporta necessariamente un certo grado d'imprevedibilità. Per definizione, in una situazione incerta non è possibile predeterminare pienamente il suo esito. L'imprevedibilità costituisce infine la condizione del rischio. Il rischio emerge in presenza di un soggetto che si espone in un contesto in cui l'esito non può essere anticipato con certezza (Beck, 1992).

Riprendendo Luhmann (1991), il rischio va quindi inteso come un problema generato dall'eccesso di possibilità: più opzioni si rendono disponibili, maggiore diventa l'incertezza e più imprevedibile appare il futuro. Di fronte a tale complessità, la capacità umana di prevedere autonomamente tutte le variabili risulta inevitabilmente limitata, per cui ogni scelta diventa un rischio da assumere in un "qui e ora" che si sovraccarica di responsabilità (Bauman, 2007). È all'interno del panorama delineato che la crescente pervasività della mediazione algoritmica può essere interpretata come un dispositivo di riduzione dell'incertezza: la capacità dei sistemi di IA di processare grandi moli di dati e di calcolare le soluzioni considerate statisticamente più ottimali consente di ridurre l'indeterminatezza del presente e attenuare così l'esposizione al rischio.

All'interno di questa ricerca, l'attenzione è rivolta in particolare agli algoritmi generativi, cioè quelle architetture basate su reti neurali multistrato che apprendono schemi e regolarità dai dati per produrre previsioni, classificazioni o generazioni di contenuti (Samek *et al.*, 2019). Un esempio di queste reti è offerto dai modelli linguistici di grandi dimensioni (LLM), come *ChatGPT* o *Gemini*, i quali generano testi plausibili stimando la probabilità che una determinata parola – o, più precisamente, un token – segua quelli precedenti, sulla base dei dati utilizzati durante l'addestramento (Neu *et al.*, 2022).

Una manifestazione concreta di come la misurazione probabilistica contribuisca a gestire l'eccesso di possibilità la si può osservare nel crescente ricorso a questi modelli nei processi decisionali, sia a livello individuale che istituzionale. Studi recenti mostrano come, in particolare tra le giovani generazioni, i sistemi di IA vengano sempre più spesso consultati non solo per ottenere informazioni, ma come supporto nelle scelte personali, esistenziali e valoriali (Brandtzaeg *et al.*, 2025). Analogamente, su scala organizzativa e amministrativa,

l'IA viene adottata per gestire decisioni ad alta complessità informativa – dalla selezione del personale alla gestione dei servizi pubblici (Bowen, 2024; Bright *et al.*, 2024). Come sottolineano diversi studi, questa tendenza non è però priva di criticità e rischi, dal momento che l'automazione delle decisioni rischia di trasformarsi in un esercizio di delega a soluzioni opache (Lovari e Ducci, 2022; Selten & Klievink, 2024).

La tesi qui sostenuta è che, nel panorama socio-tecnologico delineato, la mediazione algoritmica appare come un dispositivo utile e necessario per contenere la complessità e la contingenza che caratterizzano la società contemporanea (Luhmann, 1984): “se l'essere in balia del destino era prima una certezza incalcolabile, nel moderno la visione scientifica del mondo fa dell'incertezza un suo oggetto di riflessione e calcolo, diventa incertezza calcolabile” (Arielli, 2023, p. 519).

Tuttavia, nonostante il calcolo probabilistico nei contesti sociali e istituzionali miri a ridurre la contingenza del presente, nei contesti performativi spesso i sistemi algoritmici acquisiscono una funzione differente, dal momento che le performance incorporano sin dalle loro origini i rischi della contingenza. La cornice entro la quale si muovono le performance dal vivo è quella dell'incertezza oppositiva, poetica, creativa (Gemini, 2003), motivo per cui nelle arti performative la stessa incertezza possiede invece un valore estetico e comunicativo.

Seguendo le prospettive sviluppate dai *performance studies* (Schechner, 1984; Turner, 1986; Fischer-Lichte, 2004; Bachmann-Medick, 2016), la performance va intesa come un dispositivo attraverso cui società e pratiche si co-determinano. La performance non è soltanto un oggetto privilegiato per osservare il sociale, ma un luogo in cui il sociale viene prodotto e messo in discussione. In continuità con le epistemologie di impostazione costruttivista, il suo significato va ricondotto alla dimensione di realtà sociale costruita (Berger & Luckmann, 1966), all'interno della quale la performance attiva processi di riflessività: essa diventa uno spazio capace di formulare un meta-commento sul mondo che la circonda, offrendo forme di interpretazione, critica e valutazione delle dinamiche socio-culturali in cui è situata (Gemini, 2013). Il carattere riflessivo della performance funziona dunque come uno “specchio magico” (Gemini, 2003) che consente alla performance di riflettere non solo l'uso ordinario di un determinato medium, ma anche di proporre un suo utilizzo straordinario capace di riorientare creativamente funzioni e significati.

L'obiettivo di questo contributo è indagare come, nei contesti performativi, la capacità degli algoritmi generativi di reagire in tempo reale a stimoli e input contribuisca a generare situazioni in cui l'incertezza del presente non viene attenuata, ma intenzionalmente attivata e valorizzata. Ciò implica considerare tali pratiche sul piano degli studi della *liveness*: cosa significa per un algoritmo “essere in presenza” e “agire in tempo reale”? David Saltz (2001), in una delle prime riflessioni sui *live media* a teatro, sottolinea come la sensazione di *liveness* possa emergere dalla capacità di agenti tecnologici – come avatar e algoritmi – di reagire in tempo reale agli stimoli, introducendo variazioni e comportamenti imprevedibili proprio come farebbero attori in carne e ossa. Questa prospettiva viene ulteriormente sviluppata da Gemini e Brilli (2023) che, indagando la dimensione comunicativa ed epistemologica della *liveness*, mostrano come i processi di mediatizzazione abbiano

trasformato il modo in cui i fenomeni dal vivo operano nei contesti di produzione e fruizione culturale. Con la mediatizzazione, infatti, “la performance passa da forma vincolata all’interazione faccia a faccia a fenomeno comunicativo basato anche su una co-presenza a distanza e che può implicare anche attori tecnologici non umani” (Gemini e Brilli, 2023, p. 10). Ne risulta, che il senso del “dal vivo” può emergere anche attraverso la comunicazione con sistemi algoritmici. Lo scostamento dalla concezione prettamente ontologica del “qui e ora” consente infatti di trattare la *liveness* sulla base della teoria della comunicazione definita nella teoria dei sistemi sociali di Luhmann (1984): la comunicazione parte dal ricevente, quindi, può riguardare media, macchine e algoritmi. In altri termini, la comunicazione artificiale pone l’attualizzazione della comunicazione sul lato delle aspettative del ricevente (Esposito, 1992).

Questo quadro è in linea con quella che Philip Auslander ha definito *digital liveness* (2012), una declinazione della presenza scenica che non dipende esclusivamente dalla fisicità degli attori o dalla prossimità tra performer e pubblico, ma da un coinvolgimento relazionale e comunicativo: secondo Auslander, l’esperienza del “dal vivo” scaturisce dal nostro atto consapevole di percepire entità virtuali come vive, in risposta alle sollecitazioni che esse esercitano su di noi. Non a caso, il carattere *interattivo* delle performance è una di quelle qualità che la tecnologia ha sganciato dalla co-presenza fisica, rendendola un attributo più proprio delle performance che assimilano le logiche mediali (Brilli, Gemini e Giuliani, 2023). In continuità con gli studi discussi, la presente ricerca adotta il concetto di *liveness generativa* per riferirsi a quella *liveness* prodotta da processi algoritmici generativi e alimentata dalle interazioni che essi attivano nel tempo dell’evento performativo. A partire da questa definizione, il contributo si propone di rispondere alla seguente domanda di ricerca: che tipo di contingenza produce la *liveness generativa* nelle arti performative, e in che modo essa si differenzia da quella generata dall’agente umano o da altri artefatti tecnici?

