



Journal for Social
Science Archives

Journal for Social Science Archives

Online ISSN: 3006-3310

Print ISSN: 3006-3302

Volume 4, Number 1, 2026, Pages 282 – 296

Journal Home Page

<https://jssarchives.com/index.php/Journal/about>



Regulating Synthetic Realities: A Comparative Analysis of Global Legal Responses to AI-Generated Deepfakes

Akbar Ali Datoo¹, Alishah Aziz Gulzar² & Waleed Ahmed Khan³, Dr. Rana Shahzad Qaiser⁴

¹M.Phil. in Criminology | AI, Digital Criminology, Intelligence & Security, Department of Criminology, University of Karachi, Email: akbarali.datoo@gmail.com

²Criminologist & Cyber Security Analyst, Founder & President SARG (Strategic Academic Research Group), Senior Vice President - Diplomatic Forum for Socio-Economic Foundation, Email: alishahgulzar@live.com

³M.Phil. Scholar, Teaching Assistant, Department of Criminology, University of Karachi, Email: waleedahmedkhanofficial@gmail.com

⁴Director General, Science and Information Technology Department, Government of Sindh, Email: dr.rana@sindh.gov.pk

ARTICLE INFO

Article History:

Received: December 10, 2025
Revised: January 05, 2026
Accepted: January 20, 2026
Available Online: February 02, 2026

Keywords:

Deepfakes, Comparative law, Regulatory Frameworks, AI Governance, Cybercrime, Synthetic Media, Legal Reform

Corresponding Author:

Akbar Ali Datoo

Email:

akbarali.datoo@gmail.com

ABSTRACT

The proliferation of AI-generated deepfakes has prompted diverse regulatory responses across jurisdictions, revealing fundamental tensions between technological innovation, individual rights, and societal protection. This article provides a comparative analysis of legal and policy frameworks addressing deepfake-related harms in the United States, European Union, China, South Korea, and Pakistan, examining how different legal traditions and institutional contexts shape regulatory approaches. It identifies three emerging regulatory paradigms: application-based regulation targeting specific high-harm uses; life cycle-based regulation addressing synthetic media at different phases of creation and dissemination; and subject-based regulation focusing on platform and service provider responsibilities. The analysis reveals that while no jurisdiction has fully resolved the challenges posed by deepfakes, proactive legal frameworks, specialized institutional capacity, and preventive strategies significantly enhance response effectiveness. Conversely, reliance on technology-neutral cybercrime provisions, fragmented institutional approaches, and limited forensic capacity substantially constrains regulatory efficacy, particularly in developing jurisdictions. The article concludes by identifying best practices and transferable lessons for jurisdictions seeking to strengthen their legal and institutional responses to deepfake-related harms.



1. Introduction

The rapid proliferation of AI-generated deepfakes has created a regulatory challenge that transcends national boundaries, legal traditions, and institutional capacities. Synthetic media audio, video, and images generated through deep learning models can be created, disseminated, and consumed across jurisdictions with unprecedented speed and scale, exposing fundamental limitations in existing legal frameworks that were not designed to address AI-generated content. The challenge is compounded by the distinctive characteristics of deepfake-related harms: unlike traditional forms of fraud, defamation, or harassment, deepfake offenses often involve synthetic content that may have no corresponding real-world event, complicating evidentiary standards and burden of proof (Chesney & Citron, 2019). The transnational nature of digital content distribution means that offenses may span multiple jurisdictions, creating jurisdictional fragmentation and enforcement gaps. Moreover, the rapid evolution of generative AI technologies means that legal frameworks risk obsolescence almost as soon as they are enacted (Sandoval et al., 2024).

The regulatory landscape for deepfakes has evolved significantly since synthetic media first emerged as a mainstream concern. What began as a niche technological curiosity has transformed into a pervasive societal challenge with implications for democratic processes, individual privacy, financial systems, and national security. Deepfake production is increasingly being offered "as a service," powered by autonomous AI systems capable of executing multi-step fraud schemes from synthetic job candidates passing live video interviews to romance scams depleting retirement accounts. The numbers are stark: Gartner projects that one in four job candidate profiles globally will be fake by 2028, while Deloitte estimates \$40 billion in US fraud losses from generative AI by 2027.

