©NIJCRHSS
Publications

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789

NEWPORT INTERNATIONAL JOURNAL OF CURRENT RESEARCH IN HUMANITIES AND SOCIAL SCIENCES (NIJCRHSS)

Volume 6 Issue 1 Page 59-64, 2026

https://doi.org/10.59298/NIJCRHSS/2025/61.5964

Deepfakes and Democratic Trust: Social Impacts and Policy Responses

Kakembo Aisha Annet

Faculty of Education, Kampala International University, Uganda

ABSTRACT

Deepfake technology, powered by advances in artificial intelligence, has emerged as a transformative yet disruptive force in contemporary information ecosystems. While it offers innovative applications in entertainment, education, and digital communication, its capacity to fabricate highly realistic audio-visual content raises serious concerns for democratic governance. This paper examines the social impacts of deepfakes on democratic trust, focusing on their role in amplifying misinformation, undermining informational integrity, and eroding confidence in public institutions. It analyzes the technical foundations of deepfakes, the challenges associated with their detection, and their implications for electoral processes, media credibility, and civic engagement. The study further evaluates existing policy responses across jurisdictions, including regulatory measures, platform governance strategies, and international initiatives, highlighting both their potential and limitations. Emphasis is placed on the importance of media literacy, civil society engagement, and multi-stakeholder collaboration in mitigating the risks posed by synthetic media. The paper argues that safeguarding democratic trust requires a balanced approach that combines technological innovation, legal regulation, and ethical accountability, while avoiding excessive restrictions that could undermine freedom of expression. Ultimately, it concludes that resilient democratic systems must adapt to the realities of synthetic media by strengthening transparency, accountability, and public awareness.

Keywords: Deepfakes; Democratic trust; Misinformation; Media literacy and Platform governance.

INTRODUCTION

The term "deepfake" refers to computer-generated audio-visual media in which certain characteristics of a person or object are altered, typically performed by artificial intelligence algorithms [1]. The word originated in 2017 when a user on the social media site Reddit began uploading and sharing pornographic videos that replaced the face of the performer with that of a celebrity [2]. The technology has advanced rapidly since its inception and has since been applied to other formats. It has enabled new forms of creative expression in the arts, entertainment, and education, while also raising profound ethical concerns [3]. Deepfake capabilities have spread beyond mere image manipulation to include audio-only alterations of speech that allow one person to impersonate another [4]. Concerns have arisen about the use of deepfake technology in political and electoral contexts, particularly regarding democracy, media, hate speech, and disinformation [1]. Deepfakes also provide opportunities for creative expression and education by allowing any text, comic strip, or short story to be animated. Below is a video demonstrating the operation of a Text-To-Video Generative AI system [2].

The Phenomenon of Deepfakes

Deepfakes are increasingly accessible, and their use appears to be on the rise. Consequently, they have become a potential tool for disinformation, including in political contexts, and may threaten democratic processes by spreading misleading or false narratives [1]. Deepfakes exacerbate existing challenges to informed decision-making and democratic trust by reducing the shared factual basis needed for debate, deliberation, and engagement with democratic processes [2]. Notably, while many fear that deepfakes might be leveraged for disinformation, especially during electoral contests, the evidence for such use to date is limited, with only a small number of

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789

verified cases [3]. Yet, similar concerns are warranted for other types of misinformation, highlighting the importance of fostering broader approaches to information quality, digital literacy, and public trust in institutions [3].

