AI in a Class-Diverse India: Rights, Representation, and Regulation

Soumya Singh Chauhan

Jindal Global University, Sonipat, India

DOI 10.3217/978-3-99161-062-5-008, CC BY 4.0

https://creativecommons.org/licenses/by/4.0/deed.en
This CC license does not apply to third party material and content noted otherwise.

Abstract. The integration of Artificial Intelligence (AI) into governance frameworks is accelerating across the Global South, and India stands at the forefront of this transformation. From biometric welfare systems and predictive policing to algorithmic surveillance, AI is increasingly embedded in public service delivery and state infrastructure. However, this technological expansion occurs within a socio-political landscape deeply shaped by caste, religion, and economic class. This paper critically interrogates how AI systems intersect with India's entrenched hierarchies, revealing the representational, regulatory, and ethical gaps that threaten to reproduce and entrench structural injustice.

Drawing from interdisciplinary frameworks in AI ethics, critical data studies, and postcolonial science and technology studies, the paper engages with concepts such as sociotechnical imaginaries, algorithmic discrimination, and data colonialism. It explores how digital systems often erase class-based identities, resulting in opaque decision-making, discriminatory surveillance, and the erosion of privacy and agency for marginalized communities. Through case studies of facial recognition, welfare exclusion, and predictive policing, the paper demonstrates how caste, religious, and economic markers are indirectly encoded into algorithmic governance.

The analysis reveals that India's techno-solutionist regulatory model prioritizes innovation and efficiency over rights, accountability, and inclusion. The Digital Personal Data Protection Act, 2023, fails to address algorithmic discrimination, ensure transparency, or mandate oversight. In response, the paper proposes a rights-based, class-conscious AI governance model rooted in India's constitutional commitments to equality, justice, and fraternity. It calls for participatory design, disaggregated data practices, and robust accountability mechanisms to ensure AI serves as a tool of inclusion rather than oppression.

1 Introduction

Artificial Intelligence (AI) is rapidly transforming the landscape of governance and public service delivery in India, mirroring the global phasing in of the same. From biometric identification systems to algorithmic credit scoring and AI-driven surveillance, the state is increasingly embedding AI technologies into the architecture of administration and control.[1] This integration, however, is not occurring in a social vacuum. It is unfolding within a deeply stratified society, where caste, religion, gender, and class shape not only individual lives but also institutional logics and state practices.[2] [3]

India's socio-political fabric is characterized by vast and persistent inequalities. The intersection of caste-based exclusion, religious discrimination, and economic marginalization continues to structure access to rights, recognition, and resources.[4] In such a context, the presumed neutrality of AI technologies—frequently celebrated for their efficiency and objectivity—must be interrogated. AI systems do not merely automate decisions; they encode values, histories, and exclusions.[5] When deployed without attention to the complexities of Indian society, they risk reproducing and even amplifying existing hierarchies under the guise of modernization.[6]

This paper asks three central research questions:

[1] Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

[2] Teltumbde, A. (2018). *Republic of caste: Thinking equality in the time of neoliberal Hindutva*. Navayana.

[3] **Positionality Statement:**

The author acknowledges her social position outside the lived experiences of caste-based and religion-based discrimination, algorithmic violence, and surveillance targeting marginalized communities. This paper engages with questions of caste, data disaggregation, surveillance, and digital exclusion by drawing upon the work of Dalit, Bahujan, Adivasi, Muslim, and working-class scholars, activists, and civil liberties organizations. The author does not claim epistemic authority over these perspectives, nor can she fully convey the affective and material weight of being profiled, misrecognized, or erased by technological systems. Calls for caste-conscious data governance, rights-based regulation, and participatory frameworks are made here with humility, and with the recognition that even well-intentioned interventions risk reproducing extractive logics if they are not guided by those most affected. Discussions of disaggregated data, algorithmic representation, and techno-legal reform must centre the voices and leadership of communities historically excluded from knowledge production and decision-making. This paper therefore positions itself as a contribution to and not as a substitute for broader, community-led critiques of epistemic and infrastructural injustice in AI governance in India. It is an invitation to continued engagement, correction, and collaborative transformation.

