Global AI Regulation Frameworks: A Comparative Review of Efforts to Control Deep fakes and Misinformation

Ms. Priyanka Tandilkar[1], Dr. Mahesh Singh Gautam[2], Dr. Amol Kumbhare[3], Mrs. Manisha Vishwakarma[4]

Corresponding Author priyankatandilkar@aku.ac.in

Abstract— The proliferation of generative artificial intelligence has transformed synthetic media, commonly termed—deep fakes,‖ from a niche technical curiosity into a mainstream vector for fraud, non-consensual imagery, and electoral misinformation. Documented incidents rose from an estimated twenty-two cases acrossthe2017–2022periodtooveronehundredandfiftyin2024 alone, prompting a wave of regulatory activity across major jurisdictions. This review synthesizes and compares the regulatory instruments adopted by the European Union, the United States, India, and China between 2023 and 2026, examining their definitional approaches, enforcement mechanisms, and penalty structures. Drawing on primary legal texts, government notifications, and industry incident data, the review finds substantial convergence around mandatory labeling and platform-level take down obligations, alongside marked divergence in enforcement philosophy: the European Union and China favor ex-ante, provider-side technical mandates, whereas the United States relies on a narrower, harm-triggered criminal frame work layered over a fragmented patch work of state statutes. India's 2026 amendments occupy an intermediate position, combining statutory definitions with intermediary-centered due-diligence obligations. The review concludes that while labeling and takedown regimes represent a necessary first step, persistent gaps in cross-border enforcement, detection-technology reliability, and protection for satire and legitimate expression continue to constrain the practical effectiveness of these frameworks.

Keywords: deep fakes, synthetic media, misinformation, artificial intelligence governance, AI Act, content labeling, platform regulation, comparative law

I. INTRODUCTION

Generative artificial intelligence systems capable of synthesizing convincing images, audio and video of real individuals have moved within a period of roughly five years, from research demonstrations to freely accessible consumer tools. This democratization of synthetic-media production has coincided with a marked increase in its malicious application, spanning non-consensual intimate imagery, financial fraud through executive impersonation, and the manipulation of electoral discourse. Industry incident-tracking data indicate that reported deep-fake cases rose from a cumulative twenty-two between 2017 and 2022 to one hundred and fifty in 2024, with the first quarter of 2025alonesurpassingtheentiretyofthepreceding year[19]. The regulatory response to this phenomenon has been neither uniform nor synchronized. Some jurisdictions, notably the European Union have embedded deep-fake specific transparency obligations with in a broader horizontal AI statute; others, such as the United States, have proceeded through narrowly targeted criminal legislation layered over an expanding patch work of state law; while China and India Have opted for labeling –centered administrative regimes enforced primarily through intermediary liability. This divergence raises important questions of comparative regulatory design: which mechanisms are proving durable, which are attracting constitutional or operational challenge, and to what extent do the resulting frame works meaningfully reduce the incidence or harm of synthetic-media misuse.

This review addresses these questions through a comparative synthesis of the principal regulatory instruments adopted between 2023 and 2026 in four jurisdictions that, together, govern a majority of global internet users and host the largest concentration of generative-AI developers. The review is structured as follows. Section 2 outlines the methodology used to identify and synthesize sources. Section 3 present a jurisdiction-by-jurisdiction analysis of the European Union, the United States, India and China. Section 4 offers a comparative synthesis, supported by a summary tableandincident-growthdata.Section5discussescross-cutting enforcement challenges. Section 6 concludes with recommendations for future regulatory design.