Technical Foundations and Capabilities

Deepfakes result from machine-learning methods that recognize and reproduce audio-visual patterns. Generative adversarial networks (GANs) create deepfake content through an adversarial iterative process involving two competing neural networks, a generator and a discriminator [3]. The generator develops examples that approximate an authentic training dataset, while the discriminator assesses whether samples originate from the dataset or the generator [3]. The generator evolves until it produces outputs indistinguishable from genuine content. Other machine-learning approaches for deepfakes employ face-swapping or style-transfer techniques. Trainees use readily available datasets to replicate styles, enabling scholars to produce usable educational restrictions and conditionally chosen animations before GAN training [2]. The technology produces moving images featuring individuals performing scripts written by others, allowing videos of public figures to artificially generate fabricated messages [1]. The potential applications of deepfakes include political satire, parody, or educational content aimed at preventing manipulation. Despite this, deepfakes convey messages that viewers misconstrue as authentic [1]. For example, in face-to-face translations in the Indian context, unaccompanied visuals may mislead audiences into incorrectly assuming that the speaker communicates in a specific language. Instances of deepfakes used for malicious purposes also exist, notably a fabricated video of Ukrainian President Zelenskyy. By disseminating misleading imagery, deepfakes are used strategically to manipulate information and influence public perception [2]. Deepfake technology poses a challenge to the regulated marketplace of ideas that democracy requires, as the scheme simultaneously fails to inform or enlighten an audience. The efficacy and reliability of messages transmitted through such channels become compromised, and citizens cannot discern between veracity and fabrication. Through its potential to degrade the milieu of common facts or shared truths needed for deliberative discourse, deepfake technology holds the capacity to disrupt institutions critical for democratic governance by eroding trust. Such institutions encompass the media, law enforcement, and the judiciary [4]. Deepfakes can divert attention away from essential issues as agents introduce decoy topics or controversies to engage the public. Consequently, the susceptibility of democratic practices, political behaviour, civic participation, and public policies to deepfake technology has emerged as a point of concern for scrutiny [5].

Detection Challenges and Advancements

First-order detection methods rely on signal processing techniques to discriminate between the expected characteristics of real and fake videos [4]. Convolutional neural networks exploit differences in representation between real and altered videos. Second-order methods investigate temporal dynamics by analysing variations in frame-to-frame information, e.g. head pose or blink events, typically absent in fakes. Both of these are simple, efficient, and top-performing detection types, though datasets quickly become outdated as deepfake generation progresses [3]. Adversarial networks generate indistinguishable samples; intelligent and unknown manipulations introduce simultaneous challenges. Generative adversarial networks (GANs) and deep-learning architectures primarily convolutional networks and recurrent neural networks remain valuable, yet barely increase detection capability [4]. Robustness varies: compression and noise are less convoluted, but challenging; GAN-generated and deep video remain the most demanding; these conditions hinder effective detection [5]. Massive public datasets containing diverse manipulation types enable transfer learning, boosting performance, and sparking further research. Performance also largely depends on manipulated videos; insufficient educational or entertainment materials hinder transfer success to untrained cases [5].

Implications for Democratic Trust

Concerns regarding the effects of deepfake technology on trust in institutions and democracy are often expressed within the frameworks of misinformation and are thus tied to public disinformation campaigns, especially in the context of elections [2]. Political disinformation, however, has been an issue long before deepfakes appeared. Similarly, public skepticism towards democracy and its institutions is not a recent phenomenon, and the current levels of distrust in both government and the media are unprecedented. In such a climate, deepfakes are unlikely to be a primary factor determining trust in political institutions [6]. Moreover, while deepfake media can be used for public disinformation, as noted, the number of verified instances where they were used with that goal in mind is low, particularly in election-related contexts [2]. Instead, a more general perspective on trust is warranted. As democratic deliberation aims to arrive at collectively binding norms by discussing issues towards the goal of reaching a common understanding, this process presupposes a certain minimum degree of mutual trust between participants. Hence, any significant weakening of this informational basis is predicted to hinder citizens' willingness to engage in political discussions [5]. The systematic denial of factual statements and the broad dissemination of blatant falsehoods in public discourse do not radically change the existing media landscape, as

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789

they amplify existing trends instead [6]. Technologies per se do not generate specific political regimes; their development instead occurs during particular modes of socio-political organization [6]. The notion of further undermining existing trust in democratic institutions merely articulates an opportunity structure rather than determining political character, as a lack of trust already exists and deepfakes cannot be understood as a new condition[1].

