Journal of Media Horizons
ISSN: 2710-4060 2710-4052 Volume 6, Issue 3, 2025

THE ETHICS OF DEEPFAKE TECHNOLOGY IN MEDIA: A REVIEW STUDY

Misbah Ilyas[*1], Dr. Shazia Hashmat[2], Faiza Bibi[3]

[*1,3]*MPhil Student, Communication and Media Studies Department, Fatima Jinnah Women University, Rawalpindi, Pakistan*
[2]*Assistant Professor, Communication and Media Studies Department, Fatima Jinnah Women University, Rawalpindi, Pakistan*

DOI: https://doi.org/10.5281/zenodo.15854030

Keywords
Deep Fakes. Ethical. Technology. Advancement. Synthesize. Government

Article History
Received: 03 April, 2025
Accepted: 25 June, 2025
Published: 10 July, 2025

Corresponding Author: *
Misbah Ilyas

Abstract
This study explores the ethical implications of the deep fake technology in media. Technological advancements have led to the rise of artificial intelligence (AI). With the development of AI, deep fake also reached heights offering advantages and harm in many ways. AI has emerged as a significant problem in the media. AI is used in making fabricated content which looks so realistic that cannot be identified easily. The methodology of this study is based on the review of the literature following the inclusive and exclusive criteria and guided by the PRISMA method. This research synthesizes findings from a total of 22 studies published between 2018 and 2024 were analysed. Deep Fakes are content that appears realistic and is difficult to distinguish from authentic media. These studies came from Google Scholar and Sci Space. The results showed that deepfake are being used to mislead people, harm reputations, and create fake news. There are also not enough laws or tools to stop the harmful use of this technology. This paper after the selected studies suggests that governments, media companies, and online platforms should work together to make stronger rules, teach people how to spot fake content, and create better ways to detect deep fake. The goal is to reduce the risks while still allowing the good uses of this technology.

INTRODUCTION

Deepfake technology is a type of synthetic media generated through artificial intelligence (AI), has emerged as a critical concern in the media industry due to its ability to create highly convincing but entirely fabricated content, including films, photos, and audio recordings. These AI-generated alterations are so realistic that they blur the lines between fact and fiction, making it increasingly difficult for audiences to differentiate between what is real and what has been manipulated (Westerlund, 2019). While deepfake technology can be harnessed for creative, educational, and entertainment purposes, its misuse has raised serious ethical concerns that demand urgent attention. These concerns range from the spread of misinformation and fake news to privacy violations, with the potential to undermine public trust in the media and further complicate the challenges of verifying digital content (Tuysuz & Kılıç, 2023).Inaccurate information spread faster through social media, and it can impact millions of users (Figueira & Oliveira, 2017).One of the most pressing issues arising from the misuse of deepfake technology is its role in the dissemination of misinformation. Deepfake videos, for example, can be manipulated to make it appear as though someone has said or done something they never actually did, creating highly convincing false narratives. This can have a profound impact on

public opinion, influence political outcomes, and tarnish individuals' reputations. The malicious use of deepfakes to spread false information has become an alarming tool in political campaigns, corporate rivalry, and social media platforms, contributing to widespread confusion and mistrust (Westerlund, 2019). The potential to shape public perception based on fabricated evidence poses a unique challenge for society in terms of information authenticity.

Another critical ethical concern is the invasion of privacy enabled by deepfake technology. Since deepfake can replicate someone's likeness, voice, or actions without their consent, they can be used to manipulate or deceive others by creating fake videos that feature real people. This has significant implications for personal security, as deepfakes have been used for identity theft, cyberbullying, and extortion. Additionally, deepfake technology can be used to make false accusations, tarnishing an individual's character or integrity without their knowledge. The ethical dilemmas surrounding deepfakes are compounded by the fact that the technology can be used to generate explicit or harmful content, often targeting individuals in vulnerable situations (Tuysuz & Kılıç, 2023).

Despite the widespread abuse of deepfakes, the legal frameworks and regulatory measures needed to address these ethical concerns remain insufficient. In many countries, laws regulating deepfakes are either outdated or non-existent, leaving individuals without clear legal recourse when their identities are compromised or when misinformation is spread using fabricated media. Furthermore, deepfake technology is evolving at a rapid pace, making it increasingly difficult to detect and prevent such content from being circulated in the first place. The continuous advancements in AI and machine learning only make the task of identifying deepfakes more challenging, complicating efforts to maintain the integrity of digital media (Patalauskaitė, 2024)

This research aims to examine the ethical challenges posed by deepfake technology in the media and entertainment industries, focusing on key issues such as the spread of false news, privacy violations, and the erosion of public trust. The study also explores potential legal, technological, and regulatory solutions to mitigate the harmful effects of deepfakes while preserving their positive applications. Through an analysis of these concerns, the research seeks to contribute to the ongoing discourse on the responsible use and regulation of AI-driven media technologies.

