International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 11 November 2025
Page No: 4070-4078
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0596

Deepfake Technology: A Innovation and Threat

Dr Pushparani MK[1], Shrilaxmi Bhat[2], Shreedhanya B[3], Pavithra[4], Maibam Yoihenba Meitei[5]
[1]Associate professor, Dept. of CSD, Alvas Institute of Engg. & Tech., Mijar, Karnataka, India.
[2,3,4,5]UG Scholar, Dept. of CSD, Alvas Institute of Engg. & Tech., Mijar, Karnataka, India.
Emails: *drpushparani@aiet.org.in[1], shrilaxmibhat16@gmail.com[2], shreedhanyab16@gmail.com[3], pavithradevraj733@gmail.com[4], maibamyoihenbaba25@gmail.com[5]*

Abstract
Deepfake technology, a product of sophisticated artificial intelligence and machine learning algorithms, has profoundly altered the landscape of digital media. Its emergence is characterized by a fundamental duality: it presents as both a groundbreaking technological innovation and a potent societal threat. This review paper provides a comprehensive analysis of this complex technology, delving into its core generative mechanisms, its wide-ranging applications, and the significant challenges it poses to cybersecurity, ethics, and democratic institutions. The analysis explores the ethical and beneficial uses of deepfakes in sectors such as healthcare, education, and entertainment, while simultaneously detailing their malicious applications in financial fraud, political disinformation, and the creation of non-consensual explicit content. A critical examination of the ongoing "arms race" between deepfake generation and detection reveals the inherent difficulties in developing effective countermeasures, exacerbated by a fundamental asymmetry in the cost and speed of creation versus detection. The paper further scrutinizes the limitations of existing legal frameworks and the nascent, fragmented global regulatory responses. This study concludes that while deepfakes offer genuine promise as a creative tool, their current and most widespread use as a weapon for deception and manipulation positions them as an urgent and systemic threat to verifiable reality and public trust.

Keywords: *Deepfake; Generative AI; Disinformation; Ethical Innovation; Cybersecurity.*

1. Introduction

1.1. Background and Context

The rapid evolution of artificial intelligence (AI) has ushered in a new era of digital media manipulation, with deepfake technology standing at the forefront of this transformation. A deepfake is a form of synthetic media, including video, photo, or audio, that has been altered using deep learning algorithms to appear deceptively authentic. The term, a portmanteau of "deep learning" and "fake," first entered the public lexicon in late 2017 with the emergence of non-consensual pornographic videos. Since its inception, deepfake technology has progressed from a niche phenomenon to a widely accessible tool, capable of producing hyper-realistic content that can mislead and deceive audiences on a mass scale. This technological advancement embodies a profound duality: on one hand, it represents a remarkable feat of innovation with applications that can benefit humanity; on the other, it poses a direct and immediate threat to public trust, personal security, and the integrity of information itself. The proliferation of deepfakes is closely tied to the "post-truth" era, a period characterized by widespread digital disinformation and information warfare. In a social media environment where false information can go viral in seconds, the ability to create convincing forgeries of people saying or doing things they never did compounds the difficulty of distinguishing between fact and fiction. This challenge is no longer confined to the domain of celebrity scandals or political parodies; it has evolved into a precision weapon for financial fraud and targeted attacks on corporate and democratic institutions [1].

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 11 November 2025
Page No: 4070-4078
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0596

1.2. Purpose and Scope of the Review

The purpose of this paper is to provide a comprehensive, multi-disciplinary review of deepfake technology, critically analyzing its dual nature as both a groundbreaking innovation and a potent societal threat. This review will navigate the complex landscape of deepfakes by examining its fundamental technical mechanisms, exploring its diverse beneficial applications, and detailing its malicious uses across various sectors. The analysis will also investigate the technological countermeasures developed for detection and the inherent challenges that hinder their effectiveness. Furthermore, the report will scrutinize the complex ethical dilemmas and the evolving legal and regulatory frameworks designed to mitigate deepfake-related harms. By synthesizing a range of sources, from academic papers and technical reports to industry insights and legal analyses, this study aims to offer a nuanced perspective on a technology that is reshaping the very fabric of digital communication and trust.

