THE FUTURE OF FREE SPEECH

THAT VIOLATES MY POLICIES

AI LAWS, CHATBOTS, AND THE FUTURE OF EXPRESSION

Directed by
Jordi Calvet-Bademunt,
Jacob Mchangama, and Isabelle Anzabi

OCTOBER 2025

Acknowledgments

The Future of Free Speech is an independent, nonpartisan think tank based at Vanderbilt University. Our mission is to reaffirm freedom of expression as the foundation of free and thriving societies through actionable research, practical tools, and principled advocacy. We envision a world in which the right to freedom of expression is safeguarded by law and strengthened by a culture that embraces diverse viewpoints.

This project was led by Jordi Calvet-Bademunt (Senior Research Fellow), Jacob Mchangama (Executive Director), and Isabelle Anzabi (Research Associate) at The Future of Free Speech. Together, they also drafted the chapters on the European Union and the United States of America.

We are grateful to Justin Hayes, Director of Communications, for overseeing the design of the report; Wendy H. Burch, Chief Operating Officer, for coordinating all administrative aspects of the project; and Sam Cosby, Director of Development, for leading the funding efforts that made this work possible.

We extend our thanks to the leading experts who contributed chapters on their respective jurisdictions: Carlos Affonso Souza (Brazil), Ge Chen (China), Sangeeta Mahapatra (India), and Kyung Sin (K.S.) Park (Republic of Korea). We are also grateful to Kevin T. Greene and Jacob N. Shapiro of Princeton University for their chapter, "Measuring Free Expression in Generative AI Tools."

We thank all the experts who contributed to individual chapters of this report; their names are listed in the relevant sections.

We are further indebted to Barbie Halaby of Monocle Editing for her careful editorial work across all chapters, and to Design Pickle for the report's design.

Finally, we are especially grateful to the Rising Tide Foundation and the Swedish Postcode Lottery Foundation for their generous support of this work, and we thank Vanderbilt University for their collaboration with and support of The Future of Free Speech.

Publication Date: October 2025

Preface

In this report, we explore the ways in which public and private governance of generative artificial intelligence (AI) shape the space for free expression and access to information in the 21st century.

Since the launch of ChatGPT by OpenAI in November 2022, generative AI has captured the public imagination. In less than three years, hundreds of millions of people have adopted OpenAI's chatbot and similar tools for learning, entertainment, and work.[1] Anthropic, another AI giant, now serves more than 300,000 business customers.[2] AI companies are valued in the hundreds of billions of US dollars[3], while established technology giants such as Google, Meta, and Microsoft are investing billions in the race to dominate the field.[4]

Generative AI refers to systems that create content — including text, images, video, audio, and software code — in response to user prompts.[5] Chatbots such as ChatGPT are the most visible examples, but generative AI is rapidly being embedded into the tools people use every day for both communication and access to information, from social media and email to word processors and search engines.

Recognizing generative AI's potential for expression and access to information, The Future of Free Speech undertook a first-of-its-kind analysis of freedom of expression in major models. In February 2024, we assessed the "free-speech culture" of six leading systems, focusing on their usage policies and responses to prompts.[6] Our findings revealed that excessively broad and vague rules often resulted in undue restrictions on speech and access to information.[7] By April 2025, when we updated this work, we observed signs of change: Some models showed greater openness.[8]

This current report builds on those foundations and pursues a more ambitious goal. Supported by leading experts, The Future of Free Speech undertakes a deeper examination of how national legislation and corporate practices shape freedom of expression in the era of generative AI. "*That Violates My Policies": AI Laws, Chatbots, and the Future of Expression* explores:

- AI legislation in Brazil, China, the European Union, India, the Republic of Korea, and the United States.[9] In this report, AI legislation refers to laws and public policies addressing AI-generated content, with particular focus on elections and political speech, hate speech, defamation, explicit content (including

1 MacKenzie Sigalos, "OpenAI's ChatGPT to Hit 700 Million Weekly Users, Up 4x from Last Year," *CNBC*, August 4, 2025, https://www.cnbc.com/2025/08/04/openai-chatgpt-700-million-users.html.
2 Hayden Field, "Anthropic Is Now Valued at $183 Billion," *The Verge*, September 2, 2025, https://www.theverge.com/anthropic/769179/anthropic-is-now-valued-at-183-billion.
3 Kylie Robison, "OpenAI Is Poised to Become the Most Valuable Startup Ever: Should It Be?," *Wired*, August 19, 2025, https://www.wired.com/story/openai-valuation-500-billion-skepticism/; Krystal Hu and Shivani Tanna, "OpenAI Eyes $500 Billion Valuation in Potential Employee Share Sale, Source Says," *Reuters*, August 6, 2025, https://www.reuters.com/business/openai-eyes-500-billion-valuation-potential-employee-share-sale-source-says-2025-08-06/.
4 Blake Montgomery, "Big Tech Has Spent $155bn on AI This Year: It's About to Spend Hundreds of Billions More," *The Guardian*, August 2, 2025, https://www.theguardian.com/technology/2025/aug/02/big-tech-ai-spending.
5 Cole Stryker and Mark Scapicchio, "What Is Generative AI?," IBM Think, March 22, 2024, https://www.ibm.com/think/topics/generative-ai.
6 Jordi Calvet-Bademunt and Jacob Mchangama, *Freedom of Expression in Generative AI: A Snapshot of Content Policies* (Future of Free Speech, February 2024), https://futurefreespeech.org/wp-content/uploads/2023/12/FFS_AI-Policies_Formatting.pdf.
7 Calvet-Bademunt and Mchangama, *Freedom of Expression in Generative AI*.
8 Jordi Calvet-Bademunt, Jacob Mchangama, and Isabelle Anzabi, "One Year Later: AI Chatbots Show Progress on Free Speech — But Some Concerns Remain," *The Bedrock Principle*, April 1, 2025, https://www.bedrockprinciple.com/p/one-year-later-ai-chatbots-show-progress.
9 To select the countries, we considered Stanford University's 2023 Global AI Vibrancy Ranking (the most recent available at the time of writing), along with factors such as geographic diversity, population size, democratic and freedom status, and the presence of existing or emerging AI-related legislation.

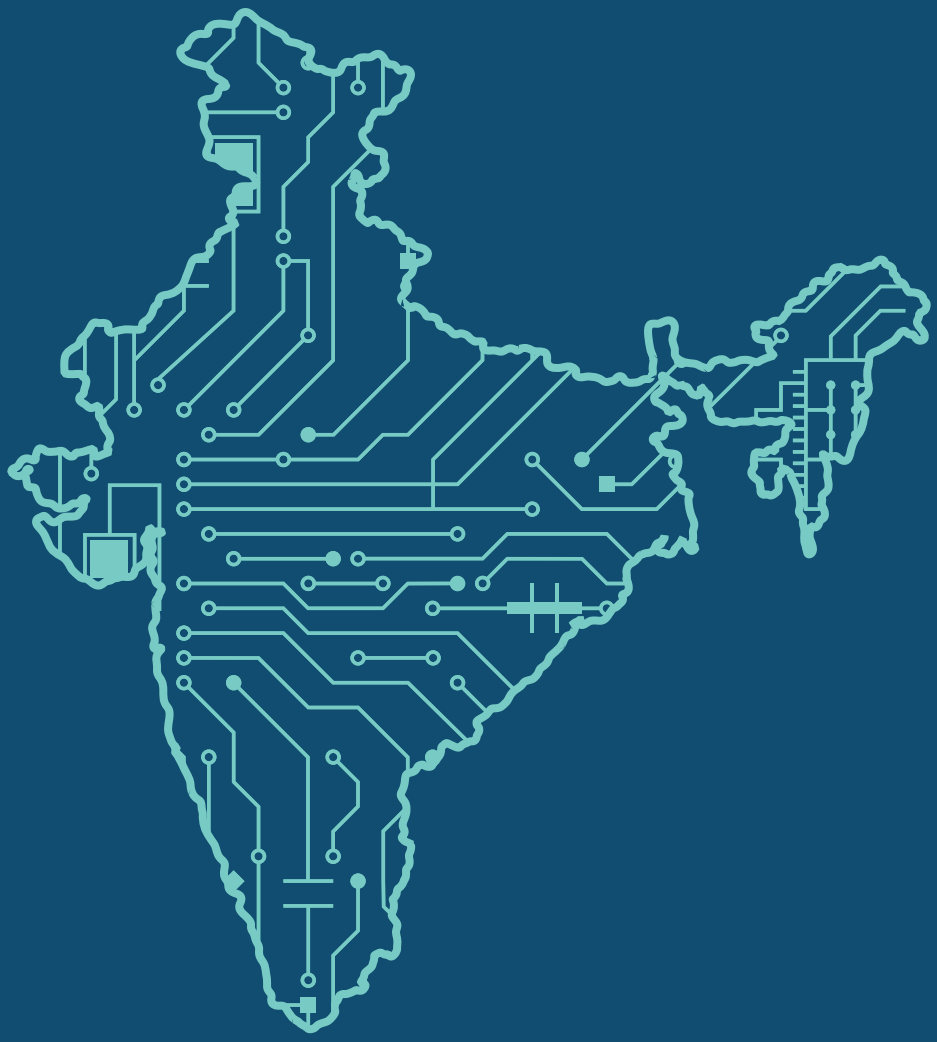
child sexual abuse material and nonconsensual intimate images), and copyright. We also consider measures that actively promote freedom of expression, such as AI literacy initiatives and policies supporting cultural and linguistic diversity.