1.1Problem Statement

Deepfake technology is the dire straits in the media because it can create fake but very realistic videos and audio. This makes it hard to tell what is real and what is fake (Pawelec,2024). Many deepfake are used to spread misinformation, fake news, and manipulate public opinion, which can mislead people and damage trust in the media (Afshari & Mohammadi, 2023). Another big concern is privacy violations, as deepfakes can be made without a person's consent, leading to serious personal and professional harm. Despite these risks, there are not enough laws and rules to control deepfake misuse and detecting them is still a big challenge (Karnouskos, 2024). This study explores the ethical problems deepfake technology creates in the media and discusses possible legal and technological solutions. (Tuysuz & Kılıç, 2023). Although researches have been carried out on this important issue but synthesized overview is required to see what are similarities and differences in the findings of all these studies. A review study is one of the approaches to analyze just of multiple studies in order to see an overall discussion regarding this problem.

1.2Research Objectives

O_1. To analyze the ethical problems caused by deepfake technology in the media, including its role in spreading misinformation, privacy violations, and public distrust.

O_2. To explore the privacy risks associated with deepfake technology, including unauthorized content creation, identity theft, and personal harm.

O_3. To examine how deepfake technology impacts trust in news, journalism credibility, and the authenticity of digital content.

O_4. To assess the legal challenges in regulating deepfake technology, identifying gaps in existing laws and policies.

O_5 To recommend policy interventions and awareness strategies for governments, media organizations, and digital platforms to minimize deepfake-related threats.

1.3 Research Questions

RQ_1. What ethical problems are caused by deepfake technology in the media?

RQ_2. What are the privacy risks of deepfake technology in the media?

RQ_3. How does deepfake affect trust in news and journalism?

RQ_4. What legal challenges exist in controlling deepfake misuse in the media?

RQ_5. How can governments and media organizations address the harmful effects of the deepfake technology?

1. Methodology

Table 1: Inclusion and Exclusion Criteria of the Study Literature.

Inclusion Criteria	Exclusion Criteria
Studies published between 2018 and 2024.	Studies published before 2018 (unless foundational or highly cited).
English language publications.	Non-English publications.
Media, journalism, entertainment, or digital platform-related studies.	Deepfake research in non-media domains.
Research addressing misinformation, privacy, trust, regulation in media.	Studies that reference deepfakes without relevant ethical analysis.

The methodology used in this study was double checked to make sure it is reliable, and it covers everything needed.

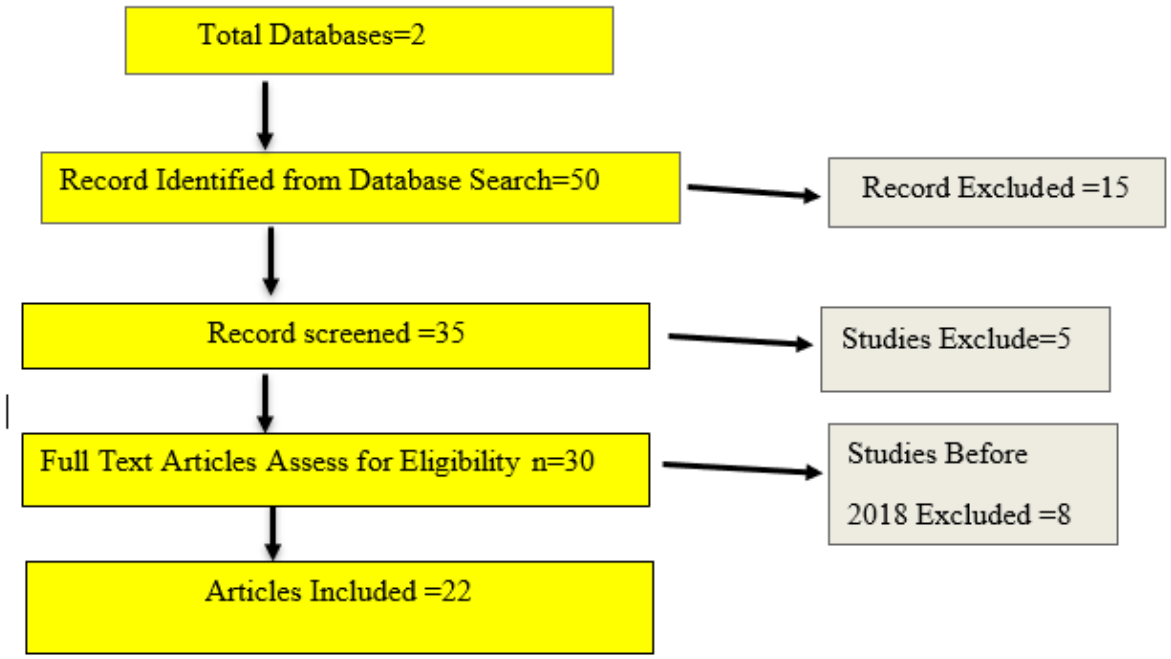

Fig. 1 PRISMA Flow Chart for the Articles Selection Process

Table 2 shows the percentage of articles selected from the database. Most articles from Google Scholar (n=17) and sci space (n=5)

Table 3 shows the percentages of selected literature according to their publication years. As it is clearly shown that the studies (n= 4) were published from 2018-2020, followed by (n=18) published from2021 and onwards.

