APPLICATIONS AND RISKS OF DEEPFAKE TECHNOLOGY: SECURITY, ETHICAL, AND LEGAL CHALLENGES

[1]Primbetov Abbaz

[1]Phd student, Tashkent University of Information Technologies named after Muhammad Al-Khwarizmi
[1]Senior lecturer, Computer engineering, Tashkent University of Applied Sciences
e-mail: abbaz0203@mail.ru

Abstract: Deepfake technology, driven by modern artificial intelligence and deep learning models, enables highly realistic manipulation of visual and audio media. While it offers benefits in film production, gaming, education, and digital content creation, deepfakes also introduce serious security, ethical, and legal risks. Malicious uses include political misinformation, identity theft, financial fraud, and non-consensual content generation, all of which threaten privacy, public trust, and social stability. Recent global and regional cases demonstrate the growing impact of deepfakes and highlight the need for effective detection methods and regulatory frameworks. Strengthening forensic analysis and developing robust AI-based detection systems are essential for ensuring media integrity and protecting individuals and institutions from deepfake-related threats.

Keywords: Deepfake, AI manipulation, security risks, ethical issues, legal challenges, misinformation, identity fraud.

1. INTRODUCTION

The rapid advancement of artificial intelligence, particularly deep learning–based generative models, has fundamentally transformed the way digital content is created and manipulated. Among these technologies, deepfakes represent one of the most powerful and controversial innovations, enabling the generation of highly realistic synthetic images, videos, and audio. By leveraging architectures such as GANs, autoencoders, and diffusion models, deepfake systems can replicate human faces, voices, expressions, and movements with remarkable fidelity.

While deepfake technology offers significant opportunities in creative industries—such as film production, virtual education, entertainment, and human–computer interaction—it simultaneously introduces serious risks. Malicious deepfakes can be used to spread misinformation, manipulate public opinion, impersonate individuals in financial or political contexts, and produce non-consensual or defamatory content. These threats pose challenges for cybersecurity, digital forensics, ethical governance, and legal regulation.

Given the dual-use nature of deepfakes, understanding their generation mechanisms, content types, and potential applications is essential for designing effective detection algorithms and protective measures. This study provides a systematic overview of major deepfake categories, their creation methods, and the associated risks, thereby contributing to the ongoing development of robust, AI-driven defense systems to safeguard media integrity and public trustn [1].

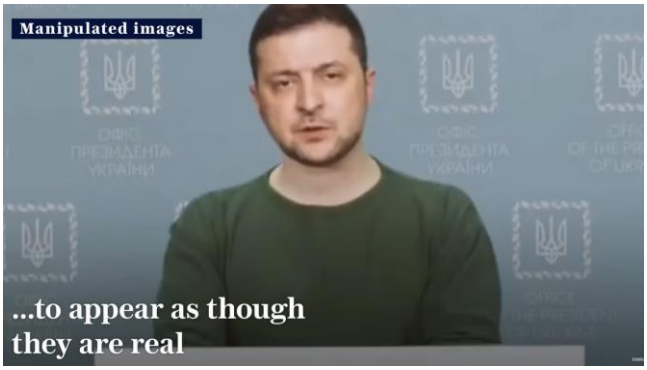
2. METHODOLOGY

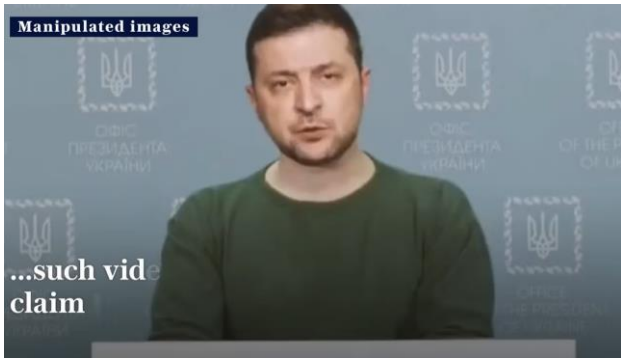
In recent years, deepfake technologies have emerged as one of the most significant global threats to the integrity of the information environment. Synthetic images, audio, and videos generated using advanced deep learning models not only endanger personal dignity but also pose serious risks to societal security, political stability, and economic systems. Deepfakes enable highly realistic falsification of a person's visual or vocal identity, making it possible to fabricate events that closely resemble real-life occurrences. This phenomenon currently manifests dangerous consequences in three major domains:

- Socio-political manipulation — the dissemination of false information through fabricated representations of state leaders, politicians, and public figures.
- Financial and personal fraud — the illegal execution of banking transactions, employment-related processes, or monetary transfers through forged voices or facial identities.
- Cultural and ethical manipulation — the distortion of moral values, social norms, and cultural perceptions through falsified media content.

