JOURNAL OF INTERNATIONAL CRISIS AND RISK COMMUNICATION RESEARCH
ISSN: 2576-0017
2024, VOL 7, No S6

Deepfakes as a threat in political crisis management and mitigation strategies

Piedad Mary Martelo Gómez[1], Raúl José Martelo Gómez[2],Annherys Isabel Paz Marcano[3]

Abstract:

Managing the political crisis caused by the use of deepfakes is becoming increasingly important in a world characterized by rapid technological advances and ethical, legal, and social challenges. This paper focuses on identifying and studying the key variables that help mitigate the risks of deepfakes in political crisis scenarios. Based on an intensive literature review and structural analysis using the MICMAC technique, ten relevant variables were identified, grouped into key, determinant, autonomous, and dependent categories. The results show the need to address these variables jointly to reduce the negative effects of deepfakes on the stability of the political and social context. Furthermore, they offer an in-depth understanding of the challenges of mitigating this threat and provide a solid basis for the design and implementation of strategies that adequately incorporate technological, legal, social, and educational perspectives.

Keywords: deepfakes, political crisis, disinformation, regulation, media literacy, cybersecurity.

Introduction

In recent decades, advances in artificial intelligence (AI) have transformed multiple areas of society, including the creation of digital content using complex tools such as deepfakes (Karnouskos, 2020). This technology allows the creation of ultra-realistic videos, images, or audios that impersonate people and raises significant concerns in areas such as disinformation, privacy, and security (Chesney & Citron, 2019). In the political context, deepfakes constitute a very specific threat as they can be used to mobilize public opinion, destabilize democracies, and exacerbate political crises (Bonfanti, 2020).

[1]Odontologist. Independent researcher. Professor of the Dentistry Program at the Universidad de Cartagena, Colombia. Email: pmartelog@hotmail.com. ORCID: https://orcid.org/0000-0002-5405-0324.
2 Specialist in Networks and Telecommunications; Master in Computer Science. Systems Engineer. Tenured Research Professor of the Systems Engineering Program at the Universidad de Cartagena. Leader of the INGESINFO Research Group. Cartagena de Indias, Colombia. E-mail: rmartelog1@unicartagena.edu.co
ORCID: https://orcid.org/0000-0002-4951-0752.
[3]Doctor in Administration. Master in Human Resources Management. Professor and researcher at the Universidad de La Guajira Colombia. E-mail: aipaz@uniguajira.edu.co. AIKA research group. Orcid: Orcid.org/0000-0001-7538-1563.

Regarding their effects on disinformation and global security, recent studies have focused on their analysis; for example, Westerlund (2019) states that deepfakes can be a tool to undermine trust in traditional media and our democratic institutions, thus generating an environment of "epistemological uncertainty". Along the same lines, Vaccari and Chadwick (2020) emphasize that although technological platforms have generated detection tools, there is no perfect system, which emphasizes the necessary regulatory and educational strategies. Floridi (2023) dedicated his research to ethical and legal aspects and emphasized that the development of emerging technologies should be applied to robust regulatory environments.

Despite efforts to explain and mitigate these risks, a lack of structural analysis of the main factors involved in their impact on political crisis management can be seen. A systemic approach has not been developed to explain which variables are most influential and which ones depend on them and should be taken into account in order to arrive at mitigation strategies. This study is relevant because it addresses an emerging phenomenon that could have devastating consequences for global governance and political stability.

The MICMAC technique allows for a systematic structural analysis, therefore, it not only helps in understanding the problem but also in defining solid policies and strategies. Thus, this work falls within the objective outlined by Godet (1994), who believes that the use of prospective approaches facilitates the possibility of addressing complex problems in situations of high uncertainty. In terms of assessment, the fundamental objective of this work is to identify and classify the variables that influence the impact of deepfakes on political crisis management, using the MICMAC technique. This allows for identifying relationships of influence as well as their dependencies, and prioritizing actions in a plan to reduce their incidence.

This work offers the following contributions: first, the novel use of MICMAC for the analysis of deepfakes in the political context; second, the systematic inventory of the technological, political, social, and regulatory factors that influence the problem; and second, the proposals based on the study to strengthen resilience against deepfakes through regulatory policies, investment in detection technologies, and educational campaigns.