- Corporate practices of major AI developers, including Alibaba, Anthropic, Google, Meta, Mistral AI, DeepSeek, OpenAI, and xAI.[10] We examine their usage policies, model performance in responding to prompts, and the limited available information on their training data and development processes.

This report seeks to provide a rigorous and timely analysis of how generative AI is reshaping the space for free expression in both the public and private spheres. Building on these insights, The Future of Free Speech is developing guidelines to help policymakers and companies ensure that generative AI protects and enhances freedom of expression and access to information, two cornerstones of democratic societies.

In an era of rapid technological change, safeguarding free expression is a matter not only of rights but of preserving the conditions for open, informed, and thriving democracies.

10 We selected major models from leading companies that are accessible through a web interface and include text-generation capabilities. In addition, we considered the geographic location of the model provider and the degree of openness of the models.



Artificial Intelligence and Freedom of Expression in India

Sangeeta Mahapatra*

*Dr. Sangeeta Mahapatra is a research fellow at the German Institute for Global and Area Studies (GIGA), Hamburg, working on artificial intelligence and internet governance, digital authoritarianism, countering disinformation, and building cyber resilience. She focuses on South and Southeast Asia, collaborates with local civic partners, and has led research projects that have academic, policy, and social impacts.

Abstract

This chapter examines how India's approach to regulating artificial intelligence (AI), including generative AI, affects freedom of speech and expression, offering comparative insights for other democracies, especially in multilingual and low-resource contexts. India provides a pivotal example, balancing its aspirations as a global AI leader with the realities of a democracy marked by deep social inequalities, strong state control, and extensive surveillance capacity. AI is positioned as both a driver of progress and a means to reduce inequality through initiatives such as "AI for All," multilingual platforms like Bhashini, and open-source model development under BharatGPT and AI4Bharat. With no dedicated AI law in India, generative AI provisions are incorporated into existing legal frameworks and government advisories, most notably the IT Act, IT Rules, and the Digital Personal Data Protection Act, which combine inclusive measures with centralized control over datasets and computing resources.

I first examine India's AI regulations in this context, then analyze the convergences between these measures and global AI governance guidelines and norms, and finally consider how these regulations apply to six critical areas: defamation, explicit content, hate speech, political content, copyright, and privacy. While the current framework advances access and participation, it can encourage over-removal of lawful speech, selective enforcement of harmful content, and fragmented protections. Recommendations include statutory guarantees of transparency in AI moderation, disclosure of takedown criteria, due process and user contestation rights, independent oversight for high-risk AI uses, and governance of datasets and foundational models through open, representative, and audited sources managed by multiple stakeholders. Strengthening privacy safeguards by narrowing government exemptions would align practice with constitutional principles and protections. India's case illustrates the challenges and opportunities of linking national priorities with a human rights baseline for transnational AI regulation while preserving dignity, participation, and autonomy.

Sangeeta Mahapatra

Dr. Sangeeta Mahapatra is a research fellow at the German Institute for Global and Area Studies (GIGA), Hamburg, working on artificial intelligence and internet governance, digital authoritarianism, countering disinformation, and building cyber resilience. She focuses on South and Southeast Asia, collaborates with local civic partners, and has led research projects that have academic, policy, and social impacts.

1. Introduction

Generative artificial intelligence (AI)[1] tests the limits of India's legal capacity to protect expressive freedoms, not just in speech but in the creation, circulation, and control of content. In the absence of dedicated legislation, regulatory responses depend on repurposing older legal frameworks, particularly those governing speech and digital media. Expressive rights[2] have, since their inclusion in the Constitution, been carefully crafted to balance liberal democratic ideals with the complex realities of a highly populous and diverse developing country. Freedom of speech and expression was seen as essential for individual empowerment and democratic participation. This approach embodies transformative constitutionalism, which views rights not merely as defenses against state power but as instruments for advancing social justice and collective progress. However, while drawing on Western liberal principles, India's legal framework on expressive rights was also tempered by its colonial legacy, with the framers making these rights not absolute but subject to reasonable restrictions to protect sovereignty, public order, and morality. The Supreme Court of India, often invoking the preservation of democratic integrity, extended the interpretation and implementation of rights and restrictions through doctrines like proportionality and imminent harm, balancing free speech with state interests.[3] It is in this context that India's AI policies emerge and operate within the domain of expressive rights. These policies not only carry forward the norms, aspirations, and restrictions embedded in India's traditional legal framework but also reflect the power structures that influence their interpretation and enforcement.

As India transitions to a digital-first country, integrating digital technologies into all facets of daily life,[4] AI becomes a crucial element in discussions surrounding expressive rights. The National Strategy for Artificial Intelligence (2018), with its "AI for All" slogan, positions AI as more than a general-purpose productivity tool; it is seen as a potential equalizer in a country where millions still lack meaningful digital access. Consequently, ensuring equitable access to accurate and accountable AI systems becomes as vital as any constitutional guarantee. Access to AI, in this sense, holds both instrumental and intrinsic value within the discourse of expressive rights: It can either facilitate or impede these rights or, in a truly digitized society impacting all modes of expression and participation in democratic life, evolve into a right in itself. The implications of this will be particularly pronounced in India's pluralistic society, where AI systems can intersect with long-standing inequalities and state power. As AI technologies shape the production, moderation, and consumption of content, India will need to develop a future-ready legal architecture to ensure that expressive rights remain protected amid technological change.

1 Generative artificial intelligence is defined by the Government of India as the capability of AI-enabled systems to use existing text, audio files, or images to generate new content. See Ministry of Electronics and Information Technology, "Generative Artificial Intelligence," Press Information Bureau, February 3, 2023, https://www.pib.gov.in/PressReleasePage.aspx?PRID=1896016.
2 Expressive rights broadly encompass the rights of individuals to express their thoughts, opinions, and beliefs without undue state interference or restriction, forming an essential component of civil liberties and political freedoms. They also include inferred rights, such as access to information and privacy.
3 Lawrence Liang, "Free Speech and Expression," in *The Oxford Handbook of the Indian Constitution*, ed. Sujit Choudhry, Madhav Khosla, and Pratap Bhanu Mehta (Oxford University Press, 2016), https://doi.org/10.1093/law/9780198704898.003.0045.
4 KPMG, "India's Digital Dividend: The Strategic Roadmap Towards Becoming a Global Digital Leader," January 2025, https://assets.kpmg.com/content/dam/kpmgsites/in/pdf/2025/01/indias-digital-dividend-the-strategic-roadmap-towards-becoming-a-global-digital-leader.pdf.

However, India currently lacks a dedicated AI law. Instead, it relies on a patchwork of policies, guidelines, and advisories — an approach best described as "regulation-on-the-go."[5] This reactive stance means India drafts its policies based on AI's performance and evolving use rather than proactively establishing a comprehensive legal framework before widespread adoption.[6] AI regulations in India often follow a sectoral approach. This implies that AI development and use cases are viewed from each sectoral perspective, leading to a fragmented regulatory approach.[7] Similarly, for expressive rights in the AI context, India primarily derives its regulatory understanding from existing, traditional legal frameworks on freedom of speech and expression. Consequently, the norms and restrictions governing AI-generated content or AI's influence on expression are largely interpreted through the lens of older legal frameworks that are often ill equipped to address AI's unique needs and challenges.[8]

The intertwined role of AI in expressive rights is further complicated by India's dual governmental approach: a light-touch regulatory stance to encourage rapid AI deployment for economic growth contrasted with a heavy-handed application of existing speech laws. The latter approach leverages legislation such as the Indian Penal Code (IPC) of 1860 (now the Bharatiya Nyaya Sanhita, 2023), the Information Technology Act of 2000, and subordinate regulations like the Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules of 2021, known as IT Rules 2021.[9]

While AI presents significant opportunities to democratize expression and enhance accessibility — for instance, through AI-powered translation tools, content creation assistance, and platforms that empower marginalized voices to reach broader audiences — its governance in India faces a unique challenge. This challenge lies in simultaneously expanding and constraining expressive freedom, potentially solidifying preexisting controls within the country's democratic framework. Further complicating matters, the 2025 AI Governance Guidelines subcommittee's report, despite championing harm minimization and regulatory capacity, emphasizes voluntary commitments, which risks diluting accountability and transparency for AI deployers.[10] Concurrently, broad exemptions granted to the government within the IT Rules 2021 and the Digital Personal Data Protection Act of 2023 (DPDA 2023) further expand executive power in regulating expressive rights.

In this complex environment, where old laws and new technologies coexist uneasily, I explore how India's evolving AI policies and regulatory choices may influence expressive rights, first outlining the constitutional and legal standards on expressive rights alongside emerging AI policies, their alignment with international standards, and applicability to AI-generated content. I then analyze specific provisions relevant to defamation, explicit content, hate speech, election and political content, copyright, and empowering speech and conclude with findings on AI policies that protect and strengthen expressive rights.