This article provides a comparative analysis of legal and policy frameworks addressing deepfake-related harms across five jurisdictions: the United States, the European Union, China, South Korea, and Pakistan. These jurisdictions were selected to represent diverse legal traditions, regulatory approaches, and levels of technological and institutional capacity. The analysis examines how different legal systems have responded to the challenges posed by deepfakes, identifying emerging regulatory paradigms, best practices, and persistent gaps.

The analysis proceeds in four parts. Part Two examines the regulatory approaches of the United States, European Union, China, and South Korea, identifying three emerging paradigms: application-based regulation, life cycle-based regulation, and subject-based regulation. Part Three analyzes Pakistan's legal and institutional framework, highlighting the challenges facing developing jurisdictions and the emergence of new institutional capacities. Part Four identifies best practices and transferable lessons for jurisdictions seeking to strengthen their regulatory responses. The article concludes by reflecting on the future of deepfake governance in an era of rapidly evolving AI capabilities.

2. Emerging Regulatory Paradigms

2.1 The United States: Application-Based Regulation

The United States has pursued a predominantly application-based regulatory approach, targeting specific high-harm uses of deepfake technology rather than adopting comprehensive synthetic media legislation. This approach reflects broader American governance patterns that privilege state-level experimentation and sectoral regulation over comprehensive federal frameworks. The result is a fragmented legal landscape characterized by significant variation across jurisdictions and persistent gaps in coverage.

State legislatures have led the regulatory response, enacting 169 laws since 2022 and introducing 146 bills in 2025 alone. Since 2022, 46 states have enacted deepfake legislation. Early legislative efforts at the state level targeted two primary areas of concern: political manipulation and non-consensual pornography. Texas and Virginia enacted statutes criminalizing deepfake videos disseminated during election periods, establishing specific penalties for political deepfakes that could influence electoral outcomes. California developed legislation addressing non-consensual synthetic pornography, creating both civil remedies and criminal liability for the creation and distribution of deepfake pornography.

At the federal level, regulatory efforts have been more targeted but increasingly significant. The TAKE IT DOWN Act (Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act), signed into law on May 19, 2025, represents the most comprehensive federal legislation addressing deepfake-related harms. The Act criminalizes the nonconsensual publication of intimate images, including "digital forgeries" (i.e., deepfakes), in certain circumstances. It also requires covered platforms to implement a "notice-and-removal" process to remove such images at the depicted individual's request within 48 hours of notification. The criminal prohibition took effect immediately upon signing, while covered platforms had one year (until May 19, 2026) to establish the required notice-and-removal process. The Act solves the problem of inconsistent, or non-existent, legislation protecting victims of deepfake pornographic images at the state level.

Other federal efforts have addressed specific applications of deepfake technology. The Federal Trade Commission has pursued enforcement actions against companies using deepfake technology for deceptive practices, while the Federal Communications Commission has addressed voice cloning in the context of robocalls. Legislative proposals have been introduced to address deepfake pornography, election interference, and identity fraud, but comprehensive federal legislation has not been enacted.

The application-based approach has significant limitations. Political deepfakes face the strictest scrutiny, but definitional inconsistencies across states complicate enforcement. Gaps remain regarding deepfake applications not explicitly addressed in existing statutes. The absence of a comprehensive federal framework creates coordination challenges for interstate investigations and cross-border prosecutions. Moreover, the reactive nature of application-based regulation means that new forms of deepfake harm may emerge before legislative responses can be developed. As of

June 2026, more than 30 states have passed some type of law regulating AI-generated media that applies to political advertising, with various exemptions for satire and parody. This patchwork of state laws creates compliance complexity for businesses operating across multiple jurisdictions.

2.2 The European Union: Life Cycle-Based Regulation

The European Union has adopted a life cycle-based regulatory paradigm, addressing deepfakes at different phases of creation, dissemination, and impact through multiple legal instruments. This approach reflects the EU's broader regulatory philosophy of comprehensive, principle-based governance that emphasizes transparency, accountability, and risk mitigation. The EU framework is notable for its ambition, its emphasis on transparency obligations, and its attempt to address synthetic media throughout the AI lifecycle.

The AI Act represents the most ambitious regulatory effort globally, classifying deepfake technologies within broader AI governance frameworks. Under the AI Act, certain applications of deepfake technology are subject to transparency requirements, including mandatory disclosure that content is AI-generated. Article 3(60) of the Act defines a "deepfake" as any image, audio, or video content created or modified by AI that imitates real people, objects, places, entities, or events in a way that could mislead someone into believing it is genuine. High-risk applications face additional obligations regarding risk assessment, data governance, and human oversight.

