THE LEGAL FRONTIER OF AI: FROM DEEPFAKES TO DIGITAL ACCOUNTABILITY

Ayantika Pal, CLATapult, CLAT Coaching, Kolkata

Atri Karmakar, Siksha 'O' Anusandhan National Institutes of Law

ABSTRACT

Across various sectors, including education, medicine, entertainment, and legislation, the adoption of Artificial Intelligence (AI) technologies has surged in recent years. Generative models like ChatGPT and Gemini have revolutionized information access, operational efficiency, and productivity, enabling individuals and institutions to operate with enhanced speed, intelligence, and efficacy. However, this rapid technological advancement also introduces critical legal and ethical challenges. AI is increasingly being misused for illicit activities such as the creation of deepfake videos and audios, celebrity impersonation in deceptive advertising, automated plagiarism, and the spread of misinformation and propaganda. These practices not only violate intellectual property rights and individual privacy but also pose a serious threat to democratic processes. This article critically examines how current legal frameworks in India address these emerging challenges. It evaluates the applicability, liability and effectiveness of key statutes such as The Information Technology Act, 2000, The Bharatiya Nyaya Sanhita, 2023, The Digital Personal Data Protection Act, 2023, The Digital Personal Data Protection Rules, 2025 and Intellectual Property Rights Laws in combating crimes associated with AI-generated content. By drawing on Indian legal frameworks, comparative international models, and landmark judicial decisions, this article identifies existing ethical challenges, legal gaps, and proposes comprehensive recommendations for responsible and ethical AI governance. When lawmakers, courts, companies, and citizens work together, India can encourage innovation without sacrificing privacy, fairness, or trust. AI shouldn't be left to run unchecked, accountability must be built in from the start.

Keywords: Artificial Intelligence, Deepfakes, Data Protection, Privacy, Technologies.

Introduction

We're living in a time when just a few lines of code can create a video that looks and sounds completely real, even if the person in it never actually said or did those things. The line between what's real and what's fake has never been thinner. Artificial intelligence (AI), which lets machines mimic human thought and behavior, has reshaped nearly every aspect of modern life. It's opened up incredible opportunities in areas like communication, automation, and creative content. One of the most controversial uses of this technology is deepfakes, which are AI-generated videos, images, or audio clips that can convincingly imitate someone's appearance, voice, and gestures. Tools like ChatGPT, Gemini, and others have numerous legitimate applications in areas such as entertainment, education, and content creation. Deepfakes, more than anything, highlight the double-edged nature of AI. On one hand, AI has great potential to make life easier by improving healthcare, education, transport, and business. It can help doctors detect diseases early, assist students in learning better, and even make online shopping smarter. But on the other hand, AI also has perils or dangers. It can spread fake news, invade privacy, and even take away human jobs. Deepfakes, biased algorithms, and misuse in surveillance are some serious risks. As this technology becomes more widespread, it also raises tough ethical questions about responsibility and fairness. If an AI system causes harm, who should be held accountable: the developer who built it, the company that deployed it, or the technology itself? This "autonomy versus accountability" dilemma makes it clear that regulation isn't as simple as banning harmful uses. It's also about balancing free expression with the need to prevent real-world damage, and ensuring that systems are transparent, fair, and respectful of privacy.

These issues are being debated worldwide, but they feel especially urgent in India. The country's rapid embrace of digital technologies has outpaced the laws meant to protect people. Deepfakes are already sparking public outrage and limited legal action, yet the loopholes are obvious. Current laws like the Information Technology Act, 2000, the Bharatiya Nyaya Sanhita, 2023, and the Digital Personal Data Protection Act, 2023 offer some protection, but they were never designed to handle the speed and scale of AI-generated content. As India steps deeper into an AI-driven future, stronger legal and ethical guardrails aren't just desirable, they're vital to safeguarding democracy, privacy, and public trust.

Understanding Deepfakes and AI Misuse

In today's digital world, it's getting trickier to tell what's real and what's not, especially with

the rise of deepfakes. What was once easy to recognize has evolved, as deepfakes can now mimic a person's voice, facial expressions, and gestures with great accuracy. This advancement raises serious concerns about public trust, the integrity of information, and democratic processes. Deepfakes are based on a machine learning technique called Generative Adversarial Networks (GANs)[1], where two AI systems, the generator and the discriminator, compete to create and detect fake content. As this technology advances, it becomes harder to identify these fakes. While deepfake technology can be used creatively and for entertainment, it is often misused for harmful purposes. For example, a realistic deepfake of a politician making offensive statements could appear right before an election, influencing public opinion and threatening democracy. Additionally, people can be targeted with fake videos used for harassment, blackmail, or defamation. As tools for creating deepfakes become more accessible, the risks of spreading misinformation and causing harm continue to grow.

India has already seen several real-life cases of AI misuse, exposing gaps in the current legal system. One of the most discussed cases was the deepfake of actress Rashmika Mandanna in 2023[2], where an AI-generated video showed her entering an elevator wearing revealing clothes. The original video belonged to a British influencer, but her face was replaced with Rashmika's using advanced deepfake technology. Although the video was quickly debunked, it highlighted the serious threat deepfakes pose to individual dignity, digital consent, and mental health.