According to their designs keeping in mind the numbers and percentages of the literature, most studies (11) were based on a review approach, while (3) studies used a survey design and (2) studies used an interview approach. Some studies (6) used case study. In terms of research methods (4) studies used quantitative approach, (16) studies used qualitative approach and (2) study used a mixed method approach.

The researchers looked at how the studies collected their data and calculated how many used each method.

Databases	Number	%
Google scholar	17	77.27%
Sci space	5	22.73

Table 2: Numbers And Percentages of Literature According to Their Database.

Table 3: Number and Percentages According to the Literature According to Publication Year.

Year	Number	%
2018-2020	4	18.18%
2021 and onwards	18	81.82

Table 4: Numbers and Percentages According to Their Paradigm Model

Paradigm model	Qualitative	Quantitative	Mixed method
Number	16	4	2

Table 5: Frequencies and Percentages Of The Literature According To Their Data Gathering Approaches.

Data Gathering Approach	Survey	Interview	Review	Other
Number	3	2	11	6

1.1. Validation of selected methodology

The method used in this review was carefully checked to make sure it was reliable and covered everything needed. Here are main steps taken to validate it:

1.2. Adherence to PRISMA Guidelines

PRISMA guidelines are a set of rules that help researchers systematically collect, analyze, and report information from different studies in a clear and organized way. They ensure that the review process is thorough, transparent, and easy to understand.

1.3. Inclusion and Exclusion Criteria

The criterion for this study is designed to select those articles that specifically mentioned about the ethics of deepfake.

1.4. Search Strategy

The strategy for searching articles was to use a specialized platform: Google scholar. I carefully choose keywords to cover it, ensuring no important studies were missed.

1.5. PRISMA Flow Chart

A PRISMA flowchart (Fig.1) is a visual way to show how researchers select, review, and include articles in their study. It helps make the process clear and proves that the researchers were done carefully and systematically.

3. Review Of the Selected Literature:

The rise of deep fake technology has brought significant challenges and ethical concerns in the media landscape. Deep fakes, which use advanced

artificial intelligence (AI) techniques such as Generative Adversarial Networks (GANs), allow the creation of hyper-realistic images, videos, and audio that can manipulate or fabricate content and are difficult to identify from authentic material. This technological advancement, while revolutionizing the entertainment and media industries, has become a double-edged sword, opening the door to significant societal risks such as misinformation, privacy violations, defamation, and the erosion of public trust.

3.1 Ethical Dilemmas and Privacy Concerns

The ability of deep fake technology is to fabricate realistic media that raises several ethical issues, particularly regarding privacy and consent. As deep fake technology becomes more accessible, individuals are increasingly at risk of having their similarities manipulated without their consent. This could result in severe privacy violations, as seen in cases where deep fake videos are used for threatening, payback porn, or the creation of defamatory content. Researchers like Bekkar Amina and Fatima Arab (2025) have examined the ethical challenges posed by AI-generated media, particularly focusing on deep fakes and their potential to infringe upon personal rights. Ethical concerns in the age of deepfakes include the defamation, manipulation and harm to anyone's privacy. These scholars argue that deeper ethical reflections are necessary, as the ease with which deepfake technology can be employed makes it difficult to safeguard personal privacy in the digital age. Moreover, deepfake technology also raises questions about the ownership of digital content. As AI-generated content becomes increasingly sophisticated, who owns the rights to digitally manipulated media? Is it the creator of the deep fake, the individual whose likeness is used.

3.2 Misinformation and Public Trust

The most distressing result of deepfake technology is its ability to spread misinformation very quickly. Deep fakes are a powerful tool for creating deceptive media that can be used to manipulate public opinion, influence elections, or damage the reputations of individuals. As deepfake videos and images become increasingly indistinguishable from real content, they present a grave threat to public trust in the media. Deepfake content can be strategically used in political campaigns, where videos of politicians or public figures can be modified to make them appear to say or do things they never actually happened. This not only creates mess among voters but can also leads to distrust in the political process. The 2020 U.S. presidential election, for instance, saw growing concerns about deepfake videos being used to manipulate voters, highlighting the need for more stringent controls and media literacy programs to address these risks (Kaur et al., 2024)

The potential for deepfakes to undermine trust in the media is also echoed by scholars like Tuysuz et al. (2023), who stress the importance of media literacy and the need for legal frameworks to safeguard against the harmful effects of deepfake technology. In their view, the inability of the public to distinguish between real and fake content exacerbates the challenge of combating misinformation, making it crucial to develop both technological and social resolves to sustain the credibility of the media.