II. METHODOLOGY

This review adopts a narrative (descriptive) comparative-law methodology rather than a systematic review protocol, reflecting the objective of synthesizing rapidly evolving primary legal instruments alongside contemporaneous industry and news reporting rather than aggregating quantitative effect-size data across controlled studies. Primary sources consulted include the consolidated text of the European Union Artificial Intelligence Act (Regulation (EU) 2024/1689) and the Digital Services Act; the United States TAKE IT DOWN Act (PublicLaw119-12) and associated Congressional Research Service legal analysis the Government of India's Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, as notified by the Ministry of Electronics and Information Technology and the Cyberspace Administration of China's Measures for Labeling of AI-Generated Synthetic Content together with the mandatory national standard GB45438-2025. These were supplemented by secondary legal commentary from law firms and academic outlets and by industry incident-tracking data sets(Resemble AI, Gartner, Sumsub, Pindrop, Keepnet Labs, and IRONSCALES) used to contextualize the scale of the underlying problem. Sources were retrieved between 2025 and mid-2026 and cross-checked against at least two independent outlets where the underlying claim concerned statutory dates, penalty thresholds, or enforcement figures. Given the pace of legislative change in this domain, this review should be understood as a snapshot of the regulatory landscape as of July 2026, with the explicit expectation that several provisions discussed, particularly the European

Union's Code of Practice on deep-fake marking and India's rules, remain subject to further guidance and amendment.

III.

3. Regulatory Landscape: A Jurisdiction-by-Jurisdiction Analysis

A. 3.1 The European Union: A Risk-Tiered, Labeling-Centered Model

The European Union's Artificial Intelligence Act classifies deep-fake generating systems as posing "limited risk," triggering transparency rather than prohibition [4]. Article 3(60) defines a deep-fake as AI-generated or manipulated image, audio, or video content that resembles an existing person, object, place, entity, or event and would falsely appear authentic to a person[4].Article50(4)imposes a disclosure obligation on deplorers of such systems, requiring that manipulated content be labeled in a manner that is clear and distinguishable at the point of first exposure [1],[2].Full enforcement of these transparency obligations is scheduled for 2 August 2026, though the European Commission has concurrently promoted voluntary early adoption through an AI Pact and a draft Code of Practice on the technical marking and detection of synthetic content, expected to be finalized by mid-2026 [5].

A distinguishing feature of the European approach is its layering of the AI Act's content-level transparency duties over the Digital Services Act's platform-level accountability regime, which governs the removal of illegal content, including non-consensual sexual deep-fakes, irrespective of whether a labeling failure has occurred [3]. This layering was illustrated in early2026when French authorities referred a case involving non-consensual sexually explicit deep-fakes generated through a major AI system to prosecutors under the Digital Services Act rather than the AI Act's labeling provisions, underscoring that the two instruments address distinct harms [3]. Legal commentary has also noted that the AI Act's scope is broader than commonly assumed by businesses, extending to AI-generated marketing avatars, synthetic voiceovers, and digitally altered corporate communications, not merely malicious political or sexual content [5].

B. 3.2 The United States: Harm-Triggered Federal Law over a Fragmented State Patchwork

Unlike the European Union's horizontal instrument, the United States has proceeded through narrowly scoped federal legislation targeting a specific harm category. The TAKE IT DOWN Act, signed into law on 19 May 2025, criminalizes the knowing publication of non-consensual intimate imagery, including AI-generated "digital forgeries," and requires covered platforms to establish a notice-and-removal process with a forty-eight-hour compliance window, requirement that becameenforceableon19May2026[6],[7].Violations carry criminal penalties of up to two years' imprisonment where the victim is an adult and three years where the victim is a minor, with enforcement of the platform-compliance provisions vested in the Federal Trade Commission [10].

Beyond this federal baseline, regulation of deep-fakes in the United States remains substantially devolved to the states. As of mid-2026, forty-five to forty-six states had enacted some form of deep-fake-specific statute, addressing sexual deep-fakes, election-related synthetic media, or voice and likeness rights, though coverage and stringency vary considerably [8], [9]. Election-focused disclosure statutes, present in thirty states ahead of the 2026 midterm elections, have faced sustained constitutional challenge; a federal court struck down key provisions of California's Assembly Bill 2839 in August 2025 on First Amendment and Communications Decency Act Section 230 grounds, and a similar Minnesota statute has faced comparable litigation [8]. This pattern suggests that, in the American constitutional context, content-based restrictions on political synthetic media face materially greater judicial scrutiny than the narrower, harm-defined prohibitions embodied in the TAKE IT DOWN Act.

