Sarhad Journal of Legal Studies

January-June 2025, Vol. 01, No. 01, [01–15]
Published: 31st January 2025
Publisher: Sarhad Institute of Legal Studies, SUIT
https://journal.suit.edu.pk/index.php/sjls

RESEARCH PAPER

AI-Generated Deepfakes and the Crisis of Legal Authenticity: Reconstructing Evidentiary Standards in the Digital Age

Osama Muhammad[1]*

ABSTRACT

The rapid proliferation of AI-generated deepfakes has created a critical challenge for contemporary legal systems by undermining the authenticity of digital evidence and exposing individuals to new forms of harm. This study examines how synthetic media destabilizes established principles of evidence, privacy, and accountability, creating a crisis of legal authenticity. The purpose of the research is to evaluate the adequacy of current legal frameworks and to propose a structured regulatory response to deepfake-driven risks. Using a comparative legal-analytical approach, the study examines regulatory developments in the United States, European Union, China, and South Asia, and applies critical discourse analysis to assess emerging doctrines on digital manipulation. The findings reveal substantial gaps in evidentiary standards, limited platform liability mechanisms, and the absence of harmonized international norms governing synthetic media. The study concludes that traditional evidentiary presumptions are insufficient for an era in which falsified content is indistinguishable from the real. It proposes a three-pillar regulatory framework—authenticity infrastructure, platform gatekeeping, and cross-border harmonization to strengthen legal protections and preserve trust in judicial and democratic institutions.

Keywords: AI Regulation; Deepfakes; Digital Evidence; Legal Authenticity; Platform Liability.

INTRODUCTION

Artificial intelligence (AI) generated deepfakes have rapidly transformed the digital information environment, challenging longstanding legal assumptions about the authenticity of audiovisual evidence. As synthetic media becomes increasingly indistinguishable from genuine recordings, courts and regulatory bodies face unprecedented difficulties in determining what constitutes reliable digital content. This development raises profound concerns regarding the protection of individual rights, the legitimacy of judicial processes, and the stability of democratic governance. Against this backdrop, the present study investigates how deepfakes destabilise traditional legal norms and seeks to identify the regulatory mechanisms required to address the resulting governance dilemmas. The purpose of this article is to evaluate the adequacy of current

[1] LLM Scholar, Khyber Law College, University of Peshawar, Pakistan. osama56640@gmail.com

legal frameworks in managing deepfake-related harms and to develop a structured model capable of safeguarding evidentiary integrity and protecting vulnerable individuals. The research covers criminal law, civil liability, evidence law, and digital governance across multiple jurisdictions, including the United States, European Union, China, and South Asia. Positioned within the broader context of rapidly advancing AI technologies, this study argues that synthetic media has created new vulnerabilities for legal systems that were developed in an era when the authenticity of audiovisual evidence could largely be taken for granted (MKPO, 2025).

The significance of this study lies in its contribution to a nascent yet increasingly urgent field of legal scholarship. As deepfakes enable identity manipulation, political disinformation, financial fraud, and non-consensual explicit content, legal systems are confronted with challenges far beyond the capacity of conventional evidentiary doctrines and regulatory tools. Although some jurisdictions have enacted partial reforms, coherent and comprehensive governance mechanisms remain absent at both national and international levels. Understanding these gaps is essential for formulating workable legal responses. The background to this problem reveals a structural tension between the speed at which AI technologies evolve and the comparatively slow pace of legal adaptation. Deepfakes undermine trust in digital media, blur the boundaries between truth and fabrication, and create opportunities for malicious actors to exploit weaknesses in legal and regulatory systems. They also produce new forms of harm that fall between established categories of privacy law, defamation, cybersecurity, and digital evidence, highlighting the need for a holistic regulatory approach (Williams, 2025; Khan et al., 2021).

This study proceeds from the hypothesis that existing legal frameworks are structurally inadequate to address the evidentiary, regulatory, and cross-border challenges posed by deepfakes. It argues that the absence of integrated authenticity mechanisms, limited platform accountability, and fragmented international norms collectively prevent effective governance of synthetic media. To test this hypothesis, the research is guided by two central questions: how deepfakes undermine foundational assumptions of evidence law, individual rights, and digital accountability; and what legal, technological, and regulatory reforms are required to govern synthetic media at national and international levels. Methodologically, the article employs a comparative legal-analytical approach to examine the statutory responses, judicial practices, and policy developments in selected jurisdictions. This is supplemented by critical discourse analysis to interpret emerging debates on digital manipulation and synthetic media governance. The anticipated outcome is the identification of core doctrinal and regulatory gaps and the development of a three-pillar governance model focused on authenticity infrastructure, platform gatekeeping, and international harmonisation (Khan et al., 2021).

The remaining structure of the article reflects the logical progression of this inquiry. The next section outlines the technological and legal nature of deepfakes and maps the broader risk landscape. Subsequent sections analyse the implications for evidentiary standards, individual rights, platform liability, and international law. A comparative survey of national responses precedes the proposal of a comprehensive regulatory framework. The article concludes with reflections on the future of legal authenticity and recommendations for strengthening judicial and democratic resilience in the age of synthetic media.