AI has also been used in politics, as seen during the 2020 Delhi Assembly elections[3]. The Delhi unit of the BJP released a deepfake video of its leader Manoj Tiwari, in which his face and voice were altered using AI to produce a campaign message in Haryanvi (a language he doesn't speak)[4]. Although meant as a communication tool, it raised serious ethical and regulatory issues about voter manipulation. With no law banning the use of AI in election propaganda, the Election Commission was largely powerless. This incident underscores the urgent need for electoral safeguards, particularly since deepfakes can mislead voters and undermine the

[1] Papastratis I, "Deepfakes: Face Synthesis with GANs and Autoencoders" (Sergios Karagiannakos, June 2, 2020) <https://theaisummer.com/deepfakes/> accessed September 1, 2025
[2] Ojha A, "Man Accused in Rashmika Mandanna's Deepfake Video Case Arrested" India Today (January 20, 2024) <https://www.indiatoday.in/india/story/man-accused-in-actor-rashmika-mandannas-deepfake-video-case-arrested-by-delhi-police-in-andhra-pradesh-2491281-2024-01-20> accessed July 28, 2025
[3] Pranav Dixit, 'Indian Politicians Are Using Deepfakes to Win Votes' *BuzzFeed News* (Delhi, 20 February 2020) https://www.buzzfeednews.com/article/pranavdixit/india-politicians-deepfakes accessed 3 September 2025.
[4] Alavi M and Achom D, "BJP Shared Deepfake Video On WhatsApp During Delhi Campaign" NDTV (February 20, 2020) <https://www.ndtv.com/india-news/in-bjps-deepfake-video-shared-on-whatsapp-manoj-tiwari-speaks-in-2-languages-2182923> accessed September 1, 2025

democratic process, which conflicts with the constitutional principles of free and fair elections.

In India, the legal system is evolving and still figuring out how to address the misuse of AI and deepfake technology properly. While there are some protections under the Bharatiya Nyaya Sanhita (BNS) 2023, the Information Technology Act, 2000, and the Indian Constitution, these provisions often fall short when dealing with the unique challenges posed by AI-generated content. For example, Section 66E of the IT Act addresses privacy violations, and Section 67 covers the publishing or sharing of obscene material online. However, these laws weren't designed to handle the complexities and rapid changes associated with AI-based impersonation and manipulation. The new Digital Personal Data Protection Act, 2023, aims to protect personal data privacy and shows promise in regulating the misuse of personal data for deepfakes. Still, its main focus is on data protection, and it does not fully cover synthetic media or AI-generated impersonations. Additionally, enforcement mechanisms are still under development. Creating new legal responses may also be necessary. Laws aimed at deterring malicious uses of deepfakes could be helpful, but they must be carefully crafted to safeguard free speech. Effective combat against deepfakes requires collaboration among governments, tech companies, and civil society. In India, it's crucial to align new laws with the constitutional protections in Articles 19 and 21. This will help ensure safety from digital harms while also protecting individual freedoms as we enter the age of AI.

Indian Legal Framework Governing AI

1. The Information Technology Act, 2000

This Act is India's main law for regulating digital transactions, data, and cybersecurity. While it doesn't explicitly mention artificial intelligence (AI), many of its provisions apply to AI systems. Section 43A[5] holds companies responsible for mishandling sensitive personal data due to negligence, which is a crucial safeguard, as AI systems often process large amounts of user information. Section 66[6] addresses cybercrimes, helping protect AI technologies from

[5] 43A. Compensation for failure to protect data.
Where a body corporate, possessing, dealing or handling any sensitive personal data or information in a computer resource which it owns, controls or operates, is negligent in implementing and maintaining reasonable security practices and procedures and thereby causes wrongful loss or wrongful gain to any person, such body corporate shall be liable to pay damages by way of compensation to the person so affected.
[6] 66. Computer related offences.
If any person, dishonestly or fraudulently, does any act referred to in section 43, he shall be punishable with imprisonment for a term which may extend to three years or with fine which may extend to five lakh rupees or with both.

threats such as hacking or intellectual property theft, like someone stealing AI-generated algorithms. The Act also provides legal recognition for electronic records and digital signatures, which is essential when AI is used in automated decision-making or digital transactions. Overall, these regulations help ensure data security, privacy, and trust in AI, making strong enforcement vital for India to remain competitive in the global AI arena.

Intermediary Liability under the IT Act -

The Information Technology Act of 2000 (IT Act) is pivotal regarding intermediary liability in India. Section 79 provides intermediaries (like social media platforms) with "safe harbour" protection, relieving them from responsibility for user-generated content, as long as they function solely as conduits and adhere to due diligence responsibilities. For instance, if a deepfake video is posted on WhatsApp or Instagram, the platform is not held directly accountable, provided it did not create or alter the content and quickly took action to remove it upon notification. This provision ensures that platforms are not unfairly weighed down with liability for every piece of content generated by users. Damaging content can proliferate extensively before platforms can identify and eliminate it, leading to irrevocable harm to personal reputations, privacy, or electoral processes. Victims contend that platforms should shoulder more responsibility, as they have the technical capability to identify and restrict deepfakes using AI filters. Conversely, platforms argue that enforcing proactive monitoring requirements infringes on free speech and is technically impractical on a large scale. Proposed Amendments The government has made efforts to limit safe harbour provisions over time. The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021 mandate platforms to delete content within 36 hours of receiving notice[7] and to allow the traceability of messages[8]. Suggested amendments propose stricter due diligence, compulsory AI-powered detection tools, and accountability for not promptly addressing harmful content. Critics warn that imposing excessive intermediary liability may lead to over-censorship, prompting platforms to eliminate legitimate content to evade liability. The struggle to find a balance between free expression and harm reduction continues to be a heated discussion.

