

Deepfakes and Misinformation: Tackling Misinformation in the Age of AI

Table of content

I. Abstract	3
II. Introduction	4
III. Understanding Deepfakes and Misinformation	5
IV. How AI Enables the Spread of Misinformation: Technological Drivers	6
V. Impacts of Deepfakes and Misinformation in India	8
A. Legal Implications	8
B. Social and Political Impact	9
C. Economic Consequences	10
VI. India's legal framework on Deep Fakes and Misinformation	11
VII. Regulatory Gaps and Challenges in the Legal sphere	12
A. Absence of a Dedicated Deepfake Law	12
B. Ineffective Enforcement of IT Rules	13
C. Lack of AI-Specific Regulations	14
D. Challenges in Identifying Deepfakes	14
E. Limited Legal Recourse for Victims	15
F. Limited Platform Accountability and Transparency	16
G. Insufficient Public Awareness and Digital Literacy	16
VIII. Comparative Analysis of Global Approaches to Combat Deepfakes	17
A. Legislative Measures: Lessons from the United States and the European Union	17
B. AI-Based Detection Systems: Learning from China	18
C. Public Awareness Campaigns: The UK Model	19
D. Media and Social Media Regulations: Insights from Singapore	19
E. Collaboration with Tech Companies: The South Korean Strategy	20
IX. Policy recommendations	21
A. Enact the Deepfake and Digital Manipulation Prevention Act (DDMPA) under the Information Technology Act, 2000	24

B. Amend the IT Rules, 2021 under the Information Technology Act, 2000	24
C. Introduce AI Risk Regulation under the Draft Digital India Act, 2025	25
D. Build AI Forensics Infrastructure under the Cybersecurity Framework of MeitY	25
E. Amend Section 67A of the Information Technology Act, 2000 and Relevant IPC Sections	25
F. Amend Section 79 of the Information Technology Act, 2000	26
G. Integrate AI Literacy into the National Education Policy (NEP)	26
H. Expand India's Cyber and AI Diplomacy under the Ministry of External Affairs	27
X. Conclusion	29

I. Abstract

The rise of AI-driven deep fakes and misinformation is reshaping how we perceive truth in the digital age. With tools like Generative Adversarial Networks (GANs), anyone can now create highly convincing fake videos, audios, or images by blurring the lines between reality and fabrication¹. This article explores how such synthetic content is being used to spread political propaganda, commit financial fraud, damage reputations, and even harass individuals, particularly women, through non-consensual content. Focusing on India, the study examines the country's legal response, including the Information Technology Act, 2000², Indian Penal Code, 1860³ and the upcoming Digital India Act, 2025⁴. Despite these frameworks, critical gaps remain, in enforcement, clarity, and AI-specific safeguards⁵. To provide perspective, the article compares India's stance with global efforts like the U.S. Deepfakes Accountability Act⁶ and China's AI content regulation⁷ offering lessons India can adapt to its context. In response, the study puts forward clear and actionable recommendations like to create a National AI Governance Authority (NAIGA), mandate watermarking for AI-generated content, and launch widespread digital literacy programs. Ultimately, this research calls for a balanced legal and technological approach, one that keeps pace with innovation while protecting public trust, democratic processes, and individual dignity.

II. Introduction

The rapid evolution of artificial intelligence (AI) has given rise to deepfakes as synthetic media created through advanced machine learning techniques, particularly Generative Adversarial Networks (GANs), which can produce hyper-realistic audio, video, and imagery that mimic real individuals. These technologies, originally developed for innovation and creative exploration, have now become tools for large-scale misinformation, political manipulation, financial fraud, and digital harassment⁸. In India, where over 850 million people actively access the internet⁹. The impact of deepfakes is particularly pronounced. The dissemination of synthetic content during election cycles, financial transactions, and social unrest has exposed severe gaps in the country's legislative preparedness. While the **Information Technology Act, 2000 includes relevant provisions such as Sections 66D (identity fraud), 67A (explicit content), and 79 (intermediary liability)**¹⁰ and the **Indian Penal Code, 1860 criminalizes forgery, misinformation, and defamation through Sections 468, 471, and 505(1)(b)**¹¹, , these statutes are not tailored to address AI-generated

¹ [Kietzmann et al., 2020](#)

² [Information Technology Act, 2000](#)

³ [Indian penal code, 1860](#)

⁴ [Digital India Act, 2025](#)

⁵ [MeitY, 2024](#)

⁶ [U.S. Deepfakes Accountability Act](#)

⁷ [China's AI content regulation](#)

⁸ [Kietzmann, Lee, McCarthy, & Kietzmann, 2020, p. 135.](#)

⁹ [MeitY, 2023, para. 4](#)

¹⁰ [Information Technology Act, 2000](#)

¹¹ [Indian Penal Code, 1860](#)

synthetic media. Furthermore, institutional enforcement remains weak due to jurisdictional ambiguity and the anonymity of cyber perpetrators¹².

This study critically investigates the technological and legal challenges posed by deepfakes in India, and highlights gaps in existing frameworks, and offers a comparative overview of international responses. The European Union's Digital Services Act mandates transparency, algorithmic accountability, and proactive content moderation by platforms¹³, while the United States introduced the Deepfakes Accountability Act to enforce watermarking, disclosure, and penal provisions for AI-generated disinformation¹⁴. These models underscore the global recognition of synthetic media as a regulatory priority. The article advocates for a comprehensive Indian strategy that includes a dedicated AI legal framework, platform-level detection mandates, institutional capacity-building, and public digital literacy campaigns. As synthetic media continues to evolve in sophistication, India's ability to preserve democratic integrity and information security will depend on the speed and scale of its regulatory response.

III. Understanding Deepfakes and Misinformation

Deepfakes are a form of synthetic media created using artificial intelligence (AI) to manipulate or fabricate visual, audio, or textual content. These hyper-realistic alterations rely on deep learning techniques, especially **Generative Adversarial Networks (GANs)**, to produce content that closely mimics real human expressions, speech, and actions. The term "**deepfake**" itself is derived from "**deep learning**" and "**fake**," emphasizing the AI-driven nature of the technology. Misinformation, on the other hand, refers to the spread of false or misleading information, whether intentional (disinformation) or unintentional. While misinformation has always existed, AI-powered content creation has significantly amplified its impact, making it more convincing and harder to detect. Deepfakes, as an advanced form of misinformation, pose a severe challenge in today's digital landscape. They have been used in **political propaganda, cyber fraud, blackmail, and even in altering historical narratives**. According to the IRJAES Journal (2024)¹⁵ AI-generated misinformation has experienced **exponential growth**, shaping public discourse and perceptions in ways that were previously unimaginable.

