Spectrum of Engineering Sciences
ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 4, 2026

REVIEW ON DEEPFAKE TECHNOLOGY AND ITS IMPACT ON SOCIAL LIFE

[1]Waqas Ahmad, [2]Zeeshan, [3]Syed M. Hassan Shah, [4]Muhammad Irfan Fareed, [*5]Uzair Iqbal

[1]CS&IT Department, Sarhad University, Peshawar
[2]ICS & IT Department the University of Agriculture Peshawar
[3]CS&IT Department Iqra National University Peshawar
[4]Department of Life Sciences, University of Management and Technology (UMT), Lahore.
*[*5]Computer System Engineering, UET, Peshawar*
[1]ahmadwaqas186@gmail.com, [*5]engr.uzairiqba@gmail.com

Keywords
Deepfake technology, synthetic media, generative adversarial networks, digital misinformation, media forensics, AI governance

Article History
Received on 17 March, 2026
Accepted on 09 April, 2026
Published on 11 April, 2026

Corresponding Author:
Uzair Iqbal

Abstract

Deepfake technology, which uses advanced artificial intelligence architectures like Generative Adversarial Networks (GANs), diffusion models, and transformer-based systems to produce highly realistic manipulated audiovisual content, is a major breakthrough in the creation of synthetic media. This study explores the technological development of deepfakes from lab testing to widely available apps, evaluating both their beneficial uses in the fields of entertainment, accessibility, education as well as its potential for detrimental abuse. This study shows how deepfake proliferation fundamentally challenges conventional assumptions regarding audiovisual evidence reliability and public trust in digital information ecosystems through an analysis of documented cases from 2025 involving corporate fraud exceeding $25 million, political disinformation campaigns, and widespread privacy violations targeting vulnerable populations.The study assesses current detection techniques, such as frequency-domain analysis, multimodal verification systems, and convolutional neural networks, while critically analyzing their limitations in generalizing beyond training datasets in the face of quickly developing generative capabilities. Additionally, this study examines ethical standards and regulatory frameworks that are developing in many jurisdictions, highlighting the need for coordinated worldwide governance structures that strike a balance between the advancement of technology and the defense of individual rights. The results show that integrated strategies combining technological breakthroughs in detection, extensive media literacy programs, cooperative governance standards, and strong legal protections are necessary for effective mitigation. In the end,In order to maintain information integrity in a time of synthetic media, this paper concludes that deepfake technology exhibits dual-use characteristics, where its societal impact is largely dependent on governance mechanisms, ethical deployment practices, and public awareness rather than inherent technological attributes. This calls for ongoing interdisciplinary collaboration.

1. Introduction:

Deepfakes are media files, like pictures, videos, and audio recordings, that have been made or changed using artificial intelligence to look like real people [31]. These outputs use advanced deep learning methods, like generative models that can find and copy patterns from huge datasets of human faces, voices, and body movements. The Generative Adversarial Network (GAN) is a key technology for making deepfakes. It uses two competing neural networks: a generator that makes fake content and a discriminator that checks its authenticity. Both networks get better through repeated adversarial training, which leads to results that are more and more convincing. Diffusion models (like those used in DALL-E 2 and Stable Diffusion) and transformer-based architectures (like StyleGAN-T and GPT-style models) have made synthetic media even better and more realistic. They let you control facial expressions, lip sync, and full-body motion in great detail [31], [32].

At first, deepfake technology was only available to AI experts in research labs. The first uses of it in the late 2010s often showed visible artifacts like uneven lighting or edge distortions. But starting around 2020, the rise of user-friendly software made it much easier for people to get started. Face Swap and DeepFaceLab are examples of desktop apps that let you swap faces in videos using only a few source images. Mobile apps like Reface made deepfake technology available to everyone, letting people put their own faces into movie clips or celebrity videos in just a few seconds. Eleven Labs and Resemble AI are two platforms that let you make very natural-sounding synthetic speech from just a few minutes of recorded audio. Also, generative video platforms like Synthesia let users make AI-generated presenters for business and educational content that look and sound real, with gestures and lip-syncing in multiple languages [32].

