Mediascapes journal 26/2025

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
*** hmmurtuza@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 Deeptrace 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 Wagensveld (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 & Wagensveld, 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 & Wagensveld, 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	DEEPFAKES 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/csis-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=202320240AB2655	Legislation	Enacted
18	USA, California	Senate Bill 942 https://leginfo.legislature.ca.gov/faces/billNavClient.xhtml?bill_id=202320240SB942	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/privacycode-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, Kasperson 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., & Wagensveld, 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