CONCEPTUAL AND THEORETICAL FRAMEWORK

This study is grounded in a multidisciplinary conceptual and theoretical framework that integrates doctrines of evidence law, theories of digital governance, and emerging scholarship on algorithmic manipulation. At the conceptual level, deepfakes are understood as a form of synthetic media produced through generative AI models that challenge the traditional legal presumption of audiovisual authenticity. Theoretically, the article draws on authenticity theory, which posits that legal systems rely on stable markers of truth to maintain institutional legitimacy, and on governance theory, which emphasizes the need for adaptive regulatory mechanisms in technologically dynamic environments. The analysis is further situated within the broader literature on platform governance, highlighting how private digital intermediaries shape information flows and influence the distribution of both harm and accountability. Additionally, the study adopts elements of risk society theory, recognizing that modern legal systems must navigate uncertainties created by emerging technologies whose consequences are neither fully understood nor easily regulated. By synthesizing these theoretical perspectives, the framework provides a basis for examining how deepfakes disrupt evidentiary reliability, create novel privacy and security risks, and necessitate a comprehensive regulatory approach capable of addressing both individual harms and systemic vulnerabilities (Khan et al., 2020).

RESEARCH METHODOLOGY

This study employs a qualitative research methodology grounded in comparative legal analysis, doctrinal examination, and critical discourse analysis. The comparative component evaluates statutory frameworks, regulatory instruments, and judicial responses to deepfakes across multiple jurisdictions—including the United States, European Union, China, and South Asia—to identify convergences, divergences, and regulatory gaps. Doctrinal analysis is used to interpret existing laws on evidence, privacy, defamation, cybercrime, and platform liability, assessing their applicability in the context of AI-generated synthetic media. Critical discourse analysis supplements this approach by examining scholarly debates, policy documents, and institutional reports to understand evolving narratives around digital authenticity, AI governance, and algorithmic harm. Primary sources include legislation, case law, regulatory guidelines, and official reports, while secondary sources include academic literature and policy analyses. The methodological approach is purposively designed to reveal structural deficiencies within current legal systems and to support the development of a holistic, three-pillar regulatory framework capable of addressing the evidentiary, technological, and cross-border challenges posed by deepfakes (Khan et al., 2020).

DEEPFAKES AS A LEGAL PHENOMENON: CONCEPTUAL AND TECHNOLOGICAL BACKGROUND

Defining Synthetic Media

Synthetic media refers to digital content audio, video, or images that is artificially generated or manipulated using advanced computational algorithms, particularly deep learning models. Deepfakes, a prominent form of synthetic media, are typically produced using Generative Adversarial Networks (GANs), which enable the creation of hyper-realistic content that can be nearly indistinguishable from genuine recordings. While initially popularized for entertainment, creative, and artistic purposes, deepfakes have increasingly been deployed for malicious objectives, including identity theft, disinformation campaigns, non-consensual sexual imagery, and political manipulation. The distinctive characteristic of synthetic media lies in its ability to simulate reality at scale and speed, undermining conventional assumptions of trust and verifiability in audiovisual evidence. This dual-use nature where the technology can serve both socially beneficial and harmful ends positions synthetic media as a critical area of legal and regulatory concern, demanding urgent attention from courts, lawmakers, and digital governance institutions to safeguard authenticity, accountability, and individual rights (Apolo & Michael, 2024; Khan et al., 2020).

The Risk Landscape

The rise of AI-generated deepfakes has introduced a complex and multidimensional risk landscape that challenges legal systems, social institutions, and individual rights. These risks can be broadly categorized into criminal law, evidence law, civil liability, election law, and national security, each presenting unique legal and regulatory dilemmas.

Criminal Law Risks

Deepfakes facilitate a range of criminal activities, most notably impersonation for fraud, extortion, and blackmail. By replicating an individual's voice, image, or gestures, malicious actors can convincingly impersonate victims to gain financial benefits, coerce compliance, or manipulate third parties. For instance, deepfake audio of corporate executives has been used in fraudulent transfers, while synthetic videos have been deployed to blackmail individuals with fabricated content. These developments challenge conventional criminal statutes, which are often ill-equipped to address crimes executed with entirely synthetic digital identities (Murray, 2025; Kahn & Wu, 2020).

Evidence Law Challenges

The integration of deepfakes into legal proceedings raises profound concerns regarding the admissibility, authenticity, and reliability of evidence. Traditional doctrines such as the chain of custody, expert verification, and the presumption that visual or audio recordings are trustworthy are under significant strain. Courts must now consider whether recorded material might be artificially generated, manipulated, or deceptively edited, creating a "liar's dividend" where perpetrators can deny wrongdoing by claiming evidence is a deepfake (Gregory, 2024; Khan, 2018).

Civil Liability Concerns

Beyond criminal implications, deepfakes expose individuals to civil harms including defamation, harassment, and invasion of privacy. Non-consensual deepfake pornography, synthetic identity fraud, and fabricated statements can damage reputations, cause psychological distress, and result in financial or social losses. Existing civil remedies, such as tort claims for defamation or privacy violations, are limited by jurisdictional boundaries, evidentiary challenges, and the rapid dissemination of synthetic content online (Williams, 2025; Hui et al., 2025).