3.3 India: Intermediary due Diligence and Statutory Recognition of Synthetic Content

India's regulatory response has developed through amendment of existing intermediary-liability rules rather than through freestanding AI legislation. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Amendment Rules, 2026, notified by the Ministry of Electronics and Information Technology on 10 February 2026 and operative from 20 February 2026, introduce a statutory definition of "synthetically generated information" as content that is artificially or algorithmically created, generated, modified, or altered using a computer resource in a manner that appears reasonably authentic [11], [12]. The amendment requires significant social media intermediaries to obtain user declarations regarding synthetic content, deploy verification measures, and ensure prominent labeling, while compressing take down time lines for high-risk synthetic content, such as non-consensual intimate imagery or impersonation, to as little as three hours from notification [12].

Non-compliance carries a distinctive consequence within the Indian framework: loss of the safe-harbor protection ordinarily available to intermediaries under Section 79 of the Information Technology Act, 2000, exposing platforms to direct liability for user-generated synthetic content [13]. Commentators have noted that this design, while administratively efficient, shifts substantial adjudicative discretion to platforms and to the executive branch, and have raised concerns regarding the absence of explicit statutory exemptions for satire, news reporting, or academic use, alongside sharply compressed compliance windows that may be difficult for smaller intermediaries to meet [14].

3.4 China: State-Mandated Technical Labeling

China's Measures for Labeling of AI-Generated Synthetic Content, jointly issued by the Cyberspace Administration of China, the Ministry of Industry and Information Technology, the Ministry of Public Security, and the National Radio and Television Administration, together with the mandatory national standard GB 45438-2025, took effect on 1 September 2025 [15], [16]. The regime imposes obligations on four categories of actor: AI content-generation service providers, content-propagation platforms, application-distribution stores, and end users, and requires both explicit labels, such as visible watermarks, captions, or

audio cues, and implicit, machine-readable metadata identifiers embedded within the content file itself [17]. A three-tier classification system, confirmed, possible, and suspected AI-generated content, governs the corresponding labeling and traceability obligations and platforms are required to detect and reinforce labels even on content originally uploaded without one [17].

China's framework is, in scope, the most comprehensive labeling regime currently in force, extending beyond image, audio and video to text and virtual assets and forming part of the Cyberspace Administration's broader "Qinglang" campaign against online misinformation and fraud [18]. Comparative legal analysis suggests that China's mandatory, upstream, provider-embedded labeling approach anticipates elements of the European Union's forthcoming Code of Practice on deep-fake marking, and may represent an early template for the technical direction of global convergence, not with standing the very different political and civil-liberties context in which it operates [17].

3.5 Emerging and Comparator Frameworks

Several additional jurisdictions merit brief note. The United Kingdom criminalized the creation of non-consensual deep-fake intimate images, with penalties of up to two years' imprisonment, effective January 2026. South Korea has confronted an acute domestic crisis involving AI-generated sexual imagery, reporting approximately two hundred and ninety-seven deep-fake sex-crime cases within a seven-month period in 2024, nearly double the 2021 figure, prompting expedited legislative and platform-enforcement responses. These developments reinforce the broader observation that regulatory intensity has tended to correlate with the visibility of a domestically salient harm event, whether an electoral incident, a celebrity-targeted scandal, or a documented fraud case, rather than proceeding from a uniform international baseline.