Deepfakes have gone from being experimental demonstrations to widely used digital tools as these tools have become easier to get. People who don't know much about technology can now make convincing fake content using pre-trained models and automated interfaces. Some of these applications even work entirely in a web browser. Deepfakes have real uses in marketing, making things more accessible (like restoring voices for people with speech problems), education (like re-enacting historical events), and entertainment (like making actors look younger or dubbing movies). However, they also bring up serious moral and social issues. Deepfake technology has been utilized to generate customized learning avatars and to maintain the voices of patients afflicted with degenerative diseases [31].

Deepfake systems are becoming more powerful and common, which calls into question long-held beliefs about the reliability of audiovisual evidence. People used to think that audio and video recordings were reliable ways to show what was really going on, as the saying goes, "seeing is believing." Deepfakes challenge this premise and disturb social information ecosystems by rendering fabricated content seemingly genuine, consequently obscuring the distinction between fact and fiction. The fast progress of these technologies has made people talk more about misinformation (like non-consensual pornographic deepfakes and political impersonations), digital trust, identity protection (especially the risks of biometric spoofing), and the future of truth in online spaces. Researchers are actively working on detection algorithms and digital watermarking techniques to protect against this, but these defensive measures are often not as good as the newest deepfake generation methods [33].

2. Technological Foundations of Deepfakes:

Deepfakes are produced using sophisticated generative artificial intelligence models that have been trained on large datasets of real-world images, audio files, and video sequences. These systems detect statistical patterns of lip synchronization, vocal tone, facial structure, head movement dynamics, and micro expressions. The most important architecture in the early development of deepfakes was Ian Goodfellow's 2014 introduction of the Generative Adversarial Network (GAN). GANs consist of two competing neural networks: a discriminator and a generator.

The discriminator evaluates the realism of the synthetic media produced by the generator. Through iterative competition, the system progressively enhances outputs to more closely resemble real human features [1].

After employing GAN-based techniques, researchers used encoder–decoder frameworks and autoencoders to enable identity swapping and facial reenactment in video streams. Large benchmark datasets such as Face Forensics++ have been used to systematically evaluate generation and detection techniques [2]. Recently, diffusion-based generative models have improved texture realism, lighting consistency, and image fidelity by outperforming earlier GAN limitations in stability and detail reproduction [3].
Transformer-based neural networks, which can simulate speech prosody and acoustic patterns in addition to visual synthesis, are used in voice cloning. These multimodal systems make it more difficult to distinguish authentic recordings from deepfakes by combining synchronized audio generation and facial synthesis.

Researchers have looked closely at both the generation and detection mechanisms. Studies published in ACM Computing Surveys and IEEE Signal Processing Magazine document new forensic detection methods and rapid advancements in synthetic media realism [9][5]. Examples of detection methods include multimodal verification systems, biological signal inconsistencies (like eye blinking or pulse detection), convolutional neural networks (CNNs), and frequency-domain analysis [6].
Even with advances in forensic AI, detection systems typically struggle to generalize outside of training datasets. As generative models advance, their outputs—particularly diffusion and large-scale foundation models—become more adaptable and less prone to observable artifacts. Consequently, there is a technological arms race, with advances in synthetic generation often outpacing defensive detection capabilities [7].

Because of this, the technological foundations of deepfakes represent a rapidly evolving dual-use innovation, where increasing generative sophistication challenges existing verification frameworks and necessitates continuous development of robust countermeasures.

3. Applications and Benefits:

Despite widespread concerns about misuse, deepfake technology has many beneficial applications across a range of industries. When used appropriately and under regulatory supervision, synthetic media systems have the potential to enhance learning environments, accessibility, and creativity.
In the entertainment and film industries, deepfake-based visual synthesis has been used to digitally recreate actors, enhance special effects, and perform age modification (de-aging or rejuvenation) without requiring a lot of manual post-production. Modern generative AI systems reduce production costs and expand creative possibilities by enabling realistic facial reenactment and seamless CGI integration [8], [9]. Diffusion-based generative models and neural rendering techniques significantly improved photorealism and temporal consistency in cinematic applications between 2020 and 2025.