Electoral and Political Manipulation

Deepfakes have become tools for manipulating political discourse and influencing voter behaviour. Synthetic videos can falsely depict politicians endorsing policies, making controversial statements, or engaging in illegal activities. Such content erodes public trust, interferes with democratic processes, and creates challenges for election law and regulatory authorities, who must contend with both the speed and the cross-border reach of misinformation campaigns (Llorente, 2024).

National Security Threats

On a macro level, deepfakes pose risks to national security through misinformation campaigns, hybrid warfare, and cyber operations. States and non-state actors can deploy synthetic media to manipulate public perception, incite social unrest, or interfere in foreign elections. This dual-use threat blurs the lines between domestic legal regulation and international law, requiring coordination between national security agencies, digital platforms, and international governance frameworks. The global spread of generative AI tools such as Sora, DALL·E, MidJourney, and open-source models has accelerated these risks by lowering technical barriers and enabling rapid, scalable production of synthetic content. This proliferation emphasizes the urgent need for adaptive legal, technological, and regulatory responses that can manage both the individual harms and systemic threats posed by deepfakes (Ghiurău & Popescu, 2024; Khan & Hussain Shah Jillani, 2019).

DEEPFAKES AND THE COLLAPSE OF EVIDENTIARY RELIABILITY

The End of Visual Trust

Historically, courts and legal systems have treated video and audio recordings as inherently reliable forms of evidence. The assumption was that, unless there was clear evidence of tampering, audiovisual material could be trusted to accurately represent reality. This principle formed the basis of many evidentiary rules, including presumptions of authenticity, the chain of custody, and reliance on visual or auditory cues to establish facts. However, the advent of AI-generated deepfakes fundamentally challenges this long-standing assumption. Advanced generative models can produce videos, images, and audio clips that are virtually indistinguishable from genuine recordings, making it increasingly difficult for judges, juries, and forensic experts to determine the veracity of evidence using traditional methods alone (Hausknecht, 2025).

As a result, a recording can no longer be automatically trusted simply because it appears authentic. Even high-definition videos, facial expressions, or voice patterns cannot provide reliable verification without additional technical analysis. This erosion of visual trust has significant implications for the justice system, as it not only complicates the process of establishing facts but also opens the door to wrongful convictions, false accusations, and manipulation of legal outcomes. Courts now face the dual challenge of distinguishing authentic evidence from synthetic fabrications and developing new procedural safeguards, including expert forensic evaluation, metadata analysis, and cryptographic verification, to restore confidence in digital evidence (Gupta & Fatunmbi, 2024).

Burden of Proof Challenges

The rise of deepfakes has created unprecedented complications in the allocation of evidentiary burdens within legal proceedings. Traditionally, the burden of proof lies with the party asserting a claim or allegation; however, synthetic media disrupts this fundamental principle by creating scenarios in which both plaintiffs and defendants face new and complex evidentiary pressures. Innocent individuals, for instance, may find themselves compelled to prove that a manipulated video or audio recording is falsified, rather than relying on the presumption of authenticity. Conversely, accused persons can exploit deepfakes to claim that otherwise authentic evidence has been fabricated, a phenomenon often referred to as the “liar’s dividend.” Both situations undermine core principles of justice: the first imposes an unfair evidentiary burden on victims, while the second allows perpetrators to evade accountability by casting doubt on legitimate evidence. The resultant uncertainty not only jeopardizes fair trial rights but also impedes the truth-finding function of the judiciary, as courts struggle to differentiate genuine material from artificially generated content. Addressing these challenges requires the development of new procedural safeguards, including advanced forensic verification techniques, standards for digital evidence authentication, and judicial guidelines for interpreting synthetic media (Makhambetsaliyev, 2025).

JUDICIAL RESPONSES IN VARIOUS JURISDICTIONS

United States

In the United States, courts are increasingly confronting the evidentiary challenges posed by deepfakes through reliance on expert testimony, digital forensic analysis, and detailed metadata verification. Traditional evidentiary frameworks, particularly Federal Rules of Evidence 901, which governs authentication of evidence, are under significant strain as synthetic media becomes more sophisticated and harder to detect. Judges and juries now frequently depend on forensic experts to identify signs of manipulation, inconsistencies in metadata, and anomalies in audio-visual content. Despite these measures, rapid technological advancement often outpaces judicial capacity, creating risks of both wrongful conviction and unwarranted exoneration. This situation highlights the urgent need for procedural reforms, judicial training in AI literacy, and the integration of technological verification tools into courtroom processes (Shirish & Komal, 2023).

European Union

In the European Union, legislative responses have focused on regulation and risk mitigation through the proposed AI Act, which requires risk assessments for AI-generated content. While this represents a significant step toward governance of synthetic media, the Act does not provide comprehensive guidance on judicial evidentiary procedures or authentication protocols in legal proceedings. Consequently, EU courts continue to navigate evidentiary challenges on a case-by-case basis, balancing principles of fair trial, privacy, and digital accountability. The lack of clear procedural standards for handling deepfakes in litigation creates uncertainty for both plaintiffs and defendants (Shi, 2025).