1.3. Originality and Contribution

This review distinguishes itself from a simple literature survey by offering a cohesive and deeply analytical framework that integrates the technical, ethical, and legal dimensions of deepfake technology. It provides a focused analysis of the most salient background information required for a comprehensive understanding of the subject. The paper addresses key questions concerning the methods of deepfake creation and detection, the overarching challenges, and the potential applications. It contributes to the contemporary discourse by demonstrating that the technology itself is not inherently good or evil, but rather its morality is determined by its application. This timely work emphasizes the urgency of a coordinated response to a threat that is rapidly outstripping society's ability to contain it.

1.4. Paper Structure

The subsequent sections of this report are structured to provide a clear and logical progression. Section 2 establishes the theoretical foundation and the core mechanisms of deepfake generation. Section 3 explores the ethical and beneficial applications of the technology. Section 4 provides a detailed account of the malicious uses and societal risks. Section 5 analyzes the current state of deepfake detection and the challenges facing it. Section 6 addresses the ethical and legal imperatives, including a discussion of existing and proposed legislation. Finally, Section 7 presents a synthesis of the findings and offers a conclusion on the duality of deepfake technology, along with recommendations for future work [2].

2. Theoretical Foundation and Generation Mechanisms

2.1. The Genesis of Deepfakes

Deepfake technology, as a category of synthetic media, is characterized by its use of advanced AI to create convincing forgeries of images, videos, and audio. The technology is broadly classified into two primary categories: deepface and deepvoice. Deepface involves the superimposition of one person's face onto another, with the goal of falsifying facial gestures and expressions. Deepvoice, on the other hand, is used to clone or imitate an individual's original voice from audio fragments, allowing for the creation of new, original sentences or entire speeches. The emergence of these technologies in late 2017 marked a turning point in digital manipulation, moving beyond simple editing to the creation of entirely new, and often untraceable, realities.

2.2. Core Generative Models

The creation of deepfakes relies on artificial neural networks, computer systems loosely modeled on the human brain that excel at pattern recognition. Two primary deep learning architectures are most notably used for this purpose: Generative Adversarial Networks (GANs) and autoencoders [3].

2.2.1. Generative Adversarial Networks (GANs)

GANs are a powerful class of generative AI composed of two competing neural networks: a generator and a discriminator. The generator's role is to create fake content, such as a synthetic image or video, from a random noise input. The discriminator's role is to act as a critic, attempting to distinguish between the real data and the content produced by the generator. This competition is an iterative process: as the generator improves at

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 11 November 2025
Page No: 4070-4078
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0596

creating more plausible forgeries, the discriminator becomes more adept at identifying subtle imperfections. This continuous, adversarial training cycle results in a feedback loop that drives the generator to produce increasingly convincing, lifelike content. While GANs generally produce more convincing deepfakes, they are also more difficult to use than other methods.

2.2.2. Autoencoders

Autoencoders are a type of neural network trained to reconstruct input from a simpler, compressed representation. In the context of deepfake creation, particularly face-swapping, two autoencoders are employed. A single encoder is trained on a vast number of images from two different subjects to learn the essential features of each face. This shared encoder creates a compressed "sketch" of a face. Two separate decoders are then used, one for each subject, to learn how to reconstruct a face from that sketch. To create a deepfake, the encoder is fed an image of the source subject, and its compressed representation is then fed into the decoder of the target subject. The result is a new image of the target subject's face mimicking the facial expressions and gestures of the source subject Shown in Figure 1.