3.3 The Technological Evolution of Deepfakes

Deepfake technology began gaining public attention in 2017 when a Reddit user began uploading videos using deep learning techniques to create explicit content by superimposing the faces of celebrities onto pornographic material. The rapid advancement of AI, particularly GANs, has since enabled deepfakes to become more sophisticated and harder to detect. According to researchers like Zobaed (2021), the complexity of these algorithms now allows for the seamless alteration of media in real-time, making it incredibly difficult to figure out between fabricated and genuine content. These advancements raise significant ethical concerns about the manipulation of digital media and its impact on the credibility of online information. Scholars such as Gupta et al. (2021) argue that deepfakes are increasingly being used in a wide array of contexts, from political manipulation to the creation of fraudulent content in business. These uses can cause harm to individuals, businesses, and entire political systems, thus warranting the need for robust

detection methods and ethical guidelines to mitigate these risks.

3.4The Role of Media and Public Awareness

In the fight against deepfakes, media literacy and public awareness plays an essential role. As deepfakes become more widespread, it is crucial to educate the public about the risk and hazards associated with synthetic media and how to recognize manipulated content. Educational campaigns that teach individuals how to verify the authenticity of digital media are essential to combate the spread of deepfakes.

According to Gregory (2023), public awareness campaigns should focus on helping people to develop critical intellectual skills and improve their ability to assess the credibility of digital content. In addition, platforms such as social media networks must take more responsibility for regulating deepfake content. The collaboration between platforms, governments, and civil society is essential to mitigate the negative effects of deepfakes and ensure the responsible use of AI in media.

3.5 Legal Implications and Regulatory Challenges

The legal landscape surrounding deepfake technology is still evolving, and existing laws are often insufficient to address the unique challenges posed by synthetic media. Deepfake technology intersects with several areas of law, including defamation, privacy, and intellectual property, each of which requires careful consideration in the context of AI-generated content.

Afshari and Mohammadi (2023) discuss the legal challenges posed by deepfakes, noting that current legal frameworks are not designed to handle the complexities of AI-generated content. They argue that there is an urgent need for updated laws that can address the harms caused by deepfake technology, such as the spread of fake news, defamation, and identity theft. Defamation laws, for instance, may not be equipped to handle cases where a person's likeness is used in a manipulated video or audio without their consent.

Another critical legal issue is the use of deepfake technology for political manipulation. The creation of false political videos, for example, can have significant repercussions for electoral integrity and public trust. According to Kaur et al. (2024), there is a growing consensus among legal scholars that the regulation of deepfake technology must be a priority for governments worldwide to protect democratic institutions from manipulation.

3.6 Detection and Countermeasures

Given the potential for harm, developing effective deepfake detection methods has become a top priority for researchers and technologists. Various machine learning models and algorithms have been proposed to detect deepfake content, including those that focus on inconsistencies in facial movements, speech patterns, and digital artifacts in videos. Zobaed (2021) highlights several deepfake detection techniques, such as those using deep learning to identify inconsistencies in the facial expressions or voice synchronization in manipulated media. As deepfake continues to rise, many methods to counter their impacts are introduced.Researchers are continually developing more advanced algorithms to keep pace with the growing sophistication of deepfakes. For instance, new tools now analyze the digital forensics of media to reveal hidden manipulation artifacts that may not be visible to the human eye. regardless, detecting deepfakes remains a significant challenge. As the technology improves, detection methods must become fractionally complex and refined to stay ahead of malicious actors. Furthermore, the development of effective countermeasures requires collaboration between technology companies, governments, and independent organizations to ensure that deepfake detection is both effective and accessible.

2. Thematic Overview

The thematic analysis of 22 reviewed papers on deepfake technology highlights a multidimensional discourse around its ethical, legal, psychological, and societal consequences. The studies were coded and categorized into major and minor themes that emerged repeatedly across the literature. Among the most dominant themes were legal and ethical challenges, misinformation and disinformation, and societal and psychological impacts. Scholars like Pawelec (2024), Afshari and Mohammadi (2023),

and Verma (2024) consistently addressed overlapping concerns such as consent violations, identity theft, trust erosion in media, and regulatory shortcomings. The frequency and intensity of themes reflect an urgent scholarly focus on mitigating the threats posed by AI-generated synthetic media. While certain studies emphasized legal solutions and policy frameworks, others explored the deeper psychological impact on individuals and societies, showcasing the complex and pervasive nature of deepfake technology in the digital age.

Table 6: Major and Minor Themes

Major themes	Minor themes	Concept Code
Legal and Ethical Challenges	Privacy concern	Misuse of personal data, Identity theft through AI, Violation of digital rights
	Consent & Autonomy	Non-consensual AI-generated media, Digital manipulation, Exploitation risks
	Misinformation & Disinformation	Fake news amplification, Election interference, Political deepfake propaganda
	Intellectual Property & AI Ethics	Copyright challenges in AI media, Misuse of open-source AI models, Ethical considerations in AI creation
Security and Financial Risks	Cybersecurity Threats,	Deepfake-based hacking, AI-generated voice scams, Biometric data vulnerabilities
	Fraud & Economic Crimes	AI-enabled phishing, Synthetic identity fraud, financial scams using deepfake technology
Societal and Psychological Impact	Trust in Media & Journalism	Erosion of credibility in news, AI-powered fake journalism, Challenges in media verification
	Psychological & Emotional Impact	Social anxiety due to deepfakes, psychological distress among victims, Fear of digital manipulation
	Political & Social Manipulation	Voter misinformation, AI-powered propaganda campaigns, Distortion of public policy through deepfakes
Regulatory and Mitigation Strategies	AI Detection & Verification	Deepfake detection algorithms, Blockchain for media authentication, Digital watermarking
	Policy & Legal Frameworks	International deepfake regulations, AI governance laws, Industry self-regulation on synthetic media.
	Media Literacy & Public Awareness	AI ethics education, Training programs on recognizing deepfakes, Strengthening media literacy initiatives