[7] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Rule 3(1)(d)
[8] The Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021, Rule 4(2)

2. Digital Personal Data Protection Act, 2023 (DPDP Act)

The Digital Personal Data Protection Act, 2023 (DPDP Act), now in force, is a critical law governing how AI systems process personal data in India. It mandates informed consent[9], data minimization, and the appointment of Data Protection Officers[10], which directly impacts AI tools involved in activities like personalized content generation and IP creation. While the Act empowers individuals to exercise greater control over their data and imposes significant penalties for non-compliance, it does not explicitly address the ownership or intellectual property (IP) rights of AI-generated outputs, especially when these outputs are derived from personal data. This legal grey area creates challenges in determining liability and authorship, making it essential to harmonize the DPDP Act with existing Intellectual Property Rights (IPR) laws to safeguard both individual rights and commercial interests in AI-driven innovation. Furthermore, with AI models relying heavily on large datasets for training, strict adherence to privacy standards is vital for maintaining public trust. The older Personal Data Protection Bill, 2019, which introduced key concepts like purpose limitation, data localization, and accountability, along with setting up a Data Protection Authority, has since been replaced by the DPDP Act, but its influence is evident in the new law's structure. However, the DPDP Act still lacks detailed provisions on automated decision-making and profiling, which were addressed more clearly in the PDP Bill. As AI increasingly affects individuals' rights through profiling or algorithmic decisions, India will need future amendments or supplementary rules to bridge these regulatory gaps and ensure responsible, transparent, and rights-based AI governance[11].

3. Digital Personal Data Protection Rules, 2025

The Digital Personal Data Protection (DPDP) Rules, 2025 do not contain an "AI infringement" chapter per se, but several provisions can be applied where AI systems misuse personal data in ways that overlap with infringement, particularly in deepfakes, impersonation, or unauthorized voice/image cloning. Rule 3(2) requires Data Fiduciaries to implement technical and organizational safeguards proportionate to the risk of harm, which covers preventing AI models from generating or disseminating manipulated personal data. Rule 5(1)(b) mandates purpose-

[9] Digital Personal Data Protection Act, 2023, S 6
[10] Digital Personal Data Protection Act, 2023. S 10
[11] Paras Sharma and Bhavya Sharma, 'Balancing AI Innovations with Privacy Laws (in light of India's DPDP Act, 2023)' (2025) 5(4) Indian Journal of Legal Review 1130

specific consent, meaning AI systems cannot repurpose collected data (including biometric and facial data) for training or content generation without fresh authorization. Rule 7(4) obligates prompt reporting of personal data breaches, relevant when AI datasets are compromised or misused. Rule 8(2) imposes accountability measures on Significant Data Fiduciaries (SDFs), including mandatory audits, impact assessments, and the appointment of a Data Protection Officer, which can help pre-empt AI misuse. However, these rules stop short of addressing non-personal data misuse, copyright/trademark infringement, or autonomous AI liability, leaving a legal vacuum where AI outputs cause intellectual property violations without directly involving personal data.

4. Intellectual Property Laws: Copyright, Patent, and Trademark

In India, the main intellectual property tools to address AI-driven infringement are found in copyright, trademark, and patent law, though each has its scope and limitations when applied to autonomous AI outputs. Under the Copyright Act of 1957, Sections 13 - 14 protect original literary, artistic, musical, and dramatic works, granting creators exclusive rights to reproduce, adapt, and communicate them. AI infringements can occur when copyrighted works are scraped for training datasets without permission or when AI-generated outputs closely imitate protected expression. Section 51 treats such acts as infringement, allowing civil remedies, while Section 57 protects moral rights, which can be invoked against deepfakes, AI-manipulated performances, or unauthorized voice cloning that distorts an author's work or harms their reputation. However, since copyright protection requires human authorship, works generated entirely by AI without significant human creative input fall outside statutory protection, creating a gap in both ownership and enforcement.

The Trade Marks Act of 1999, through Sections 29 and 30, protects registered marks against unauthorized use that confuses, dilutes brand distinctiveness, or tarnishes reputation. In the context of AI, infringement can happen when generative models produce advertisements, endorsements, or chatbots impersonating brands or their representatives without approval. For example, AI-generated product images featuring a famous logo without permission can be considered trademark infringement or passing off. Yet, current trademark provisions focus on "use in the course of trade" by natural or legal persons, leaving ambiguity in cases where AI itself autonomously generates infringing content without a directly identifiable human operator.

The Patents Act of 1970, under Section 48, grants patentees exclusive rights to prevent others from making, using, or selling their inventions without permission, which could, in theory, include AI-driven replication of patented processes or designs. However, Section 3(k) explicitly excludes "a mathematical or business method or a computer program per se" from patentability, significantly limiting direct protection for AI algorithms and models. Only AI-related inventions that produce a "technical effect" or contribute to a "technical contribution", such as improving energy efficiency or enabling new manufacturing processes, can be patented, as seen in past grants like *IPA 3323/CHENP/2012*. This exclusion means that while patents can protect certain AI-assisted inventions, they cannot safeguard AI's internal workings or prevent unauthorized duplication of algorithmic logic.