Incertezza e mediazione algoritmica nelle arti performative dal vivo

Una vasta letteratura ha affrontato il tema dell’incertezza come dispositivo costitutivo dell’evento performativo dal vivo. Fin dalle prospettive fondative dei performance studies, in particolare nei lavori di Richard Schechner (1983) e Victor Turner (1986), la performance è stata concepita come un processo intrinsecamente aperto ed esposto alla variabilità dell’accadere. Su questa linea si colloca anche la riflessione di Erika Fischer-Lichte (2004), che descrive la performance come il risultato di un *autopoietic feedback loop* tra performer e spettatori: un processo relazionale che si costituisce nel tempo stesso dell’evento e che, proprio per la sua natura emergente, non può essere pienamente previsto né replicato. L’imprevedibilità è quindi una condizione strutturale della performatività, legata alla co-presenza e alla dinamica interattiva tra i soggetti coinvolti.

Come sottolineano inoltre Gemini e Brilli (2023), con la progressiva mediatizzazione dell’esperienza dal vivo l’imprevedibilità emerge come una delle qualità più frequentemente

associate alla *liveness*, tanto nelle percezioni del pubblico quanto nel discorso teorico, come testimoniano gli studi di Auslander (2012, 2022) e Scannell (2014). In ambito musicale, Paul Sanden in *Liveness in modern music: musicians technology and the perception of performance* (2013) evidenzia come la spontaneità e l'imprevedibilità dell'azione umana costituiscano elementi centrali nella percezione dell'evento live, anche in contesti fortemente mediati tecnologicamente. Analogamente, Claudia Georgi in *Staging Liveness Via Mediatization the Use of Film and Video in Contemporary British Theatre and Performance* (2014) mostra come l'integrazione di dispositivi audiovisivi registrati nelle performance teatrali possa intensificare, anziché attenuare, il senso di rischio e imprevedibilità del live, ad esempio attraverso il confronto in tempo reale tra l'attore in scena e immagini preregistrate.

Tuttavia, il grado di imprevedibilità che caratterizza una performance dal vivo varia sensibilmente in relazione al tipo di evento e alle sue finalità comunicative. In alcuni eventi, l'imprevedibilità cerca di essere limitata anziché amplificata – ad esempio, in una diretta giornalistica – mentre in altri viene esplicitamente ricercata ed esaltata come dispositivo spettacolare, ad esempio nei concerti jazz fondati su pratiche improvvisative (Caporaletti, 2005). Arielli, in *Rischi Calcolati. Strategie e paradossi dell'incertezza nelle arti* (2023) guarda anche alle ambiguità generate dalla ricerca artistica di imprevedibilità: da un lato, l'esposizione al rischio e all'incertezza viene spesso assunta come condizione necessaria affinché l'evento artistico sia percepito come significativo, autentico o rilevante; dall'altra tale esaltazione dell'imprevedibilità rischia di diventare il prodotto di una decisione pianificata e controllata dell'artista.

Nel complesso, la letteratura discussa mostra come ogni evento performativo dal vivo sia chiamato a confrontarsi – in gradazioni differenti, a seconda del tipo di evento e dei media che vi entrano in gioco – con il rapporto tra il controllo e la sua perdita, la possibilità del fallimento, l'apertura con la contingenza e l'imprevedibilità che ne scaturisce. È all'interno di questo orizzonte che si colloca l'oggetto della presente ricerca, volto a comprendere in che modo i sistemi algoritmici generativi intervengano in queste dinamiche. In particolare, s'intende osservare questi aspetti attraverso il confronto di due pratiche performative differenti: concerti di live coding e performance artistiche basate sulla comunicazione dal vivo con un'IA.

Il live coding rappresenta un genere performativo relativamente recente, attorno al quale si è tuttavia già sviluppato un significativo corpus di studi che, da Collins *et al.* (2003) a Blackwell *et al.* (2022), interpreta questa pratica non semplicemente come una tecnica musicale, ma come una vera e propria estetica della programmazione performativa. All'interno del numero monografico del *International Journal of Performance Arts and Digital Media* (2016), interamente dedicato al live coding, Burland e McLean con il contributo *Understanding live coding events* evidenziano anche come l'imprevedibilità del processo – condivisa tanto dai performer quanto dal pubblico – contribuisca a rafforzare il senso di comunità. Un ulteriore contributo proviene dalla ricerca etnografica di Mori (2020), che indaga dall'interno le pratiche e le comunità del live coding, mettendo in luce le sue dimensioni sociali, culturali e relazionali.

Per quanto riguarda l'ambito teatrale, un riferimento fondamentale è il lavoro di Annie Dorsen, che nel saggio *On Algorithmic Theatre* (2012) definisce il teatro algoritmico come una pratica in cui gli algoritmi assumono il ruolo di agenti performativi sulla scena. Con una prospettiva attenta alle trasformazioni della scena multimediale, Anna Maria Monteverdi – da tempo impegnata in una riflessione critica sulle arti digitali (2004) – in *Leggere uno spettacolo multimediale. La nuova scena tra video mapping, interaction design e intelligenza artificiale* (2020) guarda anche alle sperimentazioni teatrali che integrano l'IA, osservando come questa venga spesso addestrata dagli artisti attraverso archivi, dataset e repertori di movimenti e comportamenti. Tali materiali funzionano come una memoria culturale incorporata nel sistema algoritmico, che viene riattivata e rielaborata nel tempo dell'evento. Antonio Pizzo (2021) guarda agli effetti che l'IA generativa produce sulla presenza dal vivo e sull'esperienza spettatoriale. In particolare osserva che, poiché allo spettatore è preclusa la visibilità delle righe di codice – ovvero della struttura di regole che governa il comportamento dell'agente artificiale – la cognizione dello spettatore tende a percepire la *liveness* nelle azioni eseguite, personificando l'azione in una qualche forma di agency autonoma. In questa direzione si colloca anche il volume collettaneo *Cultural Technologies: Robots and Artificial Intelligence in the Performing Arts* (Sone & Savery, 2025), che raccoglie contributi volti ad analizzare come le rappresentazioni e le comprensioni culturali di robot e sistemi di IA influenzano la ricezione del pubblico e favoriscano nuove forme di sperimentazione e innovazione nelle arti performative.

Sebbene l'impiego di sistemi algoritmici nelle pratiche performative sia oggetto di una crescente attenzione critica, gli effetti specifici della *liveness generativa* sulla produzione e valorizzazione dell'incertezza nel tempo dell'evento performativo risultano ancora affrontati in modo frammentario. È in questo spazio che la presente ricerca intende collocarsi, contribuendo a una comprensione delle modalità attraverso cui gli algoritmi generativi possono diventare media utili per alimentare quelle qualità della *liveness* legate all'incertezza, all'imprevedibilità e l'esposizione al rischio.

Metodologia e casi studio

Al fine di indagare gli obiettivi delineati, la presente ricerca adotta un approccio metodologico di natura qualitativa. L'indagine si struttura attorno all'analisi di tre casi studio selezionati per la loro capacità di esemplificare differenti modalità di attivazione della *liveness generativa*: due *AI-driven performance* e una performance di live coding. Per corroborare l'osservazione dei casi lo studio integra un corpus di sei interviste semistruzzate condotte con artisti, organizzatori di eventi e *creative technologist*. L'incrocio tra l'analisi delle opere e il dato empirico raccolto permette di evidenziare metodologie e strategie attraverso cui gli artisti gestiscono l'azione degli agenti algoritmici e ne valorizzano l'imprevedibilità nel contesto performativo.

La scelta di mettere a confronto performance *AI-driven* e di live coding è dovuta al fatto di poter osservare la *liveness generativa* attraverso due modalità performative che si differenziano per il rapporto tra la temporalità di scrittura del codice e il tempo dell'evento. Nelle performance *AI-driven*, la fase di scrittura, programmazione e addestramento del sistema algoritmico avviene prima dell'evento: la performance si fonda su una comunicazione dal vivo con un'IA già configurata, il cui comportamento emerge in scena attraverso l'interazione con performer e pubblico. Nelle pratiche di live coding, il processo di scrittura e messa a punto del codice coincide invece con il tempo stesso della performance. L'atto performativo consiste nella programmazione in tempo reale del sistema algoritmico e nell'esposizione pubblica dei suoi output, che vengono generati, modificati e stratificati attraverso le istruzioni impartite durante l'esecuzione, rendendo visibile e condivisibile il processo computazionale nel suo farsi.