4. Comparative Synthesis

Three patterns emerge from this comparison. First, all four jurisdictions have converged on labeling and disclosure as the baseline regulatory mechanism, suggesting an emerging international norm around transparency-by-design rather than outright prohibition of synthetic-media generation. Second, enforcement architecture diverges sharply along a spectrum from ex-ante, provider-embedded technical mandates, exemplified by China and, to a lesser degree, the European Union's forthcoming Code of Practice, to ex-post, harm-triggered criminal liability, exemplified by the United States federal approach. Third, the severity of financial exposure varies by more than two orders of magnitude, from the European Union's turnover-linked penalty of up to seven percent of global revenue to the comparatively modest statutory fines available under most United States state statutes, a disparity that is likely to shape where multinational platforms priorities compliance investment.

The scale of the underlying problem that these frameworks seek to address is illustrated in Figure 1, which charts the growth in documented deep-fake incidents from accumulative twenty-two cases across2017–2022 to one hundred and fifty in 2024 and one hundred and seventy-nine in the first quarter of 2025 alone [19].

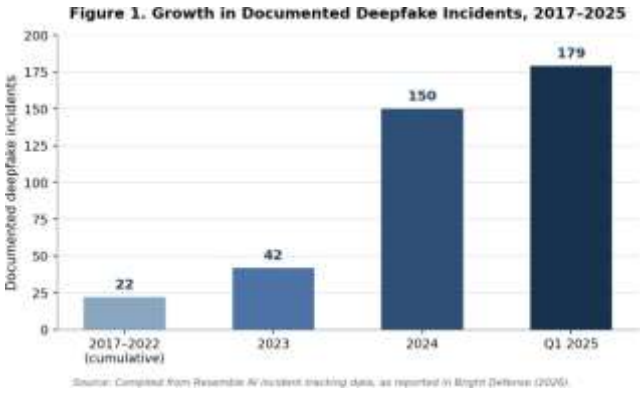

Table 2 supplements this figure with a broader set of industry-reported indicators of scale and harm, drawn from cyber security and fraud-analytics literature rather than legal sources, to contextualize the urgency underlying the regulatory responses catalogued above.

Table2. Selected Indicators of Deep-fake Prevalence and Harm, 2024–2026

Indicator	**Value**	**Source/Period**
Documented deep-fake incidents, 2017–2022 (cumulative)	22	Resemble AI incident tracking
Documented deep-fake incidents, 2024	150	Resemble AI incident tracking
Documented deep-fake incidents, Q1 2025	179 (exceeds all of 2024)	Resemble AI incident tracking
Organizations reporting a deep-fake-related incident in prior 12 months	62–85%	Gartner (2025, n=302); IRONSCALES Fall 2025 Threat Report
Rise in deep-fake fraud-attempt frequency in contact centers, 2024	>1,300%	Pin drop 2025 Voice Intelligence and Security Report
Share of non-consensual explicit imagery among all online deep-fake content	96–98%	Keepnet Labs/ Sensitivity AI
Countries reporting election-related deep-fake incidents	38	Surfshark, cited in Station X (2026)

Discussion: Enforcement Challenges and Structural Gas

Notwithstanding this convergence around labeling obligations, several structural challenges limit the practical effectiveness of the frameworks reviewed.

First, detection technology remains imperfect and adversarial contestable: audio-deep-fake detectors achieve accuracy in the high eighties under controlled conditions but degrade materially against adversarial or previously unseen generation techniques, and unaided human detection of high-quality synthetic video has been measured at under twenty-five percent accuracy [23]. Watermark-based labeling, while more robust than early implementations, remains vulnerable to removal or alteration, a limitation acknowledged even with in the European Union's own regulatory impact analysis [2].

Second, cross-border enforcement remains structurally weak. Generative-AI tools and hosting infrastructure are frequently domiciled outside the jurisdiction in which harm materializes, and none of the four frameworks reviewed establishes a binding mutual-recognition or cross-border takedown-cooperation mechanism comparable to, for example, mutual legal assistance treaties in other domains. The Digital Services Act's platform-accountability provisions offer the most developed extraterritorial reach among the frame works considered, but even this mechanism depends on the platform maintaining an establishment or significant user base within the European Union [3].