China

China has adopted one of the most proactive legal frameworks regarding synthetic media through the 2022 Provisions on Deep Synthesis Internet Information Services. These regulations mandate labeling of AI-generated content and impose clear responsibilities on digital platforms to ensure authenticity and prevent misuse. By requiring real-name registration and content auditing by platforms, the law seeks to create systemic accountability in the dissemination of synthetic media. While highly effective in controlling the spread of deepfakes, this approach has been criticized for imposing strict state oversight, raising concerns about freedom of expression and privacy (Lin, 2025).

South Asia (Pakistan and India)

In South Asia, countries such as Pakistan and India have yet to develop comprehensive legal frameworks addressing deepfakes. Courts primarily rely on under-equipped digital forensic laboratories to authenticate evidence, creating delays and gaps in judicial processing. The absence of dedicated legislation for synthetic media, coupled with limited technical expertise, leaves victims of deepfake-related crimes and litigants vulnerable. This reactive approach contrasts sharply with proactive frameworks seen in China and partially in the EU and the US, highlighting the need for region-specific reforms, investment in digital forensic capacity, and updated procedural standards. These diverse jurisdictional responses illustrate the uneven readiness of global legal systems to contend with the evidentiary, civil, and criminal risks posed by deepfakes, underscoring the need for both national and international regulatory harmonization (Paris, 2021).

THE AUTHENTICATION CRISIS

The rise of deepfakes has created a profound crisis of authentication, challenging one of the most fundamental assumptions of modern legal systems: that evidence presented in court can be trusted to accurately reflect reality. Courts are now confronted with multiple layers of vulnerability that affect the entire lifecycle of digital evidence. Before evidence is even submitted, there is a pre-submission authenticity problem, as highly sophisticated synthetic media can be introduced into proceedings without any reliable mechanism to verify its origin, integrity, or accuracy. Traditional safeguards such as chain-of-custody protocols, basic metadata checks, and standard forensic review are often insufficient to detect manipulation, leaving litigants whether plaintiffs or defendants exposed to potential harm or wrongful allegations (Ghani et al., 2025).

During trial proceedings, courts face an in-court authenticity challenge. Judges and juries must assess the credibility of audiovisual evidence that can be indistinguishable from genuine recordings. Conventional evaluation methods, including visual inspection, witness corroboration, and expert testimony, are increasingly inadequate when applied to deepfakes created with advanced generative algorithms. This gap complicates fact-finding, as litigants may strategically leverage synthetic media to create doubt, support claims, or deflect liability, giving rise to what scholars term the "liar's dividend." The situation is further exacerbated by the limited technical expertise among legal professionals and the absence of standardized protocols for authenticating complex AI-generated content (Mathlouthi et al., 2025).

Even after judicial decisions are rendered, courts are not insulated from manipulation. Post-adjudication attacks, in which fabricated media is circulated to misrepresent proceedings or discredit rulings, threaten both institutional credibility and public confidence in the justice system. Such attacks may intimidate witnesses, influence public perception, or pressure appellate processes, creating systemic vulnerabilities beyond individual cases. Collectively, these pre-submission, in-court, and post-adjudication challenges illustrate how deepfakes destabilize the evidentiary foundations of law. Addressing this crisis demands a comprehensive response that combines technological verification tools, enhanced judicial training in digital forensics, procedural reforms for evidence authentication, and proactive strategies to safeguard the integrity of both trials and judicial institutions in the digital era (Khan et al., 2025).

DEEPFAKES AS VIOLATIONS OF RIGHTS AND FREEDOMS

Deepfakes have emerged as a pervasive threat to fundamental rights and freedoms, affecting privacy, dignity, reputation, and political participation. One of the most widespread abuses involves non-consensual deepfake pornography, which disproportionately targets women and violates core rights related to privacy, bodily autonomy, and personal dignity. The creation and dissemination of such content often occur without consent, causing psychological harm, social stigma, and reputational damage. These practices also raise serious concerns regarding data protection, as personal images, videos, and biometric information are exploited without legal authorization. Beyond violations of privacy, deepfakes serve as powerful tools for defamation and identity fraud. False statements can be visually and audibly portrayed as real, while manipulated videos may depict individuals engaging in illegal or immoral acts. Such synthetic content is also employed in financial and corporate fraud; for example, CEO voice deepfakes have been used to authorize fraudulent transactions or mislead employees, exposing both individuals and institutions to substantial harm. The weaponization of speech through deepfakes undermines the traditional legal balance between freedom of expression and protection from reputational damage, creating complex challenges for civil liability and enforcement (Hui et al., 2025).