5. Bharatiya Nyaya Sanhita (BNS), 2023

The Indian Penal Code (IPC) has now been replaced by the BNS, 2023, which includes provisions that address issues related to impersonation, defamation, and obscenity, providing a foundational legal structure that can be applied to the misuse of AI. For example, impersonation is addressed in Section 419 (cheating by personation) of the IPC, which has corresponding provisions in the BNS that punish individuals for taking on another person's identity with the intent to deceive or cause harm. Defamation, previously outlined in Sections 499 - 500 IPC and now in Section 356 BNS[12], offers protection to individuals against false statements that can damage their reputations, a concern that is particularly relevant in the context of AI-generated deepfakes or falsified content aimed at public figures. Likewise, the laws governing obscenity (Section 292 IPC and Section 67 of the IT Act) target the sharing of sexually explicit or offensive materials, which can include AI-altered images or videos. Although these legal provisions provide some mechanisms for addressing these issues, they were not specifically designed for autonomous systems or the extensive, anonymous distribution of synthetic media. AI technology is capable of producing realistic impersonations or defamatory material almost instantaneously, resulting in traditional enforcement methods being reactive and sluggish. This discrepancy underscores the necessity to adapt BNS regulations to align with AI-targeted rules under the Information Technology Act, the Digital Personal Data Protection Act, and the

[12] 356. Defamation.
(1) Whoever, by words either spoken or intended to be read, or by signs or by visible representations, makes or publishes in any manner, any imputation concerning any person intending to harm, or knowing or having reason to believe that such imputation will harm, the reputation of such person, is said, except in the cases hereinafter excepted, to defame that person.

forthcoming AI governance regulations to ensure quicker take-downs, improved clarity regarding liability attribution, and more effective prevention of digital harms.

6. AI Governance Initiatives

India's AI policy ecosystem, through initiatives like the National e-Governance Plan, the New Education Policy (NEP), and AIRAWAT[13], reflects a strong governmental push towards embedding AI into governance, education, and research infrastructure. The National e-Governance Plan leverages AI to automate processes, improve decision-making, and enhance citizen service delivery, while the NEP's introduction of coding education from the 6th standard signals early skill development for building an AI-ready workforce. AIRAWAT, launched by NITI Aayog, provides a dedicated AI research and analytics platform to address India's AI infrastructure needs. Complementary efforts from the Ministry of Electronics and IT (MeitY) and NITI Aayog, such as the #AIforAll[14] and Responsible AI for All frameworks, stress fairness, transparency, and self-regulation, values directly relevant to managing intellectual property rights (IPR) in AI contexts. Similarly, the Draft National Data Governance Framework Policy (2022)[15] aims to facilitate anonymized data access for AI innovation while attempting to balance IP protection with the public interest.

[13] Somani D, "What Is Pune-Based AI Supercomputer 'AIRAWAT' Highlighted in Economic Survey?" Times Of India (July 23, 2024) <https://timesofindia.indiatimes.com/technology/tech-news/what-is-pune-based-ai-supercomputer-airawat-highlighted-in-economic-survey/articleshow/111952889.cms> accessed August 15, 2025
[14] NITI Ayog, *National Strategy for Artificial Intelligence* (June 2018)
[15] Eshani Vaidya & Sreyan Chatterjee, *Draft National Data Governance Framework Policy* (The Dialogue 2022)

Case Laws

***Arijit Singh v. Codible Ventures LLP*[16]**

Facts –

Bollywood playback singer Arijit Singh filed an ex parte ad-interim suit seeking protection of his personality rights covering his name, voice, vocal style, mannerisms, image, signature, and overall persona against unauthorized AI-based cloning and commercial exploitation by multiple defendants, including AI tools, merchandise sellers, GIF platforms, and domain registrants ("arijitsingh.com", etc.).

Decision & Reasoning –

The Bombay High Court granted a dynamic injunction restraining all defendants from using any aspect of Singh's personality, including through AI generation for commercial or personal gain, without his consent. The court recognized that -

- Personality/Publicity Rights are protectable for celebrities (identity + goodwill).
- Moral Rights under Section 38-B of the Copyright Act are infringed when performances or likenesses are distorted or misused.
- AI tools enabling voice cloning or persona misuse are not shielded by freedom of speech when driven by commercial exploitation.

Legal Impact -

- IP and Personality Rights: Affirmed that AI-mediated misuse falls squarely within existing IP and personality-right doctrines.
- IT Act: While not directly invoked, the order effectively curbs platforms misusing digital technologies, aligning with the spirit of Section 43 (unauthorized access/damage to systems).

[16] Arijit Singh v. Codible ventures LLP [2024] SSC Online bom 2445

Anil Kapoor v. Simply Life India & Others[17]

Facts –

Actor Anil Kapoor sought an interim injunction against multiple entities for using his name, image, voice, and signature catchphrase "jhakaas" in AI-generated videos, GIFs, merchandise, and domain names without authorization.