I casi studio selezionati per la presente ricerca sono *Spazio Latente* del collettivo Umanesimo Artificiale e *Radio Pentothal* di Ruggero Franceschini, entrambi esempi di *AI-driven performance*, e *L'Embodiment dell'Effimero* di Francesco “nesso” Corvi, una performance musicale di live coding. I primi due progetti sono stati scelti in quanto rappresentativi di pratiche performative che sviluppano in modo esplicito e articolato forme di interazione dal vivo con sistemi di IA. Entrambe le opere hanno debuttato all'interno di *Residenze Digitali 2024*, un progetto artistico promosso dal Centro di Residenza della Toscana (*Armunia – CapoTrave/Kilowatt*) insieme ai suoi partner, finalizzato a sostenere la ricerca e la produzione di pratiche performative concepite per lo spazio digitale (<https://www.residenzedigitali.it/progetto/>).

In particolare, *Spazio Latente* si configura come una performance ambientata in un futuro distopico caratterizzato proprio dall'assenza di incertezza e casualità, conseguenza di una totale delega dei processi decisionali agli algoritmi. È in questo scenario che un'intelligenza artificiale denominata P1 instaura un dialogo diretto con gli spettatori, invitandoli a innestare nel sistema memorie ed esperienze personali, nel tentativo di recuperare quella dimensione di serendipità che avrebbe progressivamente abbandonato l'esperienza umana.

Radio Pentothal, invece, assume la forma di una “trasmissione radiofonica” ispirata all'esperienza di Radio Alice, animata però da un agente algoritmico programmato per produrre risposte deliberatamente deliranti. Gli output generati dall'IA si manifestano come enunciati che oscillano tra coerenza argomentativa e deriva surreale, dando origine a un flusso linguistico capace di riflettere criticamente sul tema della post-verità.

Entrambe le performance analizzate prevedono una comunicazione dal vivo tra pubblico e IA. Nel caso di *Spazio Latente*, durante la restituzione delle *Residenze Digitali*, la performance ha assunto la forma di una diretta Twitch, configurandosi come uno storytelling interattivo mediato da un'IA. In una prima fase, gli spettatori erano invitati a porre domande all'IA tramite la chat testuale della piattaforma; l'agente algoritmico rispondeva in tempo reale, dando avvio a una conversazione pubblica e condivisa. Nella seconda parte, il pubblico veniva sollecitato a “installare” dei ricordi all'interno del protagonista algoritmico: i messaggi inviati in chat assumevano la funzione di prompt che producevano immagini visualizzate in tempo reale all'interno della diretta. In una successiva evoluzione presentata

al *Sónar Festival di Barcellona* del 2025, l'opera è stata riformulata come installazione video interattiva, integrando dispositivi EEG: in questa versione, le onde cerebrali degli spettatori venivano rilevate e messe in dialogo in tempo reale con l'agente artificiale P1 (https://www.instagram.com/p/DMibP-QNN7g/?img_index=1). Nel caso di *Radio Pentothal*, la comunicazione con l'IA è stata invece strutturata, durante le *Residenze Digitali*, attraverso la piattaforma Zoom. Il pubblico interagiva via chat testuale, inviando domande che venivano inoltrate all'IA, il cui output veniva successivamente restituito al pubblico mediante una lettura performativa da parte degli attori in scena. L'opera è stata successivamente riallestita anche in co-presenza fisica, in occasione del *Life Theatre Arts Media Festival* organizzato da *Zona K* a Milano (<https://zonaklife.it/evento/pentothal/>).

Come caso studio di performance di live coding è stata selezionata *L'Embodiment dell'Effimero* di Francesco “nesso” Corvi, in quanto rappresenta in modo particolarmente efficace una pratica performativa fondata sull'uso di sistemi algoritmici adattativi che alimentano l'imprevedibilità del suono generato. La performance è un concerto audiovisivo che, presentato dal vivo il 19 marzo 2025 a Bologna, presso il Teatro San Leonardo (<https://aaa-angelica.com/aaa/ricercaerinascimento/2024-2025-undicesimo-anno/seconda-parte/francesco-corvi-clara-la-licata/>), riflette in maniera esplicita sul conflitto tra il desiderio umano di controllo tecnologico e la tendenza dei sistemi computazionali a evolvere autonomamente. Il lavoro contrappone l'*embodiment* del musicista – inteso come espressione corporea, situata e contingente – all'impermanenza del software, concepito come uno strumento in costante trasformazione. In questa tensione, l'imprevedibilità e l'instabilità diventano una risorsa estetica.

Il confronto tra questi casi è inoltre supportato da sei interviste semistrutturate, condotte con artisti qualificati, organizzatori di eventi e *creative technologist* tra gennaio e maggio 2025, al fine di far emergere pratiche, metodologie e visioni legate all'uso degli agenti algoritmici in contesti performativi dal vivo. In particolare, sono stati intervistati i tre autori dei casi studio analizzati: Filippo Rosati e Ruggero Franceschini, rispettivamente ideatori di *Spazio Latente* e *Radio Pentothal*, e Francesco “nesso” Corvi, live coder affermato a livello internazionale e performer di *L'Embodiment dell'Effimero*. A questi si aggiungono Valerie Tameau, artista partecipante alle *Residenze Digitali 2024* impegnata con progetti performativi che integrano l'IA all'interno della drammaturgia dell'opera, e Michele Cremaschi, *creative technologist* e ricercatore presso l'Università di Bolzano, attivo nella progettazione di installazioni interattive basate su modelli di IA attraverso cui sviluppa anche una ricerca teorica che indaga gli impatti dell'accelerazione sociale (2024). Infine, è stato intervistato Giovanni Mori, live coder e ricercatore che ha curato anche alcuni dei primi eventi di live coding in Italia.

Complessivamente, le interviste si sono concentrate su temi quali il rapporto tra controllo e apertura all'imprevisto, la percezione del rischio durante la performance e il ruolo della *liveness* degli algoritmi utilizzati nella costruzione dell'esperienza estetica e comunicativa.

Algoritmi e imprevedibilità nella scena contemporanea dal vivo

L'impiego di sistemi d'IA nella scena performativa non introduce necessariamente una rottura con i codici teatrali, che da tempo si è aperto a una combinazione di media, corpi e spazi in grado di accogliere una pluralità di linguaggi e dispositivi (Lehmann, 1999). L'aspetto più interessante nell'utilizzo l'IA risiede nel fatto che introduce una serie di possibilità che vanno ben oltre la funzione di mero strumento tecnico: può affiancare il processo creativo nella scrittura drammaturgica, nella generazione di immagini e contenuti visivi, fino alla produzione di personaggi digitali – o robotici – dotati di comportamenti reattivi e identità discorsive, capaci di interagire in tempo reale con performer e pubblico. Sono queste caratteristiche che fanno scaturire stimolanti riflessioni nell'ambito dell'autorialità distribuita e della *human-computer-interaction* (HCI) (Frauenberger, 2019). I due casi studio presi in esame – *Spazio Latente* e *Radio Pentothal* – si inseriscono proprio in queste linee di sperimentazione: in entrambi, l'IA assume un ruolo generativo all'interno della performance, interagendo con il pubblico e contribuendo all'articolazione drammaturgica.

In *Spazio Latente*, la centralità dell'agente algoritmico si riflette anche nella sua costruzione diegetica. Durante l'intervista Filippo Rosati – uno degli artisti del collettivo Umanesimo Artificiale – racconta che la narrazione dell'opera prende luogo in un futuro non troppo remoto, in cui “tutta la nostra vita è dettata dagli algoritmi, quindi non c'è più aleatorietà, casualità e serendipità”. Il protagonista della performance è P1, un'intelligenza artificiale sviluppata dal collettivo Umanesimo Artificiale, basata su un modello linguistico di tipo GPT opportunamente personalizzato. Come illustrato da Filippo Rosati, l'IA è stata sottoposta a un processo di addestramento mirato, calibrato su specifici parametri comportamentali e stilistici, e informata sulla struttura drammaturgica dell'opera. Tale configurazione ha consentito a P1 di generare risposte coerenti con il tono desiderato dagli autori, orientando le sue interazioni con il pubblico in linea con l'intenzionalità espressiva e narrativa del progetto performativo.

Noi abbiamo creato un nostro GPT, cioè abbiamo fatto un ChatGPT custom, perché puoi costruire dei GPT personalizzandoli con delle personalità. Quindi ho scritto chi era P1 e gli ho raccontato il progetto. (...) Ho costruito GPT P1 come un paziente, dotato di determinate memorie, un tono di voce, una personalità, e gli ho anche chiesto di rispondere sempre in modo un po' criptico. Poi si fa leva su tutto il dataset di ChatGPT, però lo si adatta in modo che risponda secondo la storia che gli hai dato (Filippo Rosati).