One of the most concerning aspects of deepfakes is their **high level of realism**, which makes them incredibly deceptive. AI-generated videos and images can replicate real-life facial expressions, body movements, and speech patterns, making it increasingly difficult to differentiate between authentic and manipulated content. This has led to the weaponization of deepfakes in politics, media, and finance, where they are used to manipulate elections, create fake evidence, and carry out financial fraud. Another alarming characteristic of deepfakes is their **ease of access and low barrier to entry**. Open-source tools and AI models have made it possible for individuals with minimal technical expertise to create highly realistic deepfakes, further increasing their misuse. Additionally, **the rapid evolution of AI has made deepfake detection increasingly difficult**. While early deepfake videos contained visible flaws such as unnatural facial movements or mismatched audio,

¹² [MeitY, 2024, pp. 7-8](#)

¹³ [European Commission, 2022](#)

¹⁴ [U.S. Congress, 2019](#)

¹⁵ [IRJAES JOURNAL, 2024](#)

modern versions have significantly improved. A study published in SSRN (2023)¹⁶ highlights how deepfake detection tools are struggling to keep pace with AI-driven forgeries, raising concerns about their potential role in disinformation campaigns, cybercrimes, and identity theft.

The **evolution of deepfakes** has been driven by AI advancements and increased accessibility. Before 2017, AI-based morphing techniques were rudimentary and primarily confined to research environments. However, the introduction of GANs revolutionized the field, making deepfake technology widely available and increasingly sophisticated. In recent years, deepfake technology has been weaponized in political misinformation campaigns, with AI-generated videos influencing elections and distorting public opinion. For instance, the 2024 elections in multiple countries witnessed a surge in deepfake-driven misinformation, proving how AI-powered deception can have severe consequences for democracy. A report in IEEE Xplore (2024)¹⁷ explains how deepfakes have been used to spread false narratives, erode trust in legitimate sources, and manipulate political discourse. As AI continues to evolve, deepfake capabilities will become even more advanced, posing ethical, legal, and security challenges. The widespread accessibility of this technology underscores the urgent need for robust detection mechanisms, stronger legal frameworks, and greater public awareness to combat the rising threat of AI-generated misinformation.

IV. How AI Enables the Spread of Misinformation: Technological Drivers

Artificial intelligence (AI) has fundamentally changed the way misinformation is created, disseminated, and consumed, making false content more believable, scalable, and difficult to detect. Key technological advancements like **deep learning, generative AI, automated social media manipulation, real-time misinformation, and AI-driven microtargeting** have significantly contributed to this growing problem.

One of the most alarming developments is **deep learning and generative AI**, particularly through **Generative Adversarial Networks (GANs)**. These models enable the creation of hyper-realistic deepfakes manipulated videos, images, and audio that make fabricated events appear real. Deepfakes have been widely used in political campaigns, financial scams, and social engineering, making it increasingly difficult to trust digital media. During the 2024 European elections, deepfake videos misrepresenting politicians went viral, misleading voters before fact-checkers could respond¹⁸. The growing ease of creating high-quality fakes has led to a trust deficit in news, government statements, and even live broadcasts, as people struggle to distinguish between real and AI-generated content¹⁹.

Beyond deepfakes, AI-powered **text generation tools** have facilitated the **mass automation of fake news production. Large Language Models (LLMs)**, such as GPT-based systems, can generate entire news articles, fabricated interviews, and misleading reports with human-like fluency. Misinformation now spreads at an unprecedented scale, as AI-generated text can be customized to resemble credible sources. Studies show that AI-generated misinformation spreads up to ten times faster than human-written falsehoods, overwhelming

¹⁶ [SSRN, 2023](#)

¹⁷ [IEEE Xplore, 2024](#)

¹⁸ [Tandfonline, 2020](#)

¹⁹ [IEEE Xplore, 2024](#)

fact-checkers and amplifying confusion among readers²⁰. AI doesn't just create misinformation, it optimizes its believability and impact, making it harder for the public to discern fact from fiction²¹.

Another major contributor to the spread of misinformation is **AI-powered social media manipulation. Automated bots flood platforms with false information**, artificially increasing engagement through likes, shares, and comments to boost the visibility of misleading content. Social media algorithms, designed to prioritize viral content over accuracy, further amplify disinformation, often pushing false narratives over verified news. AI-driven misinformation campaigns are designed to provoke emotional responses, making people more likely to engage with and share misleading content. Political propaganda, conspiracy theories, and divisive content are particularly vulnerable to algorithmic amplification, shaping public opinion in dangerous ways²².

Perhaps the most **disturbing development is real-time AI-generated misinformation**. AI-powered tools can now manipulate live video and audio streams in real time, making it possible to alter speeches, financial reports, and news broadcasts instantly. This has already been used in financial fraud, where fake live-streamed press conferences misled investors, resulting in significant stock market losses²³. The ability to alter reality in real time means that misinformation can influence real-world decisions before it can be debunked, posing a significant challenge for governments, journalists, and cybersecurity experts.

Finally, **AI-driven microtargeting has made misinformation more personalized and persuasive**. Unlike traditional misinformation campaigns that relied on mass distribution, AI now enables the customization of false narratives for specific individuals or groups. Social media algorithms analyze user behavior, preferences, and biases, ensuring that misinformation is delivered to those most likely to believe it. This creates echo chambers where false narratives are reinforced, making individuals more susceptible to manipulation. In political campaigns, micro targeted misinformation has been used to suppress voter turnout, influence election outcomes, and deepen social divisions²⁴.

These technological advancements have made misinformation more believable, scalable, and difficult to counteract, fundamentally altering how people engage with information. As AI continues to evolve, addressing its role in misinformation requires proactive intervention from policymakers, technology companies, and civil society to preserve public trust and democratic integrity.

V. Impacts of Deepfakes and Misinformation in India

Deepfakes, a product of advanced artificial intelligence, have emerged as a significant threat to digital integrity in India. These AI-generated synthetic media have been exploited for various malicious purposes, from misinformation campaigns to financial fraud. The rapid rise of deepfake technology has triggered legal

²⁰ [SSRN, 2024](#)

²¹ [ProQuest, 2024](#)

²² [Cambridge, 2024](#)

²³ [Springer, 2024](#)

²⁴ [WEF, 2024](#)

challenges, social disruption, and economic concerns. This paper examines the multifaceted impact of deepfakes in India, exploring their legal, social, and economic consequences.

A. Legal Implications

One of the most prominent legal implications of deepfakes lies in the realm of **defamation and false light**. Deepfakes can be engineered to make it appear as if a person has said or done something they never did often with a high degree of realism. This can result in serious reputational harm, character assassination, and social ostracism. Traditional defamation laws in India, such as Sections 499 and 500 of the IPC, address the publication of false statements that damage an individual's reputation. However, these statutes were not conceived with AI-generated impersonations in mind. Furthermore, Indian jurisprudence does not explicitly recognize "false light", a privacy tort recognized in jurisdictions like the United States, which protects individuals from misleading public portrayals that are offensive or embarrassing²⁵. The absence of this framework limits legal recourse for those whose identities are synthetically misrepresented without direct defamatory claims.