Methodology

This study is an exploratory-descriptive study with a qualitative design (Sampieri, 2018), based on structural analysis using the MICMAC (Cross-impact matrix, multiplication by a classification) technique. Exploratory-descriptive studies are relevant for understanding emerging or understudied phenomena, such as deepfakes in political crisis management (Deckert & Wilson, 2023). MICMAC allows for identifying and classifying the key variables that structure a complex system and studying their influence and dependency (Godet, 1994). The qualitative perspective is ideal as it

allows for a deep understanding of the determinants that influence the problem through the perception and knowledge of experts, which is essential for addressing a highly complex and uncertain topic (Creswell & Creswell, 2017).

The participants were 12 experts selected through purposive sampling. The inclusion criteria were: proven experience in areas related to digital technology, political crisis management, digital content regulation, or computer security, as well as publications or participation in important projects on disinformation, deepfakes, or emerging technologies. The number of participants (12) is justified as being adequate for structural analysis techniques, such as MICMAC, where the interest lies in the quality and depth of contributions rather than a large sample size (Godet, 1994); furthermore, similar studies suggest that a group of 10-15 experts is sufficient to reach consensus on complex issues (Delbecq & Van de Ven, 1971).

A systematic literature review was conducted, which allowed for extracting preliminary variables related to deepfakes in political crisis management. These variables were classified into four dimensions: technological, political, social, and regulatory, as shown by previous systemic studies (Yin, 2018). Once the variables were identified (from the literature), a first list was generated, which was subsequently approved by the participants and, with the support of data processing software (Softprosp), evaluated by experts in terms of their influence and dependency using the cross-influence matrix. Finally, the variables were classified as determinants, key, autonomous, or dependent; this allowed for determining the priority variables for designing mitigation strategies for the potential risks associated with deepfakes.

Regarding the limitations of this work, the following are presented: considering a small number of experts, its results cannot be generalized to other contexts, although this limitation is excused by the nature of exploratory qualitative studies, which aim for depth rather than breadth (Creswell & Creswell, 2017). Likewise, the categorization of variables is conditioned by the experience, opinions, and knowledge of the experts, which is susceptible to generating bias. To limit this, the diverse profile of the participants was guaranteed, and a systematized methodology was chosen. On the other hand, the speed of advances in deepfakes is very rapid, which can render these findings obsolete, which can be corrected by applying longitudinal studies.

Results

The use of the MICMAC analysis methodology has provided a detailed view of the interactions of technological, political, social, and regulatory factors in the context of the application of deepfakes in political crisis management. The main results of the analysis are presented below.

During the variable identification phase, which was based on a thorough review of the scientific literature, the variables associated with the most notable aspects of deepfakes were revealed. These variables include advances in deepfake technology, detection

tools, current regulatory frameworks, impact on public opinion, collaborative work between public and private actors, and dissemination on networks. Table 1 shows the ten variables identified for the structural analysis using MICMAC, along with their respective descriptions. All variables are coded to facilitate their use in subsequent phases of the analysis.

Table 1. Key variables for MICMAC analysis

Code	Variable Name	Description
V1	Advances in deepfake technology	Continuous development of tools that allow for generating fake content with greater realism and lower cost.
V2	Deepfake detection tools	Technological systems that are designed to identify and prevent the spread of manipulated content.
V3	Regulations and legal frameworks	Legislation and regulations focused on the creation, distribution, and criminalization of malicious deepfakes.
V4	Level of media education	Citizens' ability to critically analyze information and detect misinformation.
V5	Public-private collaboration	Cooperation between governments and technology companies to address the threat of deepfakes.
V6	Impact on public opinion	Influence of deepfakes on public trust in the media and institutions.
V7	Government response capacity	Resources and political strategies to mitigate the impact of deepfakes in crisis situations.
V8	Proliferation on social media	The speed and scope with which deepfakes spread on digital platforms and social media.
V9	Ethical and social perceptions	Opinions and attitudes of society regarding the use of technologies related to deepfakes.
V10	Investment in research and development	Resources allocated to innovations to detect and mitigate the risks associated with deepfakes.

Source: Authors

The identified variables were organized according to their level of influence and dependency within the system, which examines the impact of deepfakes on political crisis management. Some variables, such as V1 (Advances in deepfake technology) and V3 (Regulations and legal frameworks), were identified as having a significant weight within the system, while others, such as V8 (Proliferation on social media), have a lower degree of influence. This is illustrated in the Cross-Impact Matrix, where a value of zero (0) indicates a non-existent relationship, one (1) indicates a low influence, two (2) represents a moderate influence, and three (3) indicates a high influence.

