

Original Scientific Paper/Original naučni rad
Paper Submitted/Rad primljen: 31.12.2025.
Paper Accepted/Rad prihvaćen: 20.01.2026.
DOI: 10.5937/SJEM2600054B

UDC/UDK: 004.8:004.94]:17

AI deepfake tehnologija

Lazar Bezbradica¹, Ana Kosanović², Borjana Georgiou Sekuloski³, Ratko Stajić⁴

¹ Belgrade School of Engineering Management; email: lazar.bezbradica5@gmail.com

² Belgrade School of Engineering Management; email: anakosanovic05@gmail.com

³ Belgrade School of Engineering Management; email: borjanasekuloski@gmail.com

⁴ Belgrade School of Engineering Management; email: ratko.stajic184@gmail.com

Sažetak: Razvoj veštačke inteligencije doveo je do značajnih tehnoloških prednosti, ali je istovremeno stvorio nova etička i bezbednosna pitanja, među kojima se posebno izdvaja pojava deepfake tehnologije. Ova tehnologija, zasnovana na neuronskim mrežama, omogućava generisanje izuzetno realističnih, ali lažnih audio-vizuelnih sadržaja, čime se značajno ugrožava sigurnost i integritet informacija, kao i poverenje javnosti. Ovaj rad, zasnovan je na sistematskoj analizi relevantne naučne i stručne literature, koja obuhvata najnovije tehnologije u oblasti veštačke inteligencije, sa posebnim osvrtom na deepfake sadržaj. Koristili su se izvori iz baza podataka kao što su: PubMed, Google Scholar, IEEE Xplore i relevantni časopisi iz oblasti informacionih tehnologija i etike. Rezultati istraživanja tehnologija 2025. godine ukazuju na drastičan porast i kompleksnost deepfake tehnologije, što značajno utiče na bezbednost internet korisnika. Naime, deepfake fajlovi se procenjuju da će do kraja 2025. godine dostići 8 miliona, što je enorman skok sa pola miliona u 2023. godini. Još alarmantnija činjenica je ta, da je efikasnost ljudskog otkrivanja visoko kvalitetnog deepfake sadržaja jako niska, i iznosi svega 24,5%. Cilj ovog rada jeste da predstavi ulogu veštačke inteligencije u stvaranju i detekciji deepfake sadržaja, kao i da istraži već postojeće modele, metode, algoritme i strategije za njihovo razotkrivanje i suzbijanje.

Ključne reči: veštačka inteligencija, dezinformacije, deepfake, etika.

AI deepfake technologies

Abstract: The development of artificial intelligence has led to major technological advancements but has simultaneously created new ethical and security challenges, most notably the emergence of deepfake technology stands out. Based on neural networks, this technology enables the creation of highly realistic yet fabricated audio-visual content, threatening information integrity and public trust. This paper presents a systematic analysis of recent scientific and professional literature, focusing on deepfake technologies and their ethical implications. Sources include databases such as PubMed, Google Scholar, IEEE Xplore, and leading journals in information technology and ethics. Researches from 2025 indicate a drastic rise in both the volume and sophistication of deepfake content, posing growing risks to online security. Deepfake files are projected to reach 8 million by the end of 2025, compared to half a million in 2023. Furthermore, human detection accuracy for high-quality deepfakes remains low, averaging only 24.5%. This study aims to examine models, algorithms, and strategies for their identification and mitigation.

Keywords: artificial intelligence, disinformation, deepfake, ethics.