In the context of **copyright infringement**, deepfakes pose a significant challenge to intellectual property rights. These synthetic creations often rely on unauthorized use of copyrighted content, including video clips, voice recordings, performances, and biometric data to generate manipulated media. Under the Indian Copyright Act, 1957, the creator of an original work holds exclusive rights over reproduction and derivative use. However, the Act remains silent on AI-generated works, particularly when the content is altered using machine learning models trained on thousands of protected materials. Additionally, there is no clear legal doctrine for assigning liability, whether to the tool developer, the user, or the AI system itself²⁶. As a result, victims of deepfake misuse may struggle to establish ownership and control over their digital likeness or intellectual creations.

Equally critical are the **privacy rights** threatened by deepfakes, especially those involving non-consensual image-based abuse. The Supreme Court's verdict in *Justice K.S. Puttaswamy v. Union of India* (2017)²⁷ affirmed the constitutional status of privacy under Article 21, setting a strong legal precedent for challenging unauthorized data use and identity manipulation²⁸. Yet, the current data protection and IT laws fall short in addressing synthetic identity fabrication. The recently enacted Digital Personal Data Protection Act, 2023, provides general provisions for consent and personal data processing, but does not directly tackle the reproduction of facial features, voices, or body movements by AI²⁹. This leaves victims vulnerable to digital invasion without a targeted legal framework to claim redress for deepfake-enabled privacy violations.

²⁵ [El-Garhy, 2024](#)

²⁶ [Gautam, 2024](#)

²⁷ [K.S. Puttaswamy v. Union of India \(2017\)](#)

²⁸ [Supreme Court of India, 2017](#)

²⁹ [MeitY, 2023](#)

The issue becomes even more acute in the domain of **cybercrimes and harassment**, where deepfakes are increasingly used as instruments of abuse. Women, in particular, are disproportionately affected by non-consensual pornographic deepfakes, which are shared online to intimidate, shame, or extort. Although Section 67A of the Information Technology Act, 2000, criminalizes the publication of sexually explicit content in electronic form, the provision does not account for synthetic media or the nuanced absence of real consent in deepfakes³⁰. Moreover, victims often face institutional apathy, limited investigative capacity, and delayed justice. Reports by the National Commission for Women indicate that deepfake complaints have risen sharply in recent years, but the response mechanisms remain fragmented and under-resourced³¹.

Given the complexity and severity of these challenges, there is an urgent need for new laws and regulations that specifically address deepfake technology. India's proposed Digital India Act (2023) is a step in the right direction, aiming to replace the outdated IT Act and impose stricter accountability on digital platforms. However, the draft legislation does not yet outline obligations for watermarking, AI content disclosure, or explicit consent in media synthesis³². In contrast, the **European Union's Artificial Intelligence Act** mandates transparency and categorizes AI systems based on the level of risk they pose, with deepfakes considered high-risk and subject to disclosure requirements³³. Similarly, the **U.S. DEEPFAKES Accountability Act** proposes watermarking, source attribution, and criminal penalties for non-consensual or malicious synthetic content³⁴. These international efforts provide valuable models for India, underscoring the necessity for legislation that anticipates the ethical and legal dilemmas posed by generative AI.

B. Social and Political Impact

Deepfakes have significantly eroded public trust in media, especially during election cycles. The **2019 General Elections** saw instances of AI-generated political misinformation, prompting the Election Commission of India (ECI) to raise concerns about their influence on voter behavior³⁵. A **2021 Reuters Digital Report** found that 63% of Indians struggle to differentiate real news from fake, underscoring the growing challenge of combating misinformation³⁶. This erosion of trust weakens democratic institutions, as manipulated content can mislead voters and fuel political polarization.

Another alarming social consequence of deepfakes is their disproportionate impact on women's safety. AI-generated **non-consensual deepfake pornography** has become a tool for harassment, blackmail, and reputational damage. The **National Commission for Women (NCW)** reported a significant rise in such cases, emphasizing the severe psychological toll on victims³⁷. However, the legal system has been slow to

³⁰ [Information Technology Act, 2000](#)

³¹ [NCW, 2022](#)

³² [MeitY, 2023](#)

³³ [European Commission, 2023](#)

³⁴ [U.S. Congress, 2023](#)

³⁵ [ECI Report, 2019](#)

³⁶ [Reuters Digital Report, 2021](#)

³⁷ [NCW Report, 2022](#)

address these violations, with many victims struggling to seek justice. Strengthening laws against AI-enabled sexual harassment and enhancing digital safety measures is crucial to tackling this issue effectively.

C. Economic Consequences

Beyond legal and social disruptions, deepfakes pose significant economic risks, particularly in the financial sector. AI-driven **identity fraud and banking scams** have become increasingly sophisticated, as deepfake technology enables cybercriminals to bypass biometric security systems. The **2020 RBI Cyber Security Report** highlighted multiple cases of deepfake-related financial fraud, warning about the vulnerability of digital banking platforms³⁸. In 2023, **CERT-In (India's cybersecurity agency)** issued alerts regarding deepfake scams targeting Indian banks and digital payment platforms, emphasizing the need for stronger AI-based fraud detection systems³⁹.

Deepfakes have also negatively affected corporate reputations and market stability. The **2022 Business Ethics Survey** by FICCI found that 47% of Indian companies faced brand damage due to AI-generated misinformation, with fake videos of CEOs making false announcements causing stock market fluctuations⁴⁰. Misinformation campaigns driven by deepfakes can result in consumer distrust, stock manipulation, and corporate crises, demonstrating the urgent need for AI-driven content verification systems in business environments.

Deepfakes and misinformation pose a serious threat to India's legal, social, and economic stability. While existing laws such as the IT Act, 2000, and IPC provisions provide partial protection, they remain insufficient to fully tackle AI-generated content abuse. The Digital India Act, 2023, if implemented, could bridge these gaps by introducing stronger AI regulations and platform accountability measures. However, beyond legal frameworks, India must also focus on public awareness, advanced AI detection systems, and corporate vigilance to combat the deepfake crisis effectively. Addressing this issue requires a **multi-pronged approach that balances innovation with ethical responsibility**, ensuring that digital technologies serve society without compromising truth and security.