The transparency requirements set out in Article 50 of the EU AI Act are designed to make it clear when content has been generated or altered by AI, including deepfakes. Article 50(2) requires providers of generative AI systems to ensure that the outputs of the AI system are marked in a machine-readable format and detectable as artificially generated or manipulated. Article 50(4) requires deployers of AI systems that generate or manipulate image, audio, or video content constituting a deepfake to disclose that the content has been artificially generated or manipulated, except where the use is authorized by law for criminal investigation or prosecution.

The transparency obligations will apply from 2 August 2026. The European Commission published a First Draft Code of Practice on Transparency of AI-Generated Content on 17 December 2025, designed to aid compliance with the Article 50 obligations. The Code consists of two sections: the first covers marking and detection rules applicable to providers of generative AI systems, while the second covers rules for deployers regarding labelling deepfakes and AI-generated text on matters of public interest. A further draft was expected in March 2026, with a final code anticipated in June 2026 ahead of the entry into force of Article 50 in August 2026.

The AI Act has undergone revisions through the Digital Omnibus process. The list of prohibited AI practices will be expanded to include sexualized deepfakes: effective as of December 2, 2026, AI systems that generate non-consensual sexual or intimate content, as well as systems that generate depictions of child sexual abuse or assist in their creation, will be banned. For existing AI systems, the deadline for implementing the labelling of AI-generated content has been extended to December 2, 2026. For other transparency requirements and new AI systems, the original effective date of August 2, 2026, remains unchanged.

The Digital Services Act complements the AI Act by establishing platform accountability measures for harmful content, including synthetic media. Platforms are required to implement content moderation systems, provide transparency about algorithmic processes, and cooperate with law enforcement in responding to illegal content. The Code of Practice on Disinformation incorporates synthetic media considerations into broader information integrity frameworks.

The life cycle-based approach offers several advantages. By addressing deepfakes at multiple phases, the EU framework can respond to both creation and dissemination of synthetic media. The emphasis on transparency and accountability creates positive obligations for technology developers and platforms, shifting responsibility upstream. However, implementation challenges persist, particularly regarding the distinction between prohibited synthetic media and permissible satire or artistic expression. The rapid evolution of AI technologies also means that regulatory frameworks risk obsolescence, though the Omnibus process demonstrates the EU's commitment to adaptive governance.

2.3 China: Subject-Based Regulation

China has adopted a subject-based regulatory paradigm, focusing on service providers and platforms as the primary subjects of regulation. This approach reflects China's broader governance model, which emphasizes state oversight of digital platforms and content moderation. The Chinese framework is notable for its directness, its emphasis on mandatory labelling, and its integration with broader content governance mechanisms.

In 2019, China enacted regulations requiring platforms to label AI-generated content and prohibiting the use of deepfakes for defamation, destabilization, or social harm. These regulations require platforms to implement technical detection systems, report synthetic content to authorities, and comply with content removal orders. The regulatory framework establishes positive obligations for platform gatekeepers, shifting the burden of content moderation from the state to private actors.

The regulatory architecture was significantly strengthened with the implementation of the Measures for Identifying AI-Generated Synthetic Content on September 1, 2025. This regulation requires AI-generated content published online to include both visible labels for audiences and invisible metadata for tracing responsibility. The measures require AI content to be prominently labelled with an "AI-generated" mark within the first three seconds of video content, with the mark size not less than 5% of the screen width. This system establishes labelling as a mandatory threshold for publication. The accompanying mandatory national standard, "Cybersecurity Technology Methods for Identifying AI-Generated Synthetic Content" (GB 45438-2025), and the service provider coding rules (TC260-PG-20252A) were simultaneously issued, marking a new stage in China's AI content governance.

Research indicates that the AI labelling policy has had measurable effects. A sampling study conducted by an AI governance team at a western university in the fourth quarter of 2025 found that users' "questioning awareness" of content from unknown sources increased by nearly 40% after the implementation of the AI labelling policy. In a cross-border AI false news tracing case,

the accountability cycle was shortened from an average of 72 hours to 12 hours due to the ability of implicit identifiers to quickly lock onto the generation tools and dissemination nodes of content. Testing by the China Software Testing Center showed that the full compliance rate of mainstream multimodal large models in text-to-image and text-to-video scenarios had reached 80% to 90%.