Deepfakes also pose a serious threat to political processes and electoral integrity. Synthetic media can falsify speeches, simulate endorsements of political candidates, or incite communal tensions, thereby influencing voter perceptions and public discourse. Several recent elections worldwide have already experienced the disruptive effects of AI-fabricated content, raising urgent questions about democratic accountability, regulatory oversight, and the role of platforms in mitigating misinformation. Children and other vulnerable groups face additional risks, as they are increasingly targeted through deepfake-enabled cyberbullying, identity misuse, and harassment. These threats intersect with international standards, such as the United Nations Convention on the Rights of the Child (CRC), and national child protection laws, highlighting the need for specialized safeguards to prevent exploitation, ensure accountability, and provide redress. Collectively, these developments underscore how synthetic media can infringe upon a broad spectrum of rights and freedoms, demanding a comprehensive legal and regulatory response to protect individuals and uphold the rule of law in the digital age (Khan & Ullah, 2024).

PLATFORM LIABILITY AND REGULATORY GAPS

Social media and digital platforms play a central role in the rapid dissemination of deepfakes, effectively shaping both the scale and impact of synthetic media harms. Legal approaches to platform liability, however, vary significantly across jurisdictions. In the United States, platforms benefit from broad immunity under Section 230 of the Communications Decency Act, which shields them from most legal responsibility for user-generated content. In contrast, the European Union's Digital Services Act (DSA) introduces due diligence obligations, requiring platforms to assess and mitigate risks associated with harmful content, including AI-generated manipulations. China mandates proactive monitoring, content labeling, and watermarking of synthetic media, holding platforms directly accountable for ensuring authenticity. Meanwhile, South Asian jurisdictions, such as Pakistan and India, rely primarily on ad hoc takedown requests, leaving gaps in timely enforcement and consistent protection against the spread of harmful deepfakes (Khan, 2024).

A critical challenge arises from the absence of mandatory disclosure or watermarking mechanisms, which deprives courts and regulatory authorities of reliable reference points for distinguishing authentic content from synthetic fabrications. Without built-in authenticity signals, legal actors are forced to rely on forensic analysis or expert testimony, both of which may be limited by technical complexity, resource constraints, and the rapid pace of AI development. Compounding these issues is the role of algorithmic amplification, whereby recommendation and ranking systems on social media platforms can inadvertently magnify the reach of harmful synthetic media. Despite the outsized influence of such algorithms, accountability measures remain largely unregulated, allowing malicious or misleading content to propagate widely before legal or technical intervention is possible. Together, these factors highlight the urgent need for harmonized platform liability frameworks, transparency measures, and algorithmic governance standards to reduce the risks posed by deepfakes and ensure that platforms act as responsible intermediaries in the digital ecosystem (Khan, 2024).

INTERNATIONAL LAW DIMENSIONS

The global proliferation of deepfakes raises significant questions for international law, particularly as synthetic media increasingly becomes a tool for hybrid warfare and cross-border influence operations. States and non-state actors are employing AI-generated content to manipulate public opinion, disrupt social cohesion, and interfere with political processes in other countries. These developments intersect with the principles of the United Nations Charter, particularly Article 2(4), which prohibits the threat or use of force against the territorial integrity or political independence of states. Additionally, international humanitarian law (IHL) and human rights treaties—especially those protecting freedom of expression—must be reconciled with emerging norms that address misinformation and the harmful impacts of synthetic media on civilian populations. A major challenge in regulating deepfakes internationally is the lack of cross-border coherence. The speed and global reach of AI-generated content make unilateral state regulation largely ineffective, as harmful deepfakes can easily bypass national borders and legal jurisdictions. Currently, no binding international treaty specifically addresses synthetic media, leaving states to navigate the governance of deepfakes through fragmented national laws, soft-law instruments, and voluntary platform measures (Khan & Jiliani, 2023).

Emerging debates on responsibility focus on both states and private actors. Questions of state responsibility revolve around accountability for cross-border disinformation campaigns, including whether a state can be held liable for using synthetic media to interfere in another country's domestic affairs. Parallel debates consider the corporate responsibility of technology companies under soft-law mechanisms such as the United Nations Guiding Principles on Business and Human Rights (UNGPs), which emphasize due diligence and the prevention of human rights violations. Together, these issues highlight the pressing need for international coordination, norms, and regulatory frameworks that address the transnational and hybrid nature of deepfakes while balancing sovereignty, human rights, and technological innovation (Khan & Usman, 2023).

COMPARATIVE LEGAL RESPONSES

Legal responses to deepfakes vary significantly across jurisdictions, reflecting differences in technological capacity, regulatory philosophy, and societal priorities. In the United States, regulation remains fragmented, with several state-level laws targeting specific harms. California, for example, prohibits the dissemination of political deepfakes close to elections, while Virginia criminalizes non-consensual deepfake pornography. Despite these initiatives, there is currently no comprehensive federal law that addresses the full spectrum of deepfake-related risks, leaving significant gaps in accountability and enforcement (Khan et al., 2023).

The European Union has taken a more proactive approach through the proposed AI Act and the Digital Services Act (DSA), which impose labeling obligations, transparency requirements, and risk classification for synthetic media. These regulations aim to mitigate harms by compelling platforms and developers to adopt preventative measures and provide users with verifiable information about AI-generated content. Nevertheless, loopholes persist, particularly in addressing private harms such as defamation, identity fraud, and evidentiary challenges within judicial proceedings, leaving victims partially unprotected (Khan, 2023).