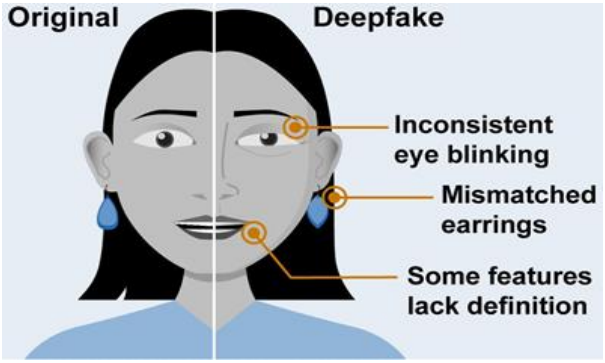

Figure 1 Key Visual Cues for Detecting Deepfake Manipulation

2.3. The Deepfake Ecosystem

The barrier to entry for deepfake creation has fallen dramatically. What once required a high degree of technical skill and significant computational power is now democratized through openly available computer applications and tutorials. The creation of a convincing deepfake, while still requiring a substantial number of training images of the faces to be swapped, is now achievable by anyone with basic computer skills and a home computer. The rapid evolution of these tools has lowered the bar even further, with convincing video deepfakes now creatable in as little as 45 minutes, and voice cloning requiring just 20 to 30 seconds of audio. B This widespread availability of user-friendly software has a profound consequence for the threat landscape. The malicious potential of this technology is no longer limited to state-sponsored actors or highly skilled cybercriminals. The accessibility of these tools allows a much wider array of individuals, from hobbyists and disgruntled employees to social agitators, to create and disseminate highly deceptive content. This pervasive access to powerful generative technology fundamentally changes the scale and scope of the threat, making it exponentially more difficult for society and law enforcement to contain. The threat is not just a high-profile attack but a constant, decentralized deluge of forged information [4].

3. The Innovation Frontier: Ethical And Beneficial Applications

Despite their notoriety for misuse, deepfakes hold significant potential as a tool for ethical innovation, particularly in creative and highly technical domains. The same core generative AI technologies used to deceive can also be harnessed to create compelling, beneficial, and ethically sound applications.

3.1. Revolutionizing Entertainment and Media

In the entertainment industry, deepfake technology offers a range of possibilities for enhanced visual effects, deepening audience immersion in movies, television shows, and video games. It can be used to achieve creative effects or maintain story continuity when an actor is unavailable. For example, a deceased actor's likeness could be digitally recreated to complete a film, or their appearance could be modified to portray them at a different age, a practice known as "digital de-aging". This technology, when used with transparency and consent, allows for new forms of storytelling and preservation of artistic legacies.

3.2. Advancing Healthcare and Medicine

The applications of deepfakes in healthcare are particularly promising. Generative models can be

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 11 November 2025
Page No: 4070-4078
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0596

used to create realistic simulations of patient cases for medical training, allowing students and professionals to enhance their diagnostic skills and bedside manner in a controlled environment. These simulations can imitate a variety of medical histories and patient personalities, providing a rich training ground for high-stakes decision-making. Beyond training, the technology can be leveraged for therapeutic and diagnostic purposes. Deepfakes can be used in personalized therapy, such as virtual exposure therapy, which offers patients with PTSD, phobias, or anxiety disorders a controlled environment to confront their fears. They can also be applied to medical imaging and radiodiagnostic techniques to create more accurate diagnoses in oncologic pathologies. Another ethical application involves anonymizing video records of patients by altering their facial features while preserving valuable expressions for diagnostic purposes, thereby ensuring privacy while facilitating deeper research [5].

3.3. Enhancing Education and Accessibility

The potential of deepfakes to enhance education and accessibility is considerable. The technology can be used to create educational videos and simulations that are more engaging and interactive for students. For individuals with disabilities, deepfakes can be transformative. They can be used to generate sign language interpretations for videos, ensuring equal access for deaf or hard-of-hearing users. Additionally, the technology can provide natural-sounding voices for communication devices, offering individuals with speech disabilities a more expressive means of verbal communication. It can also generate audio interpretations of visual content for those with visual impairments. This capacity to adapt and personalize content holds the potential to make education and media more functional and accessible for a wider audience. The applications described above demonstrate a fundamental paradox of deepfake technology. The very same generative models (GANs and autoencoders) that are used for the most insidious and damaging purposes are also the engines driving profound positive change in medicine, education, and entertainment. The technology itself is agnostic; its designation as an "innovation" or "threat" is determined entirely by the intent of its user. This paradox underscores a critical point: a complete ban on the technology would stifle significant advancements that could benefit humanity, compelling the need for a nuanced, harm-focused regulatory approach that balances the potential for good with the imperative to prevent misuse

4. The Malicious Threat: Exploitative Uses and Societal Risks

While the innovative applications of deepfakes are compelling, their current and most widespread use is for malicious purposes. The technology has become a powerful weapon for deception and manipulation, posing a grave and urgent threat to individuals, organizations, and the foundational pillars of society