[4] Jaffrelot, C. (2021). *Modi's India: Hindu nationalism and the rise of ethnic democracy*. Princeton University Press.

[5] Benjamin, R. (2019). *Race after technology: Abolitionist tools for the new Jim code*. Polity.

[6] Noble, S. U. (2018). *Algorithms of oppression: How search engines reinforce racism*. NYU Press.

1. How do AI systems interact with and potentially reinforce class-based hierarchies in India?
2. What representational and regulatory gaps exist that allow for the perpetuation of bias and exclusion through AI technologies?
3. What would an equitable and context-sensitive model of AI governance look like in a society as structurally unequal and diverse as India?

To answer these questions, the paper adopts an interdisciplinary approach that draws from legal studies, critical data science, and postcolonial science and technology studies. It engages with conceptual frameworks such as sociotechnical imaginaries, algorithmic discrimination, and data colonialism, and situates them within India's historical and contemporary structures of inequality. The analysis weaves together case studies, policy critique, and normative frameworks to assess the risks of uncritical AI adoption in state functions.

The paper is structured as follows: it begins with the theoretical framework that informs the critique, followed by an exploration of India's socio-historical context of inequality. It then analyses how AI systems reproduce social disparities, challenges the myth of technological neutrality, and evaluates the surveillance architecture and its disproportionate effects on marginalized communities. The regulatory and legal landscape is then examined, highlighting both domestic gaps and international best practices. Finally, the paper outlines a rights-based, inclusive vision for AI governance in India, one that centres justice, representation, and structural reform over technocratic efficiency.

2 Theoretical Framework

To critically evaluate the role of AI in reproducing class-based hierarchies in India, this paper draws from interdisciplinary frameworks in AI ethics, critical data studies, and postcolonial science and technology studies. These perspectives reveal how ostensibly neutral technologies are socially and politically embedded and often reinforce historical structures of inequality.

One key concept is sociotechnical imaginaries, which refer to collectively held visions of the future that are enacted through technological development.[7] In India, these imaginaries are often shaped by aspirations of 'smart governance,' 'Digital India,' and 'technological sovereignty.' AI becomes a symbol of progress and modernity, even when

[7] Jasanoff, S., & Kim, S.-H. (2009). *Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea. Minerva*, 47(2), 119–146. https://doi.org/10.1007/s11024-009-9124-4

implemented without adequate attention to justice, consent, or social difference.[8] These imaginaries obscure the fact that technologies are not neutral tools, but instruments embedded in political and institutional contexts.

The concept of data colonialism provides a useful lens to critique the extractive logics of AI in postcolonial societies. Even when developed domestically, AI systems in India are often built on paradigms that commodify human life through digital data.[9] The absence of community consent, the aggregation of behavioural and biometric data without oversight, and the transnational flow of such data to private or state actors mirror colonial forms of dispossession, only now digitized.

A related concept is algorithmic bias, which refers to the ways in which AI systems inherit and magnify social inequalities encoded in historical data.[10] In the Indian context, where caste, religion, and economic status shape access to services and opportunities, datasets often reflect these structural disparities. Yet AI systems built on such data are presented as objective or efficient, hiding their potential for exclusion.

Finally, the notion of surveillance capitalism helps explain the state's increasing reliance on AI technologies for monitoring, profiling, and regulating populations.[11] These systems extract behavioural data to classify individuals into risk categories, often without transparency or accountability. When used by the state, such surveillance is not only a matter of privacy but of political control, with disproportionate impacts on already marginalized groups.

Together, these concepts challenge the assumption that AI technologies can be separated from the societies in which they are developed and deployed. In a nation marked by enduring hierarchies, a class-neutral AI policy is not merely inadequate, it risks becoming an instrument of automated injustice.

[8] Ghosh, B., & Arora, S. (2019). *Smart as democratically transformative? An analysis of 'Smart City' sociotechnical imaginary in India.* IDS/Steps Centre Working Paper 109.

[9] Couldry, N., & Mejias, U. A. (2019). *The costs of connection: How data is colonizing human life and appropriating it for capitalism*. Stanford University Press.