Informational Integrity and Trust in Institutions

Deepfake disinformation threatens democratic trust by disrupting factual deliberation and fostering a "post-fact" society [1]. This erosion of trust extends to individuals, media outlets, and institutions, jeopardizing vital processes such as elections, which rely on citizen participation and dialogue [2]. Although skepticism toward political entities is not new, deepfakes amplify this distrust by degrading the very ecosystems that underpin reliable information. Actual instances of deepfake exploitation in electoral contexts remain rare; verified examples include doctored videos involving politicians, but the potential threat to democratic systems may be overstated [2].

Electoral Processes and Public Opinion

The rise of deepfakes raises concerns about their potential use during electoral processes to influence voter behaviour or undermine trust in electoral integrity [2]. Political deepfakes could target political candidates by making false statements during campaigns, casting doubt on candidates' credibility or integrity. Moreover, deepfakes intensify existing disinformation tactics and politics of fear. Concerns that deepfakes might fuel the spread of false narratives and provoke violent acts have escalated globally, with fears of political violence connected to candidate videos featuring poorly imitated bodies and voices [3]. The potential for deepfake audiovisual media to drown out citizen voices, convey false or misleading instructions on voting procedures, or debase candidates poses further threats to inclusion. Other forms of deepfake disinformation harm democracy by fostering distrust in information related to electoral procedures [3]. Detached from reality and lacking pivotal information concerning the location or timing of votes, such audiovisual content amplifies uncertainty and confusion and impedes informed participation. Absent collective deliberation on issues, candidates, processes, outcomes, and the political community, democratic government cannot take root. Deepfake disinformation that escalates distrust consequently deepens candidate-based fixation, reduces issue consideration, and clusters citizens around like-minded social media connections, which aggravates polarization [2]. Political deepfakes heighten these risks by drifting toward physically specific, highly dynamic simulated media engaging the likeness of political actors. Social media dissemination of digitally simulated audiovisual artifacts featuring the bodies or voices of prominent political figures exacerbates the threat [6].

Media Literacy and Civic Engagement

The information disorder exacerbated by deepfakes extends beyond the obvious threat to public trust in information and institutions [3]. Public trust in one another, often unfavorably labeled "social capital," is a vital underpinning of all democratic governments; when that trust decays, democracies and their vital core functions, electoral participation, freedom of assembly, the respect of laws, and the legitimacy of non-violent political action are jeopardized [5]. Deepfakes are similarly detrimental to societal cohesion and stability within democracies by contributing to citizens' parochialism. Citizens are increasingly disconnected from political issues and institutions. Public engagement with policies, politicians, and local issues has plummeted [6]. Deepfakes' impacts on people's perceived credibility and characterization further contribute to this election and institutional estrangement and related democratic disenfranchisement, coupled with increasing barriers to information dissemination [3]. Deepfakes distort videos of high-profile politicians, celebrities, and government officials, warranting active examination. Deepfake videos in foreign languages can disrupt national politics, particularly in linguistically diverse communities [2]. Mass media consumption transitioned to social media and unmonitored platforms; deepfake dispersion rates surged. Deepfake-induced trust decay and the normalization of distrust threaten equal access to public information, fostering ignorance [5]. Deepfakes undermine citizens' investment in the democratic process and communal welfare, propelling them towards a solitary, survival-centric approach [6]. An all-encompassing collective awareness and concern for societal welfare is crucial for a robust regime, enabling the emergence of collective redistribution and pro-civil behavior. Data underscores a growing citizenry resistance to issues of collective significance beyond survival, essential to functioning, serving democracy [2].

