World Journal of Advanced Research and Reviews
eISSN: 2581-9615 CODEN (USA): WJARAI
Cross Ref DOI: 10.30574/wjarr
Journal homepage: https://wjarr.com/

(Review Article)

Deepfakes, Cybersecurity, and the Fragile Chain of Trust in Public Institutions

Tomilola Ayeni *

University of Northampton, England, United Kingdom.

World Journal of Advanced Research and Reviews, 2026, 29(02), 645-648

Publication history: Received on 03 January 2026; revised on 10 February 2026; accepted on 12 February 2026

Article DOI: https://doi.org/10.30574/wjarr.2026.29.2.0358

Abstract

Deepfakes have been widely discussed as a media ethics issue, a free speech dilemma, or a potential threat to democratic institutions. This paper argues that deepfakes should also be understood as a cybersecurity problem, because they exploit the same trust relationships that underpin secure systems and public governance. Deepfakes are not merely "fake videos", they are tools for identity manipulation that can bypass technical defenses and disrupt decision-making, public communication, and institutional legitimacy. The paper examines how deepfakes intersect with existing legal frameworks, including data protection, defamation, broadcast regulation, and cyber incident response, and proposes that public institutions should treat synthetic media as an emerging risk within cybersecurity governance.

Keywords: Deepfakes; Cybersecurity Governance; Institutional Trust; Synthetic Media

1. Introduction

Deepfakes are often discussed as a media ethics problem or a free speech headache. Increasingly, though, they are a cybersecurity issue. Not in the narrow sense of malware or network intrusion, but in a broader, more dangerous sense. Deepfakes attack trust itself. They exploit the same assumptions on which cybersecurity frameworks depend: that a familiar face signals authenticity, that a known voice signals authority, and that official-looking content is likely legitimate [2].

* Corresponding author: Tomilola Ayeni

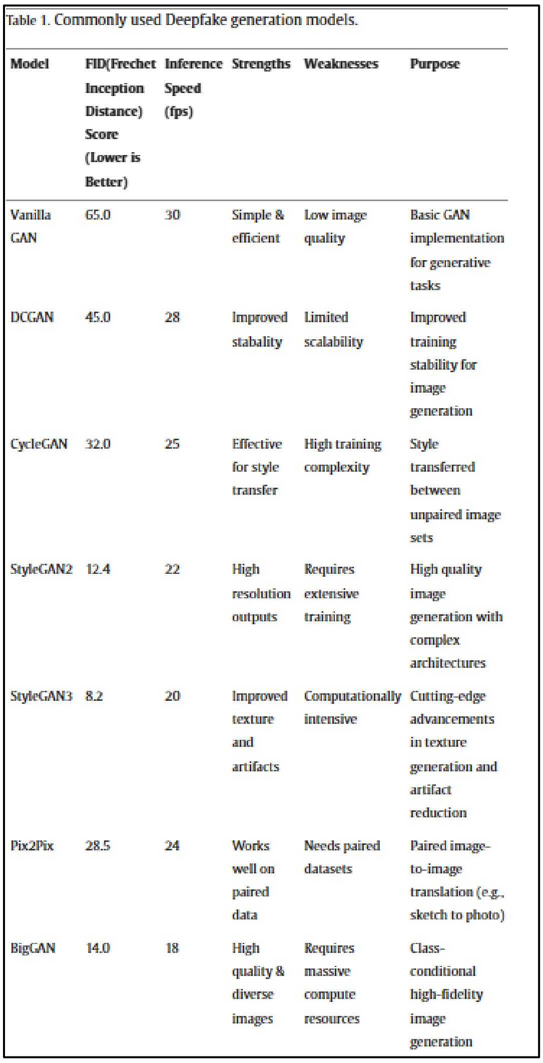

Table 1. Commonly used Deepfake generation models.

Model	FID(Frechet Inception Distance) Score (Lower is Better)	Inference Speed (fps)	Strengths	Weaknesses	Purpose
Vanilla GAN	65.0	30	Simple & efficient	Low image quality	Basic GAN implementation for generative tasks
DCGAN	45.0	28	Improved stabality	Limited scalability	Improved training stability for image generation
CycleGAN	32.0	25	Effective for style transfer	High training complexity	Style transferred between unpaired image sets
StyleGAN2	12.4	22	High resolution outputs	Requires extensive training	High quality image generation with complex architectures
StyleGAN3	8.2	20	Improved texture and artifacts	Computationally intensive	Cutting-edge advancements in texture generation and artifact reduction
Pix2Pix	28.5	24	Works well on paired data	Needs paired datasets	Paired image-to-image translation (e.g., sketch to photo)
BigGAN	14.0	18	High quality & diverse images	Requires massive compute resources	Class-conditional high-fidelity image generation

Figure 1 Commonly used Deepfake generation models (Singh and Dhumane, 2025)

For public lawyers advising government agencies, this shift matters. Cybersecurity has traditionally focused on protecting systems and data. Deepfakes target people instead. They turn identity into an attack surface, and once identity is compromised, technical defenses often fail.

This is already happening. Deepfake audio has been used to impersonate executives and public officials in sophisticated fraud schemes, convincing staff to authorize wire transfers or disclose sensitive information. Video deepfakes have been deployed to promote financial scams, spread medical misinformation, and undermine confidence in public messaging [4]. These are not speculative threats. They are operational risks that sit squarely at the intersection of cybersecurity, governance, and law.