5 Sangeeta Mahapatra, "Ethical AI Governance to Prevent Digital Authoritarianism: Insights from South and Southeast Asia, with a Focus on India and Singapore," *DigiTraL Policy Papers,* GIGA, April 2025, https://www.giga-hamburg.de/en/publications/contributions/ethical-governance-prevention-digital-authoritarianism-south-southeast-asia-studies-india-singapore.
6 Amlan Mohanty and Shatakratu Sahu, "India's Advance on AI Regulation," Carnegie Endowment for International Peace, November 21, 2024, https://carnegieendowment.org/research/2024/11/indias-advance-on-ai-regulation?lang=en; Sriya Sridhar, "India's AI Governance Guidelines Report: A Medley of Approaches," *Tech Policy Press,* January 16, 2025, https://www.techpolicy.press/indias-ai-governance-guidelines-report-a-medley-of-approaches/.
7 Mahapatra, "Ethical AI Governance to Prevent Digital Authoritarianism."
8 Mohanty and Sahu, "India's Advance on AI Regulation."
9 Sangeeta Mahapatra, Janjira Sombatpoonsiri, and Andreas Ufen, "Repression by Legal Means: Governments' Anti-Fake News Lawfare," *GIGA Focus Global,* Number 1 (2024), https://doi.org/10.57671/gfgl-24012.
10 Sridhar, "India's AI Governance Guidelines Report."

2. Substantive Analyses

2.1. General Standards of Freedom of Expression

Expressive rights in India comprise the constitutional, statutory, and jurisprudential protections that enable individuals to express opinions, access information, dissent, and engage in public discourse. To offer analytical clarity for examining the intersection of expressive rights and AI governance, this section posits a three-tiered classification: autonomy-based rights, participation-based rights, and dignity-based rights,[11] which are mutually reinforcing:

1) Autonomy-based rights protect individual agency and self-determination — the capacity to form, hold, and express personal beliefs essential for judgment and participation in a democratic society. These rights include the freedom of speech and expression,[12] the right to remain silent,[13] and the right to receive information.[14] This last right was subsequently codified as a statutory right under the Right to Information Act of 2005.[15]

2) Participation-based rights facilitate democratic engagement by enabling collective action and discourse. These include the freedom of the press[16] and the right to peaceful assembly and association.[17] In the online domain, it would include calls to action and individual/collective speech and activism.

3) Dignity-based rights emphasize the protection of personal identity, bodily integrity, and informational privacy, including the right to privacy.[18]

This tripartite framework, grounded in constitutional jurisprudence, provides an integrated view of expressive rights, underscoring their interconnectedness. As such, AI governance of expressive rights must incorporate each of these dimensions when addressing freedom of speech and expression.

This framework also integrates crucial safeguards that collectively define the relationship between expressive rights and competing interests. For instance, the right to privacy extends beyond bodily and informational integrity to establish boundaries on state surveillance, algorithmic profiling, and data practices in AI governance.[19] The protection against defamation, although not a fundamental right, is recognized as a constitutionally valid restriction on freedom of speech and expression, falling under the guideline ensuring that the right to reputation is balanced with expressive liberty, especially in digital contexts. Copyright protection,

11 This classification is the author's own derivation, based on an analysis of the articles in the Constitution of India 1950, court rulings, and legislative acts mentioned in the text that are related to expressive rights
12 Indian Const. art. 19(1)(a).
13 Bijoe Emmanuel and Others v. State of Kerala and Others (1986), https://indiankanoon.org/doc/1508089.
14 Secretary, Ministry of Information and Broadcasting v. Cricket Association of Bengal & ANR. (1995), https://indiankanoon.org/doc/539407.
15 People's Union for Civil Liberties v. Union of India (2023), https://indiankanoon.org/doc/15059075.
16 Indian Express Newspapers v. Union of India (1984), https://indiankanoon.org/doc/223504/.
17 Indian Const. art. 19(1)(b)-(c).
18 Right to privacy is recognized as a fundamental right under Article 21 in *Justice K.S. Puttaswamy v. Union of India* (2017).
19 Sangeeta Mahapatra, "Digital Surveillance and the Threat to Civil Liberties in India," GIGA Focus Asia, Number 3 (2021), https://www.giga-hamburg.de/en/publications/giga-focus/digital-surveillance-and-the-threat-to-civil-liberties-in-india.

codified under the Copyright Act 1957, reinforces the value of individual authorship while ensuring democratic knowledge dissemination, reflecting a balance between creator rights and public access. The rights to information and press freedom, derived from the fundamental right to freedom of speech and expression,[20] further illustrate how expressive rights are inherently linked to the public's right to know and participate in democratic discourse. Together, these expressive rights, whether fundamental or derivative, need to remain responsive to the evolving challenges of AI, including algorithmic regulation, content moderation, and data governance.

However, expressive rights in India are not absolute. The "reasonable restrictions" permitted under Articles 19(2) to 19(6) of the Constitution allow the government to curtail expressive freedoms on grounds such as public order, decency, and sovereignty. Indian courts have increasingly subjected these restrictions to a proportionality test, requiring that any limitation be lawful, necessary, and minimally impairing.[21]

In practice, statutory and executive measures often create ambiguities, especially in the context of online speech and expression (including generative AI content) governed by information technology laws. Section 69A of the Information Technology Act of 2000 (IT Act 2000) empowers the executive to block online content in the interests of sovereignty, public order, or decency; this power was upheld by the Supreme Court in *Shreya Singhal v. Union of India* (2015), subject to procedural safeguards. Yet the IT Rules 2021 have expanded this regulatory authority to digital platforms in ways that incentivize censorship.[22] Rule 3(1)(b) of the IT Rules 2021 requires intermediaries to make reasonable efforts to ensure users do not host or share content that is defamatory or obscene or threatens public order, among other categories. Rule 3(1)(d) obligates intermediaries to remove such content, generally 36 hours, upon receiving a court order or a government agency notification, creating pressure on platforms to err on the side of removal to avoid liability. As a result, platforms often resort to proactive monitoring, using AI-driven automated systems to flag or take down content preemptively, even before formal complaints arise. This enables an environment where AI-based moderation systems, trained on datasets that can encode biases, determine what speech is permissible. This dual dynamic of incentivized over-removal and selective enforcement results in both excessive censorship of legitimate speech and inadequate moderation of harmful content, reinforcing the imbalance in freedom of expression protections. The opacity of these systems means that users have limited avenues to challenge such a determination or understand the basis for content removal.

This is problematic, as legal scholars have argued that India's free speech regime frequently prioritizes state interests over individual liberties — particularly in matters of political dissent, hate speech, and media regulation.[23] Others have drawn attention to the uneven and discretionary enforcement of expressive rights, highlighting deep regional and social disparities.[24]

Automated systems — especially in content moderation, surveillance, copyright enforcement, and speech recognition — are altering how expression is regulated and experienced. Generative AI challenges the existing

20 Subramanian Swamy v. Union of India (2016), https://indiankanoon.org/doc/80997184/. The Supreme Court upheld the constitutionality of criminal defamation (§§ 499–500, Indian Penal Code), citing the right to reputation under Article 21 as a reasonable restriction on free speech under Article 19(1)(a) of the Constitution of India.
21 For instance, in *Modern Dental College and Research Centre v. State of Madhya Pradesh* (2016), the Supreme Court articulated the proportionality test as a standard for assessing restrictions on fundamental rights. See https://indiankanoon.org/doc/70187318/.
22 Janjira Sombatpoonsiri and Sangeeta Mahapatra, "Regulation or Repression? Government Influence on Political Content Moderation in India and Thailand," Digital Democracy Network, Carnegie Endowment for International Peace, July 31, 2024, https://carnegieendowment.org/research/2024/07/india-thailand-social-media-moderation?lang=en.
23 Gautam Bhatia, *Offend, Shock, or Disturb: Free Speech under the Indian Constitution* (Oxford University Press, 2016). The argument is discussed throughout the book.
24 Aparna Chandra and Gladson J. Haokip, "Hate Speech Laws in India: A Complex Legal Terrain," in *Law and Politics of Religious Offense in India,* ed. Niraja Gopal Jayal (Oxford University Press, 2022), 119–142.

rules on authorship, intellectual property, misinformation, and speech control. The next section turns to India's emerging AI policy landscape and its broader implications for expressive rights.

2.2. AI-Specific Legislation and Policies

The Government of India's framing of AI policies is noteworthy: It projects AI as a "public good" and not just as private innovation; it is integral to both democratic participation and development across linguistic and socioeconomic divides. This framing has profound implications for AI's relationship with expressive rights, expanding the scope of AI beyond technical implementation to fundamental democratic values. Citizens' equitable right to access government developmental services is inextricably linked to their right to freely express demands for these services, provide feedback, and seek information about them. The principles of responsible and ethical AI thus govern the terrain of these intertwined rights, ensuring that technological progress does not undermine democratic freedoms. Significantly, in the foreword to NITI Aayog's Responsible AI for All (2021) document,[25] then Vice Chairman Rajiv Kumar explicitly connected India's AI principles to fundamental constitutional rights.

Alongside Responsible AI for All (2021), India's official AI policy documents also include the National Strategy for AI (2018), the IndiaAI program (2024), the DPDP Act (2023), the draft AI Governance Guidelines (2025) by the Ministry of Electronics and Information Technology (MeitY), and the proposed Digital India Act (DIA).[26] NITI Aayog's policies emphasize democratizing information access and enhancing both collective and individual speech rights in public spaces and local languages; in contrast, the ministry-level acts and guidelines risk granting excessive power to the state and private platforms, potentially constraining citizens' expressive freedoms.