VI. India's legal framework on Deep Fakes and Misinformation

³⁸ [RBI Cyber Report, 2020](#)

³⁹ [CERT-In, 2023](#)

⁴⁰ [FICCI, 2022](#)

Year	Legislation/ Case	Key provisions
1860	Indian Penal Code	<ol style="list-style-type: none"> Section 468 & 471: Criminalizes forgery and fraudulent use of digital content. Section 505(1)(b): Penalizes misinformation inciting violence. Section 500: Covers defamation, including deepfake-related reputation harm.⁴¹
2000	Information Technology (IT) Act	<ol style="list-style-type: none"> Section 66D: Punishes identity fraud via deepfakes. Sections 67 & 67A: Prohibit sexually explicit deepfake content. Section 79 (Amendment, 2008): Grants "safe harbor" to platforms unless they fail to remove flagged deepfake content.⁴²
2015	Shreya Singhal Case (Free Speech & Intermediary Liability)	<ol style="list-style-type: none"> Struck down Section 66A of the IT Act, ensuring free speech protections. Upheld platform accountability, restricting harmful misinformation.⁴³
2017	Puttaswamy Judgment (Right to Privacy)	<ol style="list-style-type: none"> Recognized right to privacy under Article 21. Strengthens legal grounds against unauthorized deepfake use.⁴⁴
2023	Bharatiya Nyaya Sanhita (BNS), 2023	<ol style="list-style-type: none"> Modernized criminal law, replacing IPC. Recognizes malicious deepfake use as a criminal offense.⁴⁵
2023	Digital Personal Data Protection (DPDP) Act, 2023	<ol style="list-style-type: none"> Requires user consent before processing personal data. Establishes penalties for deepfake-related identity fraud.⁴⁶
2024	Government Initiatives on Deepfake Regulation	<ol style="list-style-type: none"> AI-based deepfake detection mandated for platforms. Introduced misinformation tracking systems for elections. Awareness programs launched to educate citizens.⁴⁷
2025 (Upcoming)	Digital India Act (DIA), 2025	<ol style="list-style-type: none"> Replaces IT Act, 2000 with modern AI regulation. Criminal penalties for deepfake creation & distribution. Establishes Digital Safety Authority for oversight.⁴⁸
2025 (Expected Implementation)	Deepfakes Regulation Measures	<ol style="list-style-type: none"> Mandatory removal of flagged deepfake content within a timeframe. Fines and imprisonment for deepfake-related crimes. AI-driven misinformation monitoring by regulatory bodies.⁴⁹

Table 1 : India's Legal Framework on Deepfakes and Misinformation

⁴¹ [IPC, 1860](#)

⁴² [IT Act, 2000](#)

⁴³ [Supreme court, 2015](#)

⁴⁴ [Supreme Court, 2017](#)

⁴⁵ [CPO](#)

⁴⁶ [PIB](#)

⁴⁷ [Indian Express, 2024](#)

⁴⁸ [Mondaq, 2025](#)

⁴⁹ [The Print, 2025](#)

VII. Regulatory Gaps and Challenges in the Legal sphere

The rapid advancement of artificial intelligence (AI), particularly deepfake technology, has significantly outpaced India's existing legal frameworks. While digital regulations have evolved, the misuse of AI-generated content poses severe threats, including misinformation, identity theft, and non-consensual exploitation. Despite legislative measures, the legal system lacks a comprehensive approach to addressing these challenges. The following are key regulatory gaps and enforcement challenges in India's legal landscape.

A. Absence of a Dedicated Deepfake Law

India currently lacks specific legislation to regulate deepfake technology, resulting in a significant regulatory vacuum. The **Information Technology Act, 2000** provides a general legal framework to address cybercrimes but does not explicitly recognize or regulate AI-generated synthetic content, including deepfakes⁵⁰. Similarly, while **Sections 499 and 500 of the Indian Penal Code (IPC)** define and penalize defamation, they were enacted in the 19th century and do not contemplate the technological sophistication or psychological impact of AI-driven impersonations⁵¹. This lack of a dedicated legal structure allows deepfake creators to exploit loopholes, often escaping accountability by operating anonymously or across borders. What complicates enforcement even further is that **deepfake content frequently exists in a grey area between parody and malicious misinformation. A synthetic video that mimics a public figure might be presented as political satire or creative commentary, thus falling under the protection of free expression.**

However, the same techniques can be used to fabricate inflammatory speeches or explicit content that cause tangible harm, all while evading legal classification due to the ambiguous nature of intent and interpretation⁵². Indian law currently lacks the nuance or legal tests necessary to differentiate between protected expression and harmful manipulation in the context of deepfakes. Although provisions related to forgery, impersonation, or identity theft may sometimes be applied, they do not address the full range of ethical and societal harms posed by synthetic media, nor do they provide clarity on platform responsibilities or user consent in such cases.

B. Ineffective Enforcement of IT Rules

The **Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules, 2021** place a legal obligation on intermediaries particularly social media platforms to remove unlawful content, including misinformation, within a stipulated time frame⁵³. However, enforcement of these rules has remained inconsistent. Many platforms operate under broad or poorly defined categories of **"harmful AI-generated content,"** which leads to uneven moderation practices. Some platforms remove deepfake videos promptly, while others fail to act until the content has already gone viral, by which time the

⁵⁰ [Information Technology Act, 2000](#)

⁵¹ [Indian Penal Code, 1860](#)

⁵² [El-Garhy, 2024](#)

⁵³ [MeitY, 2021](#)

reputational and psychological harm is often irreversible. A key reason for this **ineffectiveness is the absence of standardized AI content detection mechanisms**.

Most platforms rely on user reporting rather than proactive detection, and the lack of government-mandated technology audits or compliance protocols means enforcement is largely self-regulated. This is in sharp contrast to frameworks like the **European Union’s Digital Services Act**, which requires platforms to proactively identify, label, and remove AI-generated disinformation and deepfakes through mandated risk assessments and transparency reports⁵⁴. Adding to the regulatory vacuum, there is **currently no centralized grievance redressal mechanism for victims of deepfake abuse in India**, making it difficult for individuals to report, escalate, and seek resolution for digital impersonation or manipulation. Unlike countries such as the **United Kingdom or South Korea**, which have structured legal pathways for synthetic media complaints, including digital tribunals or AI-specific complaint portals. India’s response remains fragmented across law enforcement units, platform-specific channels, and civil court remedies⁵⁵. This **absence of institutional clarity** exacerbates the trauma of victims and delays legal redress, further undermining the deterrence capability of the existing rules.

C. Lack of AI-Specific Regulations

India currently lacks a comprehensive regulatory framework tailored specifically to artificial intelligence, despite the rapid growth of AI-driven technologies like deepfakes. Unlike the **European Union’s Artificial Intelligence Act**, which introduces a risk-based classification system for AI applications and mandates transparency, human oversight, and safety compliance for high-risk tools⁵⁶. India has yet to implement any such centralized policy. The absence of AI-specific law creates significant ambiguity when it comes to liability, ethical use, and institutional accountability in cases involving deepfake abuse.