However, implementation challenges persist. The "last mile" of regulation has not been fully addressed, with some platforms' AI identifiers being "too small in font, low in colour contrast, and marginalized in position," rendering them ineffective. The market for "AI watermark removal" and "traceless identifier removal" services remains active. Once implicit metadata is stripped and explicit watermarks are removed, the traceability chain is broken. As one observer noted, labels "only guard against gentlemen, not villains". Yet, as commentators have argued, the significance of labels lies not in "stopping bad actors" but in reconstructing the underlying protocol of trust similar to ingredient labels on food packaging, which do not prevent counterfeiters but help consumers establish rational expectations.

The subject-based approach has several advantages. By focusing regulatory obligations on platforms, China can achieve widespread compliance without requiring individual enforcement actions. The emphasis on labelling and transparency enables users to identify synthetic content, reducing the risk of deception. However, the regulatory framework's integration with broader content control mechanisms raises concerns regarding collateral restrictions on legitimate expression. The emphasis on platform obligations may also create incentives for over-censorship, as platforms seek to avoid regulatory sanctions. The challenge of "identifier evasion" demonstrates that technological solutions alone cannot address the deepfake challenge.

2.4 South Korea: Targeted Criminalization and Emerging Comprehensive Framework

South Korea has emerged as a particularly instructive case in deepfake regulation, combining targeted criminalization of specific applications with emerging comprehensive frameworks. The South Korean experience demonstrates both the possibilities and limitations of legislative responses to synthetic media harms.

South Korea's initial regulatory response focused on non-consensual synthetic pornography, driven by widespread deepfake sexual abuse cases involving K-pop celebrities and ordinary citizens. The country amended the Act on Special Cases Concerning the Punishment of Sexual Crimes to criminalize the creation and distribution of deepfake pornography, establishing specific penalties for sexually exploitative synthetic media with sentences exceeding those for traditional pornography offenses. This legislative response was prompted by public outrage over the scale of deepfake sexual abuse, including revelations that male students at hundreds of schools and universities created Telegram channels to produce and share synthetic explicit videos of female teachers, classmates, and family members.

However, the limitations of this targeted approach became apparent in a significant 2025 court ruling. A South Korean court ruled that the distribution of AI-generated nude images cannot be punished under current deepfake pornography laws unless the depicted individual is a real, identifiable person. This decision highlighted a critical gap in the legal framework: synthetic

content that does not correspond to an identifiable real person falls outside the scope of existing prohibitions. The ruling raised fundamental questions about the effectiveness of the "Deepfake Prevention Law" and whether current legal frameworks adequately address the full range of deepfake-related harms.

In response to these limitations, South Korea has begun developing more comprehensive regulatory approaches. On December 10, 2025, the government announced a new set of measures aimed at curbing AI-driven false and misleading advertising, including deepfakes, under the AI Act scheduled to take effect in January 2026. The measures respond to the proliferation of AI-generated ads using digitally fabricated experts or deepfake videos and audios of celebrities promoting products from weight-loss pills to illegal gambling sites. The measures seek to prevent distribution of such advertisements, promptly block them once distributed, and strengthen financial sanctions against violators.

The National Assembly is also pushing for amendments to require mandatory watermarks on AI-generated content, with the possibility of imprisonment for removing or falsifying them. This represents a shift toward a more comprehensive, life cycle-based approach that addresses synthetic content at multiple stages. Proposed legislation would specify platform operators' deletion obligations, strengthen the designation of domestic agents for foreign operators and joint liability, and refine criminal punishment provisions to effectively protect citizens from digital sexual crimes.

South Korea's evolving approach demonstrates the importance of responding to specific, high-harm applications of deepfake technology while also developing comprehensive frameworks that address emerging forms of harm. The focus on sexual offenses, while important, initially left other forms of deepfake harm such as financial fraud, political manipulation, and identity theft less comprehensively addressed. The reactive nature of the legislative response also raised questions about whether legal frameworks can keep pace with emerging deepfake applications. The shift toward more comprehensive regulation suggests recognition that targeted criminalization alone is insufficient.

3. Pakistan: Legal and Institutional Challenges

3.1 The Prevention of Electronic Crimes Act (PECA) 2016 and Its Amendments

Pakistan's primary legal framework for addressing deepfake-related harms is the Prevention of Electronic Crimes Act (PECA) 2016, which provides broad coverage of cyber-enabled offenses but lacks explicit provisions for synthetic media. Several provisions of PECA are potentially applicable to deepfake-related conduct, including sections dealing with cyber harassment, unauthorized use of identity information, online defamation, and misrepresentation.