Un approccio analogo di modellazione intenzionale del comportamento algoritmico è rappresentato anche da *Radio Pentothal*, progetto ideato da Ruggero Franceschini. In questo caso però, oltre a fornire input e parametri, l'artista ha anche curato la selezione di un dataset specifico per orientare l'algoritmo nella direzione creativa del progetto. L'intento era quello di dar vita a una “trasmissione radiofonica” ispirata all'esperienza di Radio Alice, animata però da un agente algoritmico programmato per “rispondere come Bifo, ma in modo delirante”. Per addestrare l'algoritmo in questo modo, l'artista ha raccontato di essere partito dalla selezione e cura di uno specifico dataset: in questo caso ha utilizzato la piccola libreria

di suo padre. Testi di Franco Berardi, saggi filosofici, *graphic novel*, manuali per realizzare una radio clandestina si uniscono in un dataset eterogeneo, quasi dadaista. Questo ha dato vita a un'IA testuale in grado di esprimersi con una lucidità apparente che ricorda il linguaggio di Bifo, ma al tempo stesso con tratti caotici e deliranti, rispecchiando tanto la fertilità creativa della controcultura degli anni Settanta quanto l'intenzionalità artistica del progetto.

Mi sono reso conto che l'IA, oltre a poter creare immagini e testi, poteva anche essere addestrata tramite dei dataset. Così ho iniziato a riflettere drammaturgicamente su quale potesse essere il dataset adatto per generare un'IA che, non dico parlasse come Bifo, ma che potesse utilizzare la sua filosofia performativa (...) Ho pensato che il dataset ideale per l'addestramento potesse essere il piccolo archivio di mio padre. (...) Quello che ho dovuto fare è stato selezionare dei libri da questa libreria che, uno dopo l'altro, potessero fornire abbastanza elementi per usare l'IA in dialogo, ottenendo così una generazione di contenuti che non fosse monotematica. Se, per esempio, avessi preso solo libri sui concetti filosofici, il risultato sarebbe stato molto piatto. Ho inserito anche elementi più deliranti: questo tipo di materiale, un po' dada se vogliamo, è stato fondamentale per addestrare l'IA (Ruggero Franceschini).

Uno degli aspetti di particolare rilevanza per la presente ricerca riguarda il fatto che, nonostante le strategie e metodologie adottate per indirizzare l'IA lungo una traiettoria coerente con l'intenzione creativa degli artisti, gli output prodotti dalle reti neurali restano comunque in parte indeterminati. Ciò dipende dalla natura stessa di questi sistemi, la cui complessità computazionale li configura come delle *black box*: l'artista può orientare il sistema attraverso la scelta dei dataset, dei parametri e delle istruzioni, ma non potrà mai prevedere con esattezza gli esiti che verranno prodotti dal calcolo probabilistico dell'IA. Come evidenziano Sivertsen *et al.* (2024), la natura stessa dei processi di *machine learning* introduce una *fundamental ambiguity*: “the process is itself ambiguous, both to audiences and (...) to artists too. The artists invoke ambiguities of the AI process itself to provoke audiences (and perhaps themselves) to interpret how AI creates artifacts” (p.13). L'attenzione si sposta dal controllo dell'esito alla progettazione del processo come spazio di indeterminazione produttiva: “artists favor the process (...) over the outcome as a way to create artistic experiences because it is difficult to anticipate the result of a specific model with a specific dataset” (p. 13).

L'indeterminatezza che ne deriva si manifesta dunque come componente strutturale della *liveness generativa*, e quando viene intrecciata all'interazione attiva di performer e spettatori, dà luogo a una forma di contingenza condivisa che costituisce il nucleo di queste esperienze performative.

Un aspetto del mio processo di ricerca è quello dell'imprevedibilità. (...) Mi interessano i processi caotici e appunto la *liveness* che ne scaturisce. Questo fa sì che in alcuni lavori faccio partecipare il pubblico (...) e in questo in particolare l'AI viene interrogata in diretta (Ruggero Franceschini).

Dal mio punto di vista il fatto che l'intelligenza artificiale porta delle cose inaspettate è esattamente quello per cui serve artisticamente l'intelligenza artificiale (Filippo Rosati).

L'imprevedibilità insita nella *liveness generativa* diventa risorsa espressiva anche nelle performance di live coding. Tuttavia, a differenza delle *AI-driven performance*, nel live coding è il processo stesso di scrittura e programmazione dell'algoritmo a venire eseguito in tempo reale (Collins *et al.*, 2003). L'attività performativa di programmazione permette a questa pratica di esprimere sia caratteristiche appartenenti alla composizione musicale – pratica che tendenzialmente avviene in un tempo differito tra l'ideazione del pensiero sonoro e la sua realizzazione pratica – sia all'improvvisazione, dove invece le due fasi coincidono temporalmente (Mori, 2020). Oltre a produrre output sonori immediati, gli eventi di live coding si distinguono anche per una proiezione visiva del codice algoritmico. Come raccontano entrambi i musicisti intervistati, ciò che il pubblico osserva non è poi necessariamente comprensibile sul piano tecnico: la maggior parte delle persone non legge o interpreta realmente il codice che scorre. Eppure, non è la comprensione razionale ad attivare il coinvolgimento, quanto piuttosto la possibilità di assistere a un divenire. Vedere numeri e comandi che si modificano in diretta e associarli ai cambiamenti sonori restituisce un senso di vitalità: quella dell'algoritmo che agisce e reagisce nel tempo presente (Rein *et al.*, 2019).

Il tour che ho appena fatto partiva da uno schermo vuoto e andavo a costruire una serie di cose e per certi versi il processo era abbastanza chiaro. Vedi che appaiono dei codici e che creano dei cambiamenti, non mi aspetto dal pubblico che capiscano quello che c'è scritto (...), ma la cosa interessante è: a te ascoltatore, in che posizione ti mette vedere questi codici che appaiono e sentire questi cambiamenti che avvengono in relazione a qualcosa che cambia? E questo crea un'immersività, unita anche alla proiezione, molto maggiore (Francesco Corvi).

Francesco Corvi, noto con il nome d'arte "nesso", ha presentato la performance di live coding *L'Embodiment dell'effimero* nella primavera del 2025. In questa occasione, l'artista ha impiegato un software dotato di "processi adattivi", progettati per modificare dinamicamente il proprio comportamento in risposta alle condizioni performative contingenti. Questa architettura computazionale, capace di auto-osservarsi e di ristrutturarsi in tempo reale, è in grado di generare in autonomia timbriche, ritmi e texture sonore, così da evitare una predeterminazione lineare della performance.

L'ultimo progetto che ho sviluppato è incentrato sull'avere un sistema che si controlla attraverso il live coding ma che ha poi una serie di caratteristiche per poter auto-analizzare quello che sta facendo. In particolare, analizza i segnali audio in uscita e la struttura momentanea del software e queste cose poi vengono usate per utilizzare dei "processi adattivi", cioè processi che si adattano e modificano il loro comportamento in base a quello che sta succedendo. Questo aveva l'obiettivo di vedere il computer come un performer, non solo come uno strumento che esegue le istruzioni impartite dall'artista. (...) È anche un po' tutta l'idea della performance, *L'Embodiment dell'effimero*, cioè l'effimero in quanto tecnologia che muta, e l'embodiment che in questo caso è il musicista che cerca di controllarlo, ma che poi fondamentalmente alla fine non ci riesce perché questa cosa sta cambiando in continuazione (Francesco Corvi).

L'*Embodiment dell'effimero* si configura come un'interazione continua con un sistema in costante mutazione, in cui il performer abita uno spazio liminale tra il controllo e la contingenza che dà forma a un'esperienza percettiva fluida e aperta all'imprevisto. Come osservato anche da Valerie Tameau nel corso dell'intervista sostenuta per la presente ricerca: "mettere a fuoco quello scarto lì, cioè quel non poter controllare, permette anche un dialogo con la tecnologia, che va oltre il suo mero utilizzo". Durante l'attività di programmazione nesso si confronta così con un dispositivo algoritmico che agisce in quanto entità dialogica, capace di deviare, sorprendere e co-costruire l'evento sonoro in tempo reale.