Although **the Personal Data Protection Bill, 2019, now replaced by the Digital Personal Data Protection Act, 2023**, focuses on user privacy, it does **not address the technological complexity or socio-legal challenges posed by synthetic media, biometric mimicry, or AI-generated identity theft**⁵⁷. For instance, while the law protects “personal data,” it does not prohibit the creation of manipulated AI content using that data, nor does it provide remedies for those impersonated through deepfakes. This regulatory vacuum has serious implications for public safety and democratic integrity. Without clear compliance standards or risk categorization mechanisms, AI developers and platforms operating in India face no binding obligation to disclose synthetic content, watermark deepfakes, or conduct algorithmic impact assessments. As a result, crimes such as AI-generated misinformation in election campaigns, deepfake financial frauds, or fabricated political endorsements remain legally under-regulated and difficult to prosecute⁵⁸. Moreover, the **absence of an AI regulatory body, such as a**

⁵⁴ [European Commission, 2022](#)

⁵⁵ [NCW, 2022](#)

⁵⁶ [European Commission, 2023](#)

⁵⁷ [PRS India, 2019](#)

⁵⁸ [Gautam, 2024](#)

National AI Commission or Ombudsman - means there is no centralized authority to issue ethical guidelines, enforce compliance, or coordinate inter-agency responses to AI-related harms.

D. Challenges in Identifying Deepfakes

One of the most pressing hurdles in addressing deepfake-related crimes is the challenge of accurate and timely detection. **Technological limitations** significantly hinder the ability of law enforcement agencies to identify synthetic media in real-time. While some advanced AI-based detection tools exist globally, they remain underdeveloped and inconsistently deployed in India. The **absence of a standardized, nationwide strategy for AI-based forensic investigation** exacerbates the issue. Most detection still relies on manual reporting or rudimentary screening mechanisms, which are often insufficient in the face of increasingly sophisticated deepfakes.

India's forensic and cybercrime laboratories, especially at the state and district levels, are critically under-resourced and under-equipped to carry out timely verification of AI-generated content. Many lack the technical infrastructure, trained personnel, and real-time access to algorithmic forensics needed to detect deepfakes before they go viral. This delay often leads to irreversible reputational, financial, or political damage, particularly when synthetic content spreads during sensitive moments such as elections or corporate disputes⁵⁹. The **NITI Aayog National Strategy for Artificial Intelligence (2018)**, while commendable in setting a broad vision for AI in sectors like healthcare and agriculture, does not provide clear regulatory or technological directives for combating deepfake threats⁶⁰. As a result, no institutional framework currently exists to support law enforcement with algorithmic traceability, digital watermarking, or cross-platform content validation in cases involving synthetic media. Without dedicated AI forensics infrastructure, India remains heavily reactive in its approach to digital manipulation.

E. Limited Legal Recourse for Victims

Victims of deepfake-based misinformation, particularly women, face considerable challenges in accessing timely and effective legal remedies. While **Section 67A of the Information Technology Act, 2000** criminalizes the electronic transmission of sexually explicit content, it does not explicitly address AI-generated synthetic pornography or impersonation without consent⁶¹. **India's existing legal protections, including Section 67A, fail to explicitly criminalize AI-generated sexual content created without consent. This legal ambiguity often impedes justice for women victims of deepfake pornography, as courts and law enforcement struggle to interpret outdated statutory language in light of rapidly evolving technology.** The **Justice Verma Committee Report (2013)**, which was instrumental in shaping reforms to sexual violence laws in India, emphasized the importance of addressing digital sexual exploitation. However, it did not foresee the advent of AI-driven abuse, where

⁵⁹ [MeitY, 2023](#)

⁶⁰ [NITI Aayog, 2018](#)

⁶¹ [Information Technology Act, 2000](#)

victims can be virtually inserted into pornographic or degrading scenarios using machine learning techniques⁶².

As a result, the legal framework **lacks specificity on the issue of synthetic consent**, where the victim's likeness is used without any physical act, yet causes significant psychological and reputational harm. Women targeted by deepfake pornography frequently encounter procedural and legal roadblocks when attempting to report such crimes. Many cybercrime cells **lack the training and tools to investigate AI-generated abuse, and ambiguous legal definitions surrounding consent, impersonation, and digital autonomy further complicate prosecution**. Moreover, victims are often subjected to secondary trauma through delays, social stigma, and lack of institutional support. In contrast, jurisdictions like the **United Kingdom and the United States** have begun drafting deepfake-specific legislation that directly addresses non-consensual synthetic media. The UK's Online Safety Bill and the US Deepfakes Accountability Act both seek to criminalize malicious deepfakes and ensure swift platform takedown and user redress mechanisms⁶³. However, India's legal system remains largely reactive, lacking both statutory clarity and institutional capacity to safeguard victims of AI-enabled digital violence.

F. Limited Platform Accountability and Transparency

In India, the regulatory framework governing digital platforms lacks the robust accountability standards seen in other jurisdictions. For instance, the **European Union's Digital Services Act (DSA)** mandates transparency in content moderation, algorithmic decision-making, and platform responses to AI-generated content, including deepfakes⁶⁴. In contrast, India's approach under the **Information Technology Act, 2000** and its accompanying Intermediary Guidelines places only limited obligations on tech platforms. These platforms enjoy "safe harbor" protections under **Section 79 of the IT Act**, which shield them from legal liability unless they fail to remove unlawful content after receiving official notice⁶⁵. However, these intermediary rules fall short in proactively curbing the spread of harmful deepfake content. Platforms are not required to detect or moderate synthetic content unless prompted by external complaints, resulting in delayed action and inconsistent enforcement. Manipulated videos and misinformation often go viral before any moderation occurs, and responses are frequently reactive triggered only by significant media coverage or public outrage.

The absence of algorithmic transparency requirements allows platforms to avoid disclosing how their systems may amplify or suppress deepfake content, thereby obscuring the extent to which platform design contributes to the virality and influence of synthetic media. Recommendation systems, engagement-based feeds, and algorithmic boosts can inadvertently promote harmful or misleading deepfakes without users or regulators understanding how or why such content gains traction. This lack of transparency hinders both public scrutiny and legal oversight. **Moreover, Indian law does not mandate regular risk assessments, public disclosure of moderation policies, or user access to appeals,**

⁶² [Ministry of Home Affairs, 2013](#)

⁶³ [U.S. Congress, 2023](#)

⁶⁴ [European Commission, 2022](#)

⁶⁵ [Information Technology Act, 2000](#)

measures that are increasingly considered standard in global best practices. As a result, tech companies operating in India are not held to uniform content governance benchmarks and retain wide discretion over what content is flagged, downranked, or removed. In the absence of strict regulatory mandates, platform accountability in moderating deepfakes remains weak, leaving users vulnerable to the spread of deceptive content with little institutional recourse.