Policy Landscape and Regulatory Approaches

In addressing the multi-faceted dangers posed by deepfakes to trust in democratic institutions and processes, both established democracies and authoritarian regimes have responded promptly with a variety of policies[4]. These actions span a spectrum from voluntary moderation by technology platforms to internal deliberations on how best to act. In the United States, the response formally began with the issuance of an influential executive order in 2020 that accompanied a broader initiative within the Federal Bureau of Investigation to illuminate the potential risks of

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789

deepfakes at that time [2]. Elsewhere, generally aligned efforts have continued, emphasizing the space for public discourse through an exploration of secure and accurate information channels [4]. Tentative preparation appears to tie in with the absence of deepfake-related incidents that could catalyze outside action or cause publicly visible harm. Outside this wave, rapid, voluntary, and ongoing action dedicated to policy recommendations began in the context of the deepfake risks to democracy and democratic practices in late 2020[6]. Comprehensive proposals remained available without formal follow-up, and the rhythm of support appears to have gradually fallen off. Sporadic and focused, a still-visible cadre of stakeholders prepared a report with the objective of assisting bordering policy-space. Addressing emerging threats, preparation and publication of trans-national recommendations occurred in early 2021 [1].

United States Policy Measures

In June 2022, California AG Rob Bonta proposed an amendment to the state's 2018 anti-deepfake legislation to prohibit the use of synthetic media in "political campaign" content without disclosing the use of that technology, as well as any materially false information about a candidate [2]. California SB 1418 was signed on November 30, 2022. In that same month, Congress debated deepfake regulations, but involvement by the Federal Trade Commission (FTC) was suspended [3]. An analysis of these efforts observed gaps in terminology or scope of content that limit existing regulations, although politicians from Congress to state attorneys general frequently raise the risk of deepfakes to political campaigns and a democratic society [4]. The introduction of the deepfake video of President Zelenskyy during the Russian invasion is one case cited where deepfakes were widely seen as posing a threat or harmful to democracy [1].

International and Comparative Perspectives

Synthetic media derived from artificial intelligence (AI) has rapidly evolved and proliferated, offering powerful, yet dangerous, capabilities to imitate the human voice and visualize lifelike personas through video. International organizations, non-governmental organizations, and regulatory authorities endorse policy interventions addressing agenda-setting within electoral contexts and governmental trust matters [4]. Mitigating trust erosion remains a major goal of stakeholders, but highly divergent governance, legislative, and regulatory fundamentals across political jurisdictions constrain the capacity to harmonize proposals [4]. The European Union (EU) considers adopting legislation addressing synthetic media, focusing on deepfakes employed within a political mobilization context and during election times [1]. Deepfakes also present threats linked to different motives, including artistic or educational purposes. Societal responses to risks and opportunities drive the governance landscape. Contributing to the societal risk of information disorder, leading to asymmetric effects on public trust in political institutions and actors, are commercial entities[4]. Commercial organizations widely utilize shared social media to disseminate speech or images, which can result in reputational or branding challenges for others. Political communication complements the aforementioned mechanisms that extend beyond voter-related political platforms. In addition to the previously noted case of President Vladimir Zelenskyy, another example arises when public personalities or political candidates deploy deepfake software to replicate the voice of another public figure [2].

Platform Governance and Responsibility

The emergence of deepfake technology raises pressing questions about platform governance and accountability. Content that is likely to be illegal or violate community guidelines must be removed or demoted on platforms; the debates surrounding the moderation of deepfakes have not reached a conclusive framework, yet certain principles can be established [4]. Current deepfake models can make synthetically altered video and audio combinations, potentially permitting the harmful dissemination of misinformation, disinformation, and hate speech. Media outputs such as selected images, text, and backgrounds can easily be manipulated to produce deepfake videos that transform the style of public figures [5]. High-definition videos with specific pronunciations can be altered to generate a synthetic version of both an individual's appearance and real-time speech production 2. Key media platforms face a responsibility to permit the use of deepfakes while simultaneously reducing the negative social impact they can produce, particularly those with the potential to undermine democratic processes or individual rights [3]. Contemporary platform governance strategies reflect different degrees of policy intervention. One intervention would prohibit the development and/or dissemination of a substantial portion of deepfakes through the high-level tightening of safety and trust-oriented regulations, requiring platforms to algorithmically deprioritize or prohibit certain deepfakes [3]. On the one hand, content that amounts to democratic disinformation, gendered or polarized narratives, or political pornography undermines the possibility for democratic debates, respect, and trust between individuals [1]. Such democratic threats have also been recognized as distinct categories of misuse capable of triggering preventive interventions. On the other hand, transparency-related interventions reflect a medium degree of policy intervention, providing the public with contextual information to better understand the nature of the contents (repetitive cases and failed risks) [4]. A hybrid