For those who have lost their ability to speak due to diseases like ALS or damage to their vocal cords, AI-powered voice synthesis and cloning technologies offer ground-breaking accessibility solutions. Neural text-to-speech systems that can recreate distinctive voice patterns from small audio samples allow people to converse with synthetic voices that mimic their natural speech identity [10], [11]. These advancements in assistive communication technologies promote digital inclusion while preserving individual identity.

The digital reconstruction of historical figures for immersive educational experiences in the domains of education and cultural preservation is made possible by deepfake-based reenactment tools. Interactive AI-generated avatars can enhance student engagement in social science and history learning environments by mimicking speech, facial expressions, and conversational responses [12]. These applications also support efforts to protect digital heritage and virtual museums.

These beneficial applications show that deepfake technology is not intrinsically harmful. Instead, its social impact is contingent upon responsible deployment, ethical standards, and governance frameworks. Synthetic media can serve as an effective tool for innovation rather than deceit when combined with authentication methods and transparency standards [13].

3.1. Real-time world-wide cases

Case No.	Year	Location / Context	Description	Impact / Outcome	Reference
1	2025	Arup Engineering, Global	Deepfake video and voice impersonation of CFO to authorize funds transfer	US $25 million attempted fraud prevented	[14]
2	2025	Hong Kong	Executives impersonated via deepfake video and voice calls	Unauthorized fund transmissions; financial risk	[15]
3	2025	Ireland	Deepfake video of presidential candidate “withdrawing” from election	Misinformation; public confusion; official complaints	
4	2025	Global / Social Platforms	AI-generated non-consensual sexualized content targeting women and minors	Emotional trauma; ethical and legal concerns	[16]
5	2025	Pakistan	Non-consensual deepfake videos of women and public figures	Public outrage; law enforcement intervention	[17]
6	2025	India	Social media deepfake video of a state chief minister in obscene context	Cybercrime investigation; reputational damage	[18]
7	2025	Scotland, UK	Deepfake nude images of a former school friend	Legal prosecution; fine imposed	[19]

Table 1: Real-time world-wide cases

4. Impacts on Social Life:

4.1 Misinformation and Public Trust:

Deepfakes have a major effect on the public's trust in information and social institutions. By producing remarkably lifelike but phony audio or video of public figures, deepfakes can lead viewers to doubt the authenticity of political statements, speeches, or news events. This uncertainty may lead to widespread skepticism, a decrease in engagement with trustworthy media, and potential polarization of public opinion. Deepfakes' rapid social media spread increases their impact because false information often spreads faster than corrective messages. Exposure to such content has been shown to reduce trust in government communications and news sources [20].

4.2 Media Credibility and Journalism:

The credibility of journalists is particularly vulnerable in the era of deepfakes. Even outstanding reporting can be damaged if phony media is widely disseminated and mistaken for real material. The public's reliance on untrustworthy sources would rise and the efficacy of journalists' work would be diminished if viewers began to question authentic videos, interviews, or investigative reports. This has important implications for democratic discourse because a decline in media trust may reduce citizens' ability to make informed decisions and actively participate in public life [21].

4.3 Psychological and Social Consequences:

The psychological effects of deepfakes extend beyond public trust. People may experience fear and anxiety due to the possibility of their voices or photos being misinterpreted on the internet. Deepfakes, which target private individuals and can cause long-term mental suffering, social stigma, and reputational harm, include fake videos, phony endorsements, and sexualized content. The fact that even one phony film of a political figure or organization can drastically lower public trust in the related institutions illustrates the potential for broad societal mistrust [22].