G. Insufficient Public Awareness and Digital Literacy

One of the most under-recognized challenges in combating the deepfake threat is the widespread lack of public awareness and digital literacy. Deepfakes, by their nature, exploit human reliance on visual and auditory cues, making it difficult for untrained individuals to distinguish between real and manipulated content. In India, this vulnerability is especially acute in **rural and semi-urban populations**, where digital access has expanded but critical media literacy has not kept pace. Although the **Digital India Programme, launched in 2015**, aims to enhance connectivity and basic digital skills, it does not explicitly focus on educating citizens about the detection and dangers of AI-generated synthetic content⁶⁶. In contrast, countries like **Finland** have adopted a proactive approach by integrating media literacy, including deepfake recognition into school curriculums, civic education, and national teacher training. The Finnish government collaborates with educators, researchers, and media outlets to train students in identifying misinformation, including synthetic videos and altered digital content⁶⁷. This nationwide model has been widely praised as one of the most effective countermeasures against digital deception in **Europe** and stands as an example India could emulate.

Mandatory inclusion of digital media literacy in Indian school curriculums and capacity-building programs for local governance bodies, such as Panchayats, municipal schools, and district-level administrators could significantly improve grassroots resilience. This would empower citizens to critically assess digital information, respond swiftly to misinformation, and avoid being manipulated by malicious actors using AI-generated content. It would also create a network of informed intermediaries capable of educating others within their communities. Currently, most digital literacy programs in India focus on teaching basic operational skills, such as using mobile apps or accessing government portals, without fostering critical thinking or **ethical awareness** about media authenticity. In an age where deepfakes can influence elections, provoke violence, or defame individuals overnight, cultivating media literacy is not just a technological need, it is a democratic imperative.

VIII. Comparative Analysis of Global Approaches to Combat Deepfakes

In the wake of rapid technological advancements, deepfakes have emerged as a significant threat to democracy, national security, and individual privacy. Countries across the globe have developed unique strategies to combat deepfakes, ranging from stringent legislation to advanced technological interventions. India, which is

⁶⁶ [Digital India, 2015](#)

⁶⁷ [European Commission, 2022](#)

experiencing a surge in deepfake-related challenges, can adopt best practices from nations that have effectively addressed this issue.

A. Legislative Measures: Lessons from the United States and the European Union

In the **United States**, the **proposed Deepfakes Accountability Act** aims to introduce strong transparency and liability measures against the misuse of synthetic media. The Act requires that AI-generated content include clear disclosures, such as digital watermarks or labels, to alert viewers that the media has been artificially created. It also imposes legal liability on individuals who knowingly create or distribute deepfakes without consent, particularly when such content results in defamation, harassment, or electoral manipulation⁶⁸. By targeting both creators and distributors of malicious deepfakes, the legislation sets a framework that balances technological innovation with the protection of individual rights and democratic processes.

Similarly, the **European Union** has advanced deepfake regulation through the **Digital Services Act (DSA)**, a landmark piece of legislation that places strong obligations on digital platforms. Under the DSA, major online platforms are required to proactively assess and mitigate systemic risks, including the proliferation of deepfakes. The Act mandates the clear labelling of manipulated content, requires regular audits of content moderation systems, and introduces penalties for non-compliance⁶⁹. This framework recognizes the role of algorithms in amplifying misinformation and holds platforms accountable for creating safer digital environments.

In contrast, **India's** regulatory response remains limited. Although the Information Technology Act, 2000, and the Intermediary Guidelines, 2021, impose content moderation duties, they do not specifically address AI-generated synthetic media or mandate transparency in platform algorithms. To strengthen its legal framework, **India could integrate global best practices such as disclosure requirements, watermarking mandates, and proactive moderation standards. Additionally, India could explore the creation of an independent Digital Content Regulatory Authority, tasked with overseeing AI-generated media, ensuring platform accountability, conducting algorithmic audits, and providing centralized grievance redressal mechanisms.** Establishing such a body would ensure a consistent, forward-looking approach to deepfake governance while reinforcing public trust in the digital ecosystem.

B. AI-Based Detection Systems: Learning from China

China has taken an aggressive approach toward combating the spread of deepfakes by integrating AI-based detection systems directly into its cybersecurity framework. Under the **Provisions on the Administration of Deep Synthesis Internet Information Services**, platforms are legally required to implement pre-upload screening mechanisms to detect and flag synthetic content before it is disseminated

⁶⁸ [U.S. Congress, 2023](#)

⁶⁹ [European Commission, 2022](#)

to the public⁷⁰. These AI-driven filters help curb the circulation of manipulated media at the source, making the detection process faster and more efficient. China's model demonstrates the potential for AI-enabled regulatory infrastructure to proactively mitigate the societal harms caused by deepfake technologies.

India can learn from this technical model by investing in **indigenous AI research and development**, supporting public-private partnerships in AI forensics, and mandating that digital platforms integrate automated deepfake detection tools. Building such a detection ecosystem would allow harmful synthetic content to be identified and removed before it goes viral, minimizing reputational, political, and financial damages. Integrating AI-based verification at both platform and governmental levels could significantly strengthen India's ability to respond to deepfake threats without overly relying on manual reporting or post hoc interventions.

However, it is important to recognize that **China's model operates within an authoritarian governance structure**, where surveillance tools often lack adequate checks and balances. **Such mechanisms raise civil liberties and privacy concerns**, particularly when detection systems could be repurposed for broader censorship or political control. **India must ensure that any deployment of AI-driven deepfake detection systems is balanced with constitutional safeguards**, including the protection of privacy rights as upheld in **the Justice K.S. Puttaswamy v. Union of India decision**⁷¹. Transparent oversight frameworks, independent audits, and clear limitations on surveillance use would be critical to maintaining democratic accountability while enhancing cybersecurity defenses against synthetic media.

C. Public Awareness Campaigns: The UK Model

The **United Kingdom** has adopted a proactive strategy toward building societal resilience against deepfakes by launching large-scale public awareness initiatives. Programs such as the **"Don't Trust, Verify"** campaign educate citizens on the risks associated with synthetic media, teaching them how to critically assess and verify digital content before accepting or sharing it⁷². These campaigns use simple messaging, multimedia content, and partnerships with schools, media houses, and online platforms to reach diverse demographic groups, effectively making media literacy a part of everyday civic education.

Drawing inspiration from this model, **India can develop similar national education campaigns** tailored to its multilingual and socio-culturally diverse population. Existing programs under the **Digital India and Information and Broadcasting Ministry frameworks** can be expanded to specifically address the identification and reporting of deepfake content. Public education would play a pivotal role in enhancing critical digital literacy, reducing vulnerability to AI-driven misinformation, and empowering citizens to act as the first line of defense against synthetic media manipulation.

⁷⁰ [Cyberspace Administration of China, 2022](#)

⁷¹ [Supreme Court of India, 2017](#)

⁷² [UK Government, 2021](#)

India's electoral commissions, education boards, and public broadcasters can co-create civic campaigns under a unified national initiative against AI misinformation, ensuring that awareness efforts are standardized, far-reaching, and sustained. Targeted initiatives during election seasons, collaborations with private media outlets, and curriculum inclusion at school and university levels could institutionalize deepfake awareness as an essential component of democratic citizenship. A coordinated, government-led but community-driven approach would help in building long-term public resilience against the evolving threats of deepfake technologies.