1. Introduction

Rapid technological development has led to the emergence of artificial intelligence, including deepfake technology. Many still cannot precisely explain this type of technology nor accurately define deepfake. As a consequence, existing regulations and laws are often not fully adequate. Effectively addressing the risks brought by technological development and deepfake requires a clear understanding of these phenomena. Unfortunately, constant technological progress makes it difficult for regulatory frameworks to fully protect users. Innovations appear in very short time intervals, which further complicates monitoring and adapting legislation. Various articles that attempt to explain deepfake cannot confidently claim what it truly is (Sharif, Atif, Ali Nagra, 2025; Paz Sandoval, Almeida Vau, Solaas, Rodrigues, 2025; Jedličkova, 2024; Parti, Szabo, 2024; Jorgensen, Shamini Gunasekaran, Grace Ma, 2025; Birrer, Just, 2025). Everyone is aware that it exploits faces, voices, and movements of people, but it has grown into identity misuse. Initially, it found its place in entertainment and creativity, which

later evolved into identity abuse. The expansion of technology has led to digital content on the internet increasing by as much as 900% from 2019 to 2020 (Kazaz, 2024, p.1). The first appearance of deepfake was observed in the pornography industry, when faces and voices of celebrities were used. Unfortunately, deepfake did not stop there. It is also used in other spheres. It has been observed in politics, during public reporting, when many characterized deepfake as a tool for triggering false alarms and loss of trust. Because of this, regulations and laws had to be introduced to protect user privacy. Educators who sought to introduce young people to the possibilities of deepfake technology inadvertently contributed to increased privacy abuses (Sharif, Atif, Ali Nagra, 2025). Therefore, it is necessary to thoroughly analyze deepfake technology so that regulations and laws can prevent all potential abuses. Analyzing existing regulations allows identification of possible shortcomings and uncovered areas that require improvement.

2. Historical Development of Deepfake Technology

Deepfake technology refers to synthetic media created using artificial intelligence, where a person's appearance, voice, or actions are realistically altered or generated. Although the term "deepfake" is relatively new, the technological foundations behind it developed gradually over several decades (Verdoliva, L. 2020). Understanding the historical development of deepfakes is essential for evaluating their current capabilities and societal implications.

2.1. Early Image and Video Manipulation

Before the rise of artificial intelligence, visual manipulation relied on traditional digital editing techniques. Software such as Adobe Photoshop enabled static image alterations, while computer-generated imagery (CGI) allowed filmmakers to create realistic visual effects (Tolosana, R., et al. 2020). These methods, however, required significant technical expertise and manual effort, making large-scale or real-time manipulation impractical.

2.2. Emergence of Machine Learning Techniques

The development of machine learning marked a turning point in digital media manipulation. Early neural networks enabled computers to recognize patterns in data, particularly in facial recognition and speech processing (Goodfellow, I., et al. 2014). Autoencoders, a type of neural network designed to learn efficient data representations, later became a key component in face-swapping techniques. These models could encode facial features and reconstruct them onto different subjects, laying the groundwork for early deepfake systems.

2.3. Generative Adversarial Networks (GANs)

A major breakthrough occurred in 2014 with the introduction of **Generative Adversarial Networks (GANs)** by Ian Goodfellow and his colleagues. GANs consist of two neural networks: a generator that creates synthetic data and a discriminator that evaluates its authenticity. Through continuous competition, the generator improves its outputs, resulting in highly realistic synthetic images and videos. GANs dramatically improved the quality of AI-generated faces and significantly accelerated deep fake development (Goodfellow, I., et al. 2014).

2.4. Public Emergence of Deepfakes

The term "deepfake" gained public attention around 2017, when face-swapping videos began appearing on online forums and social media platforms. These early deepfakes often focused on celebrities, highlighting both the power and the ethical risks of the technology (Chesney, R., Citron, D. 2019). Open-source tools and improved hardware soon made deepfake creation accessible to non-experts, contributing to its rapid spread.