5. Findings

5.1. Legal and Ethical Challenges

One of the most spotlighted concerns in the reviewed literature is the ethical and legal complexity surrounding deepfake technology. Scholars like Pawelec (2024) and Afshari & Mohammadi (2023) emphasize how non-consensual AI-generated media infringes on individual autonomy and digital rights. Issues of identity theft, unauthorized manipulation of personal content, and the misuse of open-source AI models contribute to a growing crisis of digital trust. Legal frameworks remain underdeveloped in addressing these violations, particularly in transnational contexts where regulation is uneven.

5.1.1. Sub-theme: Consent and Autonomy

Deepfake content often involves individuals without their knowledge or approval, raising severe ethical concerns. Pawelec (2024) and Verma (2024) detail the manipulation of digital identities, where consent becomes irrelevant due to the sophistication of synthetic media creation.

5.1.2. Sub-theme: Intellectual Property and AI Ethics

Questions surrounding ownership of AI-generated content and the use of proprietary data to train deepfake systems are also central. Authors argue for clearer intellectual property rights and ethical guidelines for developers.

Table 7: Legal and Ethical Challenges

Author	Year	Key Points
Mustafa Kaan Tuysuz Ahmet Kılıç	2023	Deepfake technology violates someone's privacy and steals their information without consent.
Maria Pawelec	2024	Deepfake makers affect personal values and identity.
Regina Rini and Leah Cohen	2022	Deepfakes damage someone's autonomy by portraying falsely in situations they did not consent to.
Lu Jin	2020	Deepfakes can harm the reputation of people by spreading false information,blackmail or using their pictures without permission.
Emilija Patalauskaite	2024	In a technology driven world digital content causes major ethical issues about truth and responsibility.
Shweta Negi*, Mydhili Jayachandran, Shikha Upadhyay	2021	Deepfakes are difficult to detect because of little noticeable difference.

5.2. Misinformation and Disinformation

A recurring and critical theme is the potential of deepfakes to spread false information at scale. The meaning of misinformation is to share the wrong information without trying to lie with the public (Bekkar & Arab, 2025).

Researchers including Pawelec (2024) and Afshari & Mohammadi (2023) note that deepfakes have been weaponized for political propaganda, election interference, and the amplification of fake news. These studies highlight that the technology's ability to fabricate seemingly authentic content poses a direct threat to democracy, journalism, and public discourse.

Table 8: Misinformation and Disinformation

Author	Year	Key Points
Mustafa Kaan Tüysüz Ahmet Kılıç	2023	Deepfakes erode public trust by spreading misinformation,affecting personal reputations to democratic processes.
Mika Westerlund	2019	Software for crafting high-quality, realistic deepfakes for disinformation is increasingly available as open source. This enables users with little technical skills and without any artistic expertise to near-perfectly edit videos, swap faces, alter expressions, and synthesize speech.
Stamatis Karnouskos	2020	Fake news reduces the trust enabling misinformation or disinformation tactics.

BekkarAmina, Fatima Arab	2025	Social media spread both misinformation and disinformation very fast.

5.3. Security and Financial Risks

Several papers, such as those by Kaur, J., Sharma, K., & Singh M.P. (2024) and Gregory (2023), document the increasing use of deepfake technology in cybercrime. AI-generated voice scams, hacking attempts, and biometric data breaches represent a new frontier in digital security threats. Furthermore, fraudulent schemes using synthetic media for phishing or financial scams have prompted calls for advanced authentication and verification mechanisms.

5.3.1.Sub-theme: Fraud and Economic Crimes

Deepfakes have been employed in synthetic identity creation and business fraud. The reviewed studies advocate for regulatory measures and cybersecurity strategies tailored to address this emerging threat landscape.

Table 9: Security and Financial Risks

Author	Year	Key Points
Stamatis Karnouskos	2020	AI improves the economy and business,which effectively manages the profit as well as risk.
Sam Gregory	2023	• New technology is raising concerns about privacy and safety. People can be harmed if their activities are tracked.
Bahar Uddin Mahmud, Afsana Sharmin	2021	• Deepfake uses computer tools to make fake videos that look original. The purpose is to harm anyone through lies.