4.1. Financial Fraud and Corporate Deception

Deepfakes have evolved beyond being tools for electoral manipulation and celebrity scandals into precision weapons for corporate fraud. Cybercriminals are increasingly using deepfakes to impersonate senior executives, an attack vector sometimes referred to as "vishing" or "voice phishing". This type of social engineering exploits the human element of trust that underpins many financial transactions and internal communications. A striking example of this threat is the $25.5 million fraud case involving a Hong Kong-based multinational firm. A finance worker was deceived into authorizing 15 transfers to fraudsters who used deepfake technology to pose as the company's chief financial officer and other colleagues on a video call. The worker, initially suspicious, was convinced by the realistic appearance and voices of the deepfake counterparts. This incident, which occurred in January 2024, signals a fundamental shift in how AI threatens business operations through executive impersonation. Similar attempts have targeted high-profile executives, with fraudsters using AI-cloned voices to replicate a victim's accent and mannerisms with unnerving accuracy. The scale of this threat is staggering, with deepfake fraud cases in North America surging by 1,740% between 2022 and 2023 Shown in Figure 2 Accuracy Comparison of Deepfake Detection Models [6].

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 11 November 2025
Page No: 4070-4078
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0596

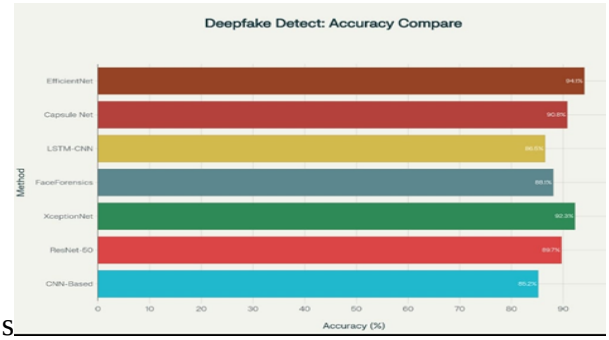

Figure 2 Accuracy Comparison of Deepfake Detection Models

4.2. Political and Social Disinformation

Deepfakes are a powerful tool for spreading misinformation and propaganda. By creating fake videos of public figures making false statements, malicious actors can manipulate public opinion, disrupt elections, or incite civil unrest. For example, deepfake videos featuring world leaders such as Donald Trump, Barack Obama, and Volodymyr Zelenskyy have been used to spread false narratives and undermine credibility. The realism of these forgeries makes it increasingly difficult for people to distinguish between legitimate news and fabricated content, thereby amplifying the "fake news" phenomenon and eroding trust in media Shown in Figure 3.

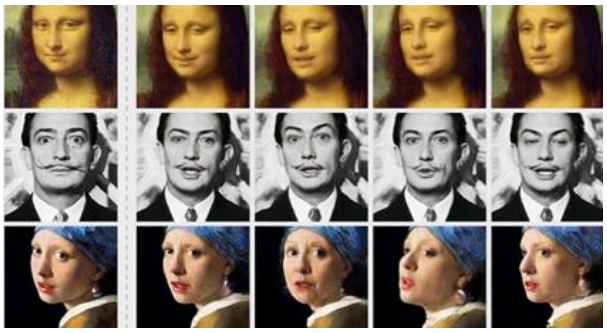

Figure 3 Sequence of Facial Transformations Generated by Deepfake Models

social media platforms, constitutes a gross violation of privacy and a tool for harassment and exploitation, causing severe emotional and psychological harm to victims. The high-profile case of non-consensual deepfake images of Taylor Swift, which sparked outrage and furthered the conversation about digital rights, demonstrates the profound violation this technology inflicts on individuals [7].

4.3. Erosion of Trust and Reality Apathy

Beyond the specific harms of fraud and defamation, the most profound threat posed by deepfake technology is its systemic effect on public trust. When deepfakes make it nearly impossible to differentiate between Shown in Figure 4.