2.2.1. India's AI Governance Framework

The National Strategy for AI (2018) emphasizes inclusive growth, implicitly supporting the right to freedom of speech and expression by aiming to empower citizens to participate more effectively in the public sphere, especially through AI-enabled services in local languages. Building on this foundation, the 2021 Responsible AI for All set out key ethical principles — safety, equality, inclusivity, privacy, transparency, and accountability — further developed in MeitY's 2025 draft AI Governance Guidelines, which stress transparency, fairness, accountability, and security. These guidelines call for user awareness in AI interactions and outcomes that uphold the rule of law, promote autonomy and informed choice, and position AI as a tool for democratic empowerment and the protection of individual liberties.

Two core tenets of India's AI vision that make the right to expression meaningful are multilingual access and AI autonomy. Integrating AI into digital public infrastructure (DPI), in everything from language platforms (e.g., Bhashini) to e-governance portals, seeks to democratize knowledge and enhance communication.[27] The IndiaAI program reinforces this by investing in AI research and computing infrastructure, emphasizing

25 NITI Aayog is the Government of India's apex public policy think tank, whose policies and strategies largely guide the governance of AI in India in the absence of an AI Act.
26 NITI Aayog, "National Strategy for Artificial Intelligence," 2018, https://www.niti.gov.in/sites/default/files/2023-03/National-Strategy-for-Artificial-Intelligence.pdf; NITI Aayog, "Approach Document for India: Part 1 — Principles for Responsible AI," February 2021, https://www.niti.gov.in/sites/default/files/2021-02/Responsible-AI-22022021.pdf; NITI Aayog, "Approach Document for India: Part 2 — Operationalizing Principles for Responsible AI," August 2021, https://www.niti.gov.in/sites/default/files/2021-08/Part2-Responsible-AI-12082021.pdf; MeitY, "Cabinet Approves Ambitious IndiaAI Mission to Strengthen the AI Innovation Ecosystem," March 7, 2024, https://www.pib.gov.in/PressReleaseIframePage.aspx?PRID=2012357; MeitY, "Report on AI Governance Guidelines Development," January 6, 2025, https://indiaai.s3.ap-south-1.amazonaws.com/docs/subcommittee-report-dec26.pdf.
27 Bhashini, India's AI-led language translation platform, was officially launched in July 2022 under the National Language Translation Mission by MeitY. It aims for digital inclusion by providing AI and natural language processing (NLP) tools for translation and digital services across Indian languages.

"Safe and Trusted AI" for the public. Sector-specific policies are evolving to address AI biases and malpractice risks, supporting safety and trust.

However, the DPDP Act has faced criticism for granting the government broad exemptions and control over data, potentially undermining individual privacy rights. Critics argue that the act's reliance on notice and consent mechanisms as primary safeguards is insufficient, particularly in a context where many individuals lack digital literacy or access to comprehensive information about data usage.[28] This reliance may exacerbate existing power asymmetries between the state, private platforms, and citizens, leaving individuals vulnerable to data exploitation without meaningful recourse. Such risks highlight the limitations of voluntary commitments as compared to binding regulations. Similarly, the 2025 AI Governance Guidelines, while advocating for ethical AI practices, have been critiqued for their reliance on voluntary compliance and the lack of enforceable mechanisms, raising concerns about their effectiveness at protecting citizens' rights.[29]

The government has also leaned on advisories, which project protection of citizens' rights while giving the state latitude to keep obligations at the level of soft compliance or selectively enforce them. This dynamic is evident in the government's advisories, such as those on algorithmic discrimination and deepfakes,[30] which project rights protection but leave enforcement contingent on state discretion.

While the government's dual approach, promoting open-source AI development alongside centralized control over data and computing resources, may appear contradictory at first glance, it actually reflects a deliberate policy trade-off between technological innovation and regulatory sovereignty. The MeitY has issued advisories reminding intermediaries to comply with existing IT Rules;[31] instructing platforms to prevent AI models from enabling unlawful content, bias, or discrimination; and mandating the labeling of AI-generated content.[32] Additionally, the Indian Computer Emergency Response Team (CERT-In) published advisories on minimizing AI-based risks and on deepfake threats, providing measures for protection.[33]

This regulatory approach acknowledges that existing laws may not fully address the unique risks posed by generative AI and its potential misuse of personal data, particularly impacting rights to privacy and protection against deepfakes. While industry favors self-regulation, there is also a recognition of the need for additional regulations for high-risk AI use cases. This highlights the need for clear ethical standards and enforceable legal rights to safeguard expressive freedoms from AI-mediated harms, such as disinformation and manipulation. Civil society organizations further emphasize the importance of representing marginalized groups in AI regulation discussions, given these groups' heightened vulnerability to negative impacts related to privacy and discrimination.[34]

The forthcoming Digital India Act, expected to replace the Information Technology Act of 2000, aims to establish a comprehensive legal framework for the modern digital ecosystem, including provisions for AI governance. While the DIA is anticipated to address high-risk AI systems and algorithmic accountability

28 Sriya Sridhar, "Data Protection Rules and Act, a Net Negative for Privacy Rights," *The Hindu,* February 13, 2025, https://www.thehindu.com/opinion/op-ed/data-protection-rules-and-act-a-net-negative-for-privacy-rights/article69212801.ece.
29 Sridhar, "India's AI Governance Guidelines Report."
30 MeitY, "Government of India Taking Measures to Tackle Deepfakes," Press Information Bureau, April 4, 2025, https://www.pib.gov.in/PressReleasePage.aspx?PRID=2119050.
31 MeitY, "MeitY Issues Advisory to All Intermediaries to Comply with Existing IT Rules," Press Information Bureau, December 26, 2023, https://www.pib.gov.in/PressReleaseframePage.aspx?PRID=1990542.
32 MeitY, *Due Diligence by Intermediaries/Platforms under the Information Technology Act, 2000 and Information Technology (Intermediary Guidelines and Digital Media Ethics Code) Rules,* 2021, March 15, 2024, https://www.meity.gov.in/static/uploads/2024/02/9f6e99572739a3024c9cdaec53a0a0ef.pdf; MeitY, "Government of India Taking Measures to Tackle Deepfakes."
33 Computer Emergency Response Team (CERT-In), Security Implications of AI Language-Based Applications, Government of India, May 9, 2025, https://www.cert-in.org.in/s2cMainServlet?pageid=PUBVLNOTES02&VLCODE=CIAD-2023-0015.
34 Mohanty and Sahu, "India's Advance on AI Regulation."

through measures such as algorithmic transparency and periodic risk assessments, specific details are still under deliberation. As of this writing, the Government of India has not released a draft of the DIA for public consultation, although it has claimed that multiple rounds of pre-draft consultations were conducted with stakeholders.

In contrast to the EU's General Data Protection Regulation (GDPR 2018) and AI Act (2024) — which institutionalize individual data rights, mandate independent regulatory authorities, and impose tiered obligations based on systemic risk — India's approach to AI governance concentrates decision-making power within the executive, with limited external accountability or statutory safeguards to assess how foundational models may shape public discourse, reinforce structural biases, or restrict expressive freedoms. This regulatory asymmetry raises concerns about transparency and recourse, especially given that generative AI is increasingly deployed in sensitive contexts like elections, content moderation, and linguistic representation. However, even the push for openness comes with strong government control. For example, the government prioritizes using Indian datasets to build foundational models.[35] Open-source initiatives such as BharatGPT, BharatGen, Sarvam-M, and AI4Bharat reflect ambitions to democratize AI tools, reduce reliance on foreign providers, and lower barriers to access via public platforms and multilingual design. Yet these models are being developed through state-led compute, proposal, and oversight mechanisms under the IndiaAI Mission, signaling deliberate and managed expansion of the country's AI infrastructure. This is meant to ensure that India does not get locked into relying on foreign companies. Initiatives like Bhashini, which supports many languages using open-source tools, ask citizens to donate their language data through "Bhasha Daan." But all this data is stored and controlled by the government on a central platform. Another example, the IndiaAI Kosh platform, as part of the bigger IndiaAI Mission, aims to make high-quality, India-specific data and tools available for local AI development.[36] While this looks like it supports open access, the government still controls key resources like central data stores and subsidized GPU access. This means the government has the power to guide how AI is developed and used, deciding which datasets and tools are prioritized.

Thus, a contradiction: While the government talks about openness and innovation, it also keeps tight control over the most important parts of AI development. This control may affect privacy, information sharing, and the direction of AI in India, raising questions about whether the technology is truly open or government led.

India's approach to AI regulation, relying on existing frameworks and non-binding advisories, leaves significant gaps, especially concerning AI's impact on speech rights. The IT Act 2000 lacks provisions for algorithmic decision-making or AI-generated content. Meanwhile, MeitY's advisories, though timely on issues like deepfakes, serve as guidance rather than enforceable rules, making compliance voluntary and leaving room for misuse. The government's reluctance to introduce an AI act until the implications of AI are fully understood creates uncertainty for developers and investors. This fragmented regulatory landscape often forces courts to stretch outdated laws to cover AI, inviting inconsistencies and potential legal challenges. Ultimately, the lack of enforceable rules and transparency behind AI decision-making risks undermining public trust and individual autonomy in a meaningful sense.