Furthermore, in the first row corresponding to variable V1 (Advances in deepfake technology), strong relationships with values of three (3) are highlighted, as in the cases of variables V2 (Deepfake detection tools), V6 (Impact on public opinion), and V8 (Proliferation on social media). This analysis allows for interpreting the interactions reflected in the Cross-Impact Matrix shown in Figure 1.

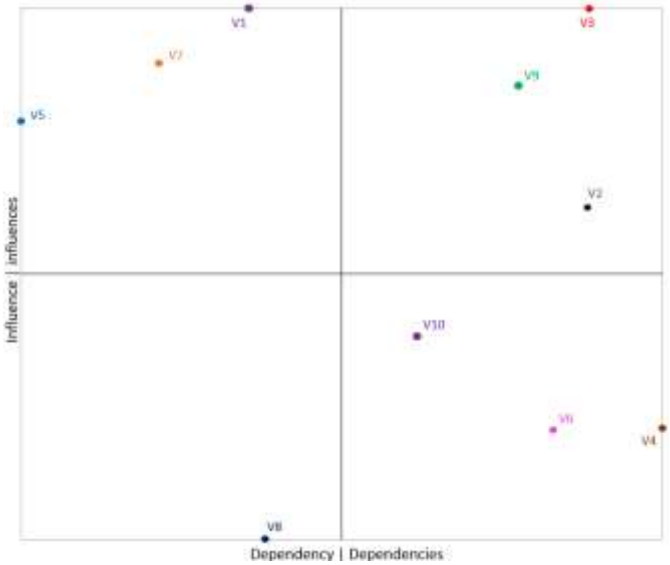
Figure 1. Cross-Impact Matrix.

Influence →	V1	V2	V3	V4	V5	V6	V7	V8	V9	V10
V1	0	3	2	2	2	3	2	3	2	3
V2	2	0	3	2	2	2	3	2	2	1
V3	3	3	0	3	3	2	2	2	2	2
V4	1	2	3	0	1	2	1	1	2	3
V5	1	3	3	2	0	2	2	2	3	2
V6	3	1	1	3	1	0	1	2	2	2
V7	2	3	2	2	3	2	0	2	3	2
V8	2	1	1	2	1	3	1	0	1	2
V9	2	3	3	3	0	3	3	2	0	2
V10	1	2	3	3	1	2	1	1	3	0

Source: Authors

The development of the cross-impact matrix facilitated the identification of interactions between the detected variables. It was highlighted that some variables, such as V2, V3, V9, V1, V5, and V7, significantly influence several aspects of the system, while others have more restricted effects. From the analysis of the cross-impact matrix, the map of direct influences presented in Figure 2 was constructed.

Figure 2. Plane of direct influences



Source: Authors

From the analysis of the Plane of direct influences, the relational variables were distributed in the quadrants that reflect their respective roles in the analyzed system.

The quadrant of the key variables (top right) corresponds to variables V2 (Deepfake detection tools), V3 (Regulations and legal frameworks), and V9 (Ethical and social perceptions) which are characterized by high levels of influence and dependency, which indicates that, on the one hand, they comprise the elementary variables to be able to analyze the interactions that occur in the system and that, on the other hand, they also represent challenges that must be taken into account.

The quadrant of determinant variables (top left) includes variables V1 (Advances in deepfake technology), V7 (Government response capacity), and V5 (Public-private collaboration). These variables have high influence and low dependency, indicating that they play a structuring role in the system and have a significant capacity to drive the system to a different state.

The quadrant of autonomous variables (lower left) presents only the variable V8 (Proliferation on social media), since this variable is of little influence and little dependency, which implies that it operates independently and has a low effect on the system in general.

The result variables quadrant (bottom right) consists of variables V4 (Level of media education), V6 (Impact on public opinion), and V10 (Investment in research and development). These variables represent the system's effects or outcomes, since they are highly dependent on the interaction between the key, determinant, and autonomous variables. The analysis of the result variables plays a fundamental role in analyzing the system's effectiveness and expected outcomes.

Table 2 provides detailed information on the results in a clear and organized manner. The table shows the variables analyzed along with their classifications, making the interpretation of the results easier and more precise.