Decision & Reasoning –

Justice Prathiba M. Singh granted the injunction, acknowledging that AI-enabled distortions of his persona, including the iconic catchphrase, could unjustly tarnish his reputation and livelihood. The court stressed that these are protectable under personality/publicity rights and that AI makes exploitation easier at scale.

Legal Impact -

- Reinforces that persona-based rights extend to catchphrases and voice attributes.
- Serves as a precedent for other celebrities, demonstrating that Indian courts recognize the gravity of AI-enabled impersonation.

ANI Media Pvt. Ltd. v. OpenAI [18]

Facts –

ANI, a leading Indian news agency, sued OpenAI, alleging that ChatGPT used its copyrighted content (both public and paywalled) without a license to train AI models. ANI also claimed that the model generated fake interviews falsely attributed to them. ANI cited unauthorized scraping & storage, and potential harm to economic value. OpenAI responded that ANI had opted out but remained accessible via syndication and claimed fair-use protections and jurisdictional immunity.

Decision & Reasoning –

[17] Anil Kapoor v. Simply Life India [2023] SCC Online Del 6914
[18] Ani Media (P) Ltd. v. Open AI Inc [2024] SCC OnLine Del 8120

Delhi HC issued summons to OpenAI and appointed an *amicus curiae* to assist on issues involving copyright infringement via AI training and misattribution. The court acknowledged the novelty of "AI training as infringement" and questions around territorial jurisdiction, dataset legality, and fair use.

Legal Impact -

- Copyright Act, 1957: Puts a spotlight on whether scraping for AI training constitutes unauthorized reproduction (Section 14) or fair dealing.
- IT Act, 2000 (esp. Section 43): ANI's claims on unauthorized access or use of digital content may invoke IT provisions, though the suit centers on copyright.
- Broader Implications: This could catalyze statutory reforms or new jurisprudence on data mining, TDM exceptions, and IP in AI.

Ethical Challenges in AI

AI's swift integration has brought notable advantages but it also introduces important ethical and regulatory concerns. The main challenges focus on striking a balance between autonomy and accountability, safeguarding free speech while reducing potential harm, and adhering to ethical standards such as transparency, fairness, non-discrimination, and auditability. Within this larger context, a significant legal discussion centers on whether the injuries caused by AI should be classified under product liability, i.e. viewing AI as a faulty product in cases of malfunction, or developer accountability, where responsibility stems from mistakes in design, programming, or biased training data. Finding this equilibrium is essential, as it determines how legal frameworks will distribute responsibility among creators, implementers, and users in a time when AI decisions increasingly influence human experiences.[19]

Product Liability, Developer Responsibility, and the Autonomy - Accountability Dilemma

A key issue in AI law involves figuring out who should take responsibility when AI systems cause damage; should it fall under product liability, developer accountability, or another legal

[19] Bheema Shanker Neyigapula, 'Ethical Considerations in AI Development: Balancing Autonomy and Accountability' (2024) 10.18178/JAAI.2024.2.1.138-148

framework? Under product liability, manufacturers are held liable for flaws in their products. In India, the Consumer Protection Act, 2019 holds producers responsible for design defects, manufacturing flaws, or insufficient warnings. When this is applied to AI, it suggests that companies could be held liable if an AI-driven product, like a self-driving vehicle or a medical diagnostic tool, fails. However, unlike conventional products, AI systems are dynamic they "learn" and "adapt" after they are put into use. A chatbot that works safely at first may later produce harmful misinformation. This leads to a challenging question: can a product be labeled as "defective" if its actions change over time in ways that the manufacturer couldn't predict? Another perspective is to assign responsibility to developers, claiming that they should be accountable if damage results from biased datasets, flawed algorithm design, or insufficient testing. For instance, a facial recognition system that inaccurately identifies individuals from certain groups highlights the shortcomings in the choices developers made regarding training data. Nevertheless, imposing blanket liability on developers creates complications. Many developers create foundational tools that others later adjust or utilize for different purposes. For example, if a company launches a general AI model and a user modifies it to create deepfakes, it becomes ambiguous whether responsibility lies with the original developer, the user implementing it, or the end user. Some academics even suggest granting AI "electronic personhood," making it responsible like a corporation, with reparations coming from insurance or liability funds. However, this idea remains contentious since AI does not possess consciousness, intent, or assets. Most legal systems, including India's, dismiss this concept and continue to concentrate accountability on human individuals who design, deploy, or profit from AI. This discussion is linked to the broader dilemma of autonomy versus accountability. Advanced AI systems often make forecasts or decisions independent of human oversight, sometimes in ways even their creators cannot fully elucidate, such as in medical diagnostics, credit scoring, or assessments of criminal risk. When harm results from such opaque decision-making, the critical query is: who is responsible? Existing Indian law, including the Digital Personal Data Protection Act, 2023, offers protections against the misuse of personal data but does not directly tackle liability for autonomous AI actions leading to financial losses, harm to reputation, or violations of rights. Across the globe, regulatory frameworks like the EU AI Act are striving to address this issue by requiring human-in-the-loop systems and redistributing responsibility among creators, deployers, and users. In conclusion, effectively addressing the dispute between product liability and developer responsibility requires legal models that acknowledge AI's changing nature while ensuring that accountability remains focused on humans. A well-rounded framework must establish clear lines of liability among developers,

deployers, and intermediaries, while incorporating safeguards to prevent harm from opaque, autonomous AI systems.