India's constitutional framework recognizes online content (including memes, videos, and satire) as protected forms of speech under Article 19(1)(a) of the Constitution. This encompasses various mediums such as speech,

35 Debarshi Dasgupta, "India Joins Global Race to Develop AI Models," *Straits Times*, February 1, 2025, https://www.straitstimes.com/asia/south-asia/india-joins-global-race-to-develop-ai-models.
36 IndiaAI, "Now Open: Expression of Interest (EOI) to Contribute Datasets to AIKosh," March 25, 2025, https://indiaai.gov.in/article/now-open-expression-of-interest-eoi-to-contribute-datasets-to-aikosh.

writing, printing, visual representations, and digital communication. The Supreme Court, in the landmark case of Shreya Singhal v. Union of India (2015), affirmed that online speech is entitled to the same constitutional safeguards as offline expression. In its decision, the court struck down Section 66A of the 2000 IT Act, which had criminalized sending "offensive" messages online, citing its vagueness and potential for misuse. This judgment underscored the importance of protecting digital expression, including satirical and critical content, from arbitrary censorship. However, the rise of AI-generated content introduces new challenges. For example, AI algorithms, often operating as "black boxes," can inadvertently censor legitimate speech or propagate biased content, thereby impacting the constitutional rights of individuals. The IT Rules 2021 address some of these concerns by mandating that significant social media intermediaries implement appropriate human oversight when deploying automated tools for content moderation to prevent infringement of users' rights to free expression.

Despite these measures, ambiguity persists regarding the classification of AI developers and deployers within the existing intermediary liability framework. The IT Rules primarily impose due diligence obligations on intermediaries, but it remains unclear whether AI developers and deployers fall into this category. Traditionally, Section 79 of the IT Act of 2000 offers "safe harbor" protection to intermediaries who do not initiate the transmission, select the receiver, or modify the information contained in the transmission, thereby implying a largely passive role. In contrast, AI models, especially generative AI, actively generate or influence content, challenging the applicability of safe harbor provisions under existing intermediary liability frameworks. This legal uncertainty creates a significant gap in accountability, as AI developers and deployers may not be clearly held responsible for harmful content generated by their systems. The MeitY's continuous advisories highlight these challenges. There is an urgent need for an updated legal framework to address the risks posed by AI technologies.

The 2025 AI Governance Guidelines prioritize harm mitigation as the central regulatory principle. The guidelines advocate for a "whole-of-government" approach, establishing an inter-ministerial coordination committee to harmonize sectoral laws and streamline AI governance, ensuring legal clarity across domains, and a technical secretariat to oversee risk assessments, develop metrics for AI accountability, and maintain an AI incident database. The guidelines propose both entity-based and activity-based regulatory frameworks.[37] However, critics argue that without clear definitions and safeguards, such regulations could inadvertently suppress legitimate expression, including political speech.

2.2.2. Alignment with Global AI Standards

India's approach to AI governance is shaped by its aspiration to be a leader in global frameworks on AI ethics. India actively endorses the Principles on AI (2019) developed by the Organisation for Economic Co-operation and Development (OECD) and aligns its MeitY 2025 draft guidelines with rights-respecting values. It is a signatory to UNESCO's Recommendation on the Ethics of AI (2021), holding stakeholder consultations with UNESCO for aligning its AI ecosystem with UNESCO guidelines on transparency, fairness, and inclusiveness. The country positions itself as the voice of low- and middle-income countries (LMICs) and a steward of democratic AI governance grounded in human rights and cultural diversity, shaping an inclusive AI ecosystem tailored to the region's needs.[38] India is a signatory to the Bletchley Declaration 2023, affirming its

37 Sakshi Sadashiv K., "Analysing MEITY's Report on Development of AI Governance Guidelines," *Medianama,* January 8, 2025, https://www.medianama.com/2025/01/223-analysing-meitys-report-on-development-of-ai-governance-guidelines/.
38 Anupama Vijayakumar, *AI Ethics for the Global South: Perspectives, Practicalities, and India's Role,* Research and Information System for Developing Countries, New Delhi, October 2024, https://ris.org.in/sites/default/files/Publication/DP-296-Anupama-Vijayakumar.pdf.

commitment to global cooperation on safe and responsible AI. It also shares concerns with the EU, whose AI Act 2024 imposes risk-based regulations on high-risk systems. India similarly plans to implement oversight of high-risk AI uses like facial recognition without imposing blanket bans that might stifle innovation or speech. During its G20 presidency in 2023, India emphasized inclusive, human-centric AI and the importance of countering misinformation.

India leverages forums like the Global Partnership on AI (GPAI) to champion AI governance rooted in democratic values and speech rights. At the 2023 GPAI Summit in New Delhi, leaders reaffirmed their commitment to trustworthy stewardship of AI aligned with the OECD Principles and to protecting rights, dignity, and privacy. Similarly, at UNESCO's global summits, India advocated balancing innovation with ethical safeguards, ensuring that expressive freedoms are protected in the AI age. Prime Minister Narendra Modi co-chaired the Paris AI Action Summit in February 2025, emphasizing the need for global AI governance that ensures equitable access, particularly for LMICs. He highlighted India's commitment to responsible AI development, with the country signing the Statement on Inclusive and Sustainable Artificial Intelligence for People and the Planet of February 11, 2025.

Yet, while projecting itself as a champion of democratic values and LMIC concerns, India's domestic record reveals contradictions. Critics cite the government's digital repression via the Information Technology Act 2000 and the IT Rules 2021, enabling content takedowns and data collection with limited oversight.[39] India is right behind Myanmar in internet shutdowns; these are often justified on security grounds but still criticized for undermining free expression.[40] Such domestic practices stand in tension with India's global advocacy for democratic AI governance.

India's strategic positioning seeks to bridge global divides on AI norms while highlighting democratic freedoms. However, its domestic record complicates this narrative, raising questions about the consistency of its commitments to speech rights and democratic values. This inconsistency also risks undermining India's credibility in international forums, potentially weakening its influence in shaping global AI governance norms.

2.3. Defamation

India's legal framework addresses defamation through established statutes, holding AI systems and their deployers accountable for harmful outputs. The Bharatiya Nyaya Sanhita (BNS) of 2023, specifically Section 356, which replaces the IPC's Sections 499 and 500, defines and penalizes defamation. This new legislation continues to extend liability to individuals or entities responsible for disseminating defamatory content. Consequently, if an AI model generates and disseminates content that harms a person's reputation, the developers or deployers of that AI could face charges under Section 356, depending on their level of control and knowledge. The IT Rules 2021 and Section 79 of the IT Act 2000 mandate that intermediaries exercise due diligence to prevent the hosting or transmission of unlawful content, including defamatory material, making AI operators potentially liable if their systems facilitate such content and they fail to remove it upon notice. The IT Rules 2021 — specifically Rule 3 — impose due diligence obligations on intermediaries to prevent unlawful content, including defamation. MeitY's March 2024 advisories, while not legally binding, extend these expectations to AI model deployments by urging platforms to label unreliable outputs, embed

39 Sombatpoonsiri and Mahapatra, "Regulation or Repression?"
40 Access Now, "Emboldened Offenders, Endangered Communities: Internet Shutdowns in 2024," February 2025, https://www.accessnow.org/wp-content/uploads/2025/02/KeepItOn-2024-Internet-Shutdowns-Annual-Report.pdf.

traceable metadata in synthetic content, and prevent the creation or dissemination of unlawful material. Although only advisory in nature, such directions shape regulatory expectations and industry norms, and noncompliance can be cited by authorities or courts as evidence of inadequate due diligence under existing law. The government's proposed remedies compel compliance, including mandatory content moderation filters within AI models, grievance redressal mechanisms for users to report defamatory AI-generated content, and clear terms of service that explicitly outline liability for AI-generated outputs. There has not yet been a definitive legal or policy ruling on defamation. However, AI-generated defamation can be inferred from some cases. For example, in November 2024, on the eve of the Maharashtra Legislative Assembly elections, AI-generated audio clips featuring politician Supriya Sule of the Nationalist Congress Party (Sharad Pawar faction) discussing alleged illicit funding from a bitcoin fraud case went viral. Sule alleged that these manipulated clips, shared by the ruling Bharatiya Janata Party's official X account (and amplified by pro-government legacy media), aimed to damage her reputation and influence voters during a critical political period.[41] She filed a cyberfraud complaint with the Election Commission and the police and also sent a defamation notice to the Bharatiya Janata Party (BJP). There has been no ruling on her case yet. The incident highlights how audio deepfakes, which are often harder to detect than visual deepfakes, can manipulate public opinion, particularly during sensitive periods like elections, and the legal remedies are slow while reputational damage is fast. It also demonstrates that while the government sets forth AI principles and guidelines, with provisions to penalize platforms and individuals for violations, the enforcement of these standards appears inconsistent when ruling party members are involved.