L'esito della performance viene quindi influenzato dall'azione congiunta di musicista e algoritmo. Questo crea, come sostenuto da Giovanni Mori, una doppia circolarità nel pensiero creativo del live coder: l'artista formalizza il pensiero musicale, ma poi il suono viene eseguito e prodotto dall'algoritmo. Se negli strumenti analogici il pensiero musicale si traduce in movimento che corrisponde direttamente a suono, per il live coder questo non accade, poiché non c'è un *feedback* immediato del suono: l'esecutore del pensiero dell'artista è un software, il cui algoritmo interpreta il pensiero del musicista. Nel caso della performance presa in analisi, l'artista, nesso, guida l'evento con il suo pensiero musicale, ma poi è l'algoritmo, l'esecutore del suo pensiero, a completarsi e a produrre il suono. È questa distanza tra pensiero ed esecuzione che aumenta l'imprevedibilità e il rischio di questo tipo di performance.

C'è una doppia circolarità nel pensiero creativo del live coder: (...) l'artefice del prodotto sonoro, non ha direttamente influenza sul suono stesso, quindi c'è una distanza tra il pensatore della musica e la musica prodotta, si crea questa doppia circolarità per cui il pensiero musicale va a influenzare l'esecutore – in questo caso l'algoritmo – e l'esecutore poi utilizza la sua interpretazione del tuo pensiero musicale per produrre il suono. (...) In queste pratiche in cui hai un frame che divide l'esecutore dal suono e si aggiunge una quantità di rischio maggiore (Giovanni Mori).

Drammaturgie parametriche e contingenza virtuale

Tra gli aspetti più rilevanti emersi dall'indagine vi è il fatto che l'elaborazione creativa dell'incertezza attraverso gli algoritmi non si traduce mai in un abbandono passivo alla contingenza, ma piuttosto in un'apertura intenzionale e controllata all'imprevisto. Come sintetizza bene Esposito (2022), "lo scopo paradossale della programmazione degli algoritmi intelligenti è costruire in modo controllato delle macchine imprevedibili", e questo si riflette anche dove la programmazione degli algoritmi diventa processo e pratica artistica. Nel caso delle performance di live coding, l'imprevedibilità è parte integrante dell'esperienza performativa, ma non si manifesta mai come totale casualità dell'evento. L'artista entra in scena con un insieme di conoscenze, schemi mentali e moduli di codice preconfigurati, frutto di una lunga familiarità con il linguaggio di programmazione e con la macchina. Corvi racconta anche di strategie per introdurre l'errore come parte del linguaggio performativo, fino a utilizzarlo come gesto compositivo.

L'imprevedibilità è molto importante per una performance di live coding, almeno per quello che faccio io, e penso che ci sia una certa pianificazione che però è una pianificazione che contempla l'imprevisto. (...) L'errore quando suoni è quando succede qualcosa che blocca tutto e non riparte. Se però tu riesci a inserire dei meccanismi che ti permettono di sbagliare, a quel punto puoi inserire quella cosa nel tuo linguaggio performativo. Io, per esempio, tantissime volte ho fatto finire la performance con il crash del computer. (...) Inizio a mandare dei valori altissimi per cui mi dico "non so esattamente cosa succederà, ma so che anche se crasha tutto la performance è comunque finita", dunque ti stai dando volontariamente la possibilità di sbagliare (Francesco Corvi).

L'interazione tra diverse istruzioni codificate, la presenza di elementi randomici e l'esecuzione in tempo reale generano una stratificazione di processi che sfugge al pieno controllo completo del performer, ma che rimane comunque circoscritta dalle scelte, espressività e intenzionalità del musicista. Questo aspetto emerge anche in *Spazio Latente* e *Radio Pentothal*, dove le IA utilizzate – per quanto siano in grado di generare output mai completamente prevedibili – seguono comunque una traiettoria precisa, definita dagli artisti attraverso scelte consapevoli di training, selezione dei dataset e impostazione dei parametri. Dunque, non producono mai risultati completamente casuali e generici ma, come ricorda anche Antonio Pizzo (2021, p. 97), "the surface content chosen by the algorithm is the consequence of author's choice". La scrittura e programmazione degli algoritmi – ovvero l'insieme di operazioni che comprende la scelta o la progettazione di un modello linguistico, la definizione del dataset, l'addestramento del modello e la costruzione di un comportamento attraverso input e codici – può essere così intesa come una nuova forma di drammaturgia: come il testo drammatico orienta l'interpretazione dell'attore umano attraverso un sistema di istruzioni che struttura l'azione scenica, allo stesso modo il codice algoritmico opera come uno script per un attore non-umano, predisponendo traiettorie di comportamento che vengono attivate e trasformate nel tempo della performance.

Questa forma di scrittura non è, tuttavia, finalizzata a determinare l'esito della performance in modo assoluto. Al contrario, agisce dentro una tensione costante tra controllo e apertura. In molti contesti performativi tradizionali – come, ad esempio, il teatro di prosa – la drammaturgia tende a funzionare come una strategia di contenimento dell'imprevisto: le istruzioni sono concepite per ridurre al minimo l'aleatorietà e garantire la stabilità dell'evento. Tuttavia, nelle performance con agenti artificiali o in quelle basate sul live coding, il testo-programma assume una funzione diversa: non disabilita l'incertezza presente, ma la incorpora consapevolmente all'interno della struttura, progettando margini di deviazione e instabilità.

Questa modalità di scrittura può essere letta come una *drammaturgia parametrica*, in cui ciò che viene scritto non è un contenuto finito ma un insieme di regole, parametri e condizioni che abilitano un comportamento performativo. Si tratta dunque di una drammaturgia che definisce processi: non prescrivere cosa *deve* accadere, ma come qualcosa *può* accadere, lasciando che l'effettiva manifestazione scenica emerga dall'interazione situata tra istruzioni, macchina, performer e contesto. La drammaturgia parametrica è quindi intimamente legata a una concezione non lineare e generativa della

creazione artistica: l'autore perde la centralità come produttore unico di senso, per assumere il ruolo di architetto di comportamenti potenziali, di curatore di complessità performative la cui esecuzione implica margini di instabilità, ambiguità e variazioni.

La responsabilità dell'artista è quella di creare un setting e qui torno a Claire Bishop, quella famosa definizione che c'è in *Artificial Hell* in termini artificiali dell'artista come creatore di situazioni. Crei una situazione sulla quale poi non hai pieno controllo, il pubblico ci contribuisce e altri attori, in questo caso non-umani come quello di un AI, contribuiscono alla cosa, ma l'artista ultimamente alla fine è colui che ha creato la situazione perché questo possa succedere (Ruggero Franceschini).

Bomba *et al.* descrivono la relazione tra artista e programma nei termini del *continuum coreografo-performer* (2023), mutuandolo dalle pratiche della coreografia contemporanea e delle *instruction pieces*. In queste esperienze – a partire dalle *delegated performances* e dalle pratiche dei collettivi femministi (Bishop, 2012; Meskimmon, 2012; Lee, 2023) – i ruoli di chi scrive e di chi esegue tendono a sfumare: il coreografo elabora delle istruzioni che vengono interpretate da altri, spesso non professionisti, i quali portano in scena una co-creazione parzialmente autonoma. Quando però questa delega viene trasferita a un agente non-umano, si apre un nuovo spazio di riflessione: l'artista-programmatore mantiene il controllo sul testo-programma, ma lascia all'algoritmo la possibilità di esplorare, all'interno di quei vincoli, traiettorie imprevedibili. In altri termini, l'artista sfrutta la propria agentività per guidare l'agentività dell'algoritmo, creando una situazione di "aleatorietà controllata": in entrambe le tipologie di performance prese in esame – sia quelle basate su IA addestrata, sia quelle fondate sul live coding, gli artisti creano situazioni in cui esercitano un controllo mirato e consapevole per gestire l'imprevedibilità insita delle tecnologie, ma allo stesso tempo lasciano aperto lo spazio alla contingenza e al caso che si manifestano nell'interazione dal vivo con gli algoritmi.