5.4. Societal and Psychological Impact

The reviewed research consistently acknowledges the broader psychological and social consequences of deepfake proliferation. Authors like Karnouskos (2024) and Pawelec (2024) Kaur, J., Sharma, K., & Singh M.P. (2024) explore how manipulated content erodes public trust in media, leads to anxiety among potential victims, and generates a climate of skepticism in digital interactions.According to Khimi, Albarqi, Saif, and Elhag (2024) deepfake hurt society and people causing -0confusion resulting8 distress.

Sub-theme: Trust in Media and Journalism

False audiovisual content undermines traditional media credibility. The difficulty in differentiating between real and fake has created significant challenges in news verification processes.

Sub-theme: Psychological and Emotional Impact

The emotional trauma inflicted on victims of non-consensual deepfakes—particularly in cases of deepfake pornography—has been highlighted as a severe consequence, often accompanied by fear, humiliation, and long-term psychological distress. (Pratap Singh, Goswami, & Garg, 2024)

Table 10 Societal and Psychological Impact

Author	Year	Key Points
Stamatis Karnouskos	2020	● Disinformation via deepfakes reduces the level of trust in news on social media. ● Social media in the modern context are mass media that have the potential to interfere with societal actions, e.g., social movements
Jaspreet Kaur, Kapil Sharma, M. P. Singh	2024	● Deepfakes affect a person's mental health causing stress, depression, and hurt victim's feelings as their fast spread can lead to psychological trauma.
Avadhesh Pratap Singh Madhav Goswami and Mugdha Garg	2024	● People with broken privacy feel upset which develop depression.
M.S. Yessimova T.V. Shevyakova	2024	● Ethical standards must be followed by journalists to stop deepfake.
Sm Zobaed, Md Fazle Rabby,Md Istiaq Hossain,Ekram Hossain, Sazib Hasan, Asif Karim, and Khan Md. Hasib	2021	● Deepfake technology uses advanced AI like GANs which helps to create fake videos and damage someone's privacy.
Weeam Khimi Kholood Albarqi and Kendah Saif Salma Elhag	2024	● Society can be harmed by deepfake,spreading lies and damaging the trust.

5.5. Political and Social Manipulation

Deep fakes have played a crucial role in altering political narratives. Studies reviewed (e.g., Pawelec, 2024; Afshari & Mohammadi, 2023) illustrate how these technologies are being used to sway public opinion, distort public policy, and undermine trust in political figures. The absence of prompt detection techniques renders these attacks difficult to counteract in real-time.

. According to Pratap Singh, Goswami, & Garg (2024), the 2019 study by Deep Trace Labs found that 96% of the deep fake online porn videos are without consent.

Table 11 Political and Social Manipulation

Author	Year	Key Points
Narges Afshari Ahmad Mohammadi	2023	● Manipulation of politicians in which fake videos are used for defamation, and in the creation of content without consent, which has raised alarming concerns about privacy.
Maria Pawelec	2024	● In trust decay, political manipulation and abuse, mirroring

		the focus of the respective public and the media.
Avadhesh Pratap Singh Madhav Goswami and Mugdha Garg	2024	• Deepfakes mainly target politicians, celebrities and even ordinary people.
Nitin Verma	2023	• During election, deepfake can trick people and create political confusion.
Sarfraj Ahmed Dr.Mohd Akbar Shaun	2022	• Deepfakes are a danger because they are difficult to detect.

5.6. Regulatory and Mitigation Strategies

Several scholars propose solutions aimed at curbing the abuse of deepfake. Kietzmann, Lee, McCarthy, and Kietzmann (2020), Pawelec (2024) discuss the advancement of detection algorithms, such as blockchain verification and digital watermarking. Moreover, they emphasize the role of policy and law in combating synthetic media misuse.

Sub-theme: Policy and Legal Frameworks

Afshari & Mohammadi (2023) argue for international collaboration on regulations, proposing AI governance laws and stricter platform accountability to ensure ethical AI use.

Sub-theme: Media Literacy and Public Awareness:

Education initiatives aimed at improving digital literacy and ethical awareness have been identified as crucial strategies. Empowering users to recognize deepfakes and engage critically with digital media can reduce vulnerability to manipulation.

Table 12 Regulatory and Mitigation Strategies

Author	Year	Key Points
Mustafa Kaan Tuysuz Ahmet Kılıç 2	2023	• Laws must be developed to promote media literacy to counter deepfakes. • Legal and ethical experts must work on a solution to detect depfakes.
Mika Westerlund	2019	• Corporate policies and voluntary action may provide more effective tools against deepfake. • Education and training, • Anti-deepfake technology.
Stamatis Karnouskos	2020	• Several countries have laws and a regulatory framework dealing with digital media and their processes.
Jan Kietzmann Linda W. Lee Ian P. McCarthy TimC. Kietzmann	2020	• Victims deserve protection but, in this regard, laws are very limited.

Discussion and **Conclusion**

The rulings from the data review reveal significant insights into the legal, ethical, and societal challenges posed by deepfake technology. The best ways to stop the spread of deepfake is through a support of legal, educational, and socio technical advances.