2.1. Political Manipulation and Threats to Information Security

The use of deepfakes in political propaganda has significantly amplified their global impact. One of the earliest high-profile incidents occurred in 2022, when a fabricated video depicting Ukrainian President Volodymyr Zelensky calling on soldiers to "lay down their arms and surrender" was broadcast online. Although the visual quality of the video was not flawless, it spread rapidly across social media platforms and undermined public trust in real political events.





Similarly, in 2019, several videos of U.S. House Speaker Nancy Pelosi were artificially slowed down to make her appear as if she were speaking under the influence of alcohol. The manipulated clips generated widespread discussion on Facebook and Twitter, bringing renewed attention to the responsibility of digital platforms in combating misinformation [2].

Another notable example involves a well-known deepfake of Mark Zuckerberg, in which he appears to boast about "controlling users' data." This widely circulated media artifact further intensified global debates surrounding the ethical limits of deepfake technology and the broader consequences for digital governance and public trust [3].

2.2. Financial Fraud and Economic Risk

Deepfake technologies have also emerged as a significant catalyst for financial fraud, creating unprecedented vulnerabilities within corporate and banking systems. In 2024, a major incident occurred at the London-based firm Arup, where cybercriminals successfully orchestrated a fraudulent transfer totaling 25.6 million USD. The perpetrators used AI-generated video calls that convincingly replicated the company's Chief Financial Officer and other executives, thereby obtaining authorization for the transaction [4].

A similar case was reported in 2019 in the United Kingdom, when the director of an energy company transferred 243,000 USD after receiving a deepfake audio call impersonating the organization's German-based chief executive. In 2020, another large-scale fraud involved a bank manager who approved a 35 million USD transfer based on instructions from a deepfake version of the CEO [5].

These incidents collectively demonstrate that AI-generated audio and video content has reached a level of realism capable of deceiving experienced professionals and bypassing institutional security protocols. As a result, financial sectors worldwide are increasingly exposed to sophisticated forms of identity manipulation and high-impact economic attacks.

3. CULTURAL AND ETHICAL MANIPULATION

Deepfake videos generated using the likeness of celebrities have become one of the most widespread and potentially harmful applications of artificial intelligence technologies. A notable example emerged in 2024, when a fabricated promotional video featuring Taylor Swift began circulating on social media platforms. In the manipulated clip, the singer appears to announce a collaboration with the cookware brand Le Creuset, claiming that the company was distributing free kitchenware sets due to a "packaging error." In reality, the video was entirely synthesized using AI and was designed to lure viewers into a phishing scheme. Users who clicked the link accompanying the video were redirected to fraudulent websites that harvested personal information and executed unauthorized financial transactions. This incident further demonstrates how deepfakes have evolved into sophisticated instruments of financial exploitation [6].

Another high-profile case involves the popular TikTok account showcasing hyper-realistic deepfake videos of actor Tom Cruise, which amassed more than 3.6 million followers. These clips depict the actor performing various activities—such as performing tricks or playing golf—with a level of realism that blurs the boundary between authentic and synthetic media. Such examples illustrate how deepfakes erode the distinction between real and fabricated content, contributing to widespread cultural and perceptual distortion [7].

In 2020, South Korea's MBN television channel broadcast a deepfake-generated version of news anchor Kim Joo-Ha, marking one of the first instances of deepfake technology being employed in mainstream media. This event signaled the emergence of AI-generated personas within official news environments and intensified debates over ethical responsibility, transparency, and media trustworthiness [8].

2.4 Deepfake Threats in the Context of Uzbekistan

In recent years, the rapid global diffusion of deepfake technologies has also significantly affected Uzbekistan's information landscape. Advanced forms of visual and audio manipulation powered by artificial intelligence now pose serious risks to political, social, and legal security. Several cases recorded during 2024–2025 provide clear evidence of these emerging threats.