However, these provisions were drafted before the widespread emergence of AI-generated synthetic media and do not specifically address issues such as synthetic manipulation, algorithmic fabrication, or the intent behind deepfake creation and dissemination. Prosecutors and courts must engage in considerable interpretive work to apply these provisions to deepfake cases, creating uncertainty and inconsistency in legal outcomes. Digital rights advocates have noted that "data

protection laws don't fully cover AI-related issues" and that "Pakistan needs to pass the Personal Data Protection Bill and clearly define AI-related offences, as outdated laws like Peca no longer suffice".

In January 2025, Pakistan's parliament enacted amendments to PECA, criminalizing the spread of "fake or false information" with penalties of fines or up to three years in prison. The amendment introduced a criminal offense against those perpetrating so-called "false and fake information". However, the amendments faced significant backlash from opposition, media, and civil society. The broad and vague definition of "fake or false information" raised concerns about potential misuse and restrictions on legitimate expression. The amendments were presented in and passed by the National Assembly on January 23, 2025, and referred to the Standing Committee the next day by the Senate after backlash.

The limitations of PECA are evident in several areas. The Act does not recognize synthetic or AI-generated content as a separate evidentiary category, making prosecution difficult when authenticity is contested. Sections addressing harassment and modesty require proof of malicious intent, which becomes complex when content appears hyper-realistic. Voice deepfakes, increasingly used in financial scams, do not fit neatly into existing statutory definitions. Courts have yet to develop standardized criteria for evaluating whether deepfake material meets PECA's threshold of "false information." A critical legal analysis examining the efficacy of Pakistan's existing and emerging legal frameworks in combating the privacy-invasive harms of deepfakes scrutinized the provisions of PECA 2016 and the draft Personal Data Protection Bill to assess their applicability and sufficiency.

3.2 Institutional Capacity Constraints

Beyond legal framework limitations, Pakistan's institutional capacity to address deepfake-related harms has been significantly constrained. The National Cyber Crimes Investigation Agency (NCCIA), the primary investigative body responsible for cybercrime enforcement, faces substantial challenges in detecting, investigating, and prosecuting deepfake-related offenses.

Forensic and technical capabilities have historically been severely limited. The NCCIA lacked specialized tools capable of detecting and verifying deepfake audio-visual content, relying instead on basic digital forensics infrastructure that was inadequate for synthetic media analysis. Training programs on artificial intelligence, machine learning, and deepfake detection technologies were minimal or non-existent, leaving investigators without the expertise required to identify synthetic content or preserve digital evidence properly.

Judicial capacity has been similarly constrained. Most judges lacked specialized training in digital forensics and AI-generated media, creating uncertainty about how to assess contested synthetic evidence. The absence of standardized protocols for evaluating deepfake authenticity led to inconsistency in how different courts approached such evidence. Courts often exercised substantial caution when presented with contested synthetic media evidence, which could lead to delays in judicial proceedings or reluctance to rely on contested evidence even when its authenticity had been reasonably established.

A dedicated Deep Fake Lab was established to analyze fabricated videos and audio clips circulating online. However, questions remained about its capacity, independence, and the admissibility of its findings in court. The NCCIA warned citizens against sharing unverified content, urging a zero-tolerance approach towards false or harmful content targeting state institutions. Social media platforms were issued notices to remove fake content within 24 hours, with repeated violations resulting in permanent account suspensions.

3.3 Emerging Institutional Responses

Despite these constraints, Pakistan has begun to develop institutional responses to deepfake-related threats. The most significant development is the establishment of the National Forensic Agency (NFA), which became fully operational in April 2026. The NFA represents Pakistan's first modern, state-of-the-art forensic agency, providing internationally standardized services and analysis within Pakistan.

The National Forensics Agency Bill, passed by the Senate in December 2024, stated that "the National Forensics Agency will integrate digital and cyber forensics to combat crimes involving electronic devices, deepfakes, and other electronic offences". The agency's digital forensics division is equipped with modern tools to handle deepfake detection, computer and mobile analysis, audio-video examination, network forensics, and drone investigations. In the field of digital forensics, facilities such as deepfake forensics, computer and mobile forensics, audio-video analysis, network forensics, and drone forensics are being provided. Over the past year, 1,500 digital forensic cases have been reported and processed.