Table 2. Classification of factors by direct influences and dependencies

Variable Type	Variable	Code
Key, strategic, or challenge variables	Deepfake detection tools	V2
	Regulations and legal frameworks	V3
	Ethical and social perceptions	V9
Determinant or influential variables	Advances in deepfake technology	V1
	Public-private collaboration	V5
	Government response capacity	V7
Autonomous or excluded variables	Proliferation on social media	V8
Dependent or result variables	Level of media education	V4
	Impact on public opinion	V6
	Investment in research and development	V10

Source: Authors

Discussions

The presentation of the results in Figure 2 and Table 2 outlines the elements considered to be of high influence and high dependency (key variables): Deepfake detection tools, Regulations and legal frameworks, and Ethical and social perceptions.

First, to address the risks associated with deepfakes, the development of detection tools is a fundamental countermeasure. Deepfake detection tools attempt to identify patterns of manipulation in audiovisual content, such as visual artifacts or motion-related inconsistencies. For Vaccari & Chadwick (2020), the capacity to respond to deepfake

threats in crisis contexts is determined by the degree of detection. Furthermore, this variable depends on investment in R&D and collaboration between governments and technology companies.

For their part, regulations and legal frameworks outline what actions can be taken against the malicious or benign use of deepfakes. As expressed by Yadlin-Segal and Oppenheim (2021), today, many countries do not have laws regulating the use of this technology, creating a notable absence in this regard. In this sense, legal frameworks can deter the malicious use of deepfakes by applying sanctions and restrictions that incur clear classification. According to Bjola (2019), the absence of regulations hinders the institutional response to the threats posed by these.

Finally, the variable has to do with ethical and social perceptions, which define how society accepts or rejects deepfakes. According to Li and Wan (2023), it is related to the degree of tolerance towards digital manipulation, since if a certain negativity is felt, it incites to regulate laws or detection. However, this variable depends on cultural factors, social values, and the media narratives that are in force.

On the other hand, variables classified as having high influence and low dependency (determinants) were: Advances in deepfake technology, Public-private collaboration, and Government response capacity.

Transformations in deepfake technology have drastically changed the way political crises operate. This is vital because it produces high-quality audiovisual content with minimal technical effort. According to Arora and Soni (2021), generative adversarial networks continue to develop, making deepfakes much more realistic and harder to detect. In this sense, advances allow malicious individuals to manipulate information for the purposes of disinformation or discrediting and/or influencing public opinion, which causes, among other things, very negative effects on political and social stability. However, this progress depends on the pace of progress in artificial intelligence, the funding available, and the availability of advanced tools.

Regarding the collaborative relationship between the private and public sectors, it is also essential to implement appropriate mitigation strategies. The latter are carried out by technology companies, such as Meta or Google, with a key role in the development of detection tools and in the proactive detection of deepfakes. According to Burton (2021), synchronization between both sectors is essential to address the technical and social challenges posed by deepfakes. Even so, such action will depend on cooperation agreements, funding, and common goals.

On the other hand, the government's response capacity will determine how effective measures to limit the negative effects of deepfakes can be, since this action includes, among other strategies, communication, official rebuttal, and legal actions to curb the effects caused by adversaries. In this sense, a proactive government response is capable of mitigating the crisis before it spreads. According to Martens et al. (2018), such a

response requires prior preparation and coordination between institutions, but depends on financial resources, experience, and international relations.

Continuing with the results, only one variable was classified as autonomous, in this case: Proliferation on social media, which makes sense since social platforms are the main means of disseminating deepfakes. The ease with which they spread content makes controlling and mitigating consequences complicated, so this variable is considered to have low dependency and low influence (autonomous). Furthermore, the virality of deepfakes on social media depends mainly on recommendation and moderation algorithms, which operate autonomously with respect to other variables analyzed.

Their direct influence on political crisis management is limited, although their impact on content amplification is relevant at an operational level. As Krishna (2020) states, virality on social media can exponentially increase the effects of deepfakes, mainly in political scenarios, and depends on recommendation algorithms, content moderation, and platform policies.

Finally, the variables that were classified as having low influence and high dependency (dependent or result) were: Level of media education, Impact on public opinion, and Investment in research and development.

Media education is an important factor in empowering audiences to recognize misinformation, including deepfakes. According to Shearer & Mitchell (2021), an educated population is less sensitive to manipulation. As a result, good media education could reduce the effects of deepfakes by enhancing citizens' critical thinking skills, given their high dependency on educational policies, awareness campaigns, and access to quality educational materials.