Freedom of Expression vs. Harm Reduction

The capability of AI to produce various forms of content, including text, images, and videos, has broadened the scope of free expression while also facilitating negative uses, such as deepfakes, misinformation, and slander. On one side, constitutional safeguards like Article 19(1)(a) in India protect free speech; however, they come with reasonable limitations as outlined in Article 19(2), which refers to public order, decency, morality, and the prevention of defamation. AI influences this equilibrium because harmful content can be created en masse, often anonymously, making it challenging to trace back to its originator[20]. Given the absence of regulations specifically addressing AI in Indian law, platforms mainly rely on self-regulation, supported by broader provisions under the IT Act and the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021. Nevertheless, these measures are primarily reactive. In the absence of proactive standards specifically aimed at AI-related content risk, there is a possibility that harm reduction efforts might distort into censorship, thereby limiting permissible speech. Therefore, regulation must distinguish between lawful yet offensive communication and speech that leads to tangible harm, utilizing AI-specific criteria.

Comparative International Approaches

Nations across the globe are hurrying to establish regulations to address the rapid advancement of AI technologies. While some focus on specific issues like facial recognition or data privacy, others are creating comprehensive AI regulatory frameworks. Many countries are opting to start with national strategies or policy guidelines rather than diving directly into strict laws. Although there is no one-size-fits-all approach, certain common themes are becoming evident. The most significant challenge is figuring out how to manage AI risks without stifling innovation. Most governments begin with broad ethical principles or strategic goals before implementing detailed legislation, due to the fast-paced development of AI and its substantial impact.

[20] Channarong Intahchomphoo and Christine Tschirhart, 'The Evolution of Data and Freedom of Expression and Hate Speech Concerns with Artificial Intelligence' (2022) 22(1) Legal Information Management 45.

EU's AI Act (2024)

The European Union's Artificial Intelligence Act (2024), officially passed in mid-2024, marks the first extensive legislation worldwide concerning AI regulation. It took effect on August 1, 2024, featuring a phased implementation[21] - the main provisions have become effective starting February 2025, requirements for general-purpose AI has come into force in August 2025, and full compliance for high-risk systems is expected by August 2027. This Act utilizes a risk-based approach, categorizing AI systems into four primary classifications - unacceptable, high, limited, and minimal risk, with a distinct category for general-purpose AI (foundation models)[22]. At the highest level, unacceptable-risk AI systems are completely prohibited. This group includes practices like social scoring, real-time biometric surveillance in public areas, predictive policing based solely on profiling data, manipulative AI aimed at vulnerable populations, and emotion recognition technologies in educational or workplace settings[23]. High-risk AI systems, such as those utilized in healthcare, education, law enforcement, or critical infrastructure, are subject to rigorous regulation. Providers are required to meet strict criteria, such as data quality controls, human oversight, transparency responsibilities, and periodic conformity assessments[24]. Conversely, limited-risk AI systems are governed by lighter regulations, mainly centered around transparency mandates like informing users when they interact with a chatbot or AI-generated content. The most inclusive category, minimal-risk AI (including spam filters and video games), does not have any mandatory legal obligations apart from voluntary guidelines. In 2023, a new aspect was introduced for general-purpose AI (GPAI), which encompasses foundation models like GPT-4. These systems must disclose summaries of their training data and provide technical documentation, with additional responsibilities for "high-impact" models that surpass certain computational thresholds. This requirement aims to enhance accountability for the most powerful AI systems that can be utilized in various applications. To enforce these regulations, the Act establishes a European AI Office, a Scientific Panel, and national supervisory authorities, thereby creating a unified governance framework across Member States. The Act also imposes significant penalties for non-compliance, akin to the EU's GDPR. Companies that implement prohibited AI practices

[21] Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 [2024] OJ L168/1 (EU AI Act 2024)
[22] ibid, arts 5–7
[23] ibid, art 5
[24] ibid, arts 8–15

may incur fines of up to €35 million or 7% of their global revenue, while other infractions can result in penalties of up to €15 million or 3% of revenue[25]. These sanctions underscore the EU's commitment to establishing global standards for AI safety and accountability. Internationally, the EU AI Act has begun to create a rippling effect, often referred to as the "Brussels Effect," where EU regulations shape global regulatory strategies. Numerous multinational corporations are adjusting their AI practices to align with the Act's guidelines to ensure compliance in various jurisdictions[26]. By integrating bans, stringent oversight for high-risk AI, transparency mandates for consumer-facing applications, and governance structures for foundation models, the EU AI Act presents a comprehensive regulatory framework that reconciles innovation with the safeguarding of fundamental rights.

USA

In June 2025, Senators Ron Wyden, Cory Booker, and Representative Yvette Clarke brought back the Algorithmic Accountability Act of 2025, marking the latest attempt in the U.S. to oversee the use of artificial intelligence and automated decision-making systems. This legislation is an evolution of earlier proposals made in 2019 and 2022, reflecting the increasing urgency surrounding AI regulation as tools such as generative AI, facial recognition, and predictive algorithms become commonplace in daily life. The primary aim of the bill is to ensure that companies are held responsible when they utilize AI for critical decisions like hiring, loan approvals, medical treatment recommendations, or targeted advertising, particularly if these systems exhibit unfairness, discrimination, or pose safety risks. Under this Act, organizations employing "high-risk" AI systems will be mandated to conduct comprehensive impact assessments. These evaluations must assess the accuracy of the AI system, examine potential biases or discrimination, evaluate its privacy protections, and identify any security threats it may present. The companies will then be required to communicate to regulators, namely the Federal Trade Commission (FTC), the measures they are implementing to mitigate those risks. The FTC will be empowered to enforce these regulations, classify infractions as "unfair or deceptive practices," and initiate legal proceedings against companies that do not comply. A notable addition to the 2025 version is an enhanced emphasis on transparency.