The NFA's first phase has been completed, while work on the second phase featuring 25 additional specialized laboratories will begin soon. The Interior Minister directed that sub-offices be established in Khyber Pakhtunkhwa, Balochistan, AJK, and Gilgit Baltistan. The agency is providing state-of-the-art services that were previously unavailable with other institutions. The NFA will support all provinces and law enforcement agencies.

The Personal Data Protection Bill 2023, although still awaiting enactment, has significant relevance for deepfake regulation. The bill emphasizes consent, lawful processing of personal data, and protection against misuse of personal information principles that directly apply to deepfake creation and dissemination. While it does not directly mention deepfakes, its provisions related to biometric data, facial images, and sensitive personal information could offer substantial indirect protection against non-consensual deepfake creation. The bill is applicable to data controllers and processors not incorporated in Pakistan operating digitally or non-digitally in Pakistan and involved in commercial or non-commercial activity.

However, these institutional developments remain nascent. The NFA has only recently become operational, and the Personal Data Protection Bill has not been enacted. Coordination between regulatory authorities and law enforcement agencies remains limited, and public awareness campaigns targeting deepfake risks are minimal. Without sustained investment and institutional reform, these gaps will continue to undermine Pakistan's ability to respond effectively to deepfake-related harms. The case of journalist Benazir Shah, who was targeted in a deepfake video in

November 2025, illustrates the challenges: while the information minister condemned the attack, Shah made it clear she would not pursue a case under PECA through the NCCIA, reflecting a lack of confidence in the legal system's capacity to address such offenses effectively.

4. Best Practices and Transferable Lessons

4.1 Clear Legal Definitions and Explicit Criminalization

The comparative analysis reveals that jurisdictions with clear legal definitions of deepfakes and explicit criminalization of high-harm uses demonstrate greater regulatory effectiveness. Explicit statutory language reduces interpretive ambiguity, enhances consistency in enforcement, and provides clearer guidance for prosecutors, judges, and law enforcement.

The EU AI Act's definition of deepfakes as "image, audio, or video content created or modified by AI that imitates real people, objects, places, entities or events in a way that could mislead someone into believing it is genuine" provides a useful model. This definition balances breadth with sufficient specificity to distinguish between harmful misuse and legitimate applications. The TAKE IT DOWN Act's inclusion of "digital forgeries" within the definition of intimate images demonstrates how existing legal categories can be extended to address synthetic content.

Best practices include: defining deepfakes and synthetic media with sufficient precision to distinguish between harmful misuse and legitimate applications; criminalizing specific high-harm uses, including non-consensual synthetic pornography, election interference, and identity fraud; establishing proportionate penalties that reflect the distinctive harm of deepfake offenses; and providing civil remedies for victims, including rights to content removal and compensation.

4.2 Institutional Specialization and Capacity Building

Effective deepfake governance requires specialized institutional capacity. Jurisdictions that have established dedicated cyber or AI-focused units within law enforcement agencies, equipped with advanced forensic tools and specialized training, demonstrate greater investigative effectiveness.

The establishment of Pakistan's National Forensic Agency with dedicated deepfake forensics capabilities represents a significant step forward. The agency's provision of deepfake identification, computer and mobile analysis, audio-video examination, network forensics, and drone investigations demonstrate the importance of comprehensive forensic capacity. The agency's processing of 1,500 digital forensic cases in its first year indicates the scale of demand for such services.

Best practices include: establishing specialized investigative units with expertise in synthetic media detection and analysis; investing in forensic infrastructure capable of authenticating AI-generated content; developing training programs for prosecutors and judges on AI technologies and evidentiary standards; and creating protocols for inter-agency coordination and information sharing.

4.3 Preventive Strategies and Public Awareness

Beyond enforcement, preventive strategies play a crucial role in mitigating deepfake-related harms. Jurisdictions that have invested in public awareness campaigns, media literacy programs, and platform accountability measures demonstrate greater societal resilience to synthetic media manipulation.

China's mandatory labelling requirements, while facing implementation challenges, represent an ambitious preventive strategy. The requirement that AI content be prominently labelled with an "AI-generated" mark within the first three seconds of video content, with the mark size not less than 5% of the screen width, establishes labelling as a mandatory threshold for publication. Research indicating that users' "questioning awareness" of content from unknown sources increased by nearly 40% after implementation suggests that such measures can have measurable effects on public skepticism and media literacy.