A common theme among the literature is the privacy concerns and abuse of personal data, as deepfake technology enables the creation of realistic fake media without the consent of individuals. Many Studies emphasize the risks of identity theft, digital manipulation, and exploitation due to non-consensual use of AI-generated media. These concerns are compounded by the fact that deepfake technology can significantly distort public trust, particularly in political and media domains.

The ethical implications of deepfakes have been a focal point of various studies. Researchers, including Verma (2024) and Karnouskos (2024), highlight the need for stringent regulatory frameworks and AI governance to address the growing risks of digital manipulation. They emphasize the importance of legal interventions, such as copyright laws and consent frameworks, to protect individuals and organizations from the misuse of synthetic media. However, gaps in the legal system remain, especially with respect to the regulation of AI-generated media in rapidly evolving technological landscapes.

Additionally, the literature reveals societal and psychological impacts, such as trust erosion in digital content. Studies like those by Pawelec (2024) and Afshari & Mohammadi (2023) suggest that deepfakes contribute to a decline in media credibility, making it increasingly difficult for the public to distinguish between authentic and manipulated content. This issue presents a considerable obstacle to the credibility of journalism and the media's function as a purveyor of truth within society. The psychological effects of deepfake technology are also noted, with victims experiencing emotional distress and social anxiety due to the manipulation of their images or voices.

Moreover, deepfakes have been recognized as a tool for political manipulation, where fake videos and audios are used to spread misinformation, interfere in elections, or distort public opinions. Studies by Verma (2024) and Pawelec (2024) identify political deepfakes as a growing concern, necessitating enhanced detection techniques and stronger media literacy programs to equip the public with the skills to critically evaluate digital content.

Another key finding in the literature is the exploration of technological solutions such as deepfake detection algorithms and blockchain technology. These innovations hold promise in providing solutions to the ethical dilemmas associated with deepfakes. However, while such solutions are promising, their widespread implementation remains a challenge, especially in the context of legal and ethical barriers that delay the adoption of these technologies.

The literature highlights the multifaceted challenges of deepfake technology, particularly concerning privacy, ethical issues, misinformation, and the psychological and societal impacts. The findings suggest a pressing need for robust legal frameworks and technological solutions to combat the growing threats of AI-generated media. While there is an increasing awareness of the negative consequences of deepfakes, especially in the areas of political manipulation and public trust, further research and innovation are required to fully address these concerns. The regulatory frameworks around deepfakes remain underdeveloped, with significant gaps in international and national laws. As the technology continues to advance, policymakers must act swiftly to implement laws that govern the ethical use of AI-generated content. In addition to legal measures, there is a strong call for media literacy education and AI ethics training to help the public recognize and critically assess deepfake content.

The technological advancements in detecting and preventing deepfakes provide hope for mitigating their harmful effects, but these solutions are still in their early stages. More research is needed to improve the accuracy and efficiency of detection methods and to ensure that they can be implemented on a large scale. Additionally, it is crucial to examine the long-term societal effects of deepfake, particularly on trust in digital content and public perception of media.

References:

Afshari, N., & Mohammadi, A. (2023). The legal implications of deepfake technology: Privacy, defamation, and the challenge of regulating synthetic media. *Google Scholar*.

Ahmed, S., & Shaun, M. A. (2022). Impact of deepfake technology on digital world authenticity: A review. *SciSpace*.

Anwar, S., Khan, S. R., Nasir, T., & Azeema, N. (2025). The AI revolution in media, redefining journalism education and professional practice from classroom to newsroom in Pakistan. *Annual Methodological Archive Research Review, 3*(4), 340–354. https://doi.org/10.63075/4t3nmg72

Bekkar, A., & Arab, F. (2025). Ethical challenges in artificial intelligence generated media content. *Google Scholar*.

Fazal, I., Nasir, T., & Mahmood, S. (2025). The impact of climate change news on mental health of youth: The rise of eco-anxiety in Pakistan. *Annual Methodological Archive Research Review, 3*(5), 269–283. https://doi.org/10.63075/m78hhp49

George, A. (2024). Defamation in the time of deepfakes. *SciSpace*.

Gregory, S. (2023). Fortify the truth: How to defend human rights in an age of deepfakes and generative AI. *Google Scholar*.

Hussain, S. A., Anwar, S., Iqbal, N., & Nasir, T. (2025). From newsrooms to algorithms: AI's role in the future of mass communication in Pakistan. *Annual Methodological Archive Research Review, 3*(4), 355–368. https://doi.org/10.63075/76hgfq50

Jin, L. (2020). Cheated by deepfakes? Deepfake detection ability, people's reactions, and ethical implications. *Google Scholar*.

Karnouskos, S. (2020). Artificial intelligence in digital media: The era of deepfakes. *Google Scholar*.

Kaur, J., Sharma, K., & Singh, M. P. (2024). Exploring the depth: Ethical considerations, privacy concerns, and security measures in the era of deepfakes. *Google Scholar*.