2.4. Explicit Content

India is strengthening its legal framework to address the misuse of generative AI in creating explicit content, especially child sexual abuse material (CSAM) and nonconsensual intimate imagery (NCII), through its existing legal measures. Section 67B of the IT Act prohibits any electronic publication or transmission of sexually explicit material depicting children. The Protection of Children from Sexual Offences (POCSO) Act of 2012 punishes creation, storage, and circulation of child pornography.[42] Adult obscene content is outlawed under the IT Act, and the BNS 2023 likewise bans the sale and distribution of obscene material, with higher penalties if minors are involved.[43] To address NCII, the IT Act targets privacy violations, including sharing private images without consent.[44] Laws on sexual harassment and defamation may also apply. In 2023, the Delhi High Court ruled that platforms must remove illegal NCII within 24 hours of notice or lose safe-harbor immunity under the IT law.[45] This is reinforced by obligations under the IT Rules 2021, which require intermediaries to act swiftly and use tools to detect and remove CSAM or intimate sexual imagery. A Cyber Crime Reporting Portal and the Indian Cyber Crime Coordination Centre (I4C) both facilitate citizen reporting and investigation.

Current statutes did not anticipate AI-synthesized imagery, creating ambiguity when no actual child or real photo is involved. Courts and policymakers are closing this gap: In 2024, the Supreme Court urged replacing the phrase "child pornography" with "Child Sexual Exploitation and Abuse Material (CSEAM)" in law, categorizing AI-generated depictions as criminal offenses.[46] Following the verdict, merely downloading or viewing CSAM, even AI-generated, is illegal. On the adult side, AI deepfake porn cases are being prosecuted.

41 Tanishka Sodhi and Azeefa Fathima, "'Bitcoin Bomb': How Legacy Media Played Up Supriya Sule's Fake Audio Clips on Election Eve," *The News Minute*, November 27, 2024, https://www.thenewsminute.com/news/bitcoin-bomb-how-legacy-media-played-up-supriya-sules-fake-audio-clips-on-election-eve.
42 Protection of Children from Sexual Offences Act, 2012, §§13–15 (India).
43 Bharatiya Nyaya Sanhita, 2023 (draft Indian Penal Code replacement), §§294–295. See also IT Act, 2000, §67A (India).
44 Information Technology Act, 2000, §66E (India).
45 X v. Union of India & Ors (2023), https://indiankanoon.org/doc/105980506/.
46 Just Rights for Children Alliance v. S. Harish (2024), https://indiankanoon.org/doc/37078038/.

In mid-2025, Assam police arrested a man for creating pornographic videos from a single photograph of an acquaintance,[47] while Delhi Police have charged offenders circulating AI-morphed nudes of ex-partners under stalking and "outraging modesty" provisions.[48]

Enforcement capacity is expanding. In March 2025, Zero Defend Security launched Vastav AI, a detection platform provided free to law enforcement to identify AI-generated or AI-altered media. MeitY's 2024 revised advisory requires transparency labels for AI content, consent mechanisms for image use, and embedded metadata for deepfake identification.[49] The IT Rules 2021 and subsequent advisories by MeitY propose platforms use AI filters and watermarking and remove notified deepfakes within 36 hours to retain safe-harbor protection. The forthcoming Digital India Act is expected to ban tools for generating CSAM. While India's legal regime is bolstered by recent rulings, improved detection tools, and stricter platform obligations, enforcement remains uneven as many police officers lack specialized AI training and access to advanced investigative resources. This gap is significant because AI-generated CSAM and NCII can distort identity and silence victims, meaning weak enforcement directly undermines dignity, privacy, and expressive freedoms. The government's cyber commando program is seeking to address this by training select officers to detect AI-enabled offenses such as deepfakes, to trace synthetic media sources, and to use AI-based tools for rapid evidence analysis.[50]

2.5. Hate Speech

In India, the regulation of hate speech is primarily governed by the IPC and its successor, the BNS. Under the IPC, Sections 153A and 295A criminalize acts that promote enmity between different groups or deliberately outrage religious feelings, focusing on content that incites violence or threatens public order. However, these provisions have been criticized for their colonial origins and potential to suppress legitimate expression.[51] The BNS includes Sections 196, 197, 298, and 353, which address offenses such as promoting hostility, harming national integration, and spreading misinformation.

The advent of generative AI and chatbots has amplified the dissemination of hate speech and disinformation, posing significant challenges to existing legal frameworks. Social media platforms' lenient moderation policies concerning content from ruling party affiliates exacerbate this issue. For example, during the 2024 elections, pro-BJP pages were reported to have spread rampant hate speech and disinformation across various platforms.[52] The lack of effective moderation allows such content to reach vast audiences, fueling communal tensions and societal divisions.

In response to the surge in harmful content, social media companies have implemented AI-based content moderation systems. These systems aim to detect and remove hate speech. However, the effectiveness of these measures is limited, especially in a linguistically diverse country like India. AI moderation tools often struggle with regional languages and dialects, leading to inconsistent enforcement. Moreover, platforms like Meta faced criticism for approving political ads containing hate speech and conspiracy theories during the 2024 elections. This highlights the challenges in relying solely on AI for content moderation without adequate

47 The Federal, "Babydoll Archi's Deepfake Case Exposes Disturbing AI Identity Theft," July 22, 2025, https://thefederal.com/category/states/north-east/assam/babydoll-archi-deepfake-case-exposes-disturbing-ai-identity-theft-198189.
48 Indian Express, "Delhi Man Arrested for AI-Generated Obscene Images of Ex-Girlfriend," July 2, 2025, https://indianexpress.com/article/cities/delhi/delhi-man-arrested-ai-generated-obscene-images-ex-girlfriend-10101399/.
49 MeitY, *Advisory on Due Diligence by Intermediaries*, March 15, 2024, https://www.meity.gov.in/static/uploads/2024/02/9f6e99572739a3024c9cdaec53a0a0ef.pdf.
50 Sandip Dighe, "Cyber Commandos to Spot and Prevent AI-Driven Offences," *Times of India,* August 9, 2025, https://timesofindia.indiatimes.com/city/pune/cyber-commandos-to-spot-and-prevent-ai-driven-offences/articleshow/123195850.cms.
51 Sangeeta Mahapatra, Janjira Sombatpoonsiri, and Andreas Ufen, "Repression by Legal Means: Governments' Anti-Fake News Lawfare," *GIGA Focus Global*, Number 1 (2024), https://doi.org/10.57671/gfgl-24012.
52 Astha Rajvanshi, "How Modi's Supporters Used Social Media to Spread Disinformation During the Elections," *Time,* June 3, 2024, https://time.com/6984947/india-election-disinformation-modi/.

human oversight. For instance, during recent tensions between India and Pakistan, AI-generated videos of political leaders, manipulated battlefield footage, and cloned voices flooded platforms, spreading false narratives at an unprecedented pace and often dramatizing real events to fit political agendas.[53] The ability of AI to generate content in various languages, including regional Indian languages, amplifies its reach and impact, enabling malicious actors to target specific communities with divisive content more effectively.

AI algorithms on social media platforms (designed to prioritize user engagement) inadvertently create feedback loops that reinforce confirmation bias, leading to the formation of echo chambers where users are exposed primarily to content aligning with their existing beliefs. This isolation from opposing views can deepen misperceptions and harmful stereotypes, as observed when users engaging with controversial figures find themselves exposed to more hateful material.

While platforms like Meta increasingly rely on AI algorithms to moderate vast amounts of content, research indicates that these algorithms often disproportionately restrict free expression in low- and middle-income regions due to Western-centric AI frameworks, limited financial investment, inadequate language training, and political and corporate biases.[54] This selective enforcement problem highlights how platforms might over-remove critical or dissenting speech while under-removing hate speech, creating a fragmented and uneven landscape of enforcement. Despite platforms' efforts, criticisms against them and against the IT Act 2000 and IT Rules 2021 are rampant. Critics argue that these legal frameworks are used to suppress dissent and remove speech critical of the ruling party and regime while allowing hate speech by regime supporters to go viral. As mentioned earlier, the IT Rules 2021 mandate that platforms remove "illegal" information within 36 hours. But specific examples illustrate a double standard when it comes to pro-regime accounts: During the 2024 general elections, a report by the Center for the Study of Organized Hate found that senior BJP leaders delivered 266 anti-minority hate speeches that were live-streamed across YouTube, Facebook, and X. Facebook removed only three of these videos, leaving 98.4% of the reported content accessible.[55]

AI-driven content moderation, while technologically advanced, can thus be influenced by political biases, leading to an inconsistent application of policies and a disproportionate impact on freedom of expression. This refers to the idea that any law that is too narrow or too wide provides room for interpretation and implementation in a way that benefits the rulers and entrenches the existing power structures. AI can enable such power to an unprecedented level and scale in India.

2.6. Election and Political Content

Indian constitutional doctrine treats political expression as the very foundation of all democratic organization. In *Romesh Thapar v. State of Madras* (1950), the Supreme Court struck down a pre-publication ban and located the right to political critique at the heart of Article 19(1)(a). This includes the right to criticize the government and its policies without fear of reprisal. As political debate migrates online, the 2025 AI Governance Guidelines extend this logic into the algorithmic age, classifying content moderation and recommender systems as "high-risk" and insisting on transparency, accountability, fairness, and public incident reporting across the AI life cycle. However, this right is subject to reasonable restrictions under Article

53 Shivani Kava, "Deepfakes, Voice Clones and AI Images Amplified Disinformation on India-Pak Conflict," *The News Minute*, June 4, 2025, https://www.thenewsminute.com/news/deepfakes-voice-clones-and-ai-images-amplified-disinformation-on-india-pak-conflict.
54 Soorya Balendra, "Meta's AI Moderation and Free Speech: Ongoing Challenges in the Global South," *Cambridge Forum on AI: Law and Governance* 1 (2025): e21, https://doi.org/10.1017/cfl.2025.5.
55 Center for the Study of Organized Hate, "Social Media and Hate Speech in India," February 10, 2025, https://www.csohate.org/wp-content/uploads/2025/02/Report-Social-Media-and-Hate-Speech-in-India.pdf.