China's 2022 Deep Synthesis Regulations represent one of the world's most comprehensive legal frameworks for governing synthetic media. These regulations mandate watermarking of AI-generated content, real-name verification for creators, and content auditing by platforms, effectively combining algorithmic accountability with strict state oversight. While highly effective in limiting misuse, this approach raises human rights concerns, particularly regarding freedom of expression and privacy, due to extensive government monitoring and control of online content (Liu et al., 2023).

In South Asia, regulation remains largely reactive and underdeveloped. Pakistan relies on provisions under the Prevention of Electronic Crimes Act (PECA 2016) and the Qanun-e-Shahadat Order, which are insufficient to address the evidentiary and privacy challenges introduced by deepfakes. India has proposed amendments under the Digital India Act to regulate synthetic media, but enforceable standards and comprehensive implementation mechanisms remain absent. These comparative insights highlight a global patchwork of legal approaches, underscoring the need for harmonized frameworks that balance innovation, platform accountability, and individual rights while addressing both public and private harms (Khan & Ximei, 2022).

A THREE-PILLAR FRAMEWORK FOR DEEPFAKE REGULATION

Addressing the multifaceted challenges posed by deepfakes requires a comprehensive regulatory framework that integrates technological, legal, and international measures. The first pillar focuses on establishing a robust authenticity infrastructure. Legal systems must develop technological backbones capable of verifying the origin and integrity of digital content. This includes mandatory watermarking for AI-generated media, cryptographic signatures for all recordings used as evidence, and the creation of national authenticity registries akin to digital evidence vaults. Real-time verification tools should be made accessible to courts, legal practitioners, and regulatory authorities to ensure that evidence can be assessed quickly and reliably, restoring trust in audiovisual content (Khan et al., 2022).

The second pillar emphasizes platform gatekeeping and liability reform. Social media and technology platforms must adopt notice-and-action regimes for harmful deepfakes, while clear legal responsibility should be established for algorithmic amplification of manipulated content. Platforms should be obligated to implement de-amplification mechanisms, provenance tools, and transparent reporting standards that facilitate independent research and regulatory oversight. By assigning explicit duties and accountability measures, this pillar aims to mitigate the rapid dissemination of harmful synthetic media and incentivize proactive governance by intermediaries (Lin & Khan, 2021).

The third pillar addresses cross-border legal harmonization, reflecting the transnational nature of deepfakes. States should negotiate a Multilateral Convention on Synthetic Media and Digital Authenticity to create consistent international standards covering watermarking, evidentiary authentication, and platform responsibilities. In addition, cyber norms developed under the United Nations Group of Governmental Experts (GGE) and Open-Ended Working Group (OEWG) should explicitly prohibit state-sponsored deepfake attacks targeting democratic processes or elections. By combining technological, domestic, and international strategies, this three-pillar framework provides a holistic blueprint for safeguarding evidence integrity, protecting individual rights, and preserving democratic institutions in the era of synthetic media (Khan, 2022).

IMPLICATIONS FOR THE FUTURE: TOWARD A NEW LEGAL EPISTEMOLOGY

The rise of deepfakes presents a profound challenge to the epistemological foundations of law, which have long relied on the assumption that objective truth can be established through reliable evidence. Historically, visual and audio recordings were treated as inherently trustworthy unless there was clear evidence of tampering. However, the emergence of highly realistic synthetic media destabilizes this principle, creating a legal environment in which appearances can no longer be equated with authenticity. This shift compels the legal system to reconceptualize the very nature of evidence, truth, and trust in judicial processes (Khan & Wu, 2021).

Adapting to this new reality requires transformative changes in judicial practice, legal education, and legislative frameworks. Courts will need to integrate proactive authenticity verification mechanisms into procedural norms, including real-time forensic tools, cryptographic verification, and standardized protocols for evaluating AI-manipulated content. Judges and lawyers must receive specialized training in AI forensics and digital literacy, enabling them to identify manipulation, interpret complex algorithmic evidence, and make informed rulings in cases where traditional markers of authenticity are insufficient. Legal education must also evolve, embedding digital competencies and AI ethics into curricula to prepare future practitioners for the evidentiary challenges posed by synthetic media (Abdelrehim Hammad et al., 2021).

Legislatures and policymakers must revise evidentiary rules to accommodate the unique properties of AI-generated content. Traditional doctrines, such as the presumption of authenticity, chain of custody, and expert witness standards, require reform to ensure that the burden of proof and procedural safeguards remain fair and effective in the deepfake era. Additionally, new regulatory mechanisms, such as mandatory watermarking, cryptographic signatures, and platform accountability measures, must be harmonized across jurisdictions to address the cross-border nature of synthetic media (Usman et al., 2021).

Ultimately, the legal system must transition from a paradigm of “seeing is believing” to one in which trust is verifiable, evidence is demonstrably authentic, and the integrity of the judicial process is actively protected. This represents not only a technical challenge but a philosophical and epistemological shift: the law must embrace a more nuanced understanding of truth in the digital age, acknowledging that perception alone is insufficient and that technological, procedural, and institutional safeguards are essential for sustaining justice in an era dominated by AI-generated content (Khan & Wu, 2021).