[10] Eubanks, V. (2018). *Automating inequality*. St. Martin's Press; Noble, S. U. (2018). *Algorithms of oppression*. NYU Press.

[11] Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs.

3 Literature Review

A growing body of interdisciplinary scholarship has raised critical concerns about the intersection of artificial intelligence (AI), surveillance, and systemic inequality, particularly in societies where legal safeguards are minimal, transparency is lacking, and historic social hierarchies remain entrenched.

Virginia Eubanks (2018) offers one of the most influential accounts in this area. In *Automating Inequality*, she demonstrates how algorithmic decision-making in welfare systems reproduces poverty and discrimination, disproportionately harming low-income, racialized, and otherwise marginalized populations.[12] Kate Crawford (2021), in *Atlas of AI*, extends this critique by tracing how AI systems are embedded in extractive logics, mining not just data, but also human labour, environmental resources, and social hierarchies, ultimately reinforcing global and historical asymmetries of power.[13]

Technical audits by Inioluwa Deborah Raji and Joy Buolamwini (2019) have shown that commercial facial recognition systems exhibit significant accuracy disparities across skin tone and gender.[14] Their research highlights that darker-skinned individuals and women face higher error rates, a risk particularly relevant in India's caste-stratified and class-divided context, where facial recognition is increasingly used in welfare delivery, law enforcement, and exam surveillance. This aligns with findings by Obermeyer et al. (2019), who demonstrated that a healthcare risk algorithm trained on cost-based proxies systematically underestimated the health needs of Black patients in the United States.[15] These insights suggest parallel dangers in India, where predictive analytics in welfare and public health may encode structural exclusions through proxy indicators.

Within the Indian context, Anand Teltumbde (2018) and Suraj Yengde (2019) argue that caste is routinely erased in data collection and yet reappears implicitly through variables such as education, geographic location, and surnames.[16] These proxies are often used in algorithmic decision-making and resource allocation, reinforcing caste hierarchies

[12] Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.

[13] Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence*. Yale University Press.

[14] Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 429–435. https://doi.org/10.1145/3306618.3314244

[15] Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. *Science*, 366(6464), 447-453. https://doi.org/10.1126/science.aax2342

[16] Teltumbde, A. (2018). *Republic of caste: Thinking equality in the time of neoliberal Hindutva*. Navayana; Yengde, S. (2019). *Caste matters*. Viking.

without formal recognition of caste. Their work highlights the unique challenge of caste-blind AI systems that replicate social bias under a veneer of neutrality.

At the global level, international human rights frameworks have increasingly sought to place normative limits on the deployment of AI in governance. The *United Nations High Commissioner for Human Rights Report on the Right to Privacy in the Digital Age* (2021) underscores that AI systems must comply with principles of legality, necessity, and proportionality, especially when used in surveillance or law enforcement.[17] These standards serve as critical benchmarks for evaluating India's expanding AI surveillance architecture, which currently lacks strong procedural safeguards, transparency, or independent oversight.

Taken together, this literature reflects a growing consensus that AI governance must be grounded in human rights, social justice, and structural reform, especially in postcolonial democracies like India, where technologies are being introduced into historically unequal infrastructures. The need for disaggregated data, participatory governance, and regulatory frameworks attuned to local contexts is urgent and well established in this emerging field.

4 Social and Historical Context of Inequality in India

To understand the risks posed by AI systems in India, it is necessary to situate them within the country's deeply stratified social order. Caste, religion, and class are not peripheral identity markers but enduring structures that shape institutional access, state power, and socio-economic mobility. AI systems introduced into this terrain do not operate neutrally; rather, they inherit and can reinforce these embedded hierarchies.

Caste remains among the most resilient systems of stratification in India, governing access to education, housing, employment, and justice.[18] Despite constitutional protections and affirmative action, caste-based discrimination persists in both explicit and invisible forms. As Anand Teltumbde and Suraj Yengde argue, neoliberalism has not dismantled caste, it has merely privatized and obscured it.[19] When digital infrastructures ignore caste, they risk reproducing its logic in algorithmic form.