19(2) on grounds like public order, security, and decency, which the state has interpreted broadly to curtail dissent.[56] The government has also used similar vagueness of defining restrictions in IT Act 2000 and IT Rules 2021 to police and criminalize political speech by critical journalists and civil society actors, defining such speech variously as disinformation, hate speech, or anti-national speech.[57]

Political speech becomes especially complicated when generative AI and chatbots act as intermediaries for political content, effectively becoming themselves agents of political speech. MeitY's March 2024 advisory mandated that platforms obtain explicit approval before deploying AI models considered "unreliable," and it requires clear labeling of AI-generated content to mitigate misuse. To bolster traceability, the advisory also emphasized embedding metadata within AI-generated outputs, facilitating the identification of content origins. This move was partly in response to an incident involving Google's Gemini chatbot, which controversially described Prime Minister Narendra Modi as "fascist," igniting debates over AI's role in political narratives. Further scrutiny arose with Elon Musk's AI chatbot, Grok, which produced unfiltered and occasionally offensive remarks about Prime Minister Modi and the BJP.[58] The Indian government examined Grok's outputs for potential breaches of decency laws and the IT Rules 2021, specifically Rule 3(1)(b), which obligates intermediaries to prevent the dissemination of prohibited content. Further, chatbots like ChatGPT are increasingly being used by Indian courts, including in the Manipur, Punjab and Haryana, and Delhi High Courts, to assist with legal research and case deliberations.[59] Despite this growing integration, India lacks formal guidelines governing chatbot use, unlike the UK, which issued guidelines in December 2023 restricting AI usage to basic tasks. The only explicit limitation in India is that ChatGPT cannot be used to decide legal or factual issues in a court of law, leaving significant ambiguity around ethical and procedural considerations.[60] Some expect the DIA to introduce more stringent provisions concerning generative AI.

Another ethical dilemma arises when generative AI is used for political messaging. It becomes difficult to regulate generative AI when politicians themselves utilize deepfake technology, such as voice cloning and resurrecting deceased leaders, for campaign purposes. In India's 2024 general elections, political parties extensively employed generative AI tools for voter outreach, engaging constituents with AI-generated robocalls, personalized messages, and deepfake videos. For example, AI was employed to recreate deceased political figures like M. Karunanidhi and J. Jayalalithaa, allowing them to "endorse" candidates from beyond the grave, a tactic that raises ethical concerns about authenticity and manipulation.[61] This happened despite the Election Commission urging parties to avoid deepfakes. While AI technologies enhanced campaign reach and constituent personalization, they also blurred the lines between genuine political communication and synthetic content, challenging the integrity of democratic discourse. No comprehensive guidelines exist to govern the ethical and legal use of generative AI in elections.

The 2025 AI governance framework emphasizes transparency, accountability, and risk management. While the final guidelines do not explicitly mandate periodic audits, MeitY's March 15, 2024, advisory removed a prior-permission requirement and instead requires platforms to label synthetic media, deploy consent pop-ups for unreliable AI outputs, and embed metadata or unique identifiers to trace deepfakes. Separately, Rule 4(4)

56 Bhatia, *Offend, Shock, or Disturb.*
57 Mahapatra, Sombatpoonsiri, and Ufen, "Repression by Legal Means."
58 Meghna Bal, "AI Regulation Gets Trickier with Grok: India Needs Adaptive, Not Reactionary Policies," Esya Centre, April 21, 2025, https://www.esyacentre.org/perspectives/2025/4/21/ai-regulation-gets-trickier-with-grok-india-needs-adaptive-not-reactionary-policies-ybz994.
59 Rajinder Kumar Vij, "Why India Urgently Needs a Legal Framework to Regulate Artificial Intelligence," NatStrat, December 24, 2024, https://www.natstrat.org/articledetail/publications/why-india-urgently-needs-a-legal-framework-to-regulate-artificial-intelligence-173.html.
60 Vij, "Why India Urgently Needs a Legal Framework."
61 Nilesh Christopher, "How AI Is Resurrecting Dead Indian Politicians as Election Looms," Al Jazeera, February 12, 2024, https://www.aljazeera.com/economy/2024/2/12/how-ai-is-used-to-resurrect-dead-indian-politicians-as-elections-loom.

of the IT Rules 2021 obliges significant social media intermediaries to use automated filters — with human oversight and periodic bias, accuracy, and fairness reviews — and to operate grievance redressal systems that notify users of takedown decisions and allow reinstatement requests.

Current AI-based content moderation systems, trained on opaque and potentially biased datasets, can inadvertently or deliberately flag critical political speech as "harmful" or "misinformation." For example, Twitter/X challenged government takedown orders in the Karnataka High Court in 2022, arguing that many targeted tweets and accounts contained legitimate political speech, including posts by opposition parties and journalists critical of government policies. Twitter claimed that such blanket blocking orders violated constitutional rights and lacked transparency. The court upheld the Indian government's powers under Section 69A of the IT Act in 2023, imposing a fine on Twitter for noncompliance. The government has tried to police criticism against itself. Comedian Kunal Kamra filed a writ petition challenging the IT Rules 2023, which mandate that platforms remove "fake or false or misleading" news regarding the "business of the Central Government" flagged by the government itself, arguing this violates freedom of speech and enables the government to unilaterally censor its critics.[62]

Mandatory traceability and proactive filtering obligations further erode anonymity and incentivize over-removal by risk-averse platforms. For example, platforms like YouTube and Meta have engaged in proactive content removal and algorithmic downranking of politically sensitive speech to avoid government scrutiny, particularly during election cycles. This dynamic is exacerbated by AI moderation tools that prioritize risk avoidance, sometimes over legitimate democratic discourse. On the other hand, an Access Now/Global Witness test found that YouTube approved all 48 dummy ads carrying blatant election disinformation in English, Hindi, and Telugu, spotlighting the limits of automated review.[63] Such divergent, AI-mediated responses, overzealous in some cases, under-zealous in others, will further consolidate government control over political speech, facilitated by AI systems that are neither transparent nor accountable. Without rigorous safeguards and algorithmic audits, AI regulations risk normalizing the suppression of political dissent.

2.7. Copyright

The Indian Copyright Act 1957 forms the backbone of intellectual property protection for creative works. Although the act mentions "computer-generated" works, it does not appear to cover works made using generative AI tools, since natural persons are typically considered authors.[64] This means that a person using AI to create a work could be considered the author or rights holder if they exercise sufficient human creative control.

India's copyright law also prohibits the unauthorized reproduction or publication of someone else's work. This principle now applies to AI as well. Recognizing the challenges AI poses to copyright, the Indian government has acknowledged the need to update its IP policies.

For instance, several Indian news publishers — including *The Times of India, Hindustan Times, Dainik Bhaskar, and The Hindu* — blocked OpenAI's web crawler GPTBot to protect their content from unauthorized scraping. And in November 2024, Asian News International (ANI), a major news agency in India, purportedly

62 Kunal Kamra v. Union of India (2024), https://indiankanoon.org/doc/172701335/.
63 Global Witness, "Votes Will Not Be Counted: Indian Election Disinformation Ads and YouTube," April 2, 2024, https://globalwitness.org/en/campaigns/digital-threats/votes-will-not-be-counted-indian-election-disinformation-ads-and-youtube/.
64 Tanisha Khanna and Gowree Gokhale, "Emerging Legal Issues with Use of Generative AI," Nishith Desai Associates, October 27, 2023, https://www.nishithdesai.com/NewsDetails/10818.

pro-BJP, sued OpenAI in the Delhi High Court. *ANI Media Pvt. Ltd. v. OpenAI* is India's first court case addressing AI training data, alleging that ChatGPT was trained on ANI's copyrighted news articles and produced excerpts or summaries without permission. ANI claimed that OpenAI refused to obtain a lawful license for the content and argued that training AI with ANI's reports infringes on copyright.[65]

The Delhi High Court acknowledged the importance of this case and outlined critical legal issues: whether storing copyrighted works for AI training violates copyright law; whether AI-generated outputs that rely on such data also constitute infringement; whether these uses can be justified as fair use under Section 52 of the act; and whether Indian courts have jurisdiction if the AI company's servers are located outside the country. The court appointed two amici curiae — an IP lawyer and a law professor — who advised that storing copyrighted material for training qualifies as "reproduction" under Section 14(a)(i) of the 1957 Copyright Act and thus amounts to infringement under Section 51.[66] In early 2025, India's largest publishing industry body joined the suit, highlighting how the outcome could affect not only news agencies but also book publishers and other content owners across the country. Another example involves popular singer Arijit Singh. In 2024, an app developer, Codible Ventures LLP, cloned his voice without permission. The Bombay High Court ruled in Singh's favor, making it the first Indian judgment on generative AI misuse in music.[67]

Given that India has not yet introduced a dedicated AI copyright law, the Copyright Office and the courts handle these cases individually. Courts have issued injunctions (as in Singh's case) and carefully evaluated fair use claims (as in the OpenAI case), thus gradually developing an Indian jurisprudence on AI and copyright.