CONCLUSION

This study has explored the multifaceted challenges posed by deepfakes to legal systems, evidentiary processes, human rights, and international governance. Deepfakes destabilize traditional assumptions of trust in audiovisual evidence, expose individuals to privacy violations, defamation, and identity fraud, and create systemic risks to electoral integrity and national security. Comparative analysis reveals a fragmented global response: while China has implemented stringent state-led regulations, the European Union emphasizes platform accountability and risk assessment, the United States relies on expert testimony within a fragmented legal framework, and South Asian jurisdictions remain largely reactive and under-resourced. These disparities underscore the urgent need for harmonized domestic and international regulatory measures. The proposed three-pillar framework encompassing authenticity infrastructure, platform gatekeeping and liability reform, and cross-border legal harmonization—provides a roadmap for mitigating the harms of synthetic media while preserving innovation and freedom of expression. Courts must integrate proactive authentication mechanisms, judicial training, and revised evidentiary rules, while legislatures and regulators should mandate watermarking, transparency, and algorithmic accountability. International cooperation is essential to address the transnational nature of deepfake dissemination, including through multilateral treaties and cyber norms that prohibit state-sponsored manipulative campaigns.

Future research should focus on developing standardized forensic methodologies for deepfake detection, evaluating the effectiveness of platform liability reforms, and examining the social and psychological impacts of synthetic media on vulnerable populations. Additionally, interdisciplinary studies integrating law, AI, ethics, and governance will be critical to inform adaptive legal frameworks that can respond dynamically to technological evolution. By emphasizing verifiable trust, accountability, and cross-border collaboration, the legal system can evolve to safeguard both individual rights and institutional integrity in the rapidly advancing era of synthetic media.

REFERENCES

Abdelrehim Hammad, A. A., Khan, A., & Soomro, N. E. (2021). Digital Economy Barriers to Trade Regulation Status, Challenges, and China's Response. *International Journal of Social Sciences Perspectives*, *8*(2), 41-49.

Apolo, Y., & Michael, K. (2024). Beyond a reasonable doubt? Audiovisual evidence, AI manipulation, deepfakes, and the law. *IEEE Transactions on Technology and Society*, *5*(2), 156-168.

Ghani, H. U., Gong, X., Haider, A., & Khan, A. (2025). The Indian concern on China-Pakistan Economic Corridor (CPEC) and justification from the perspective of international law. *China and WTO Review*, *11*(2 in press), 25-39.

Ghiurău, D., & Popescu, D. E. (2024). Distinguishing reality from AI: approaches for detecting synthetic content. *Computers*, *14*(1), 1.

Gregory, S. (2024). *From Social Media to Deepfakes: Participatory human rights witnessing and advocacy using audiovisual media, incorporating the emerging impacts of deceptive AI and technologies for authenticity and trust (2007-22)* (Doctoral dissertation, University of Westminster).

Gupta, D., & Fatunmbi, T. O. (2024). Generative AI and Deep fake s: Ethical Implications and Detection Techniques. *Journal of Science, Technology and Engineering Research*, *2*(1), 45-56.

Hausknecht, A. (2025). The Impact of Deepfakes on Trust in User-Generated Evidence. *Deepfakes and the Law: Challenges, Responses, and Critique (Taylor and Francis)*.

HUI, Z., HAIDER, A., & Khan, A. (2025). BIODIVERSITY VS. DEVELOPMENT: SUPREME COURT'S VERDICT IN MK RANJITSINH V. UNION OF INDIA.

Hui, Z., Haider, A., & Khan, A. (2025). International trade and plastic waste in oceans: legal and policy challenges. *Frontiers in Marine Science*, *12*, 1627829.

Kahn, A., & Wu, X. (2020). Impact of digital economy on intellectual property law. *J. Pol. & L.*, *13*, 117.

Khan, A. (2018). Autonomous Weapons and Their Compliance with International Humanitarian Law (LLM Thesis). *Traditional Journal of Law*.

Khan, A. (2022). E-commerce Regulations in Emerging Era: The Role of WTO for Resolving the Complexities of Electronic Trade. *ASR Chiang Mai University Journal Of Social Sciences And Humanities*.

Khan, A. (2023). Rules on Digital Trade in the Light of WTO Agreements. *PhD Law Dissertation, School of Law, Zhengzhou University China*.

Khan, A. (2024). The Emergence of the Fourth Industrial Revolution and its Impact on International Trade. *ASR: CMU Journal of Social Sciences and Humanities (2024) Vol*, *11*.

Khan, A. (2024). The intersection of artificial intelligence and international trade laws: Challenges and opportunities. *IIUMLJ*, *32*, 103.

Khan, A. S. I. F., Amjad, S. O. H. A. I. L., & Usman, M. U. H. A. M. M. A. D. (2020). The Evolution of Human Rights Law in the Age of Globalization. *Pak*

Khan, A., & Hussain Shah Jillani, M. A. (2019). Killer robots and their compliance with the principles of law of war. *JL & Soc'y*, *50*, 55.

Khan, A., & Jiliani, M. A. H. S. (2023). Expanding The Boundaries Of Jurisprudence In The Era Of Technological Advancements. *IIUMLJ*, *31*, 393.