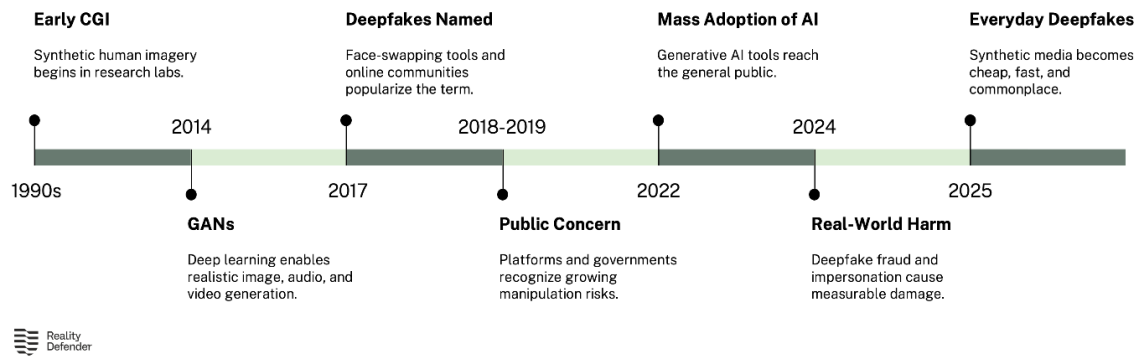
2.5. Expansion to Audio and Real-Time Media

Following advancements in video synthesis, researchers extended deepfake techniques to audio generation. Voice cloning systems trained on small speech samples were able to replicate tone, pitch, and speaking style with high accuracy. More recently, real-time deepfake systems have emerged, enabling live facial and voice manipulation during video calls and streaming. This evolution significantly increased the potential impact of deepfake technology.

In conclusion to this chapter, the historical development of deepfakes reflects broader progress in artificial intelligence and deep learning. From manual image editing to advanced generative models, deepfake technology

evolved rapidly within a short time frame. While these advancements enabled creative and beneficial applications, they also introduced serious ethical and security concerns that continue to shape ongoing research and regulation.

Figure 1: Timeline of deepfake technologies



Source: (<https://www.realitydefender.com/insights/history-of-deepfakes>, Gabe Regan 2025.)

3. Case Studies and Real World Incidents

The principal aim of this paragraph is to state and explain crimes which are usually performed by using deepfake technologies. The crimes which will be discussed further are: misinformation; identity theft; harmful content; criminal implications.

3.1. Misinformation Distribution

The most prominent risk associated with deepfakes is the distribution of misinformation. AI-generated videos and audio recordings have been used to falsely depict public figures by making statements they never made, often during politically sensitive periods (Vaccari & Chadwick, 2020).

CNN states that one of the main activists against deepfake technologies is in fact, famous actress Scarlett Johansson. She has been a victim to numerous AI-made misinformations, videos etc. which led her to become the main supporter of deepfake legislation.

3.2. Fraud and Identity Impersonation

Deepfake technologies has been heavily used in financial scams (up to 43% of which are successfully conducted), where attackers impersonate executives or family relatives to request urgent money transfers (Jennifer Gregory IBM, 2025). Victims often comply to the accuracy of the synthetic voice of their trusted individuals.

3.3. Non-consensual and Harmful content (Non-consensual Deepfake Pornography)

The ease with which this content can be created has led to 98% of all deepfake videos online being pornographic in nature, while an astonishing 99% of people victimized by deepfake pornography are women. (Gabe Regan, 2025). Case analysis underline the need for stronger legislative methods in AI use.

3.4. Legal and Criminal implications

Deepfakes have also appeared in criminal investigations, including fabricated audio or video evidence intended to damage reputations or incite conflict. Many countries such as India, Croatia, Serbia, Bosnia and Herzegovina and USA have been a victim to highly realistic, AI generated, threats of bombing educational institutions in the past 5 years.

3.5. Consequences of AI Crimes

The widespread of AI technology crimes threatens the public trust in information systems. This leads to a phenomenon called "The liars dividend". This is common in the majors such as journalism, law, politics, where

certain individuals that committed a real-life crime with real audio-video proof, can evade the verdict, by counter-claiming that the real video is AI- generated. Furthermore, this leads to obstruction of criminal investigations in general.

4. Detection and Mitigation Strategies

4.1. Detection Strategies

The rapid advancement of deepfake generation techniques has intensified the need for robust detection mechanisms. Current detection approaches primarily rely on machine learning-based classifiers trained to identify visual, auditory, and temporal inconsistencies in synthetic media (Mirsky & Lee, 2021). These include convolutional neural networks (CNNs) designed to detect artifacts in facial expressions, eye blinking patterns, and head movements that differ from natural human behavior (Li et al., 2018).