[17] United Nations High Commissioner for Human Rights. (2021). *The right to privacy in the digital age* (A/HRC/48/31). https://www.ohchr.org/en/documents/thematic-reports/ahrc4831-right-privacy-digital-age-report-united-nations-high [accessed June 15th, 2024]

[18] Thorat, S., & Neuman, K. (2012). *Blocked by caste: Economic discrimination in modern India*. Oxford University Press.

[19] Teltumbde, A. (2018). *Republic of Caste*; Yengde, S. (2019). *Caste Matters*.

Religion, especially Islam, has become another axis of algorithmic risk. The securitization of Muslim identity through laws, surveillance, and social media monitoring has intensified in recent years.[20] AI tools used in predictive policing or facial recognition often reflect and amplify this bias, especially when trained on data shaped by communal profiling.

Economic inequality intersects with both caste and religion. Despite welfare schemes and digital inclusion initiatives, India's rapid digitization has often increased exclusion at the margins. Errors in biometric authentication (e.g., Aadhaar), lack of mobile access, and opaque algorithmic assessments have disproportionately harmed Dalits, Adivasis, Muslims, and informal workers.[21] These harms are not anomalies, they are systemic outcomes of technologies designed for a universal subject who rarely reflects India's socio-economic majority.

A key challenge in analysing these exclusions lies in the absence of disaggregated data[22]. Most Indian digital governance systems collect minimal or no data on caste, religion, or class, citing neutrality or efficiency. Yet this omission leads to statistical erasure, preventing any meaningful visibility into how AI systems impact marginalized groups.[23] Calls for disaggregation are thus not just technical demands but political claims to recognition.

This socio-historical context makes clear that AI systems in India do not simply fail by accident. When implemented without attention to caste, religion, and class, they succeed in precisely the terms they were designed: to serve the dominant social order while rendering marginality invisible.

[20] Jaffrelot, C. (2021). *Modi's India*; Jamil, G. (2017). *Muslim Women Speak: Of Dreams and Shackles*.

[21] Panigrahi, S. (2022). *Marginalized Aadhaar: India's Aadhaar biometric ID and mass surveillance. *ACM Interactions*, 29*(2), 16–22.; Frontline. (2024, December 12). *Mandatory Aadhaar authentication leads to exclusion of the marginalised from PDS.*; The Hindu. (2017, February 18). *Aadhaar no standout performer in welfare delivery.*

[22] Disaggregated Data: data that has been broken down by detailed sub-categories, for example by marginalised group, gender, region or level of education. Disaggregated data can reveal deprivations and inequalities that may not be fully reflected in aggregated data. https://www.right-to-education.org/monitoring/content/glossary-disaggregated-data

[23] Vaidehi, R., Reddy, A. B., & Banerjee, S. (2021). Explaining caste-based digital divide in India. arXiv.

Kumar, A. (2022). Ignoring caste and denying development. Data4SDGs.

5 Bias and Inequity in AI Systems

AI systems in India are increasingly deployed across critical sectors such as welfare, policing, employment, and credit scoring.[24] These systems are often presented as neutral, objective, and scalable, promising efficient governance and rational decision-making. However, when built on biased or incomplete data, AI does not eliminate discrimination; it automates it.[25]

The first and most pressing issue is the absence of disaggregated data. Most AI systems in India do not collect or analyse information based on caste, religion, or socio-economic status. This creates an epistemic gap that conceals how marginalized groups are affected. Facial recognition systems, for instance, may perform poorly on darker-skinned individuals, many of whom are Dalits, Adivasis, or Muslims, but without disaggregated error reporting, this harm remains undocumented.[26] Similarly, hiring algorithms that use proxies like educational background or location may indirectly filter out candidates from historically oppressed communities.[27]

Predictive policing tools trained on biased crime data are another key concern. Studies in India and globally have shown that algorithmic policing tends to over-surveil poor, minority, and politically active populations.[28] In India, this translates into the overrepresentation of Muslims, Dalits, and urban poor as potential threats. Historical policing records, which already reflect decades of communal bias and caste-based targeting, become the foundation for AI systems that perpetuate that bias at scale.