Third, the tension between synthetic-media regulation and protected expression remains unresolved and jurisdiction-specific. United States courts have repeatedly narrowed content-based political-deep-fake restrictions on First Amendment grounds, while critics of India's 2026 amendment have raised comparable, though doctrinally distinct, concerns regarding the absence of explicit carve-outs for satire, journalism, and academic commentary, and the concentration of interpretive authority within a single executive ministry [14]. These divergent constitutional postures suggest that a uniform global standard for distinguishing harmful deception from protected commentary is unlikely to emerge in the near term. Fourth, the economic incidence of harm continues to outpace regulatory implementation timelines. Industry-reported financial losses attributable to deep-fake-enabled fraud exceeded two hundred million United States dollars in the first quarter of 2025 alone, with average per-incident losses among affected organizations approaching five hundred thousand dollars[19],[20].Given that several of the frameworks reviewed, including the European Union's full enforcement date and India's implementation window, phase in obligations over periods of one to two years, a substantial compliance gap persists between the documented scale of harm and the operative reach of the corresponding legal remedy.

Conclusion and Recommendations

This review has traced a period of rapid and largely uncoordinated regulatory activity directed at deep-fakes and AI-enabled misinformation across four major jurisdictions. Despite differing constitutional traditions and enforcement philosophies, all four frameworks now rest on a common foundation of mandatory or strongly incentivized labeling, coupled with expedited platform-level takedown obligations for the most harmful categories of content, principally non-consensual intimate imagery. This convergence suggests the emergence of an informal international baseline, even in the absence of a binding multilateral instrument.

Three recommendations follow from this analysis. First, policymakers should priorities interoperable technical standards for content provenance, building on efforts such as the Coalition for Content Provenance and Authenticity, to reduce the compliance burden on multinational platforms operating across the divergent labeling formats currently mandated by the European Union, China and India. Second, legislatures in jurisdictions with strong speech protections, notably the United States, should consider narrowly tailored, harm-specific statutes of the kind embodied in the TAKE IT DOWN Act as a more constitutionally durable path than broad content-based restrictions on political synthetic media. Third, all jurisdictions reviewed would benefit from dedicated cross-border enforcement cooperation mechanisms, given that both the infrastructure used to generate deep-fakes and the harm they cause routinely span multiple regulatory territories. As detection technology, generative capability, and legislative response continue to co-evolve, sustained comparative monitoring of the kind undertaken in this review will remain essential to assessing whether the current generation of regulatory instruments achieves its stated objective of preserving public trust in digital information.

ACKNOWLEDGEMENTS

The authors express their sincere gratitude to all the researchers, academicians, policy makers and regulatory organizations whose studies, legal frameworks and published reports have significantly contributed to the preparation of this review paper, "Global AI Regulation Frameworks: A Comparative Review of Efforts to Control Deep fakes and Misinformation." Their valuable work has provided the foundation for understanding the evolving landscape of artificial intelligence governance and regulatory approaches across different jurisdictions.

The authors are also grateful to their institution and faculty members for providing the academic environment, guidance, and encouragement necessary to complete this review.

Special appreciation is extended to the library staff and digital resource providers for facilitating access to relevant books, research articles, government publications, and legal documents.

Finally, the authors acknowledge the support of family members, friends, and colleagues for their continuous encouragement and motivation throughout the preparation of this manuscript. Their unwavering support played an important role in the successful completion of this work.

References

[1] European Parliament and Council of the European Union, Regulation (EU) 2024/1689 Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act), Art. 50, 99, 2024.

[2] M. Matias, “EU's AI Act Advances with Deep-fake Transparency Rules,” Get Clarity.ai, 2025.

[3] Columbia Journal of European Law, “Deep-fake, Deep Trouble: The European AI Act and the Fight Against AI-Generated Misinformation, ”Columbia CJEL Preliminary Reference,2024

[4] “What Constitutes a Deep Fake? The Blurry Line between Legitimate Processing and Manipulation under the EU AI Act,” arXiv: 2412.09961, 2024.