In addition to visual cues, audio deepfake detection focuses on spectral anomalies and inconsistencies in speech patterns using deep learning models (Wang et al., 2020). Multimodal detection systems that combine audio and visual analysis have shown improved accuracy, particularly against high-quality deepfakes (Tolosana et al., 2020).

Despite these advances, detection systems face challenges due to the continuous improvement of generative models, which often outpace detection capabilities, leading to an ongoing arms race between generation and detection technologies (Verdoliva, 2020).

4.2. Mitigation Strategies

Mitigation strategies extend beyond technical detection and encompass legal, platform-level, and societal approaches. At the platform level, social media companies have begun integrating automated detection tools and content labeling mechanisms to reduce the spread of malicious deepfakes (Kietzmann et al., 2020). Watermarking and cryptographic provenance systems have also been proposed to verify the authenticity of digital content at the source (C2PA, 2022).

From a regulatory perspective, policymakers advocate for legal frameworks that criminalize malicious deepfake use while protecting legitimate applications such as satire and artistic expression (Westerlund, 2019). Educational initiatives aimed at improving media literacy are equally important, as informed users are better equipped to critically evaluate digital content and resist manipulation (Vaccari & Chadwick, 2020).

Collectively, effective mitigation requires a multidisciplinary approach that integrates technological innovation, legal regulation, and public awareness to address the societal risks posed by deepfake technology.

5. Regulations and Laws

5.1. Regulations and Laws in the EU

The primary goal of their regulation is to improve the internal market for reliable artificial intelligence, as well as to ensure the protection of privacy, safety, and health, while also protecting the environment and encouraging innovation.

Key provisions in the EU include:

- Harmonized rules related to the AI systems market in the EU, supporting special requirements for high-risk systems.
- Prohibited practices, which include various segments of bans such as manipulative practices, subliminal techniques, bans on exploiting vulnerabilities based on age or economic situation, and creating AI systems that manage or expand facial recognition databases through indiscriminate data collection.
- Provisions for high-risk AI systems introduce criteria for identifying them, which must meet strict compliance conditions, including data management, human oversight, and most importantly risk management.
- Transparency provisions requiring the availability of all information about the functioning of AI systems, so that users are aware they are interacting with artificial intelligence.
- Data protection provisions that ensure compliance with all EU data protection requirements when processing personal data.

- Liability provisions introducing monetary penalties for non-compliance depending on the severity of the violation.
- EU database, which represents a centralized registry for high-risk AI systems, for better compliance and oversight.

Unfortunately, alongside these provisions, many shortcomings have been observed regarding the risks posed by deepfake.

Regulation of deepfake content cannot fully keep up, and although it advocates for high transparency in the use of high-risk AI systems and their labeling, there is no detailed provision addressing risks related to identity theft and misinformation caused by deepfake technology. Privacy issues refer to data protection regulations, but there is still a risk of misuse by AI systems. Regulation includes informed consent, but does not cover cases of unauthorized data use, nor special protective measures against false representation. Liability for misuse is also a vulnerable aspect of these provisions, as obligations of deepfake technology providers are not sufficiently defined to ensure responsibility for harmful consequences of identity misuse or reputation damage. This shortcoming may lead to further legal ambiguities (Mauro Fragale and Valentina Grilli; 2024).

The conclusion based on the analysis and brief review of these provisions in the EU is that, although regulations represent strong foundations for managing AI systems, there are significant shortcomings in addressing specific challenges posed by deepfake technology.

5.2. Regulations and Laws in the USA

The DEEPFAKES Accountability Act in the USA primarily aims to protect national security from misuse of deepfake technology, while also providing legal protection for all victims of deepfake content (from sexual exploitation, political manipulation, or various frauds to which deepfake technology is prone) (<https://www.congress.gov/bill/118th-congress/house-bill/5586/text>, 2025).