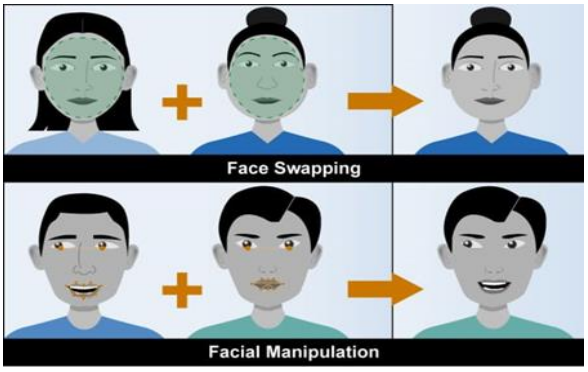

Figure 4 Illustration of Face Swapping and Facial Manipulation Techniques

4.4. Non-Consensual Explicit Content

The most pervasive and damaging use of deepfake technology is the creation of non-consensual explicit content. Studies and reports indicate that this application accounts for a staggering majority—up to 96%—of all deepfakes on the internet. This practice disproportionately victimizes women and celebrities, using their faces to create explicit material without their consent or knowledge. The widespread dissemination of this content, often on authentic and fabricated content, the credibility of legitimate news sources and media is undermined. This creates a climate of pervasive doubt where the public is increasingly unsure what to believe. This environment of distrust leads to a more insidious outcome: a phenomenon referred to as "reality apathy." In an information ecosystem saturated with forgeries, the constant bombardment of manipulated content can lead people to feel that much of the information they consume, including video and audio, simply cannot be trusted. This can result in a dangerous form of desensitization, where individuals may even dismiss genuine, verifiable footage as fake simply because it contradicts their existing beliefs or

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 11 November 2025
Page No: 4070-4078
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0596

because they have become entrenched in the notion that everything is potentially deceptive. The greatest threat is not that people will be deceived by a single deepfake, but that they will come to regard everything as deception, fundamentally eroding the shared factual basis required for a functioning democratic society.

5. The Detection Arms Race: Countermeasures and Challenges

The rapid advancements in deepfake generation have necessitated the parallel development of sophisticated detection technologies. This struggle has become an ongoing "arms race" between creators and detectors, with each side continuously evolving in response to the other's innovations. Deepfake detection is fundamentally a challenge in computer vision and signal analysis.

5.1. Technological Approaches to Detection

The primary objective of deepfake detection is to identify subtle, artificial patterns that betray the manipulation of digital media. These methods often leverage deep learning models trained to spot inconsistencies that are nearly imperceptible to the human eye [8].

5.1.1. Computer Vision and Deep Learning

Deep learning models, particularly Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), are the workhorses of deepfake detection. CNNs are highly effective at automatically learning spatial features and are trained to identify visual artifacts such as inconsistent lighting, unnatural blurring, or discrepancies in skin texture at the pixel level. They can also detect inconsistencies in facial features that AI generation often struggles to replicate accurately, such as unnatural blinking patterns or mismatched facial landmarks. RNNs, often used in conjunction with CNNs, are designed to analyze temporal patterns across video frames, enabling them to spot inconsistencies in motion or flickering that occur over time.

5.1.2. Biometric and Audio Analysis

Beyond visual artifacts, detection systems can analyze subtle biological and physiological signals. Biometric analysis focuses on unique physiological characteristics that are difficult to replicate, such as eye movements, micro-expressions, and blinking patterns. For example, Intel's FakeCatcher uses a technique called Photoplethysmography (PPG) to detect subtle changes in blood flow from video pixels, a biological signal that is nearly impossible for current deepfake generators to imitate convincingly. Similarly, AI-based audio analysis models can distinguish between natural and synthetic voices by analyzing factors such as pitch, cadence, and breath patterns. These models perform spectral analysis to identify inconsistencies in the frequency and amplitude of the audio signal, which are often present in deepfake recordings.

5.2. A Comparison of Deepfake Detection Tools

The commercial and academic landscape for deepfake detection is populated with a variety of specialized tools, each employing unique technological approaches. The following table provides a comparative overview of some of the leading solutions Shown in Table 1.