4.4 Privacy Violations and Non-Consensual Use:

One of the most detrimental social effects of deepfakes is the invasion of privacy. Identity theft, harassment, and non-consensual deepfake pornography disproportionately target women and other vulnerable groups, which can result in psychological anguish and reputational damage. Malicious actors can create convincing content without specialist knowledge because to the widespread availability of deepfake technologies, often leaving victims helpless. The fact that many countries' legal and regulatory institutions are only now beginning to adapt to these challenges shows how urgently stronger protection structures are needed.

4.5 Disinformation in Health and Public Safety Contexts:

Deepfakes are increasingly being used to spread false information in sensitive sectors like health and safety. AI-generated videos of medical practitioners have been published on websites like YouTube and TikTok to bolster unsupported medical claims or treatments. Such misleading information has the potential to undermine public confidence in healthcare systems, encourage unhealthy choices, and complicate emergency response. Deepfakes can directly affect public safety and wellbeing, as seen by the rapid dissemination of false health information.

4.6 Broader Social Implications:

Deepfakes are a complicated danger to social cohesiveness in general. They can accelerate the dissemination of false information, damage institutional credibility, and undermine public trust in information systems in addition to having psychological consequences on individuals. They also raise ethical questions about accountability, permission, and the boundaries of digital manipulation. By addressing these problems through a combination of technological solutions (detection and verification tools), governmental measures (rules against misuse), and public awareness efforts (media literacy programs), the detrimental societal consequences of synthetic media can be mitigated.

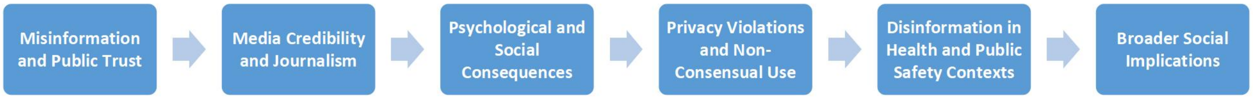

Fig.1: Impacts on Social Life

5. Regulatory and Ethical Considerations:

Academics and policymakers are emphasizing the necessity for ethical standards and legislative frameworks that strike a balance between technological growth and the defense of people's rights and the general welfare due to the rapid development of deepfake technology. Deepfakes provide dual-use challenges: they enable creative and instructional uses while also facilitating disinformation, privacy infringement, and harassment. Therefore, regulatory strategies must minimize harm without obstructing the development of legal technology [23].

5.1 Legal Frameworks:

To combat deepfakes, several countries have started passing laws and regulations. The United States, for example, has passed state-level legislation that makes it illegal to create non-consensual sexual deepfakes, while the European Union is considering AI governance regulations under the proposed Artificial Intelligence Act, which classifies high-risk AI systems, including synthetic media with potential for harm [24]. Other tactics include establishing guidelines for the authenticity of digital information and requiring media creators to employ verifiable metadata or cryptographic watermarks to distinguish original content from altered versions. These regulations aim to protect people's privacy and public trust, reduce malicious use, and enhance accountability.

5.2 Ethical Guidelines:

To ensure that deepfake technology is used appropriately, ethical paradigm development is just as important as legal regulation. Scholars choose theories that consider cultural, societal, and religious variables because misuse can have varying outcomes depending on the circumstances. Ethical norms often emphasize consent, transparency, and justice: artists should obtain consent from those whose photos are used, reveal synthetic content in a clear and intelligible way, and avoid creating content that could mislead or harm others [25].

5.3 Global Coordination and Policy Challenges:

Due to the global reach of social media platforms and the dissemination of digital information, isolated country prohibitions are often insufficient. Effective governance requires collaboration, international coordination, and the implementation of cross-border enforcement actions by governments, IT companies, and civil society. Multilateral agreements can help detect dangerous deepfakes, respond to cybercrime, and create minimum standards for the ethical use of AI while safeguarding freedom of expression and innovation.

In In conclusion, the proper application of deepfake technology requires a combination of technological tools, ethical standards, and legal protections. Policymakers must carefully strike a balance between fostering innovation and defending individual liberty, public trust, and societal values. In the lack of consistent law and ethical control, it is anticipated that the issues that deepfakes pose to society—such as misinformation, harassment, and a loss of trust—will worsen.