Key measures of the law in the USA:

- Mandatory transparency: This measure applies to all deepfake content (videos, images, and audio content). It is necessary to indicate that the content is artificially generated or altered and also include metadata ensuring transparency of the content's origin (article).
- Criminal and civil penalties: Citizens face up to 5 years in prison for offenses such as election manipulation, sexual deepfake content, fraud, and identity theft. There are also fines of \$150,000 per deepfake content. Victims have the right to sue and seek damages as well as prohibition of further distribution (<https://www.congress.gov/bill/118th-congress/house-bill/5586/text>, 2025).
- Special victim protection: Possibility of anonymous court proceedings, with special coordinators in prosecutors' offices for various types of deepfake fraud.
- Obligations of technology companies: Companies must ensure their tools include metadata and labels, along with warnings about legal obligations. Online platforms must also contain systems for deepfake detection.
- National security: A Deepfakes Task Force was formed in the Department of Homeland Security, developing systems for deepfake detection, as well as an annual Congressional report on threats.

Shortcomings and problems with regulations and laws in the USA:

- Reliance on labels in deepfake content is unrealistic. It assumes malicious actors will voluntarily comply with the law. In practice, labels can easily be removed and content uploaded without metadata.
- Limited international application, as laws and regulations cannot control or punish foreign actors. Deepfake technology is global, while laws are national.
- Detection technology lags behind. Deepfake technology advances so quickly that detection tools cannot keep up, leading to major problems.
- Does not resolve psychological and reputational damage. Deepfake content can be removed, but reputational harm is already done and cannot be compensated (<https://www.congress.gov/bill/118th-congress/house-bill/5586/text>, 2025).

The conclusion for regulations and laws in the USA is that they are very ambitious as a first step in combating deepfake technology. However, the laws are more reactive than preventive, technologically lagging behind the speed of AI system development, and rely on goodwill of actors who are often malicious.

5.3. Overview of Laws and Regulations for Deepfake in Serbia

The article by Žunić Law (2025) describes Serbia and its laws, noting that it has no specific regulations for deepfake. Serbia refers to existing regulations such as the Criminal Code, copyright laws, and personal data protection. It emphasizes that these regulations are quite ineffective and insufficiently precise, as technology advances faster than laws (Žunić, 2025). PC Press describes how deepfake technology already has significant impact in cybercrime and fraud, representing the danger deepfake technology carries, especially if there are no adequate regulations to protect potential victims. Gecić Law (2025) explains that Serbia, due to EU accession, must align its regulations and laws with the EU. However, the EU already has introduced regulations, which Serbia is still considering how to implement (Gecić, 2025). As noted above, the EU has many shortcomings in its regulations that need improvement, but for Serbia, as for any other country, regulations are necessary.

Based on all analyses, it is concluded that uncontrolled development, not only of deepfake technology but also of all high-risk AI systems, can lead to major problems. Laws must be aligned with technological development; regulations must not be outdated, they must keep pace with technology and be effective.

6. Public Education and Media Literacy

Media literacy refers to the ability to critically analyze and evaluate media content (Hobbs, 2017). In the context of deepfake technologies, this allows individuals to recognize AI-generated audios, videos and manipulation in general.

Research suggests that even short educational interventions can significantly improve individual's ability to detect deepfake content (Guess et al., 2020). These interventions include: teaching how deepfakes are generated, identifying audio-visual inconsistencies, and verification of real-life videos.

Schools, and government play a crucial role in mitigation of damaging deepfake technologies. Integrating this kind of educational courses in formal education will help society to accommodate more efficiently when it comes to deepfake.

To conclude this chapter, this paper argues that public education and media literacy are one of the most effective methods for mitigating risks of deepfake, and implementing educational strategies will lead to long-term results in terms of public trust.

7. Positive Applications and Responsible Use of AI Deepfake Technology

7.1. Deepfakes in Film, Gaming, and Entertainment

Deepfake technology has demonstrated significant potential in the film, gaming, and entertainment industries by enabling advanced visual effects, realistic character synthesis, and digital restoration of audiovisual content. In film production, deepfake-based face replacement and de-aging techniques allow creators to maintain narrative continuity while reducing production costs and time (Verdoliva, 2020). Similarly, in the gaming industry, deepfake-driven character animation contributes to more immersive and emotionally expressive non-player characters, enhancing user engagement (Tolosana et al., 2020).