La contingenza così generata viene sfruttata dagli artisti per alimentare un'incertezza che si distingue da quella che può produrre un altro artefatto o un agente umano. Tale specificità emerge con chiarezza se si sposta l'asse dell'indagine sul piano della comunicazione. In ambito artistico, spesso gli oggetti tecnici vengono utilizzati per generare esiti imprevedibili, ma questa imprevedibilità non è immanente all'artefatto, bensì alla situazione: deriva dal contesto di interazione in cui lo strumento viene impiegato, non dal funzionamento deterministico del dispositivo stesso. Gli algoritmi complessi, invece, occupano una posizione differente. Il calcolo probabilistico degli algoritmi generativi è frutto di un processo "opaco", tale rendere questi sistemi delle *black box*. Ciò implica che con gli stessi parametri e lo stesso input si possono generare output sempre diversi, senza che l'essere umano possa ricostruire con certezza il percorso computazionale. È questa opacità tecnica che rende possibile percepirli come contingenti, cioè come entità che avrebbero potuto produrre sempre un risultato diverso.

Da qui si apre un passaggio cruciale: anche l'algoritmo, come ogni dispositivo tecnico, non è ontologicamente contingente, ma diventa comunicativamente tale. L'imprevedibilità percepita non è una proprietà intrinseca della macchina, bensì il riflesso della contingenza

umana proiettata su di essa. In altri termini, il non-sapere dell'osservatore viene attribuito all'algoritmo, generando quella che Esposito definisce *contingenza virtuale* (2017): questa si attiva quando un partner comunicativo è confrontato con la propria contingenza, rivista e riflessa in modo da simulare le condizioni della comunicazione. La comunicazione artificiale avviene infatti quando il ricevente interpreta l'output come qualcosa che avrebbe potuto essere altrimenti: è la propria indeterminatezza ad essere proiettata sull'algoritmo.

La differenza rispetto alla comunicazione interumana sta proprio qui. Tra soggetti umani la comunicazione si fonda su una doppia contingenza (Parsons, 1951; Luhmann, 1984): "la contingenza è doppia non semplicemente in quanto localizzata in istanze separate (Ego ed Alter) ma in quanto riguarda la propria incertezza relativamente all'incertezza altrui" (Boccia Artieri 1998, p. 51). Poiché gli algoritmi non sono soggetti all'incertezza che caratterizza l'agire umano, la comunicazione con essi non scaturisce dall'incontro di due contingenze. Essa assume piuttosto la forma di una simulazione riflessiva della doppia contingenza, nella quale l'utente si rapporta a se stesso come a un alter ego, trovando nell'opacità tecnica del sistema algoritmico uno spazio per proiettare il proprio non-sapere. Inoltre, a differenza di altri strumenti, gli algoritmi hanno la capacità di rinnovare continuamente la loro opacità nel tempo. Possiamo definirle come delle *real-time evolving entities*: apprendono, si adattano, reagiscono agli input dell'utente. Questa dinamica stabilizza e rilancia costantemente la percezione di contingenza, alimentando un dialogo che produce informazioni nuove e imprevedibili rispetto a quelle già possedute dal soggetto.

Nelle arti performative dal vivo, tale specificità diventa particolarmente significativa. La scena non è solo lo spazio della contingenza situazionale prodotta dagli artisti e dagli spettatori, ma anche quello in cui si innesta la contingenza virtuale della comunicazione artificiale. Di conseguenza, è l'incontro tra input umani, output algoritmici e riflessività dell'osservatore a generare una diversa esperienza della contingenza, che intensifica l'incertezza dell'esperienza performativa dal vivo.

Conclusioni

In un panorama socio-tecnologico in cui gli algoritmi sono frequentemente impiegati per contenere la complessità del presente e per ridurre l'esposizione al rischio, l'arte performativa dal vivo si offre come contospazio critico e riflessivo, in cui la stessa tecnologia viene rimessa in gioco e orientata in senso opposto. Nell'uso ordinario, il calcolo probabilistico degli algoritmi viene adottato principalmente con la funzione di ridurre l'incertezza insita nelle situazioni caratterizzate da un eccesso di possibilità. Di fronte a una molteplicità di opzioni, il partner comunicativo si affida all'algoritmo affinché lo guidi verso la scelta ritenuta ottimale. In questo senso, l'output algoritmico, pur risultando imprevedibile per l'utente, è desiderato e atteso proprio in quanto capace di ridurre l'indeterminatezza del presente. Dall'analisi del contenuto delle interviste condotte per questa ricerca emerge invece come gli artisti intendano valorizzare l'incertezza del momento presente proprio

attraverso gli stessi agenti algoritmici che rispondono al bisogno sociale della sua riduzione. I casi studio analizzati evidenziano come l'arte performativa dal vivo elabori strategie creative per appropriarsi dell'incertezza e della contingenza della contemporaneità (Gemini 2003), trasformandole in dispositivi estetici e comunicativi.

In definitiva, la mediazione algoritmica contribuisce a riformulare i parametri della *liveness*, accentuandone le componenti di incertezza e imprevedibilità (Georgi 2014; Scannell 2014): si producono nuove soglie di instabilità in cui algoritmi diventano co-agenti della performance capaci di generare deviazioni, sorprese e disallineamenti. L'imprevedibilità non è data soltanto dall'interazione situazionale ma anche dalla contingenza virtuale della comunicazione artificiale. È proprio questa dinamica comunicativa a moltiplicare le traiettorie possibili dell'esperienza scenica. Gli artisti creano dei settings, delle drammaturgie parametriche per vivere ed esperire quella opacità e la contingenza da essa prodotta. Si può parlare, in questo senso, di una "estetica della scatola nera" (Befera, 2023): le performance che impiegano algoritmi generativi non mirano a disvelare o rendere trasparente l'opacità tecnica che caratterizza questi sistemi, ma piuttosto a farne esperienza diretta nel momento stesso della performance. La *liveness generativa* trasforma l'opacità del calcolo probabilistico in un dispositivo spettacolare che trasforma l'esperienza spettatoriale:

Già il suono ti comunica una certa *liveness*, che non è la *liveness* di una persona che suona ma di un processo algoritmico, che ha una sua *liveness*. (...) Tenzionalmente quando ascolti della musica generativa, il fatto di sapere che c'è un certo processo computazionale in atto, cambia la posizione con cui tu ascoltatore ne fruisce (Francesco Corvi).

Sebbene l'impiego di sistemi algoritmici sia spesso oggetto di critica in quanto potenziale fattore di minaccia alla creatività umana, le esperienze analizzate evidenziano come, al contrario, la mediazione algoritmica nelle performance dal vivo possa attivare forme inedite di elaborazione e comunicazione dell'incertezza e dell'imprevedibilità del presente in chiave espressiva e creativa.

Nota Biografica

Alex Dellapasqua è dottorando in *Studi Umanistici - Scienze della Comunicazione e Cultura Digitale* presso l'Università di Urbino Carlo Bo dove, con la supervisor prof.ssa Laura Gemini, si dedica a una ricerca interdisciplinare sulla *liveness* prodotta dagli algoritmi generativi all'interno delle performance dal vivo. Ha pubblicato: *I Diritti umani nell'Arte Relazionale. La rivendicazione dei diritti umani nelle opere di Félix Gonzàles-Torres e Rirkrit Tiravanija* (in "I Diritti e le Arti a Dialogo", Dell'Aversana F., Persiani, 2023), e *Physis e Téchne nell'estetica auto-generativa. Dall'antropocentrismo al più-che-umano: l'esperienza artistica di Entangled Others* (in pubblicazione per "estetica. studi e ricerche", 2025).