[25] ibid, art 99
[26] Anu Bradford, *The Brussels Effect: How the European Union Rules the World* (OUP 2020)

The Act proposes the establishment of a publicly accessible database where individuals can access information about the types of algorithms utilized by companies and the findings of their impact assessments. This initiative aims to provide the public, researchers, and policymakers with greater understanding of how AI systems function in real-world applications, rather than allowing them to remain as "black boxes." Proponents of the bill contend that this oversight is essential as algorithms increasingly shape individuals' opportunities and rights, and without proper supervision, they could perpetuate inequality or facilitate covert discrimination. For instance, previous research has indicated that automated hiring processes may disadvantage women and minority candidates, while credit-scoring models can unjustly affect certain communities. By implementing accountability measures, the Act aspires to promote responsible AI usage without stifling innovation. Conversely, critics caution that the legislation may impose significant burdens on businesses, particularly smaller enterprises, and that excessive regulation could hinder technological advancements. Nonetheless, the Algorithmic Accountability Act of 2025 signifies a major move toward establishing clear national guidelines for AI in the U.S., contrasting with the existing fragmented state-level regulations such as New York City's local algorithm auditing requirements.

UK's voluntary AI safety agreements

The UK has chosen a different approach than the EU and the US regarding AI regulation. Rather than implementing strict new laws, the UK government is prioritizing voluntary agreements with major AI firms[27]. In November 2023, the AI Safety Summit at Bletchley Park gathered global leaders and significant tech companies, including Google, OpenAI, and Meta. These firms consented to allow the UK's newly established AI Safety Institute early access to their most advanced AI models for risk assessment prior to and following their release[28]. This evaluation addresses issues such as bias, misinformation, security threats, and the risk of misuse. However, since these agreements are not legally enforceable, companies are not legally obliged to adhere to them; instead, they are anticipated to act in good faith[29].

In 2024, this strategy had a global impact when 16 prominent AI corporations from the U.S.,

[27] UK Government, *A Pro-Innovation Approach to AI Regulation* (Policy Paper, March 2023) https://www.gov.uk/government/publications/ai-regulation-a-pro-innovation-approach
[28] UK Government, *AI Safety Summit 2023: Bletchley Declaration* (1–2 November 2023) https://www.gov.uk/government/publications/ai-safety-summit-2023-bletchley-declaration
[29] UK Department for Science, Innovation and Technology, *Establishment of the AI Safety Institute* (2023) https://www.aisi.gov.uk

China, Europe, and other regions committed to a deal stating that they would refrain from releasing AI models that pose risks too hazardous to manage[30]. The UK also initiated collaborations with other nations, such as Singapore and the United States, to exchange research and testing methodologies, forming a network of AI Safety Institutes worldwide[31]. This aims to foster international collaboration on AI safety, as AI technology transcends national boundaries.

The UK's strategy is often referred to as the "Bletchley Effect," a more gentle, cooperative approach that seeks to strike a balance between innovation and safety without hastily adopting stringent regulations. Proponents believe this adaptable model will position the UK as a leading center for AI safety research globally[32]. However, detractors contend that voluntary agreements might lack sufficient strength. Given that companies are not legally required to comply, these commitments could potentially serve more as a means of enhancing public image rather than ensuring true accountability. Recently, the UK even rebranded its institute as the AI Security Institute, indicating a shift in focus towards national security challenges such as cyber threats, rather than broader ethical considerations[33].

Regulatory Solutions

While artificial intelligence has brought about incredible potential, it has also brought about previously unseen risks. Autonomous systems were not considered when creating India's current legal frameworks, which include the IT Act, BNS, DPDP Act, and intellectual property legislation. They find it difficult to deal with issues of accountability, transparency, and liability. For instance, who should be held accountable in the event of a collision involving a self-driving car the software developer, the manufacturer, or the passenger? In a similar vein, who bears responsibility for a deepfake that goes viral the person who made it, the website that hosted it, or the AI tool that made it possible? The only post-facto remedies available under the current systems are takedown notices and legal litigation, which are sometimes too delayed to reverse electoral manipulation or reputational harm. This gap demonstrates the urgent need for a

[30] UK Government, *AI Seoul Summit 2024: International Commitments* (May 2024) https://www.gov.uk/government/publications/ai-seoul-summit-2024-international-commitments
[31] ibid
[32] Matt Clifford, 'The "Bletchley Effect" and the Future of Global AI Governance' (Tony Blair Institute for Global Change, December 2023) https://institute.global
[33] UK Government, *AI Safety Institute Rebranded as AI Security Institute* (July 2024) https://www.gov.uk/government/news/ai-safety-institute-rebranded-as-ai-security-institute

proactive, layered regulatory strategy.