D. Media and Social Media Regulations: Insights from Singapore

Singapore has adopted one of the most structured and rapid legal frameworks for addressing the spread of synthetic media through the **Protection from Online Falsehoods and Manipulation Act (POFMA)**. This legislation empowers designated government authorities to swiftly flag, correct, and, if necessary, remove online content deemed to be false or misleading, including deepfakes⁷³. The law also mandates platforms to publish correction notices and imposes penalties for non-compliance. POFMA's rapid response mechanism ensures that misinformation is addressed before it can cause significant public harm, particularly during politically sensitive periods such as elections.

Drawing from Singapore's model, **India can refine its Information Technology Act and intermediary guidelines to create faster and more transparent procedures for flagging and removing AI-generated deceptive content.** While the current takedown process under Indian law often requires prolonged notice and verification steps, a more agile system, especially for verifiable deepfake content would enhance the country's ability to combat digital misinformation without unduly burdening free speech.

India's Model Code of Conduct, administered during elections by the Election Commission of India (ECI), could also integrate rapid digital response norms to specifically counteract election-related synthetic content. Embedding deepfake identification and takedown protocols into the electoral regulatory framework would protect the integrity of democratic processes while ensuring that emergency interventions are limited, targeted, and judicially reviewable. This would allow India to strike a necessary balance between technological governance and the constitutional guarantee of freedom of expression.

E. Collaboration with Tech Companies: The South Korean Strategy

South Korea has adopted a forward-looking strategy by fostering close collaboration between government agencies and technology companies to counteract the risks associated with AI-generated synthetic media. Working together, they have developed **watermarking techniques and content authenticity protocols** that allow authorities and platforms to trace the origins of deepfake material, making it easier to verify and

⁷³ [Government of Singapore, 2019](#)

authenticate digital content⁷⁴. These cooperative efforts help ensure that technological safeguards evolve alongside emerging threats, rather than lagging behind them.

India can adopt a similar model by promoting **structured collaboration between the government, digital platforms, and AI developers** to create traceability mechanisms for synthetic media. Rather than relying solely on post-distribution takedowns, embedding watermarking and verification at the point of content creation would allow early detection and attribution, reducing the societal damage caused by malicious deepfakes. **Watermarking and hashing protocols can be enforced through technical standards notified by bodies like the Bureau of Indian Standards (BIS) or the Telecom Regulatory Authority of India (TRAI), in partnership with the Ministry of Electronics and Information Technology (MeitY).** This would standardize the technological infrastructure for AI content traceability, making compliance clear for platforms and facilitating faster regulatory responses.

Building such an industry-government ecosystem would not only strengthen India's ability to counteract synthetic media but also signal its leadership in ethical AI governance. Encouraging voluntary compliance initially, and moving toward mandatory standards as technology matures, would provide a balanced approach that supports innovation while protecting citizens' rights in the digital space.

By learning from these global strategies, India can craft a holistic approach to combat deepfakes. Strengthening legislation, leveraging AI for detection, enhancing public awareness, regulating media, and collaborating with tech companies will ensure a robust defense against the deepfake threat. Implementing these measures will help India balance technological progress with security and integrity in the digital space.

IX. Policy recommendations

A. Enact the Deepfake and Digital Manipulation Prevention Act (DDMPA) under the Information Technology Act, 2000⁷⁵

1. Introduce Dedicated Legislation: A standalone chapter under the IT Act should criminalize the creation, distribution, and amplification of malicious AI-generated content, such as deepfakes involving pornography, political manipulation, or impersonation. This will provide specific legal tools to prosecute deepfake crimes, addressing current loopholes and ensuring such acts are punishable under clearly defined laws.
2. Define “Synthetic Media” and “Deepfake”: The law must clearly differentiate between harmful and benign synthetic content by establishing intent-based categories (e.g., satire or education vs. harm). This ensures that only malicious content is penalized, while protecting freedom of expression in artistic and educational spaces.
3. Mandate Explicit Consent: Prior written or digital consent should be required before generating or distributing AI-based representations of individuals, especially if they use a person's biometric

⁷⁴ [Ministry of Science and ICT, South Korea, 2023](#)

⁷⁵ [Ministry of Electronics and Information Technology. \(2000\).](#)

features like voice or face. This empowers individuals to control their digital identities and protects them from unauthorized manipulation.

4. Introduce Graded Offences: Penalties must vary depending on the seriousness of the offence, stronger for crimes like electoral interference and lighter for minor, non-harmful use. This will ensure balanced, proportional justice while deterring high-impact misuse.
5. Victim-Centric Remedies: Victims must have legally backed rights to content takedown, financial compensation, and confidentiality during legal proceedings. This will make legal recourse more accessible and less intimidating, especially for survivors of non-consensual synthetic abuse.

B. Amend the IT Rules, 2021 under the Information Technology Act, 2000⁷⁶

1. Mandate 24-Hour Takedown: Intermediaries must be legally required to detect, label, and remove harmful deepfake content within 24 hours of notification. Quick takedown can significantly reduce the damage caused by viral circulation of malicious content.
2. Institutionalize Algorithmic Audits: Platforms should undergo annual third-party audits of their recommendation algorithms to check whether they amplify synthetic misinformation. These audits will ensure accountability in how content spreads and pressure platforms to design safer systems.
3. Transparency Reports and Appeal Mechanism: Platforms must publish regular content moderation reports and offer clear, accessible grievance redressal systems. This will promote openness and give users confidence that their concerns will be heard and addressed.
4. Centralized Grievance Portal: A single portal should be created where victims can file complaints, connect to local cybercrime units, and access legal aid. This integration will streamline redressal, reduce reporting friction, and speed up response mechanisms.

C. Introduce AI Risk Regulation under the Draft Digital India Act, 2025⁷⁷

1. Constitute NAIGA: The National AI Governance Authority (NAIGA) should be created under MeitY to regulate, monitor, and enforce ethical AI development and use across sectors. This will institutionalize AI oversight and bring centralized expertise to manage complex AI risks.
2. Implement Risk Categorization: AI systems must be classified into risk tiers, such as high-risk for political deepfakes, with corresponding regulatory obligations. This targeted approach will ensure strict scrutiny of harmful applications while allowing innovation in low-risk areas.
3. Watermarking and Disclosure Obligations: All synthetic content must carry visible or traceable indicators to show it is AI-generated. This will enhance public awareness, help detect misinformation, and maintain transparency in digital content.
4. Algorithmic Impact Assessments (AIA): Developers must conduct risk and safety assessments for AI tools before deploying them, especially in high-risk areas. This ensures that safety and ethics are prioritized during development, preventing harmful tools from reaching users.