5.3. The Asymmetric Challenge

Despite the development of these advanced detection tools, the "arms race" remains fundamentally asymmetrical. Deepfake generation technology is evolving at an accelerating rate, with the number of deepfake videos increasing by 900% annually. Detection capabilities consistently lag behind. The challenge is compounded by the "generalization gap"—a detector model trained to spot visual patterns from one AI generator may fail when confronted with fakes from a new, unseen generator. This means that any detection solution is, by its very nature, temporary. Moreover, while automated systems are improving, they still experience significant accuracy drops when faced with "in-the-wild" deepfakes compared to laboratory conditions. Even more concerning is the human inability to reliably identify deepfakes, with accuracy hovering at just 55-60%. The asymmetry lies in the fact that it is far cheaper and easier for malicious actors to create a new type of deepfake than it is for the defensive community to develop, train, and deploy a robust detection system that can recognize it. This perpetual race requires continuous investment in research and development and underscores the urgency of creating more generalizable, real-time detection solutions that

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 11 November 2025
Page No: 4070-4078
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0596

can operate effectively outside of controlled laboratory settings.

Table 1 Leading Deepfake Detection Solutions

Tool/Platform	Core Technology	Key Features	Stated Accuracy	Use Case/Target Audience
OpenAI's Deepfake Detector	Deep Learning (CNNs)	Identifies AI-generated images produced by DALL-E 3.	98.8% for DALL-E 3 images	Content creators, researchers
Hive AI	Deep Learning API	Content	High accuracy	Digital platforms,
Tool/Platform	Core Technology	Key Features	Stated Accuracy	Use Case/Target Audience
		moderation API, classifies content as "yes_deepfake" or "no_deepfake" with a confidence score.		social media, content moderation
Intel's FakeCatcher	Photoplethys mogr aphy (PPG)	Real-time deepfake detection by analyzing biological signals (blood flow).	96% (controlled conditions)	Real-time video authentication, security, intelligence
Sensity AI	Multimodal Deep Learning	Detects deepfakes in videos, images, audio, and identities at scale. Real-time monitoring of over 9,000 sources.	95-98% accuracy	Cybersecurity firms, law enforcement, government agencies, digital forensics
FaceForensics++	Public Dataset	A widely used benchmark dataset for training and evaluating deepfake detection models.	Not a tool, but a foundational resource	Academic researchers, developers

6. Legal, Ethical, And Regulatory Imperatives

The emergence of deepfakes has exposed significant gaps in existing legal and ethical frameworks. The profound harms caused by the technology demand a reassessment of how society balances innovation with the protection of individuals and democratic institutions.

6.1. The Ethical Conundrum

The ethical debate surrounding deepfakes can be analyzed through various philosophical frameworks. From a consequentialist perspective, which assesses the morality of an action based on its outcomes, deepfakes are largely unethical. The adverse consequences, such as disinformation, financial loss, psychological harm, and the erosion of trust, far outweigh the beneficial outcomes in entertainment or medicine. From a deontological perspective, which focuses on the inherent morality of an action regardless of its consequences, the issue of deception is central. A deepfake, by its very nature, distorts reality. A deepfake created with the intention of deceiving or spreading falsehoods is morally suspect. Under this view, a deepfake is only ethically acceptable when its deceptive qualities are removed through transparency and consent, ensuring all involved parties are aware that the media is not real.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 11 November 2025
Page No: 4070-4078
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0596

This ethical complexity challenges the notion of technology's "value neutrality," the argument that a given technology is neither good nor bad on its own. While a tool like an ax can be used to build a home or commit a crime, the inherent nature of a deepfake is to create a forgery of a person or event. This places a unique ethical burden on its creators, particularly engineers, who are often bound by codes of ethics to prioritize the safety and welfare of the public. The unpredictability of a deepfake's misuse complicates assigning full accountability to its creators, yet the lack of transparency about its deceptive nature is a core ethical failing.

6.2. Limitations of Existing Legal Frameworks

Current legal systems, which were not designed for the unique harms of synthetic media, often prove insufficient in addressing deepfake-related abuses. Traditional legal frameworks, such as defamation, libel, and privacy laws, face significant limitations. For a defamation claim to succeed, proving intent to cause harm can be difficult, and the law does not always cover the emotional distress or broader societal impact. Similarly, while copyright laws may apply if a deepfake uses copyrighted material, they do not address the core harm of misrepresentation or the violation of a person's likeness. The difficulty in proving direct, measurable harm from misinformation or psychological distress makes traditional legal recourse a "tall order" for victims.