Fig.2: Regulatory and Ethical Considerations

6. Detection and Mitigation Strategies:

It takes a multifaceted approach that combines technological innovation, public awareness, and coordinated governance to effectively combat the negative effects of deepfake technology. Each component contributes differently to lessening the negative effects of synthetic media on society, but in order for these strategies to be successful, they must all overcome certain obstacles.

6.1 Technical Detection:

The detection of artificial content is mostly dependent on technical techniques. These systems use visual, auditory, and temporal features to distinguish between real and modified media. Common techniques include convolutional neural networks (CNNs), multimodal frameworks that evaluate both audio and video cues, and progressively deep architectures that combine identity-based,

ISSN (e) 3007-3138 (p) 3007-312X Volume 4, Issue 4, 2026

temporal, and spatial data to improve accuracy [26], [27]. For example, because multimodal detection frameworks simultaneously examine facial movements and audio synchronization, they are more accurate than single modal systems [28].

Even with these developments, it is still very challenging to generalize technical detection outside of carefully selected datasets. Most detection models, such as FaceForensics++ or DFDC, only perform well on the datasets they were trained on; when tested on new or real-world deepfakes created by cutting-edge platforms or approaches, their performance significantly falls [28]. This poor generalization is caused by overfitting to specific artifact patterns in training data rather than learning robust forgery cues that transfer across contexts. Emerging research is exploring methods like foundation models and identity aware detection to improve cross-dataset performance, although more work is needed [29].

6.2 Media Literacy and Public Education:

Deepfakes cannot be totally prevented by technology alone. Increasing media literacy, or people's ability to critically assess the dependability of digital content, is essential. Campaigns for public awareness, educational programs, and seminars on critical thinking help viewers spot manipulation and cultivate a healthy distrust of untrustworthy media. Studies show that understanding popular deepfake techniques and learning which signals to take into account (e.g., lip synchronization, uneven lighting, atypical eye movement) might improve human detection ability and reduce sensitivity to misleading information [30].

6.3 Collaborative Governance and Standards:

Governance has a major influence on the context in which deepfakes are created and propagated. Policymakers, IT companies, researchers, and civil society must work together to build frameworks and norms that control the creation, dissemination, and identification of deepfakes. This includes regulatory measures that require transparency for information generated by artificial intelligence, as well as provenance tracking standards (such as C2PA) and metadata watermarking that provide verifiable proof of origin [4]. Additionally, cross-sector collaboration helps ensure that best practices in detection and mitigation are widely adopted and accelerate the development of shared technologies.

6.4 Ongoing Challenges and Future Directions:

Despite these tactics, the fight against deepfakes is still ongoing. The rapid development of generating models and the emergence of evasive strategies often make it difficult for detection systems to adapt to new generation approaches. The generalization gap highlights the need for context-aware frameworks that can learn from a wide range of continuously updated media samples. When models have trouble with invisible or real forgeries, this gap appears. Using decentralized authentication methods like blockchain and combining automatic detection with human monitoring via human-in-the-loop systems are possible next stages [4].

In general, mitigating the negative effects of deepfakes necessitates a multipronged strategy that combines cutting-edge detection tools with cooperative legislative initiatives and public awareness campaigns. To protect the integrity of information and uphold public trust in an era of synthetic media, coordinated efforts across technical, social, and institutional domains are necessary because a single answer is insufficient.