⁷⁶ [Ministry of Electronics and Information Technology. \(2021\). IT Rules, 2021.](#)

⁷⁷ [Draft Digital India Act, 2025.](#)

D. Build AI Forensics Infrastructure under the Cybersecurity Framework of MeitY⁷⁸

1. Establish a National Centre for AI Forensics: A centralized forensic body should be created to detect, verify, and provide legal coordination for deepfake-related crimes. This will significantly boost India's capacity to investigate and act on synthetic media misuse with technical accuracy.
2. Equip Local Cybercrime Labs: Cybercrime labs at state and district levels should be equipped with certified detection tools and trained personnel. This will ensure rapid, localized responses to deepfake cases, especially in areas with limited resources.
3. Collaborate with Academia: Research institutions like IITs and IIITs should be involved in developing open-source, scalable detection technologies. This will encourage innovation, reduce reliance on foreign solutions, and make tools more accessible.
4. AI Threat Monitoring Hub: A national system should be established to track deepfake trends, issue real-time alerts, and coordinate early responses. This will enable proactive rather than reactive responses to emerging threats, helping contain mass misinformation.

E. Amend Section 67A of the Information Technology Act, 2000 and Relevant IPC Sections⁷⁹

1. Amend IT Act Section 67A: Expand the law to include AI-generated, non-consensual pornography and impersonation under obscene digital content. This will directly criminalize deepfake sexual content and provide a clear legal basis for prosecution.
2. Introduce "Synthetic Consent Violation": Create a legal offence for impersonation using AI-generated content, even if no physical contact is involved. This acknowledges the psychological and reputational harm caused by digital impersonation and fills a critical legal gap.
3. Fast-Track Legal Recourse: Special courts should be designated for speedy trial and resolution of deepfake-related cases. This will prevent long delays and provide timely justice to affected individuals.
4. Gender-Sensitive Enforcement: AI-crime desks should be created in cyber police stations, staffed with female officers and trained counselors. This will make reporting more accessible, especially for women and vulnerable groups facing digital sexual abuse.

F. Amend Section 79 of the Information Technology Act, 2000⁸⁰

1. Reform Safe Harbor: Platforms that fail to act on flagged deepfake content within legal timelines should lose their immunity under Section 79. This change will hold platforms accountable and push them to prioritize user safety.
2. Mandate Algorithmic Transparency: Platforms must disclose how their algorithms recommend or amplify synthetic content. This transparency will reveal potential biases or failures in content promotion systems and allow corrective oversight.

⁷⁸ [CERT-In. \(2023\). Annual Cybersecurity Report.](#)

⁷⁹ [Ministry of Home Affairs. \(2013\). Justice Verma Committee Report.](#)

⁸⁰ [European Commission. \(2022\). Digital Services Act.](#)

3. Regular Moderation Audits: Platforms should face regular, independent reviews of their content moderation strategies. This will create a continuous feedback loop to improve enforcement and reduce harmful content.
4. Penal Provisions: Repeat violations should attract heavy monetary fines or operational restrictions. Strict penalties will ensure platform compliance and prevent negligent practices from continuing unchecked.

G. Integrate AI Literacy into the National Education Policy (NEP)⁸¹

1. School Curriculum Reforms: AI and media literacy should be introduced into NCERT and CBSE curricula at the secondary school level. This will prepare students to critically engage with digital content and spot deepfakes early in life.
2. Professional Training: Journalists, educators, and civil servants should be trained to recognize and respond to synthetic misinformation. This will build frontline expertise to counteract the spread of fake content in public institutions and media.
3. Mass Campaigns: Awareness drives using Doordarshan, AIR, and multilingual digital media should be launched nationwide. This will reach broad audiences and build societal resistance to misinformation campaigns.
4. Decentralized Outreach: Panchayats and municipal bodies should receive digital literacy kits and materials. This ensures inclusion of rural and low-digital-access populations in the fight against AI-driven disinformation.

H. Expand India's Cyber and AI Diplomacy under the Ministry of External Affairs⁸²

1. Join Global AI Governance Frameworks: India should formally join platforms like GPAI and the Partnership on AI. This will align India with international best practices and give it a voice in shaping global AI norms.
2. Bilateral AI-Crime Treaties: India should sign treaties with countries like the US, UK, and EU to enable joint investigations and enforcement in cross-border deepfake crimes. Such agreements will strengthen India's global law enforcement capabilities and improve cooperation on digital threats.
3. Public-Private Innovation Platforms: Government should support collaborative efforts with the private sector to develop watermarking, verification, and ethical AI tools. These partnerships will drive innovation while embedding accountability and safety into AI ecosystems.

X. Conclusion

The deepfake crisis is no longer a distant or theoretical concern, it is a rapidly unfolding **reality with serious implications for democracy, governance, and individual rights**. AI-generated misinformation, deepfake

⁸¹ [UK Government. \(2021\). Don't Trust, Verify Campaign.](#)

⁸² [Department of Science & Technology. \(2023\).](#)

pornography, and digital impersonation are emerging not just as technological challenges, but as urgent legal and **ethical dilemmas**. In India's vast digital ecosystem, where over 850 million people are connected online, even a single manipulated video has the potential to distort public perception, incite unrest, or destroy reputations. Yet, our legal architecture, anchored in the Information Technology Act, 2000 and the Indian Penal Code, 1860, remains outdated and insufficient to address the evolving threats posed by synthetic media. While the **proposed Digital India Act, 2025** signals a progressive shift, its impact will remain limited without explicit AI-specific safeguards, enforcement clarity, and a robust deepfake response framework. The global momentum underscores this urgency, Europe's Digital Services Act mandates real-time content moderation, the U.S. Deepfakes Accountability Act introduces watermarking and consent provisions, and China obliges platforms to proactively screen AI-generated content. These models reflect that regulation is not only possible but essential in preserving information integrity and digital trust. India must therefore move beyond piecemeal reforms and adopt a comprehensive, multi-dimensional approach. Enacting the Deepfake and Digital Manipulation Prevention Act (DDMPA) under the IT Act will provide the foundational legal authority to define, criminalize, and penalize harmful synthetic content. Amendments to the IT Rules, 2021 must mandate faster takedowns, algorithmic audits, and greater transparency. The creation of a National AI Governance Authority (NAIGA) will institutionalize ethical oversight and ensure risk-based regulation across sectors. Simultaneously, building forensic capabilities, empowering cybercrime labs, and integrating AI literacy into the National Education Policy are essential steps to develop societal resilience against digital deception. Finally, **India must engage globally through AI-crime treaties and multilateral frameworks to strengthen its cross-border enforcement capacity. Legislation alone will not suffice. The battle against deepfakes demands a coordinated effort, combining legal reform, technological preparedness, institutional will, and public awareness.** In the end, the deepfake phenomenon is a test of how societies defend truth in an age of manipulation. **For India, the choice is clear: act decisively now, or risk losing the credibility of its institutions and the integrity of its digital future.**