Conclusion

Summary of Findings

The analysis presented in this paper confirms that deepfake technology embodies a powerful and paradoxical duality. On one hand, it represents a remarkable technological innovation with the capacity to revolutionize fields as diverse as entertainment, healthcare, and education. The ability to create realistic patient simulations for medical training or provide natural-sounding voices for individuals with speech disabilities demonstrates the profound ethical and beneficial potential of this technology. On the other hand, deepfakes have been weaponized for malicious purposes on an unprecedented scale. Their low barrier to entry and the sophistication of modern generative models have enabled a surge in financial fraud, political disinformation, and, most prevalently, the creation of non-consensual explicit content. The most significant consequence of their proliferation is not a single act of deception but the systemic erosion of public trust in digital media, leading to a dangerous state of "reality apathy". The "arms race" between deepfake generation and detection is fundamentally asymmetrical, with generation capabilities consistently outpacing countermeasures. The human inability to reliably spot forgeries and the technological "generalization gap" render current defensive strategies insufficient. Furthermore, the lack of a coordinated and comprehensive legal framework, both domestically and internationally, leaves individuals and institutions vulnerable to harm. Therefore, while deepfakes undeniably represent a technological innovation, their weaponization and the current inadequacy of societal and legal countermeasures position them as an urgent and significant threat.

Recommendations and Future Work

The findings of this review necessitate a multi-faceted and proactive response. For the academic and research community, the critical need is to develop more robust and generalizable deepfake detection models that can effectively function in "in-the-wild" scenarios. This requires the creation of standardized datasets that accurately reflect the complexities of modern deepfakes, which will in turn facilitate improved cross-study comparisons and reproducibility. For policymakers, the urgent task is to move beyond a reactive, harm-by-harm approach to a more coordinated and proactive regulatory framework. Future legislation must balance the protection of innovation with clear, enforceable rules on transparency and accountability, particularly for synthetically generated content that is intended to deceive or harm. Ultimately, the fight against deepfakes is not solely a technical challenge but a societal one. The most enduring defense will require a fundamental shift toward greater media literacy and critical thinking among the public. By fostering a culture of verification and healthy skepticism toward digital media, society can build a collective resilience to the insidious threat of deepfakes and protect the integrity of information and truth.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 03 Issue: 11 November 2025
Page No: 4070-4078
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2025.0596

Acknowledgements

Acknowledgements, including information on the source of any financial support received for the work being published, go here as per the template

References

[1]. The Emergence of Deepfake Technology: A Review — Mika Westerlund (Technology Innovation Management Review, Vol. 9, No. 11, November 2019)

[2]. Deep Insights of Deepfake Technology: A Review — Bahar Uddin Mahmud; Afsana Sharmin (DUJASE Vol. 5, Issues 1 & 2, 2020; originally on arXiv May 2021, updated Jan 2023)

[3]. A Review of Deepfake Technology: An Emerging AI Threat — Mridul Sharma; Mandeep Kaur (in Advances in Intelligent Systems and Computing, Springer Nature, October 2021)

[4]. Deepfake: Definitions, Performance Metrics and Standards, Datasets and Benchmarks, and a Meta-Review — Enes Altuncu; Virginia N. L. Franqueira; Shujun Li (arXiv, August 2022)

[5]. Deepfakes: current and future trends — (no authors listed in snippet) (Artificial Intelligence Review, vol. 57, article 64, February 2024)

[6]. From deepfake to deep useful: risks and opportunities through a systematic literature review — Nikolaos Misirlis; Harris Bin Munawar (arXiv / manuscript, November 2023)

[7]. Advancements in Detecting Deepfakes: AI Algorithms and Future Prospects – a review — Laishram Hemanta Singh; Panem Charanarur; Naveen Kumar Chaudhary (Discover Internet of Things, May 7 2025)

[8]. The Deepfake Conundrum: Assessing Generative AI's Threat to Digital Reality and Proposing a Multi-Layered Defense Framework — Ketan Modi (International Journal of Computer Trends and Technology, vol. 73, no. 6, pp. 97–103, 2025)