The EU's transparency requirements, which will require clear labelling of deepfakes and AI-generated content on matters of public interest from August 2026, represent another approach to prevention through transparency. The Code of Practice on Transparency provides practical guidance for compliance.

Best practices include: developing public awareness campaigns that inform citizens about deepfake risks and prevention strategies; integrating media and digital literacy into formal education systems; establishing platform accountability mechanisms, including content moderation requirements and transparency obligations; and promoting ethical AI development through industry self-regulation and public-private partnerships.

4.4 International Cooperation and Harmonization

The transnational nature of deepfake-related crimes necessitates international cooperation and regulatory harmonization. Jurisdictions that have engaged in international norm-setting initiatives, mutual legal assistance agreements, and cross-border enforcement cooperation demonstrate greater capacity to address transnational deepfake offenses.

The EU's extraterritorial application of the AI Act, which reaches providers and deployers serving EU users regardless of where they are established, demonstrates how jurisdictions can extend their regulatory reach. However, such approaches also create compliance complexity for businesses operating across multiple jurisdictions.

Best practices include: participating in international forums on AI governance and cybercrime; developing mutual legal assistance mechanisms for cross-border deepfake investigations; harmonizing legal definitions and evidentiary standards to facilitate international cooperation; and establishing partnerships with technology companies to enable content takedown and evidence preservation.

5. Conclusion

The comparative analysis reveals that while no jurisdiction has fully resolved the challenges posed by AI-generated deepfakes, proactive legal frameworks, specialized institutional capacity, and preventive strategies significantly enhance response effectiveness. The United States' application-based approach, the European Union's life cycle-based regulation, China's subject-based paradigm, and South Korea's evolving comprehensive framework each offer valuable insights while revealing persistent gaps.

The United States demonstrates the benefits and limitations of federalism in addressing synthetic media. While state-level experimentation has produced a rich body of legislation, fragmentation creates compliance complexity and enforcement gaps. The TAKE IT DOWN Act represents progress at the federal level, but comprehensive federal legislation remains elusive.

The European Union's life cycle-based approach offers the most comprehensive framework, addressing deepfakes at multiple phases through the AI Act, Digital Services Act, and Code of Practice on Disinformation. The emphasis on transparency and accountability creates positive obligations for technology developers and platforms, shifting responsibility upstream. However, implementation challenges persist, and the rapid evolution of AI technologies means that regulatory frameworks risk obsolescence.

China's subject-based approach demonstrates the potential of mandatory labelling requirements to reshape public awareness and platform behaviour. However, implementation challenges, including identifier evasion and inconsistent compliance, highlight the limitations of purely technological solutions.

South Korea's evolution from targeted criminalization to comprehensive framework illustrates the importance of adaptive governance. The limitations of the initial approach, highlighted by the court ruling that synthetic content without an identifiable victim falls outside the scope of existing prohibitions, prompted movement toward more comprehensive regulation.

For developing jurisdictions such as Pakistan, the comparative analysis offers both cautionary lessons and actionable guidance. Reliance on technology-neutral cybercrime provisions, fragmented institutional approaches, and limited forensic capacity substantially constrains regulatory efficacy. However, emerging institutional developments including the National Forensic Agency's deepfake forensics capabilities and the Personal Data Protection Bill suggest pathways toward more effective responses.

The challenge of regulating deepfakes is not static. As generative AI technologies continue to evolve, so too will the forms of harm they enable and the regulatory responses they demand. Effective governance requires continuous adaptation, interdisciplinary collaboration, and sustained investment in institutional capacity. The choices made by jurisdictions today will shape whether artificial intelligence serves as a tool for social progress or a source of enduring injustice.

The proliferation of deepfakes-as-a-service, powered by autonomous AI systems capable of executing multi-step fraud schemes, suggests that the regulatory challenge will only intensify. As

Gartner projects that one in four job candidate profiles globally will be fake by 2028, and Deloitte estimates \$40 billion in US fraud losses from generative AI by 2027, the urgency of effective regulation cannot be overstated.

Ultimately, the regulation of synthetic realities requires a multi-faceted approach that combines clear legal definitions, specialized institutional capacity, preventive strategies, and international cooperation. No single jurisdiction has found the perfect balance between innovation, rights protection, and societal safety. But the comparative analysis presented here suggests that jurisdictions that invest in comprehensive frameworks, adaptive governance, and institutional capacity are better positioned to navigate the challenges of the synthetic age.