Khattak, M. S., Nasir, T., Usman, M., & Rahim, S. (2025). AI revolution in digital media: Opportunities, challenges, and the future of journalism in Pakistan. *Annual Methodological Archive Research Review, 3*(4), 398–413. https://doi.org/10.63075/eer8yc21

Khimi, W., Albarqi, K., Saif, K., & Elhag, S. (2024). A systematic review on deepfake image generation, detection techniques, ethical implications, and overcoming challenges. *Google Scholar*.

Kietzmann, J., Lee, L. W., McCarthy, I. P., & Kietzmann, T. C. (2022). Deepfakes: Trick or treat? *Business Horizons, 65*(2), 183–193.

Mahmud, B. U., & Sharmin, A. (2022). Deep insights of deepfake technology: A review. *SciSpace*.

Mahmood, S., Nasir, T., Fazal, I., & Razzaq, R. (2025). Juxtaposing law, media and human rights: Role of social media in promoting human rights in Pakistan. Case study of legal reforms and digital activism. *Dialogue Social Science Review, 3*(6), 639–661.https://dialoguessr.com/index.php/2/article/view/351/403

Meskys, E., Liaudanskas, A., Kalpokiene, J., & Jurcys, P. (2020). Regulating deepfakes: Legal and ethical considerations. *Google Scholar*.

Nasir, T., Azeema, N., Irum, M., & Siraj, S. A. (2025). Influence of AI and digital media trends, algorithms and big data on agenda setting and narrative building of media students: A case study of universities in Islamabad. *Social Science Review Archives, 3*(2), 335–355. https://socialworksreview.com/index.php/Journal/article/view/184/208

Nasir, T., Khan, S. A., Iqbal, N., & Ahmad, H. (2025). From awareness to action: Exploring the role of media in climate change education and engagement in Pakistan. *Annual Methodological Archive Research Review, 3*(4), 383–397. https://doi.org/10.63075/pk1e7n43

Nasir, T., Mahmood, S., Ali, K., & Abubakar, M. (2025, May 27). Climate change and media policy nexus: Opportunities, challenges, and policy recommendations. Case study of Pakistan. *Dialogue Social Science Review, 3*(5), 610–631. https://thedssr.com/index.php/2/article/view/578

Nasir, T., Qazi, U., Fazail, A., Tareen, H., & Hussain, W. (2025). Media narratives regarding climate change influencing tourism patterns in Pakistan. *Journal of Media Horizons, 6*(3), 343–360. https://doi.org/10.5281/zenodo.15835956

Nasir, T., Siraj, S. A., Hannan, F. Z. U., Hussain, W., & Javed, S. (2024). A perception of university students regarding the influence of social media on the academic performance. *Journal of Peace, Development and Communication, 8*(03), 431–450. https://doi.org/10.36968/JPDC-V07-I01-25

Nasir, T., Anwar, S. A. S., Iqbal, N., & Arif, M. (2025). The psychological impact of digital media consumption on mental health, a case study of undergraduate students in Pakistan. *Annual Methodological Archive Research Review, 3*(4), 369–382. https://doi.org/10.63075/7022md02

Negi, S., Jayachandran, M., & Upadhyay, S. (2021). Deep fake: An understanding of fake images and videos. *International Journal of Scientific Research in Computer Science, Engineering and Information Technology (CSEIT), 6*(3), 33–39. https://doi.org/10.32628/CSEIT21733

Patalauskaitė, E. (2024). Ethical aspects of content creation. *Google Scholar*.

Pawelec, M. (2024). Decent deepfakes? Professional deepfake developers' ethical considerations and their governance potential. *Google Scholar*.

Razzaq, R., Riaz, R., Nasir, T., & Hussain, W. (2025). Public opinion and policy development: A psychological approach to understanding the role of public sentiment in shaping legislation – a case study of law, psychology, media and policy development nexus. *Annual Methodological Archive Research Review, 3*(5), 296–310. https://doi.org/10.63075/8jz5fk62

Rinii, R., & Cohen, L. (2022). Deepfakes, deep harms. *Google Scholar*.

Singh, A. P., Goswami, M., & Garg, M. (2023). The ethics of deepfakes: A digital age crisis. *Google Scholar*.

Tuysuz, M. K., & Kılıç, A. (2023). Analyzing the legal and ethical considerations of deepfake technology. *Google Scholar*.

Verma, N. (2023). Deepfake technology and the future of public trust in video. *Google Scholar*.

Westerlund, M. (2019). The emergence of deepfake technology: A review. *Technology Innovation Management Review, 9*(11), 40–53.

Yessimova, M. S., & Shevyakova, T. V. (2024). Deepfakes in the digital media age: Opportunities and threats. *Google Scholar*.

Zobaed, S. M., Rabby, M. F., Hossain, M. I., Hasan, E. H. S., Karim, A., & Hasib, K. M. (n.d.). Deepfakes: Detecting forged and synthetic media content using machine learning. *University of Louisiana; Southern Utah University; Dixie State University; Charles Darwin University; Ahsanullah University of Science & Technology*.