Moreover, the technology has been used for dubbing and localization purposes, where facial movements are synchronized with translated audio tracks, improving realism in multilingual content (Kietzmann et al., 2020). When applied with consent and transparency, these uses demonstrate how deepfakes can serve as a creative tool rather than a deceptive one.

7.2. Accessibility and Assistive Technologies

Beyond entertainment, deepfake technology offers promising applications in accessibility and assistive technologies. Synthetic speech and facial animation systems enable individuals with speech impairments to communicate more effectively through personalized voice reconstruction (Le et al., 2019). Additionally, facial reenactment techniques can support sign language translation and lip-reading systems, improving communication for people with hearing disabilities (Mittal et al., 2020).

Such applications emphasize the ethical dimension of deepfake technology, as they aim to enhance quality of life and social inclusion. However, these systems require strict safeguards regarding data consent and identity protection to prevent misuse.

7.3. Education and Training Simulations

In educational contexts, deepfakes can be utilized to create realistic training simulations, particularly in medicine, emergency response, and security training. AI-generated avatars can simulate real-world scenarios, enabling learners to engage in experiential learning without physical risk (Li et al., 2020). For example, virtual patients or simulated historical figures can be used to enhance engagement and retention in educational settings.

Nevertheless, educational deployment must clearly disclose synthetic content to avoid misinformation and maintain trust in learning environments (Floridi et al., 2018).

7.4. Preservation of Cultural Heritage

Deepfake and generative AI technologies also contribute to the preservation of cultural heritage by reconstructing damaged audiovisual materials and recreating historical figures for museums and digital archives. These techniques allow institutions to present interactive and immersive experiences while preserving fragile original artifacts (Güera & Delp, 2018).

When applied responsibly, such reconstructions serve educational and cultural purposes rather than historical distortion, reinforcing the importance of ethical guidelines in their implementation.

7.5. Guidelines for Ethical and Responsible Use

Responsible use of deepfake technology requires adherence to ethical principles such as transparency, consent, accountability, and harm prevention. Scholars emphasize the necessity of clearly labeling synthetic media and obtaining explicit consent from individuals whose likeness or voice is used (Floridi et al., 2018; Westerlund, 2019). Furthermore, organizations deploying deepfake systems must implement internal governance frameworks and comply with emerging regulatory standards to ensure ethical compliance.

8. Future Trends and Challenges of Deepfake Technology

As deepfake technology continues to evolve, its influence on digital communication, security, and trust becomes increasingly significant. While early discussions focused on technical feasibility, current debates emphasize future trends and the challenges associated with widespread adoption. This section examines the expected direction of deepfake development and the major obstacles it presents.

8.1. Future Trends in Deepfake Technology

This section highlights the future trends which are currently emerging and will emerge in the near future.

8.1.1. Real-Time and Multimodal Deepfakes

One of the most significant trends is the rise of real-time deepfake systems (Stanford University 2024. AI Index Report). These systems enable live manipulation of facial expressions and voices, increasing the potential for misuse in video conferencing and social engineering attacks. Additionally, multimodal deepfakes combine video, audio, and text, creating more convincing synthetic identities.

8.1.2. Increased Accessibility

Deepfake tools are becoming more user-friendly and widely available. Mobile applications and cloud-based platforms allow users with minimal technical knowledge to generate synthetic media (Stanford University 2024. AI Index Report). While this democratization supports creative and educational uses, it also lowers the barrier for malicious actors.

8.1.3. Legitimate Applications

Deepfakes are increasingly used in legitimate contexts, such as film production, video game development, digital assistants, and accessibility tools (Stanford University 2024. AI Index Report). For example, synthetic voices can help individuals who have lost their ability to speak, while digital avatars can preserve cultural heritage or historical figures.