1. Establishing Clear Liability Standards

The cornerstone of AI regulation must be clear responsibility allocation. Presently, liability for AI-related harms is dispersed and uncertain. Borrowing from product liability principles, AI developers and deployers should be held accountable for foreseeable risks, while platforms that host AI-generated content should retain safe-harbor protections only if they respond quickly to damaging content. This ensures that responsibility is distributed proportionally throughout the ecosystem, rather than relying just on victims to prove fault. A legal framework that differentiates between primary liability (developers and deployers), secondary liability (such as platforms), and user liability (malicious actors) will enable more equitable and efficient enforcement.

2. Adopting a Risk-Based Regulatory Model

Not all artificial intelligence systems represent the same threat to society. Spam filters, language-learning bots, and AI employed in games pose negligible threats, whereas applications in healthcare, law enforcement, and election processes have a direct influence on fundamental rights and democratic integrity. Inspired by the EU's AI Act, India should divide AI systems into four risk tiers: unacceptable, high, limited, and minimum. Systems that pose unacceptable risks, such as artificial intelligence for voter manipulation or mass biometric surveillance, should be simply outlawed. High-risk systems require rigorous oversight, such as independent audits, data quality standards, and required human-in-the-loop techniques. Transparency standards can govern limited-risk applications, whereas minimal-risk systems should be left completely deregulated in order to foster innovation. This strategy provides proportionate safeguards without discouraging exploration.

3. Implementing Mandatory Safeguards

For high-risk and general-purpose AI systems, regulation should impose mandatory safety protocols. These include:

- Bias and Safety Assessments: Developers must test models for discriminatory outcomes, misinformation, and security vulnerabilities before release.

- Human Oversight: Critical decisions in healthcare, financial services, or criminal justice must always involve human review.

- Content Transparency: Platforms should label AI-generated content, while watermarking, digital signatures, and authenticity verification tools should become standard to prevent misinformation and impersonation.

- Auditability: Regulators should require organizations to maintain logs of AI decision-making processes, enabling accountability in case of disputes.

Such safeguards build public trust and reduce the risks of opaque, unregulated systems shaping people's lives.

4. Protecting Victims and Ensuring Redress

One of the most difficult difficulties in AI governance is ensuring that people harmed by AI have access to justice. Victims of deepfakes, identity theft, or algorithmic discrimination frequently struggle to identify the perpetrators or establish culpability. To remedy this, India may require insurance schemes or compensation funds for high-risk AI systems. Just as automobiles require third-party insurance, AI deployers in sensitive industries may be obliged to contribute to compensation mechanisms. This guarantees that victims receive prompt assistance while also spreading risk throughout the sector.

5. Embedding Flexibility and Adaptability

AI technologies grow quickly, rendering static laws obsolete. Effective regulation must consequently incorporate adaptability. To do this, new AI systems can be evaluated in regulated sandboxes before being deployed on a larger scale, Laws and norms should be reviewed every 2-3 years to reflect technology advancements and expert groups should offer sector-specific guidelines to ensure responsiveness without frequent legislation modifications.

This flexible approach will allow India to maintain both regulatory certainty and technological dynamism.

6. Promoting Global Cooperation and Public Awareness

The negative effects of AI extend across borders. Content created in one country might have

an immediate influence in another. India must actively participate in developing international standards for watermarking, AI-generated content disclosures, election safeguards, and cross-border liability. Bilateral and international collaborations, such as the EU-US data transfer agreements, help enable coordinated enforcement. Domestically, the government should prioritise public education about AI literacy. Citizens who understand their rights and can identify AI-generated content are less susceptible to exploitation. Whistleblowers who reveal AI-related hazards should be safeguarded, and open research into AI vulnerabilities should be encouraged to increase societal resilience.

Conclusion

Artificial intelligence has outpaced the legal and ethical frameworks meant to regulate it. Deepfakes, improper use of biometrics, and ambiguous algorithmic rulings highlight how inadequate India's current patchwork of protections under the BNS, the IT Act, the DPDP Act and Rules, and disparate intellectual property laws are. Even while these frameworks have their uses, they are still reactive and unprepared to handle the rapidity and scope of problems caused by AI. This study has demonstrated that although Indian courts have started to safeguard IP and personality rights against abuse by AI, implementation is still tardy and inconsistent. The contrast with global strategies emphasises the need for India to embrace a more proactive, risk-based regulatory framework rather than relying solely on band-aid solutions. While the U.S. and U.K. models highlight the significance of striking a balance between innovation and accountability, the EU's AI Act highlights the relevance of precise risk classification.

Four pillars must support India's regulatory strategy going forward: (1) unambiguous liability standards that assign accountability to developers, deployers, and platforms; (2) mandatory safeguards like bias testing, watermarking, and disclosure of content generated by AI; (3) more robust data protection and redress mechanisms for victims of AI misuse; and (4) international cooperation on common standards to confront the cross-border nature of AI. In the end, governments and businesses cannot handle AI governance alone. It necessitates shared accountability, with legislators crafting flexible laws, courts reinterpreting established principles to address novel issues, tech companies incorporating ethics into their designs, and citizens acquiring the literacy necessary to recognise and fend against exploitation. India can guarantee that AI enhances democracy, privacy, and trust rather than weakens them with such a collaborative structure.