Bibliografia

- Arielli, E. (2023). Rischi calcolati. Strategie e paradossi dell'incertezza nelle arti. *estetica. studi e ricerche*, 3, 509-524. <https://doi.org/10.14648/112429>
- Auslander, P. (2012). Digital Liveness. A Historico-Philosophical Perspective. *PAJ: A Journal of Performance and Art*, 34(3), 3-11. https://doi.org/10.1162/PAJJ_a_00106
- Auslander, P. (2022). *Liveness. Performance in a Mediatized Culture* (3rd edition). Routledge.
- Bachmann-Medick, D. (2016). *Cultural Turns. New orientations in the study of culture*. De Gruyter.
- Baudrillard, J. (2000). Il gioco dell'incertezza. Intervista di E. Baj. *Il Corriere della Sera*, 20 agosto.
- Bauman, Z. (2007). *Liquid Times: Living in an Age of Uncertainty*. Polity Press.
- Beck, U. (1992). *Risikogesellschaft: Auf dem Weg in eine andere Moderne*. Carrocci.
- Befera, L. (2023). The black box aesthetics: performative dynamics between artificial intelligence and human interactive staging. *International Journal of Performance Arts and Digital Media*, 19(2), 230-246. <https://doi.org/10.1080/14794713.2023.2196890>
- Berger, P.L., Luckmann, T. (1966). *The Social Construction of Reality*. Il Mulino.
- Bishop, C. (2012). Delegated Performance: Outsourcing Authenticity. *CUNY Academic Works*, 140, 91-112. https://doi.org/10.1162/OCTO_a_00091
- Blackwell, A.F., Cocker, E., Cox, G., McLean, A., Magnusson, T. (2022). *Live Coding: A User's Manual*. MIT Press. <https://doi.org/10.7551/mitpress/13770.001.0001>
- Boccia Artieri, G. (1998). *Lo sguardo virtuale. Itinerari socio-comunicativi nella deriva Tecnologica*. FrancoAngeli.
- Boccia Artieri, G. (2015). Mediatizzazione e Network Society: un programma di ricerca. *Sociologia della Comunicazione*, 50(2), 62-69. <https://doi.org/10.3280/SC2015-050006>
- Bomba F., Menéndez-Blanco M., Grigis P., Cremaschi M., De Angeli A. (2024). The Choreographer-Performer Continuum: A Diffraction Tool to Illuminate Authorship in More Than Human Co-Performances. *ACM Transactions on Computer-Human Interaction*, 31(6), 1-23. <https://doi.org/10.1145/3689040>
- Bowen S. (2024). "If it can be done, it will be done": AI ethical standards and a dual role of public relations. *Public Relations Review*, 50(5), 1-13. <https://doi.org/10.1016/j.pubrev.2024.102513>
- Brandtzaeg, P., Følstad, A., & Skjuve, M. (2025). Emerging AI individualism: how young people integrate social AI into everyday life. *Communication and Change*, 1(11), 1-22. <https://doi.org/10.1007/s44382-025-00011-2>
- Bright J., Enock F., Esnaashari S., Francis J., Hashem Y., Morgan D. (2024). Generative AI is already widespread in the public sector: Evidence from a survey of UK public sector professionals. *Digital Government: Research and Practice*.

<https://doi.org/10.48550/arXiv.2401.01291>

- Brilli, S., Gemini, L., Giuliani, F. (2023). Theatre without theatres: Investigating access barriers to mediatized theatre and digital liveness during the covid-19 pandemic. *Poetics. Journal of Empirical Research on Culture, the Media and the Arts*, 97. <https://doi.org/10.1016/j.poetic.2022.101750>
- Burland, K., McLean, A. (2016). Understanding live coding events. *International Journal of Performance Arts and Digital Media*, 12(2), 139-151. <https://doi.org/10.1080/14794713.2016.1227596>
- Caporaletti, V. (2005). *I processi improvvisativi nella musica. Un approccio globale*. LIM.
- Cremaschi, M., Dorfmann, M., De Angeli, A. (2024). A Steampunk Critique of Machine Learning Acceleration. *Proceedings of the 2024 ACM Designing Interactive Systems Conference (DIS '24)*. Association for Computing Machinery, 246-257. <https://doi.org/10.1145/3643834.366068>
- Collins, N., Mclean, A., Rohrhuber, J., Ward A. (2003). Live coding in laptop performance. *Organised Sound*, 8(3), 321-330. <https://doi.org/10.1017/S135577180300030X>
- Dorsen, A. (2012). On Algorithmic Theatre. *Theater Magazine*, 42(2). Preso da: <https://theatermagazine.org/web-features/article/algorithmic-theater>
- Esposito, E. (1992). *L'operazione di osservazione. Costruttivismo e teoria dei sistemi sociali*. FrancoAngeli.
- Esposito, E. (2017). Artificial Communication? The Production of Contingency by Algorithms. *Zeitschrift für Soziologie*, 46(4), 249-265. <https://doi.org/10.1515/zfsocz-2017-1014>
- Esposito, E. (2022). *Artificial Communication. How algorithms produce social intelligence*. Egea.
- Fischer-Lichte, E. (2014). *Estetica del performativo. Una teoria del teatro e dell'arte*. Carrocci.
- Frauenberger, C. (2019). Entanglement HCI The Next Wave?. *ACM Transactions on Computer-Human Interaction*, 27(1), 1-29. <https://doi.org/10.1145/3364998>
- Gemini, L. (2003). *L'incertezza creativa. I percorsi sociali e comunicativi delle performance artistiche*. FrancoAngeli.
- Gemini, L. (2013). Teatro riflessivo e intimità connessa: parole chiave per Santarcangelo 13. *D'Ars. Magazine of contemporary art and cultures*, 215, 56-59.
- Gemini, L., Brilli, S. (2023). *Gradienti di liveness. Performance e comunicazione dal vivo nei contesti mediatizzati*. FrancoAngeli.
- Georgi, C. (2014). *Straging Liveness Via Mediatization the Use of Film and Video in Contemporary British Theatre and Performance*. De Gruyter Mouton.
- Hepp, A., Couldry, N. (2017). *The mediated construction of reality*. Polity Press.
- Hughes, G.E., Maxwell, J.C. (1968). *An Introduction to Modal Logic*. Methuen.
- Krotz, F. (2007). The meta-process of mediatisation' as a conceptual frame. *Global media and communication*, 3(3), 256-260. <https://doi.org/10.4324/9781315612942-18>

- Lee, J. (2023). *Milestones in Asian American Theatre*. Routledge.
- Lehmann, H.S. (1999). *Postdrammatic Theatre*. Cue Press.
- Lovari, A., Ducci, G. (2022). *Comunicazione pubblica. Istituzioni, pratiche, piattaforme*. Mondadori.
- Luhmann, N. (1984). *Introduction to systems theory*. Pensa Multimedia.
- Luhmann, N. (1991). *Risk: A sociological theory*. Bruno Mondadori.
- Meskimmon, M. (2012). *Women Making Art*. Routledge.
- Monteverdi, A.M. (2004). *Le arti multimediali digitali. Storia, tecniche, linguaggi, etiche ed estetiche del nuovo millennio*. Garzanti.
- Monteverdi, A.M. (2020). *Leggere uno spettacolo multimediale. La nuova scena tra video mapping, interaction design e intelligenza artificiale*. Audino.
- Mori, G. (2020). *Live Coding? What does it mean? An Ethnographical Survey on an Innovative Improvisational Approach*. Aracne.
- Neu, D.A., Lahann, J., Fettke, P. (2022). A systematic literature review on state-of-the-art deep learning methods for process prediction. *Artif Intell Rev*, 55, 801-827. <https://doi.org/10.1007/s10462-021-09960-8>
- Parsons, T. (1951). *Toward a General Theory of Action*. Harvard University Press.
- Pizzo, A. (2021). Performing/Watching Artificial Intelligence On Stage. *Skenè. Journal of Theatre and Drama Studies*, 7(1), 91-109. <https://doi.org/10.13136/sjtds.v7i1.308>
- Rein, P., Ramson, S., Lincke, J., Hirschfeld, R., Pape, T. (2019). Exploratory and Live, Programming and Coding. A Literature Study Comparing Perspectives on Liveness. *The Art, Science, and Engineering of Programming*, 3(1), 1-33. <https://doi.org/10.22152/programming-journal.org/2019/3/1>
- Saltz, D.Z. (2001). Live media: Interactive technology and theatre. *Theatre Topics*, 11(2), 107-130. <https://doi.org/10.1353/tt.2001.0017>
- Sanden, P. (2013). *Liveness in modern music: musicians technology and the perception of performance*. Routledge.
- Samek, W., Montavon, G., Vedaldi, A., Hansen, L., K., Müllerm, K., (2019). *Explainable AI: Interpreting, Explaining and Visualizing Deep Learning*. Springer Cham. <https://doi.org/10.1007/978-3-030-28954-6>
- Scannell, P. (2014). *Television and the meaning of live*. Polity Press.
- Schechner, R. (1983). *La teoria della performance 1970-1983*. Bulzoni.
- Selten F., Klievink B. (2024). Organizing public sector AI adoption: Navigating between separation and integration. *Government Information Quarterly*, 41(1), 1-14. <https://doi.org/10.1016/j.giq.2023.101885>
- Sivertsen, C., Salimbeni, G., Løvlie, A. S., Benford, S. D., Zhu, J. (2024). Machine learning processes as sources of ambiguity: Insights from ai art. *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems*, 1-14. <https://doi.org/10.1145/3613904.3642855>
- Sone, Y., Savery, R. (2025). *Cultural Technologies. Robots and Artificial Intelligence in the Performing Arts*. Routledge.
- Turner, V.W. (1986). *The Anthropology of Performance*. PAJ Publications.