8.2. Challenges Associated with Deepfakes

This section shows challenges which are associated with the misuse of deepfakes, and underlines the mitigation strategies.

8.2.1. Misinformation and Political Manipulation

One of the most serious challenges posed by deepfakes is their potential to spread misinformation (Chesney, R.; Citron, D. 2019). Manipulated videos of public figures can undermine democratic processes, erode public trust, and amplify political polarization. As realism improves, distinguishing authentic content from fake becomes more difficult.

8.2.2. Privacy and Identity Abuse

Deepfake technology raises major concerns regarding consent and identity theft. Individuals can have their likeness used without permission, often resulting in reputational damage or psychological harm. This issue is particularly severe in cases involving non-consensual explicit content.

8.2.3. Detection and the Technological Arms Race

As generation techniques improve, detection methods must evolve accordingly. Researchers describe this as an ongoing arms race between deepfake creators and detection systems (Verdoliva, L. 2020). While AI-based detectors exist, none provide a permanent solution, especially against adaptive models.

8.2.4. Legal and Ethical Regulation

Regulating deepfakes remains a global challenge. Laws often lag behind technological innovation, and enforcement varies across jurisdictions. Initiatives such as the European Union's AI Act represent early efforts to address synthetic media, but comprehensive international standards are still lacking (European Commission 2023).

To conclude this chapter, the future of deepfake technology is characterized by both innovation and risk. While advancements promise valuable applications across multiple industries, unresolved challenges related to misinformation, privacy, and regulation threaten social trust. Addressing these challenges will require collaboration between technologists, policymakers, and educators to ensure responsible use.

9. Conclusion

The development of artificial intelligence has led to major technological advancements but has simultaneously created new ethical and security challenges. Its historical development shows rapid progress. While the use of AI has led to scientific advancements in the fields of medicine, education etc. it has also been misused for criminal activities such as non-consensual pornography, identity theft and manipulation. Furthermore, these risks have led to the awareness that deepfake security needs to be intact for future generations. Amongst many strategies, few have proven to be effective: education and media literacy, as well as the litigation strategies. When it comes to the future, data indicates that deepfake technologies will be further integrated in everyday life. For its proper realization ethical boundaries, legal regulations and public awareness are essential in ensuring that AI deepfake technologies will be used for the better of society.

Literature:

1. Andrew M. Guess, Michael Lerner, Benjamin Lyons. A digital media literacy intervention increases discernment between mainstream and false news in the United States and India 22 of June 2020. Retrieved December 2025.
2. Anetta Jedličková. Ethical approaches in designing autonomous and intelligent systems: a comprehensive survey towards responsible development August 2024. Retrieved November 2025.
3. Alena Birrer, Natascha Just. What we know and don't know about deepfakes: An investigation into the state of the research and regulatory landscape 22 of May 2024. Retrieved November 2025.