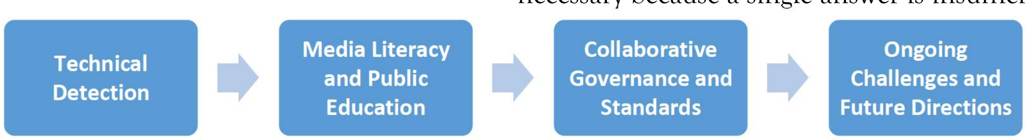

Fig.3: Detection and Mitigation Strategies

7. Conclusion:

The rapid advancement of deepfake technology, which began as a research project and is now an operational tool, provides a crucial scenario that permits both advantageous and detrimental applications. People can now create content thanks to the system's technical foundation, which makes use of advanced generative models. This is because they need to develop content that is both entertaining and instructional. The 2025 case studies demonstrate how these cutting-edge technologies allow criminals to carry out high-end frauds while influencing political decisions and, most frequently, using private photos to cause severe psychological trauma. After this incident, there is no longer any public faith in digital content because it shown that individuals will accept what they see, endangering appropriate public discourse that relies on reliable information. In order to solve this issue, cooperation is needed, which calls for more than just technical fixes to start an arms race.

Even if detection systems struggle to keep up with emerging generative technologies, the advancement of forensic detection techniques and content provenance monitoring is crucial. Three distinct approaches are needed for the solution to function: the first calls for continuous technology improvement for detection; the second calls for cooperative governance; and the third calls for the creation of worldwide transparency standards and legal frameworks. In order for people to properly evaluate synthetic media content, society must fund public education programs and media literacy training.

The constraints we set for society through our decision-making process will determine how deepfake technology advances. One of the biggest issues of the current decade is that companies must figure out how to support creative endeavors while upholding truthful information, individual privacy rights, and public trust. Whether synthetic media is used to disseminate information to the public or to deceive individuals will depend on the decisions made today.

References:

[1] I. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative Adversarial Nets," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, 2014, pp. 2672–2680.

[2] A. Rössler, D. Cozzolino, L. Verdoliva, C. Riess, J. Thies, and M. Nießner, "FaceForensics++: Learning to Detect Manipulated Facial Images," in *Proc. IEEE/CVF Int. Conf. on Computer Vision (ICCV)*, 2019, pp. 1–11.

[3] J. Ho, A. Jain, and P. Abbeel, "Denoising Diffusion Probabilistic Models," in *Proc. Advances in Neural Information Processing Systems (NeurIPS)*, vol. 33, 2020, pp. 6840–6851.

[4] "Deepfake detection in generative AI: A legal framework proposal to protect human rights," *Computer Law & Security Review*, vol. 58, 2025.

[5] L. Verdoliva, "Media Forensics and Deepfake Detection: An Overview," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 81–95, Nov. 2020.

[6] B. Dolhansky, J. Bitton, B. Pflaum, J. Lu, R. Howes, M. Wang, and C. Ferrer, "The Deepfake Detection Challenge Dataset," in *Proc. IEEE/CVF Conf. on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2020, pp. 3068–3077.

[7] R. Babaei, et al., "Generative AI and the Evolving Challenge of Deepfake Detection," *Future Internet*, vol. 14, no. 1, p. 17, 2025.

[8] L. Verdoliva, "Media Forensics and Deepfake Detection: An Overview," *IEEE Signal Processing Magazine*, vol. 37, no. 6, pp. 81–95, Nov. 2020.

[9] Y. Mirsky and W. Lee, "The Creation and Detection of Deepfakes: A Survey," ACM *Computing Surveys*, vol. 54, no. 1, pp. 1–41, 2021.

[10] J. Shen, R. Pang, R. J. Weiss, et al., "Natural TTS Synthesis by Conditioning WaveNet on Mel Spectrogram Predictions," *IEEE Int. Conf. on Acoustics, Speech and Signal Processing (ICASSP)*, 2020.

[11] A. Polyak, L. Wolf, and Y. Taigman, "Speech Resynthesis from Voice Conversion with Neural Networks," *IEEE/ACM Transactions on Audio, Speech, and Language Processing*, vol. 28, pp. 1100–1112, 2020.

[12] C. Vaccari and A. Chadwick, "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust," *Social Media + Society*, vol. 6, no. 1, 2020.

[13] R. Babaei et al., "Generative AI and the Evolving Challenge of Deepfake Detection," *Future Internet*, vol. 14, no. 1, p. 17, 2025.