Regarding the impact of deepfakes on public opinion, it is among the most immediate and evident factors because this type of content can affect voting, mobilize protests, and discredit public figures. Van der Linden et al. (2020) indicate that deepfakes can accentuate false narratives that polarize society despite the fact that this variable is highly conditioned by the ethical and social representation of deepfakes and the level of trust of the information sources.

Finally, R&D investment is a very important enabler for reducing the effects of deepfakes, as it ranges from developing detection technologies to improving media literacy. For Arora and Soni (2021), this investment allows for anticipating emerging threats, but it is highly dependent on public and private investment, as well as academic cooperation. It is a fundamental enabler of solutions, but its direct short-term impact is limited.

Mitigation strategies

Based on the classification and analysis of the variables presented, the following mitigation strategies are proposed to address the risks associated with deepfakes in political crisis management. It should be noted that these strategies are designed considering the interrelationships and hierarchies identified in the analysis.

Strategies for key variables (high influence and high dependency)

a) It is proposed to strengthen deepfake detection tools with the goal of developing and implementing advanced technologies that enable early identification of manipulated content. To achieve this, investment in research and development through public and private grants should be encouraged, in addition to creating collaborative platforms between governments, technology companies, and universities for the development of detection algorithms. Finally, international standards should be established to validate the effectiveness of detection tools.
b) Development of regulations and legal frameworks with the aim of implementing regulation standards that discourage the malicious use of deepfakes and protect social and political stability. To achieve this, national and international legislative initiatives must be promoted to criminalize the malicious use of deepfakes; establish protocols for risk assessment and applicable sanctions; and encourage global consensus through multilateral forums (UN, G20) to harmonize regulatory policies.
c) Raise awareness about ethical and social perceptions to generate awareness about the risks and ethical implications of deepfakes in society by designing educational campaigns that inform the population about the potential impact of deepfakes, disseminating real cases of malicious use to illustrate the ethical and legal consequences, and encouraging public debate about the acceptable limits of digital manipulation.

Strategies for determinant variables (high influence and low dependency)

a) Monitor and regulate advances in deepfake technology in order to oversee the development of technologies to minimize risks and promote their ethical use by requiring registrations and audits for companies that develop content generation algorithms, in addition to financing ethical AI projects that prioritize transparency and responsible use and limiting public access to advanced tools without restrictions or adequate controls.
b) Strengthen public-private collaboration to leverage resources and knowledge from these sectors to implement mitigation strategies. To achieve this, specific consortia for the detection and regulation of deepfakes must be created, integrating key actors such as governments, social media platforms, and technology companies; promoting information-sharing agreements on new threats and technological advances; and co-developing self-regulatory policies for digital platforms.

c) Preparation for government response capacity to respond quickly and effectively to crises arising from deepfakes by designing contingency plans for crisis communication, including rapid and reliable denials; training workers in the use of tools for detecting and analyzing manipulated content; and establishing specialized cybersecurity and disinformation units within government institutions.

Strategies for dependent variables (low influence and high dependency)

a) Reduce the population's vulnerability to digital manipulation through education by incorporating media literacy programs into formal education curricula; create accessible digital resources, such as guides and online courses, to identify deepfakes, and promote mass awareness campaigns in traditional and digital media.
b) Minimize the damage caused by deepfakes to public perception and social cohesion by reinforcing official narratives through trusted and verified channels; encourage digital platforms to label content suspected of manipulation and promote transparency in deepfake-related investigations.
c) Ensure the continued development of technological and educational solutions against deepfakes by allocating government and private funding to projects focused on deepfake detection and prevention, encouraging international collaboration to share research advances, and creating innovation hubs focused on cybersecurity and digital content analysis.

These strategies propose a comprehensive and structured approach to mitigate the risks associated with deepfakes, addressing technological, social, legal, and educational aspects. Their coordinated implementation can significantly reduce the impact of this threat on political crisis management.

Conclusions

This research has analyzed the variables that influence the mitigation of risks associated with deepfakes in contexts of political crisis, providing an overview of the challenges and opportunities in this area. The conclusions from the literature review, as well as the use of the MICMAC technique, allowed for identifying 10 factors that are relevant for managing the threat posed by the use of deepfakes.