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789
approach can avoid excessive restrictions while alleviating unrestricted harm. Transparency may contribute positively to individual selection of deepfakes and stimulate the development of tools to promote overarching transparency, fostering collective meaning-making and deliberation surrounding the technology [5].

Ethical Considerations and Human Rights

Deepfake technology presents serious ethical dilemmas regarding human rights and social injustices. Most notably, these challenges arise from the persistent secondary use of fake audio-visuals fitted onto the characters of others without valid consent for creation, sharing, or viewing [3]. Such usage violates the dignity of others and the experiences and rights of the principal subjects. These are particularly influential when the result is a deepfake version of a well-known speaker or leader (e.g., the example of a deepfake video showing the president of Ukraine appearing to promote false information about the invasion of his country) [5]. Another recurrent concern is the creation of sexual deepfake material, which frequently features women worldwide. Though sexual deepfake material can damage the reputation of any individual, the consequences are especially harmful in certain social and cultural contexts [4]. Therefore, it is critical to address the ethical challenges of deepfakes and assess the interplay between the employment of these technologies and the upholding of human rights [2, 1].

Civil Society and Industry Roles in Mitigation

The stakes of deepfakes are high, and the means of intervention numerous. If the exigencies of democratic societies internalised a common measure, they could motivate a fuller involvement of civil society and the industry beyond a strict regulatory framework [3]. Authoritarian states, well equipped to monitor and influence public perceptions, have a pretext to suppress the creative and critically engaged public participation essential to democracy, one increasingly carried out through digital technology [5]. The cross-cutting potential of deepfakes can undermine trust in democratic institutions while retaining the capability to foster social cohesion, mutual understanding, and trust through interventions in young people's education and communication and through examples on how to use the technology responsibly. Ultimately, if greater capacity for civil society intervention and industry cooperation were to contribute to the sustainability of democratic foundations and individual freedoms, regulation hostile to democratic foundations would be rendered unnecessary [6].

Evaluation of Policy Efficacy and Risks of Overreach

Governments and platforms have begun to address the implications of deepfakes via policy initiatives [2]. In light of the rapid technological evolution and adaptive quality associated with deepfake generation, these measures raise questions about limitations on distribution, applicable datasets, rights of platforms to limit misuse of content, safe deployment of similar tools, the influence of competitive technological changes, and new forms of information manipulation that may emerge[5]. Existing policies on audio-visual manipulation, such as a ban on misleading political advertisements in the United States, restrict tactics used in deepfake content but do not cover the technology itself. Nonetheless, such measures highlight broader themes concerning the capacitating and constraining roles of audio-visual technology [2].

Recommendations for Safeguarding Democratic Trust

Trust enables societies to carry out their business. Trust is at the same time at risk from various sources, but also needed more than ever to overcome crises. In the face of these challenges, mechanisms that secure accountability and expose threats to trust are required [3]. This hold both for the overwhelming concerns concerning deepfakes and for tackling the systematic and structural elements limiting the capacity of individuals to trust. Measures to combat deepfakes must target the phenomenon itself, removing opportunities for manipulating deepfakes to politicise the question of deepfakes and avoid other formats of false or deeply false content [4]. In a busy information ecosystem, misinformation rising from any source, including bot-generated material, is a threat to accountability. Regulatory intervention is bound to be ineffective in the extreme[5]. It would therefore be misguided to politicise the invocation of deepfakes by framing them as a particular threat. Such a measure remains nonetheless of utmost importance; the tools provided by the deep economy remain effective and even reinforce the expectation that false and misleading information can be combated effectively through counter-influence combined in time and space[6].