4. Bo Nørregaard Jørgensen, Saraswathy Shamini Gunasekaran, Zheng Grace Ma. Impact of EU Laws on AI Adoption in Smart Grids: A Review of Regulatory Barriers, Technological Challenges, and Stakeholder Benefits June 2025. Retrieved October 2025.
5. C2PA. Releases Specification of World's First Industry Standard for Content Provenance January 2022. Retrieved December 2025.
6. Cristian Vaccari, Andrew Chadwick. Deepfakes and Disinformation: Exploring the Impact of Synthetic Political Video on Deception, Uncertainty, and Trust in News February 2020. Retrieved December 2025.
7. Citron, D. K., & Chesney, R. Nonconsensual deepfakes and sexual privacy. *California Law Review*, 112(1), 1–54. 2024. Retrieved November 2025.
8. Danielle K. Citron, Robert Chesney. Deepfakes and the new disinformation war. *Foreign Affairs* 2019. Retrieved November 2025.
9. Europol. Facing reality? Law enforcement and the challenge of deepfakes. Europol Innovation Lab. 2022.
10. European Commission. AI Act and synthetic media regulation. 2023.
11. Goodfellow, I., et al. Generative Adversarial Networks. *NeurIPS* 2014.
12. Gecić Law. Da li će veštačka inteligencija biti pogubna za autorska prava? April 2023. Retrieved December 2025.
13. Gabe Regan VP of Human Engagement; Reality Defender December 2025.
14. Güera, D., & Delp, E. J. (2018). Deepfake video detection using recurrent neural networks. *IEEE AVSS*.
15. Hanan Sharif, Amara Atif, Arfan Ali Nagra. Deepfake-Style AI Tutors in Higher Education: A Mixed-Methods Review and Governance Framework for Sustainable Digital Education
16. Floridi, L., Cowls, J., Beltrametti, M., et al. AI4People—An ethical framework for a good AI society. *Minds and Machines*, 28(4), 689–707. 2018.
17. Jana Kazaz. Regulating Deepfakes: Global Approaches to Combatting AI-Driven Manipulation Decemer 2024. Retrieved November 2025.
18. Jennifer Gregory IBM. Face Card ace Card Declined: The Deepfake d Declined: The Deepfake Threat to Biometric Security in o Biometric Security in Financial Systems November 2025. Retrieved Deceber 2025.
19. Katalin Parti, Judit Szabó. The Legal Challenges of Realistic and AI-Driven Child Sexual Abuse Material: Regulatory and Enforcement Perspectives in Europe 30 october 2024. Retrieved october 2025.
20. Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. Deepfakes: Trick or treat? *Business Horizons*, 63(2), 135–146. 2020. Retrieved December 2025.
21. Le, Q. V., et al. Speech synthesis for assistive communication. *IEEE Signal Processing Magazine* 2019.
22. Li, Y., Chang, M. C., Lyu, S. In Ictu Oculi: Exposing AI created fake videos. *IEEE ICASSP* 2018.
23. Mauro Fragale and Valentina Grilli, *Columbia Journal of European Law*; 2024. Retrieved December 2025.
24. Mika Westerlund. The Emergence of Deepfake Technology: A Review November 2019. Retrieved December 2025.
25. Mirsky, Y., & Lee, W. The creation and detection of deepfakes. *ACM Computing Surveys*, 54(1). 2021.
26. Online links:
27. <https://www.congress.gov/bill/118th-congress/house-bill/5586/text>, Retrieved December 2025
28. <https://www.realitydefender.com/insights/history-of-deepfakes>, Gabe Regan 2025. Retrievedd December 2025.
29. Paz Sandoval, Almeida Vau, Solaas, Rodrigues. Threat of deepfakes to the criminal justice system: a systematic review november 2024, Retrieved November 2025.
30. Renee Hobbs. Approaches to Teacher Professional Development in Digital and Media Literacy Education. pg 28, 2017. Retrieved December 2025.
31. Stanford University. AI Index Report. 2024.
32. Tolosana, R., et al. 2020. Deepfakes and beyond: A survey of face manipulation and fake detection. *Information Fusion*, 64, 131–148. 2020.
33. Verdoliva, L. Media forensics and deepfakes. *IEEE Journal of Selected Topics in Signal Processing*. 2020.
34. Vaccari, C., & Chadwick, A. Deepfakes and disinformation. *Social Media + Society*. 2020.

35. Žunjić Law. Zunic Law proglašena za kancelariju godine u 2025. 28th of November 2025. Retrieved December 2025.
36. <https://www.thalesgroup.com/en/news-centre/press-releases/software-republique-unveils-vision-4rescue-integrated-technological> (Accessed 05.12.2025.);
37. <https://www.wri.org/insights/cop30-outcomes-next-steps> (Accessed 04.12.2025.);
38. <https://www.besnet.world/national-ecosystem-assessment/> (Accessed 04.12.2025.);
39. <https://rcel.rs/sr/cop29-kljucni-zakljucci/> (Accessed 08.12.2025.);
40. <https://pocketproject.org/> (Accessed 09.12.2025.);
41. <https://www.unoosa.org/oosa/en/ourwork/spacelaw/index.html> (Accessed 13.12.2025.).