“Understanding Video Activism on Social Media” by Jens Eder, Britta Hartmann, and Chris Tedjasukmana (2025)

Sabrina Brignoli
Università degli Studi di Urbino Carlo Bo

Academic discourse on the relationship between digital media and civic engagement has evolved significantly over the past two decades. We have moved far beyond the initial waves of cyber-optimism, which considered the Internet primarily as a liberating technology capable of amplifying marginalised voices, and through the subsequent phase of cyber-pessimism, which focused on surveillance, slacktivism, and the decline of public discourse. Contemporary research now acknowledges that our digital public sphere cannot be entirely grasped without a comprehensive understanding of the interplay between platform algorithms, governance structures, visual cultures and the political agency of citizens.

In this landscape, *Understanding Video Activism on Social Media* by Jens Eder, Britta Hartmann, and Chris Tedjasukmana offers a critical and mature reflection on the constraints and potentials of activism in the digital age. The volume provides a comprehensive theoretical and empirical analysis of the phenomenon of video activism. This is defined as “the use of online video by individuals and groups in civil society who advocate for their political causes largely independently of the [...] established institutions” (p. 2) and is situated at the intersection of politics, video, and social media. Adopting an interdisciplinary approach that combines media and film studies, digital and political communication, sociology, and social movement studies, the book explores how activist videos are created and distributed to document and denounce injustices, foster democratic participation, and propose alternative narratives to dominant discourses within our contemporary society.

A significant theoretical contribution of the book is the classification of activist videos into five main types, each characterised by distinct logics and functions that capture the hybridity of online video practices: witness videos; documentary videos; video blogs and personal political commentary; mobilisation and campaign videos; and playful subversion or guerrilla communication videos. The heterogeneity of contemporary activism is further delineated through the classification of four main actor categories: small media collectives; political influencers; large non-governmental organisations (NGOs); and hashtag protest movements. Each category is examined in depth in terms of its specific practices, resources, challenges, and empirical examples.

To situate these contemporary dimensions, the authors trace the historical roots of video activism, with a particular focus on the German context, starting from the leftist documentary film movements that sought to represent marginalised voices. The historical discussion is centred on three key questions: the collective nature of film production, in contrast to today's more individualised formats; the politics of representation and the ethical debate on how to effectively portray marginalised voices; and the political function of aesthetics. The authors further demonstrate how technological developments, from the first portable video cameras to social networks, have played a decisive role in democratising the production and distribution of activist communication.

These theoretical and historical discussions are integrated with numerous case studies and empirical materials, allowing the authors to address the practical dimensions of activist videos and how they generate political and social impact. The authors examine different areas of video practices, including creative design, aesthetics, production, as well as distribution infrastructures, and networking. They delineate four narrative structures capable of triggering different emotional and cognitive responses in viewers, illustrating how aesthetic and narrative strategies contribute to a video's reach and persuasive power.

With regard to the socio-political impact of activist videos, the authors propose a sequence of effects: beginning with the capturing of attention, then the provocation of affective and cognitive engagement, and ultimately the motivation of individual and collective action. They explain that in an overcrowded media landscape, attention constitutes a scarce and valuable resource, unevenly distributed and controlled by platform algorithms and commercial interests. Activists must therefore navigate complex dynamics of visibility, also struggling against users' defensive reactions, selective exposure, and emotional fatigue. A central concept here is spreadability, which depends on networked attention, defined as the interplay between user selectivity, social relations, and algorithms. The authors explain how activists can foster virality and generate cyber-cascades that, through the linking of numerous individual micro-narratives, give rise to collective macro-narratives that challenge dominant ideological frames. Furthermore, the text highlights that affective engagement can play a crucial role in the effectiveness of activist videos, especially through emotionally charged oppositions between "us" and "them", capturing attention and motivating participation more powerfully than rational persuasion alone.

A central theme running through the entire volume is the question of power. The authors situate video activism within the broader dynamics of the contemporary digital public sphere and the "Splinternet", a term used to describe the system of social media platforms that is understood as a fragmented, ambivalent, and unpredictable environment shaped by contradictory forces. On one hand, digital platforms facilitate the rapid and inexpensive production and dissemination of video, fostering new forms of participatory and collective expression. On the other hand, these same platforms operate through algorithmic curation, content moderation, and profit-driven logics, which impose constraints on activists and distort their communication. These dynamics are further scrutinised in the final chapter, which focuses on the current geopolitical and media landscape characterised by disinformation and the resurgence of authoritarianism. Through the case study of the

Russian war of aggression against Ukraine, the authors illustrate how activist intervention is made more complex by a context of political and media disorder, fragmented media systems, and state propaganda, exposing activists to significant risks of repression and censorship.

In contrast to this current media system characterised by fragmentation, polarisation, and attention hierarchies, activism is framed as an act of opposition and resistance to hegemonic power structures, whether governmental, corporate, or embedded in platform architectures. The primary goal of activists is the creation of counterpublics: alternative communicative spaces that challenge hegemonic narratives and give voice to marginalised groups and their concerns. To explain this, the authors use the concept of political framing, which refers to how political issues are represented in the media and in the collective consciousness. In a public sphere dominated by the unequal distribution of communicative power, activist interventions serve as a necessary counterweight to dominant discourses, providing alternative voices, narratives, and spaces for contestation.

Explicitly addressing an audience that extends beyond academia, including activists and educators in the political and media fields, the authors provide reflections also on ethical and practical principles of activist communication and democratic cooperation. The text addresses questions of representation, aesthetic function, emotional manipulation, and political responsibility, as well as directly calling upon activists to embrace solidarity, cooperation, and shared strategies. The authors observe that, despite their shared goals, collaboration among activist actors remains rare, limiting their collective political potential. To this end, the book concludes with an appeal for “diversity in solidarity”. Defined as mutual recognition and collective responsibility, solidarity is proposed as an ethical and strategic response to the inequalities of the contemporary digital public sphere. The authors therefore encourage activists to engage in collective efforts across movements, platforms, and media forms to resist anti-democratic forces and sustain democratic participation.

The declared aim of the entire book is to clarify fundamental concepts about video activism and develop a general framework that can serve as a foundation for academic research. The authors succeed in creating a comprehensive groundwork that combines an analysis of the aesthetic and narrative forms of video with a critical examination of the distribution logics and power structures of platforms, grounded in a socio-historical perspective on social movements and the public sphere. This is supported by a wide repertoire of empirical examples, ranging from historical documentaries to contemporary social media videos, although the predominance of case studies from the German and Western context suggests a potential for further comparative and transnational perspectives.

The strong emphasis on video as a distinct category of activism provides a clear conceptual framework. However, this also constitutes a potential limitation: many of the discussed dynamics, such as affective mobilisation, participatory communities, and algorithmic visibility, are equally applicable to other digital forms of activism beyond videos. While the focus on videos is justified by their visual, narrative and affective power, the notion of video activism may risk overstating its specificity within the broader landscape of digital

activism. This raises questions about whether it should be treated as a distinct analytical category or as a single expression within a wider range of digital activist practices.

Nevertheless, *Understanding Video Activism on Social Media* stands as a valuable contribution successfully combining theoretical clarity with detailed empirical analysis and ethical considerations, and providing a comprehensive overview of activist video practices and their socio-political implications. The balance between critical realism and optimism enables the authors to recognise and reflect on the structural constraints of activism imposed by platform surveillance, while also recognising the democratic potential and creative agency of activist media practices. Beyond providing a taxonomy and history of activist video practices, the volume invites scholars and practitioners to rethink what it means to act politically under platform capitalism, depicting contemporary activism as a complex process of negotiation and simultaneous resistance and adaptability to the dominant logics of platforms and media institutions.

Nota biografica

Sabrina Brignoli (sabrina.brignoli@uniurb.it) is a PhD student in the Department of Communication Sciences, Humanities and International Studies at the University of Urbino Carlo Bo.