From a legal perspective, this exposes a gap. Many cybersecurity policies are designed around unauthorized access, data breaches, and system failures [5]. Deepfakes often involve none of these. The systems work as designed. The harm happens because humans trust what they see and hear. That creates a challenge for public lawyers tasked with advising agencies on risk management and compliance. Existing cyber frameworks may technically be satisfied, even as institutional credibility is quietly eroded.

The consequences extend beyond fraud. Deepfakes complicate incident response and public communications during crises [1]. When a convincing fake video of a public official circulates during an election, a public health emergency, or a security incident, agencies are forced into a reactive mode. Time is lost verifying authenticity, countering false narratives, and reassuring the public. From a cybersecurity standpoint, this is a form of disruption. It undermines situational awareness and decision-making, which are core elements of resilience.

Media law and cybersecurity law converge here. Broadcast standards, consumer protection rules, and defamation law address downstream harm, but cybersecurity governance increasingly focuses on prevention and preparedness. Public lawyers advising agencies now need to think upstream. How do internal policies treat identity verification? Are staff trained to question audiovisual communications, even when they appear to come from leadership? Are public-facing messages protected by provenance tools or authentication mechanisms?

Consent and authorization still matter, but in cybersecurity contexts, they are insufficient. An agency may have full consent to use synthetic media for training or public outreach, yet that same content can be repurposed or weaponized outside its original context. Deepfake training videos, for example, can inadvertently provide high-quality data for malicious actors to refine impersonation attacks. Public lawyers need to consider whether internal uses of AI-generated media increase external risk.

This reframes deepfakes as part of the broader social engineering problem. Cybersecurity professionals have long warned that phishing succeeds because it mimics legitimate communication. Deepfakes raise the stakes by making those communications visually and audibly convincing. When a video appears to show a trusted official issuing instructions, the usual advice to “verify the sender” becomes much harder to follow.

Regulators are beginning to recognize this. Policy discussions in the United States, the European Union, and several African jurisdictions increasingly frame synthetic media as a security issue tied to election integrity, financial stability, and public safety [6]. Disclosure requirements, labeling rules, and AI transparency obligations are not just about ethics. They are defensive measures intended to reduce the attack surface created by synthetic identities.

Cybersecurity advice can no longer be siloed from communications policy or media regulation. Advising on a deepfake incident may involve coordinating with IT security teams, communications offices, regulators, and law enforcement simultaneously [3]. It may also require revisiting procurement rules, vendor contracts, and internal guidelines for AI tools, especially those that generate audio or video.

Importantly, this is not an argument for panic or blanket bans. Alarmism undermines credibility, and public institutions cannot afford that. The more realistic approach is institutional humility. Accept that deepfakes will exist, that they will occasionally succeed, and that legal and technical systems must be designed with that reality in mind. Cybersecurity, in this sense, becomes less about perfect prevention and more about damage control, transparency, and rapid correction.

2. Conclusion

Cybersecurity frameworks depend on trust relationships between systems, users, and institutions. Deepfakes strain those relationships. Once the public begins to doubt whether official communications are genuine, even authentic messages lose credibility. That erosion of trust is slow, cumulative, and difficult to reverse. It is also deeply relevant to democratic governance.

The challenge ahead is not merely to defend networks. It is defending credibility. Deepfakes make clear that cybersecurity is no longer just a technical discipline. It is a legal, social, and institutional one. How public lawyers respond will shape not only risk management practices, but the public’s confidence in the institutions they serve.

References

[1] Boediman, E.P. (2025). Exploring the impact of deepfake technology on public trust and media manipulation: A scoping review. Jurnal Komunikasi, 19(2), pp.313–334. https://doi.org/10.20885/komunikasi.vol19.iss2.art8

[2] Daniel, P. and Almaroof , M.T. (2025). DEEPFAKES AND CYBERSECURITY: WHEN SEEING ISN’T BELIEVING. [online] Available at: https://www.researchgate.net/publication/394400200_DEEPFAKES_AND_CYBERSECURITY_WHEN_SEEING_ISN.

[3] Kapuściński, M. (2025). Deepfake Detection Breakthrough: Universal Detector Achieves 98% Accuracy | TTMS. [online] TTMS. Available at: https://ttms.com/deepfake-detection-breakthrough-universal-detector-achieves-98-accuracy/.

[4] Kaushik, P., Garg, V., Anu Priya and Kant, S. (2024). Financial Fraud and Manipulation. Advances in business information systems and analytics book series, [online] pp.173–196. https://doi.org/10.4018/979-8-3693-6890-9.ch008.

[5] Li, Y. and Liu, Q. (2021). A comprehensive review study of cyber-attacks and cyber security; emerging trends and recent developments. Energy Reports, [online] 7(7), pp.8176–8186. https://doi.org/10.1016/j.egyr.2021.08.126.

[6] Martin-Rozumiłowicz, Dr.B. and Kužel, R. (2019). Social Media, Disinformation and Electoral Integrity. [online] Arlington, VA: International Foundation for Electoral Systems, pp.1–35. Available at: chrome-extension://efaidnbmnnnibpcajpcglclefindmkaj/https://www.ifes.org/sites/default/files/migrate/ifes_working_paper_social_media_disinformation_and_electoral_integrity_august_2019_0.pdf

[7] Singh, S. and Dhumane, A. (2025). Unmasking digital deceptions: An integrative review of deepfake detection, multimedia forensics, and cybersecurity challenges. MethodsX, [online] 15, p.103632. https://doi.org/10.1016/j.mex.2025.103632