CONCLUSION

Deepfake technology represents a double-edged development in the digital age, offering creative and communicative possibilities while posing significant risks to democratic trust and institutional legitimacy. This paper has shown that, although the current empirical evidence of widespread political misuse remains limited, the potential for deepfakes to disrupt democratic processes is substantial, particularly through their capacity to amplify misinformation, distort public discourse, and weaken the shared factual foundations necessary for collective decision-making. The erosion of trust in democratic institutions, media, and electoral systems is not solely attributable to deepfakes; rather, these technologies intensify pre-existing vulnerabilities within information ecosystems. By fostering uncertainty about the authenticity of information, deepfakes contribute to a broader

Open Access
Print ISSN: 2992-6106
Online ISSN: 2992-5789

"crisis of truth," where citizens may become increasingly skeptical of both true and false content. This dynamic risks disengagement from civic processes, heightened polarization, and reduced confidence in governance structures. Policy responses have begun to emerge at national and international levels, alongside platform-based governance mechanisms. However, these responses remain fragmented, often limited in scope, and challenged by the rapid evolution of deepfake technologies. Overly restrictive regulation risks undermining fundamental rights such as freedom of expression, while insufficient intervention may allow harmful uses to proliferate. A balanced, adaptive approach is therefore essential. Strengthening democratic resilience requires a multi-dimensional strategy. This includes investing in advanced detection technologies, enhancing media and digital literacy, promoting transparency through labeling and disclosure requirements, and establishing clear accountability frameworks for both state and non-state actors. Equally important is the role of civil society and the private sector in fostering ethical standards and responsible innovation. In conclusion, safeguarding democratic trust in the era of deepfakes demands not only technical and regulatory solutions but also a renewed commitment to the core principles of transparency, accountability, and informed public participation. By addressing both the technological and societal dimensions of the challenge, policymakers and stakeholders can mitigate risks while harnessing the positive potential of synthetic media in a manner consistent with democratic values.

REFERENCES

1. Karunian AY. *The imitation game: examining regulatory challenges of political deepfakes in the European Union* [dissertation]. 2024. Available from: OSF.
2. Pawelec M. Deepfakes and democracy (theory): how synthetic audio-visual media for disinformation and hate speech threaten core democratic functions. *Digital Society*. 2022;1(2):19. doi:10.1007/s44206-022-00010-6.
3. Kietzmann J, Lee LW, McCarthy IP, Kietzmann TC. Deepfakes: trick or treat? *Bus Horiz*. 2020;63(2):135-146. doi:10.1016/j.bushor.2019.10.004.
4. Qureshi SM, Saeed A, Almotiri SH, Ahmad F, Al Ghamdi MA. Deepfake forensics: a survey of digital forensic methods for multimodal deepfake identification on social media. *PeerJ Comput Sci*. 2024;10:e2037. doi:10.7717/peerj-cs.2037.
5. Nguyen TT, Nguyen QVH, Nguyen DT, Nguyen DT, Huynh-The T, Nahavandi S, et al. Deep learning for deepfakes creation and detection: a survey. *arXiv* [Preprint]. 2019:arXiv:1909.11573. Available from: arXiv.
6. Etienne H. The future of online trust (and why deepfake is advancing it). *AI Soc*. 2021. doi:10.1007/s00146-021-01337-0.

CITE AS: Kakembo Aisha Annet (2026). Deepfakes and Democratic Trust: Social Impacts and Policy Responses. NEWPORT INTERNATIONAL JOURNAL OF CURRENT RESEARCH IN HUMANITIES AND SOCIAL SCIENCES, 6(1):59-64. https://doi.org/10.59298/NIJCRHSS/2025/61.5964