Khan, A., & Ullah, M. (2024). The Pakistan-China FTA: legal challenges and solutions for marine environmental protection. *Frontiers in Marine Science*, *11*, 1478669.

Khan, A., & Usman, M. (2023). The effectiveness of international law: a comparative analysis. *International Journal of Contemporary Issues in Social Sciences*, *2*(3), 780-786.

Khan, A., & Wu, X. (2021). Bridging the Digital Divide in the Digital Economy with Reference to Intellectual Property. *Journal of Law and Political Sciences*, *28*(03), 256-263.

Khan, A., & Wu, X. (2021). Reforms for culmination of the deadlock in appellate body of WTO: An agenda of saving the multilateral trading system. *Journal of Humanities, Social and Management Sciences (JHSMS)*, *2*(1), 50-62.

Khan, A., & Ximei, W. (2022). Digital economy and environmental sustainability: do information communication and technology (ICT) and economic complexity matter?. *International journal of environmental research and public health*, *19*(19), 12301.

Khan, A., Abd Elrhim, A. A., & Soomro, N. E. (2021). China Perspective in Reforming of the World Trade Organization. *J. Pol. & L.*, *14*, 104.

Khan, A., Amjad, S., & Usman, M. (2020). The Role of Customary International Law in Contemporary International Relations. *International Review of Social Sciences*, *8*(08), 259-265.

Khan, A., Jillani, M. A. H. S., Abdelrehim Hammad, A. A., & Soomro, N. E. H. (2021). Plurilateral negotiation of WTO E-commerce in the context of digital economy: Recent issues and developments. *Journal of Law and Political Sciences*.

Khan, A., Jillani, M. A. H. S., Ullah, M., & Khan, M. (2025). Regulatory strategies for combatting money laundering in the era of digital trade. *Journal of Money Laundering Control*, *28*(2), 408-423.

Khan, A., Khan, A. S., & Khan, I. (2022). Responsibility Of Killer Robots For Causing Civilian Harm: A Critique Of Ai Application In Warfare Doctrine. *Pakistan Journal of International Affairs*, *5*(1), 15-33.

Khan, A., Usman, M., & Amjad, S. (2020). Enforcing Economic, Social, and Cultural Rights: A Global Imperative. *International Review of Social Sciences (IRSS)*, *8*(09).

Khan, A., Usman, M., & Amjad, S. (2023). The digital age legal revolution: taped's trailblazing influence. *International journal of contemporary issues in social sciences*, *2*(4), 524-535.

Lin, L. S. (2025). Organisational Challenges in US Law Enforcement's Response to AI-Driven Cybercrime and Deepfake Fraud. *Laws*, *14*(4), 46.

Lin, S., & Khan, A. (2021). The Concept of E-sports in Digital Era: A Case Study of China.

Liu, X., Khan, M., & Khan, A. (2023). The Law and Practice of Global ICT Standardization by Olia Kanevskaia [CUP, Cambridge, 2023, xxvi+ 361pp, ISBN: 978-1-0093-00575,£ 95.00 (h/bk)]. *International & Comparative Law Quarterly*, *72*(4), 1094-1097.

Llorente, R. V. (2024). Deepfakes in the Dock: Preparing International Justice for Generative AI. *Scitech Lawyer*, *20*(2), 28-33.

Makhambetsaliyev, D. (2025). Deepfakes And The Legal Construction Of Identity: Between Personhood And Performance. *This article has not been published anywhere yet*.

Mathlouthi, N., Haider, A., Khan, A., & Ahmad, N. (2025). The role of Hainan Free Trade Port in shaping China's WTO commitments and international trade policies. *China and WTO Review*, *11*(1), 71-82.

MKPO, D. C. (2025). Is Truth And Integrity Of Justice On Trial In The Age Of Artificial Intelligence (Ai)? Re-Assessing Digital Evidence In The Context Of Ai-Generated Evidence Such As Deepfakes. *International Review Of Law And Jurisprudence (IRLJ)*, *7*(1).

Murray, M. D. (2025). Visual Legal Rhetoric in the Age of Generative AI and Deepfakes: Renaissance or Dark Ages?. *SMU Sci. & Tech. L. Rev.*, *28*, 199.

Paris, B. (2021). Configuring fakes: Digitized bodies, the politics of evidence, and agency. *Social Media+ Society*, *7*(4), 20563051211062919.

Shi, Y. (2025). The Crisis of Deepfakes on Generation and Detection: Are Current Responses Sufficient?. *iSCHANNEL*, *19*(1).

Shirish, A., & Komal, S. (2023). A socio-legal inquiry on deepfakes. *Cal. W. Int'l LJ*, *54*, 517.

Usman, M. U. H. A. M. M. A. D., Khan, A. S. I. F., & Amjad, S. O. H. A. I. L. (2021). State Responsibility and International Law: Bridging the Gap.

Williams, B. (2025). AI And the Future Of Digital Evidence In Law And Journalism.

Williams, B. (2025). Regulatory Approaches To Ai-Generated Content And Privacy Protection.