2.8. Measures Empowering Freedom of Expression

India's AI policies, driven by the IndiaAI Mission, are designed to be inclusive and rights-based. These initiatives prioritize empowering diverse and marginalized populations by addressing digital exclusion through culturally and linguistically sensitive AI frameworks. A key component of this is the Digital India Bhashini initiative, which is developing multilingual AI to enable content creation and translation across all 22 scheduled Indian languages,[68] with full functionality already available in some languages and work ongoing in others. This is supported by the Bhasha Daan program, a crowdsourcing initiative that encourages citizens to donate language data. Through sub-schemes like Bolo India (voice recordings), Suno India (audio transcription), Likho India (text translation), and Dekho India (image annotation), citizens can contribute to building the massive open-source datasets that represent them and their needs accurately to train AI models. This approach expands opportunities for people to express themselves without a language barrier.

These national efforts are supported by proactive, state-level policies from states like Odisha, Karnataka, Telangana, and Tamil Nadu, with a particular focus on improving governance and citizen services. This includes initiatives like the Telangana Data Exchange (TGDeX) to democratize access to datasets for start-ups and academia.[69] Similarly, Tamil Nadu has launched the Tamil Nadu Artificial Intelligence Mission (TNAIM) with the philosophy of "social good by design."[70] TNAIM focuses on using AI to simplify governance and

65 Aklovya Panwar, "Generative AI and Copyright Issues Globally: ANI Media v OpenAI," *Tech Policy Press,* January 8, 2025, https://www.techpolicy.press/generative-ai-and-copyright-issues-globally-ani-media-v-openai/.
66 Sharveya Parasnis, "ANI vs OpenAI Legal Battle: Does Storing Copyrighted Content Equal Infringement?," *Medianama,* March 12, 2025, https://www.medianama.com/2025/03/223-ani-vs-openai-does-storing-copyrighted-content-equal-copyright-infringement/.
67 Dipak G. Parmar, "AI Voice Cloning: How a Bollywood Veteran Set a Legal Precedent," *WIPO,* April 17, 2025, https://www.wipo.int/web/wipo-magazine/articles/ai-voice-cloning-how-a-bollywood-veteran-set-a-legal-precedent-73631.
68 Digital India Bhashini official portal, https://www.digitalindia.gov.in/initiative/digital-india-bhashini-2/.
69 The TGDeX, while lauded for its innovative approach, is a test of India's data governance, particularly how it ensures individual consent and prevents the de-anonymization of non-personal data.
70 T. Muruganandham, "New Mission to Make Tamil Nadu Leading AI Hub in Five Years," *The New Indian Express,* November 6, 2024, https://www.newindianexpress.com/states/tamil-nadu/2024/Nov/06/new-mission-to-make-tamil-nadu-leading-ai-hub-in-five-years.

accelerate e-governance outreach. For marginalized communities, including the LGBTQ+ community, the Indian government is implementing proactive measures. As mentioned earlier, MeitY has issued advisories mandating that platforms' AI models do not perpetuate bias or discriminate based on gender, religion, or social identity. Initiatives like these are strengthened by collaborations with organizations like the United Nations Development Programme, which has used generative AI to analyze and propose recommendations for LGBTQ+ rights.[71] Complementing all of these efforts is the DPDP Act 2023 (which is awaiting implementation through official rules). The DPDP Act aims to protect sensitive personal data and reinforce the right to safe expression, though it has faced criticism for granting broad exemptions to the government.

2.9. Miscellaneous

In India, privacy is a foundational safeguard for all AI regulations and their impact on freedom of expression and other democratic rights. The *Justice K.S. Puttaswamy v. Union of India* (2017) decision recognized that intrusive surveillance could discourage free speech and democratic participation. The DPDP Act 2023 requires that organizations collect personal data — essential for AI systems — lawfully, with consent, and process it fairly. The act must ensure that AI systems cannot randomly analyze personal data to suppress lawful speech or target political opposition. Critics argue, however, that these safeguards are weakened by the Digital Personal Data Protection Rules 2025, which grant the central government sweeping exemptions. Such carve-outs may permit state surveillance and the targeting of dissent without the accountability required of private entities. This creates a fundamental tension between the Act's promise of protecting privacy and its potential role in enabling a surveillance state.

India's experience with AI-powered surveillance technologies has raised concerns about their impact on free expression. The proportionality test from the *Puttaswamy* case can help determine whether AI-enabled surveillance tools, like the Automated Facial Recognition System (AFRS), are necessary, proportionate, and legally justified. For instance, civil liberties organizations argued in a 2023 case that the Delhi Police's use of AI-powered facial recognition risked discouraging citizens from assembling and expressing their views freely, creating a climate of self-censorship.[72] AI-based AFRS also causes worries about misidentification, which could lead to the wrongful targeting of protesters or journalists.[73] Because authorities often deploy this technology without clear oversight or public consent, it risks becoming a tool for profiling and censorship, undermining both privacy and freedom of expression.

Recent developments highlight the tension between AI-driven surveillance and free expression in India. Civil society groups, like the Internet Freedom Foundation, argue that unchecked AI surveillance is a tool of digital repression. AI-based digital surveillance without strong legal safeguards, transparency, or accountability can lead to dragnet surveillance, algorithmic biases, and a lack of clear redress mechanisms, despite the Supreme Court's ruling on privacy being a fundamental right.[74]

Another major problem arises from AI chatbots. While they enhance user engagement across various sectors, these chatbots can be exploited for unauthorized surveillance and data breaches. In 2024, hackers used Telegram chatbots to leak sensitive data from Star Health, India's largest health insurer, exposing the

71 United Nations Development Programme, "Uniting Diversity: Shaping the Future of Legal Equality for LGBTQ+ in India," policy brief, October 2024, https://www.undp.org/sites/g/files/zskgke326/files/2024-11/uniting_diversity-policy_brief-final_version.pdf.
72 Anushka Jain, "The Delhi Police Must Stop It's Facial Recognition System," Panoptic Tracker, Internet Freedom Foundation, New Delhi, December 29, 2019, https://panoptic.in/case-study/the-delhi-police-must-stop-its-facial-recognition-system.
73 Mahapatra, "Digital Surveillance."
74 Mahapatra, "Digital Surveillance."

personal and medical information of millions of citizens.[75] Often, AI chatbots collect extensive personal data, including sensitive details like biometric information, without clear data protection measures or robust consent frameworks. In a country with comprehensive state surveillance, AI chatbots under the control of government entities could easily be misused for monitoring and profiling citizens. This risk heightens the need for strong safeguards.

75 Christopher Bing and Munsif Vengattil, "Hacker Uses Telegram Chatbots to Leak Data of Top Indian Insurer Star Health," *Reuters*, September 20, 2024, https://www.reuters.com/technology/cybersecurity/hacker-uses-telegram-chatbots-leak-data-top-indian-insurer-star-health-2024-09-20/.

3. Conclusion

India's regulatory approach to generative AI and expressive rights reflects the country's ambition to be a leading voice for democratic AI governance and LMICs. Initiatives such as "AI for All," multilingual platforms like Bhashini, and open-source model development under BharatGPT and AI4Bharat demonstrate how technology can advance the goal of democratizing AI by bridging linguistic divides and expanding participation in public discourse. Yet these inclusive measures operate alongside centralized control over datasets and computing resources, broad exemptions under the DPDP Act, and the application of the IT Act and IT Rules to content moderation in ways that can encourage over-removal or selective enforcement. This combination of innovation and regulatory control illustrates the complexity of safeguarding freedom of expression in a multilingual, high-surveillance democracy. India's "regulation-on-the-go" and sectoral approaches to AI governance, which rely on adapting existing laws and guidelines rather than enacting a dedicated AI act, offer a mixed lesson for expressive rights: allowing flexible, context-specific responses but possibly resulting in fragmented protections and uneven enforcement, raising questions for other jurisdictions weighing whether adaptability outweighs the clarity and enforceability of a comprehensive statute.

Building on India's experience, generative AI regulation for expressive rights would benefit from measures that link openness with enforceable protections. These could include statutory guarantees of transparency in AI moderation, mandatory disclosure of takedown criteria, and clear rights for users to contest automated decisions. Independent oversight should be introduced for high-risk AI uses, particularly in elections and political communication, to prevent uneven enforcement and protect the integrity of democratic discourse. Rules for datasets and foundational models could require open, audited, and representative sources, with governance structures that share control among multiple stakeholders, while limiting exclusive state control over their storage and deployment. Strengthening privacy protections under the DPDP Act by narrowing government exemptions and requiring judicial approval for AI-based surveillance would align regulation with constitutional standards such as the proportionality principle from the *Puttaswamy* ruling of 2017.

At the global level, India's position — promoting inclusive, rights-oriented AI in UNESCO and GPAI forums while maintaining domestic regulatory practices that can, at times, restrict expressive freedoms — underscores the need to close the gap between international commitments and national implementation. Lessons from India highlight the value of embedding freedom of expression safeguards into AI risk classifications, setting standards for open and representative datasets in foundational models, and ensuring independent oversight mechanisms that work across linguistically diverse and politically contested contexts. Such approaches would help ensure that AI governance frameworks support both innovation and the democratic values they are intended to protect.

THE FUTURE OF FREE SPEECH

OCTOBER 2025

www.futurefreespeech.org