The research results demonstrate the importance of comprehensively and coordinatedly including the identified variables in order to guarantee a response to the threat of deepfakes. In particular, several needs are highlighted, including strengthening deepfake detection tools, promoting the creation and updating of specific laws for the use of deepfake technology, building and enhancing critical ethical and social perceptions regarding the use of deepfakes, and promoting public-private collaboration to create effective mitigation strategies.

Furthermore, determinant variables were identified, such as advances in deepfake technology, government response capacity, and the level of investment in research and development. While these variables are crucial to mitigating the threat, they also depend on other factors, such as international collaboration and the amount of resources allocated. It is worth noting that media literacy emerged as a strategic factor that underscores the need to empower society to recognize, respond, and act critically and effectively against manipulated content.

Although it is sought to minimize bias through a systematic approach, interpretations may vary among experts. Likewise, the interaction between the identified variables, such as ethical perceptions, legal regulations, and technological advances, is inherently complex. This makes it difficult to fully address all interdependencies and dynamics. Furthermore, the proposed mitigation strategies are based on theoretical analyses and have not been tested in real-life scenarios, leaving open the possibility that their implementation may face unanticipated practical or contextual barriers. However, these limitations offer opportunities for future research that more specifically addresses regional contexts, adapts strategies to changing scenarios, and validates proposed solutions through applied case studies.

References

Arora, T., & Soni, R. (2021). A review of techniques to detect the GAN-generated fake images. . *Generative Adversarial Networks for Image-to-Image Translation*, 125-159.

Bjola, C. (2019). The 'dark side'of digital diplomacy: countering disinformation and propaganda. . *Policy Briefing, 5*.

Bonfanti, M. (2020). The weaponisation of synthetic media: what threat does this pose to national security?. . *Ciber Elcano, 57*.

Burton, S. (2021). Technological Digital Disruption in the Age of Artificial Intelligence: A New Paradigm for Leadership. In Cultivating Entrepreneurial Changemakers Through Digital Media Education. *IGI Global*, 1-35.

Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war: The coming age of post-truth geopolitics. *Foreign Affairs, 98(1)*, 147-155.

Creswell, J., & Creswell, J. (2017). *Research design: Qualitative, quantitative, and mixed methods approaches.* Thousand Oaks, California: Sage publications.

Deckert, J., & Wilson, M. (2023). Descriptive Research Methods. . *Research methods in the dance sciences/edited by Tom Welsh, Jatin, 153*.

Delbecq, A., & Van de Ven, A. (1971). A group process model for problem identification and program planning. . *The journal of applied behavioral science, 7(4)*, 466-492.

Floridi, L. (2023). The ethics of artificial intelligence: Principles, challenges, and opportunities. *Nature Machine Intelligence, 2(6)*, 261-263.

Godet, M. (1994). From anticipation to action: a handbook of strategic prospective. *UNESCO*.

Karnouskos, S. (2020). Artificial intelligence in digital media: The era of deepfakes. . *IEEE Transactions on Technology and Society, 1(3)*, 138-147.

Krishna, D. (2020). Deepfakes, online platforms, and a novel proposal for transparency, collaboration, and education. . *Rich. JL & Tech., 27*, 1.

Li, M., & Wan, Y. (2023). Norms or fun? The influence of ethical concerns and perceived enjoyment on the regulation of deepfake information. *Internet Research , 33 (5)*, 1750-1773.

Martens, B., Aguiar, L., Gomez-Herrera, E., & Mueller-Langer, F. (2018). The digital transformation of news media and the rise of disinformation and fake news-An economic perspective. . *JRC Digit. Econ. Work. Pap, 2*, 57.

Sampieri, H. (2018). *Metodología de la investigación: las rutas cuantitativa, cualitativa y mixta.* México.: McGraw Hill.

Shearer, E., & Mitchell, A. (2021). News use across social media platforms in 2020. *Pew Research Center*.

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)*, 1-13.

van Der Linden, S., Roozenbeek, J., & Compton, J. (2020). Inoculating against fake news about COVID-19. *Frontiers in psychology, 11*, 566790.

Westerlund, M. (2019). The emergence of deepfake technology: A review. . *Technology innovation management review, 9(11)*, 39-52.

Yadlin-Segal, A., & Oppenheim, Y. (2021). Whose dystopia is it anyway? Deepfakes and social media regulation. *Convergence, 27(1), 1*, 36-5.

Yin, R. (2018). Case study research and applications. *Sage Publication, Inc*, 319.