One of the most prominent incidents involved a deepfake video created from the genuine interview of Tanzila Narbayeva, Chairperson of the Senate of the Oliy Majlis, originally published by Kun.uz in December 2023. In the manipulated version, Narbayeva appears to encourage Uzbek citizens to sign contracts with the Russian Ministry of Defense. The fake video was constructed by re-modeling her face, voice, and lip movements using AI, synchronizing these synthesized elements with artificially generated audio. The statements attributed to her in the video were entirely fabricated and had no relation to the original interview, which addressed the issue of early marriage among girls in Uzbekistan [9].

Similar manipulated videos were also created using the likeness of other public figures, including weightlifter Ruslan Nuriddinov, various journalists, and bloggers. In the deepfake clip featuring Nuriddinov, he appears to claim that he has signed a contract with the Russian Ministry of Defense. The Ministry of Defense of Uzbekistan officially refuted these claims, describing the video as a "high-quality video deepfake" and confirming that the athlete was serving in the Armed Forces of Uzbekistan [9].

These deepfake materials were generated through high-precision replication of the individuals' appearance, voice, and facial expressions, combining text-to-speech (TTS) and lip-synchronization technologies to create highly realistic outputs.

Therefore, the creation and dissemination of such deepfake content may constitute not only an ethical and political violation but also a criminal offense under national law.

The analysis above demonstrates that deepfake technologies have become a significant threat to the global information environment, political systems, and economic security. By leveraging the capabilities of deep learning, deepfakes can convincingly falsify an individual's face, voice, and behavior, thereby distorting reality and presenting fabricated information as authentic. As a result, there is a growing need to intensify scientific research aimed at detecting deepfake media. The protection of social, political, and personal security—as well as the credibility of the information ecosystem—now largely depends on the effectiveness and robustness of AI-based detection models.

CONCLUSION

Deepfake technology has rapidly evolved into one of the most critical challenges of the modern digital era, reshaping global security, ethics, and information integrity. While AI-generated media offers innovative opportunities in entertainment, communication, and education, its misuse poses substantial risks to political stability, economic systems, and societal trust. Real-world cases from around the world—including political misinformation, multimillion-dollar financial fraud, and cultural manipulation—demonstrate that deepfakes are no longer theoretical threats but active instruments of deception. The emergence of deepfake incidents in Uzbekistan further highlights that no information ecosystem is immune to this growing danger.

These developments underscore the urgent need for advanced research in deepfake detection, forensic analysis, and regulatory frameworks. Strengthening AI-driven identification models,

promoting digital literacy, and implementing legal safeguards are essential steps toward mitigating the impact of synthetic media. Ultimately, ensuring public trust and safeguarding digital environments will require a coordinated effort between researchers, policymakers, and technology developers. By addressing the risks outlined in this study, society can better prepare for a future where authentic and synthetic content coexist, and where robust security measures are critical to maintaining the integrity of information.

REFERENCE:

1. Brooks, T., Princess, G., Heatley, J., Jeremy, J., & Scott, K. (2019). Increasing threats of deepfake identities. US Department of Homeland Security [online].
2. Popa, C., Pallath, R., Cunningham, L., Tahiri, H., Kesavarajah, A., & Wu, T. (2025). Deepfake technology unveiled: The commoditization of AI and its impact on digital trust. arXiv preprint arXiv:2506.07363.
3. Romero-Moreno, F. (2025). Deepfake detection in generative AI: A legal framework proposal to protect human rights. Computer Law & Security Review, 58, 106162.
4. Kaushik, P., Garg, V., Priya, A., & Kant, S. (2025). Financial fraud and manipulation: The malicious use of deepfakes in business. In Deepfakes and Their Impact on Business (pp. 173-196). IGI Global Scientific Publishing.
5. Basu, A. (2024, November). The Impact of Artificial Intelligence on Cybersecurity. In Abu Dhabi International Petroleum Exhibition and Conference (p. D021S077R001). SPE.
6. Miao, Q., Kang, S., Marsella, S., DiPaola, S., Wang, C., & Shapiro, A. (2022). Study of detecting behavioral signatures within DeepFake videos. arXiv preprint arXiv:2208.03561.
7. Mustak, M., Salminen, J., Mäntymäki, M., Rahman, A., & Dwivedi, Y. K. (2023). Deepfakes: Deceptions, mitigations, and opportunities. Journal of Business Research, 154, 113368.
8. “Ogoh bo‘ling! Taniqli shaxslar obrazi asosida Rossiya armiyasiga yollanishga chorlovchi dipfeyk videolar tarqalmoqda” [Elektron resurs] - Kun.uz – Murojaat sanasi: /16:05 / 25.06.2025 – kirish manzili (kun.uz/)