References

1. Chesney, R., & Citron, D. (2019). Deepfakes and the new disinformation war. *Foreign Affairs*, 98(1), 147–155.
2. European Commission. (2023). Proposal for a regulation on artificial intelligence (AI Act).
3. European Commission. (2025). First draft code of practice on transparency of AI-generated content. <https://digital-strategy.ec.europa.eu/en/library/first-draft-code-practice-transparency-ai-generated-content>
4. European Commission. (2026). Code of practice on transparency of AI-generated content. <https://digital-strategy.ec.europa.eu/en/library/code-practice-transparency-ai-generated-content>
5. Goodfellow, I., Pouget-Abadie, J., Mirza, M., et al. (2014). Generative adversarial nets. *Advances in Neural Information Processing Systems*, 27, 2672–2680.
6. Gregory, S. (2021). Deepfakes and the epistemic crisis. *Journal of Democracy*.
7. Hydera, E., Kikuchi, M., & Ozono, T. (2024). Empirical assessment of deepfake detection: Advancing judicial evidence verification through artificial intelligence. *IEEE Access*, 12, 151188–151203.
8. INTERPOL. (2023). Artificial intelligence and synthetic media crime.
9. Karras, T., Laine, S., & Aila, T. (2019). A style-based generator architecture for GANs. *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 4401–4410.
10. Masood, M., et al. (2021). Deepfake generation and detection: A survey. *IEEE Access*.
11. Rana, M. S., et al. (2022). Deepfake detection: A systematic literature review. *IEEE Access*.
12. Rewald, S., Simon, B., Irving, E., & Keator, K. (2025). When tech disrupts faster than rules adapt: Drafting emergency guidance for AI-affected evidence. *Opinio Juris*.

13. Sandoval, M.-P., de Almeida Vau, M., Solaas, J., & Rodrigues, L. (2024). Threat of deepfakes to the criminal justice system: A systematic review. *Crime Science*, 13(1). <https://doi.org/10.1186/s40163-024-00220-6>
14. Sheehan, L., & Beneke, E. (2026). Truth on trial: Deepfakes and the new battleground for evidence integrity. *ENSafrica*.
15. Sung, J. (2020). South Korea's response to deepfake pornography. *Asian Journal of Law and Society*.
16. U.S. Congress. (2025). TAKE IT DOWN Act, S. 146, 119th Congress. <https://www.congress.gov/bill/119th-congress/senate-bill/146>
17. U.S. Courts. (2024). Advisory Committee on Evidence Rules Proceedings.
18. UN Office on Drugs and Crime (UNODC). (2023). Cybercrime and emerging technologies.
19. Vaccari, C., & Chadwick, A. (2020). Deepfakes and disinformation: Exploring the impact of synthetic political video on deception, uncertainty, and trust in news. *Social Media + Society*, 6(1).
20. Artificial Intelligence: Council gives final green light to simplify and streamline rules. (2026, June 30). EEAS. <https://www.eeas.europa.eu>
21. China's AI content-labelling rules tame risks. (2026, February 10). Xinhua. <http://english.news.cn>
22. Deepfakes-as-a-Service Meets State Laws Governing Synthetic Media in a Fragmented Legal Landscape. (2026, January 15). *National Law Review*. <https://natlawreview.com>
23. Pakistan's first state-of-the-art national forensic agency becomes fully operational in record time. (2026, April 3). Associated Press of Pakistan. <https://www.app.com.pk>
24. Pakistan cracks down on fake news creators with arrests and bans. (2025, October 14). *Pakistan Today*. <https://www.pakistantoday.com.pk>
25. Senate okays bill to set up agency with unusual sway to fight digital crime. (2024, December 14). *Dawn*. <https://www.dawn.com>
26. South Korea court clears AI porn distributor, citing lack of identifiable victim. (2026, June 11). *Business & Human Rights Resource Centre*. <https://www.business-humanrights.org>
27. South Korea to require advertisers to label AI-generated ads. (2025, December 10). Associated Press. <https://apnews.com>

28. The TAKE IT DOWN Act: A Federal Law Prohibiting the Nonconsensual Publication of Intimate Images. (2025, May 20). Congressional Research Service.
<https://www.congress.gov>
29. Measures for identifying AI-generated synthetic content]. (2025). National Internet Information Office of China.