[14] "Weaponizing Reality: Evolution of Deepfake Technology," IBM X-Force Threat Intelligence, 2025. [Online]. Available: https://www.ibm.com/think/x-force/weaponizing-reality-evolution-deepfake-technology

[15] "Deepfakes Year in Review 2025," Bitmind AI Blog. [Online]. Available: https://bitmind.ai/blog/deepfakes-year-in-review-2025

[16] "Deepfake Pornography," Wikipedia, 2025. [Online]. Available: https://en.wikipedia.org/wiki/Deepfake_pornography

[17] "Deepfake Technology: A Growing Threat in Pakistan," Pakistan Today, Mar. 2, 2025. [Online]. Available: https://www.pakistantoday.com.pk/2025/03/02/deepfake-technology-a-growing-threat-in-pakistan-2/

[18] "CM Mann's AI Deepfake Video Sparks FIR; Cops Tracing Creator," Times of India, 2025. [Online]. Available: https://timesofindia.indiatimes.com/city/chandigarh/cm-manns-ai-deepfake-video-sparks-fir-cops-tracing-creator/articleshow/124725491.cms

[19] "Scots man fined for deepfake nude pics," The Scottish Sun, 2025. [Online]. Available: https://www.thescottishsun.co.uk/news/15232208/scots-creep-made-deepfake-naked-pics/

[20] C. Vaccari and A. Chadwick, "Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust," *Social Media + Society*, vol. 6, no. 1, 2020.

[21] Frontiers in Psychology, "False failures, real distrust: The psychological impact of deepfakes on institutional trust," *Frontiers in Psychology*, vol. 16, Article 1574840, 2025.

[22] Guardian, "AI deepfakes of real doctors spreading health misinformation on social media," Dec. 5, 2025. [Online]. Available: https://www.theguardian.com

[23] R. Chesney and D. Citron, "Deepfakes and the New Disinformation War: Threats and Legal Responses," *Foreign Affairs*, vol. 99, no. 1, pp. 147–155, 2020.

[24] European Commission, "Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act)," 2021. [Online]. Available: https://digital-strategy.ec.europa.eu/en/library/proposal-regulation-laying-down-harmonised-rules-artificial-intelligence

[25] S. Westerlund, "The Emergence of Deepfake Technology: A Review," *Technology Innovation Management Review*, vol. 11, no. 11, pp. 10–20, 2021.

[26] G. Lin, L. Lin, C. P. Walker, D. S. Schiff, and S. Hu, "Fit for Purpose? Deepfake Detection in the Real World," *arXiv preprint*, Oct. 2025.

[27] H. Boo, E. Lee, and J. Lee, "Referee: Reference- aware Audiovisual Deepfake Detection," *arXiv preprint*, Oct. 2025.

[28] K. Gandhi et al., "A Multimodal Framework for Deepfake Detection," *Journal of Electrical Systems*, vol. 20, no. 10s, 2024.

[29] M. Abbasi et al., "Comprehensive Evaluation of Deepfake Detection Models: Accuracy, Generalization, and Resilience," *Applied Sciences*, vol. 15, no. 3, Jan. 2025.

[30] C. Chen and D. H.-L. Goh, "Seeing, Hearing, and Knowing Together: Multimodal Strategies in Deepfake Videos Detection," *arXiv preprint*, Feb. 2026.

[31] Smith and L. Chen, "Generative adversarial networks and the evolution of deepfake technology," IEEE Trans. Inf. Forensics Security, vol. 15, no. 3, pp. 450–465, Mar. 2021.

[32] M. Johnson and R. Patel, "User-friendly deepfake tools: From FaceSwap to ElevenLabs," in Proc. IEEE Conf. Artif. Intell. Media, New York, NY, USA, 2022, pp. 112–119.

[33] K. Lee, "Detection challenges and societal impact of synthetic media," in Proc. IEEE Int. Workshop Digit. Watermarking, Beijing, China, 2023, pp. 78–84.