[5] Potomac Law Group, “EU Deep-fake Rules in AI Act Will

Affect More Businesses Than Usually Expected," 2026
[6] U.S.Congress,S.146–TAKEITDOWNAct,Public Law 119-12, 2025. Congressional Research Service, "The TAKE IT DOWN Act: A Federal Law Prohibiting the Non-consensual Publication of Intimate Images," Legal Sidebar LSB11314, Congress.gov, 2025. Recording Law, "Deep-fake & AI Voice Cloning Laws by State (2026),"2026
[7] Stack-Cyber, "Deep-fake Legislation Tracker: Federal, State Laws," 2026.
[8] Skadden, Arps, Slate, Meagher & Flom LLP, "'Take It Down Act' Requires Online Platforms to Remove Unauthorized Intimate Images and Deep-fakes When Notified," Insights, 2025.
[9] Vaish Associates Advocates, "Regulation of AI-Generated/Deep-fake Content and Synthetically Generated Information (SGI) in India – New Rules," 2026.
[10] "Explained: How India's New IT Rules Regulate AI Content and Deep-fakes," Forbes India, 2026
[11] "India Tightens Rules on Deep-fakes and AI-Generated Content," Law.asia, 2026.
[12] "India's New IT Rules on Deep-fakes Threaten to EntrenchOnlineCensorship,"TechPolicyPress, 2025
[13] Cyber space Administration of Chinaetal. Measures for Labeling of AI-Generated Synthetic Content and National Standard GB 45438-2025, 2025.
[14] Loeb &Loeb LLP, "China's AI-Labeling Measures and Mandatory National Standards Take Effect September 1," 2025.
[15] Bird & Bird, "New AI Content Labeling Rules in China: What Are They and How Do They Compare to the EU AI Act?," 2025.
[16] "China's Social Media Platforms Rush to Abide by AI-Generated Content Labeling Law," South China Morning Post, 2025.
[17] Bright Defense, "150+ Deep-fake Statistics (March 2026),"2026
[18] Station X, "Deep-fake Statistics [2026]: Growth, Fraud & Detection Data," 2026.
[19] Sumsub, "Fraud Trends 2026: AI Scams, Deep-fakes, and Emerging Threats," 2026.
[20] Keepnet Labs, "Deep-fake Statistics 2026: Verified Benchmarks & Risks," 2026.
[21] SQ Magazine, "Deep-fake Statistics 2026: The Hidden Cyber Threat," 2025.
[22] Raut Rahul Ganpat and Dr. Sonawane Vijay Ramnath **"**Constructing a User-Cantered Fake News Detection Model by Using Enhance Stacking Ensemble Classification Algorithms in Machine Learning Techniques"JIER, VOL.7,PP 9-13, ISSN (Online) : 2581–6357.
[23] C. Ramakrishna and Dr. Pramod Pandurang Jadhav, **"**Machine Learning Based Sentiment Analysis Implementation for Multiple Reviews Classification**"** JIER, VOL.7,PP 21-25, ISSN (Online) : 2581–6357
[24] Nazeer Shaik, Dr. C. Krishna Priya "Machine Learning Algorithms for Intrusion Detection Systems" JIER, VOL.7,PP 8-18, ISSN (Online) : 2581–6357.
[25] Sarkar Prasanjeet Jyotirmay and Dr. Santosh Pawar, "Advancing Early Detection and Prediction of Diabetes Mellitus: A Robust Soft Voting Ensemble Machine Learning Approach"JIER, VOL.7,PP 12-17, ISSN (Online) : 2581–6357.
[26] Praveen Sharma and Dr. Deepika Pathak,"A Novel Machine Learning Categorial Algorithm with Remote Detecting and GIS Based Decision Support System to Identify Disaster", JIER, VOL.5, PP 25-29, ISSN (Online) : 2581–6357.