The problem is not only technical but systemic: AI design and deployment in India lack transparency, accountability, and public oversight. There is no legal requirement for algorithmic audits or impact assessments. Civil society actors rarely have access to the

[24] Marda, V. (2018). *Artificial Intelligence Policy in India: A Framework for Engaging the Limits of Data-Driven Decision-Making.* Philosophical Transactions A: Mathematical, Physical and Engineering Sciences, Available at SSRN: https://ssrn.com/abstract=3240384 or http://dx.doi.org/10.2139/ssrn.3240384; Khurana, L, et. al. (2025). *Fintech And Financial Inclusion In India: A Data-Driven Analysis Of Digital Payments And Banking Access.* Journal of Informatics Education and Research. Vol 5 Issue 3

[25] Eubanks, V. (2018). *Automating inequality*. St. Martin's Press.

[26] Jain, G., & Parsheera, S. (2021). *Cinderella's shoe won't fit Soundarya: An audit of facial processing tools on Indian faces.* arXiv. https://doi.org/10.48550/arXiv.2112.09326

[27] Benjamin, R. (2019). *Race after technology*. Polity.

[28] Rina Chandran**. (2023).** *India's scaling up of AI could reproduce casteist bias, discrimination against women and minorities.* https://scroll.in/article/1055846/indias-scaling-up-of-ai-could-reproduce-casteist-bias-discrimination-against-women-and-minorities *[accessed June 12th, 2024]*

data or models used in decision-making.[29] The result is a class-blind and caste-unaware AI ecosystem that protects dominant interests while invisibilizing harm.

Moreover, even calls for bias mitigation through disaggregation must be approached critically. Scholars warn that disaggregated data, while important for detecting harm, can also entrench problematic social categories if used without community control or ethical safeguards.[30] Surveillance systems that categorize citizens by caste or religion may end up reinforcing stigma rather than promoting equity.

In sum, AI systems deployed in India today operate within—and often reproduce—inequitable social structures. Without structural reform and inclusive design, these systems risk becoming tools of 'automated inequality.'

6 Rights, Representation, and the Myth of Neutrality

One of the most insidious features of AI systems is the myth of neutrality—the claim that algorithms merely reflect data without political or ethical content. This myth legitimizes a form of technocratic governance that hides systemic exclusion behind a facade of objectivity.[31] In the Indian context, where identities like caste, religion, and socio-economic status shape access to rights and resources, this neutrality is both epistemically and politically violent.

AI systems are often designed without disaggregated representation in training data. In doing so, they commit a form of *epistemic injustice*—the marginalization of certain groups' lived realities and knowledge systems in the very tools meant to serve them.[32] Dalit, Adivasi, Muslim, and working-class communities are rendered invisible in datasets and, by extension, in algorithmic governance. Their needs are neither modelled nor prioritized, leading to exclusion that is both systematic and untraceable.

The absence of representation also impacts the framing of fairness in AI. Fairness metrics, if defined only in mathematical terms, fail to account for the historical and social context of discrimination.[33] For instance, a credit scoring model may treat all defaults equally without recognizing how structural poverty limits financial resilience in oppressed

[29] Marda, V. (2020). *Algorithmic accountability in India: A civil society perspective.* Medianama.; Chahal, V, Hooda, S. (2024). *Auditing AI: What is it and why does it matter for India?* Observer Research Foundation.

[30] Noble, S. U. (2018). *Algorithms of oppression*; Benjamin, R. (2019). *Race after technology*.

[31] O'Neil, C. (2016). *Weapons of math destruction: How Big Data Increases Inequality and Threatens Democracy.* Crown Publishing.

[32] Fricker, M. (2007). *Epistemic injustice: Power and the ethics of knowing*. Oxford University Press.

[33] Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning*. fairmlbook.org.

communities. Similarly, an exam surveillance system may apply the same facial recognition algorithm to all candidates, ignoring how Dalit and tribal students may face misrecognition or digital exclusion due to technical or infrastructural disparities.[34]

Who gets to define fairness, and whose values shape the algorithmic process, are fundamentally political questions.[35] In India, where technological design is dominated by upper-caste, urban, English-speaking actors, the perspectives of those most vulnerable to AI harms are rarely included in development or policy spaces. This asymmetry results not just in misrepresentation but in systemic non-recognition.

There is growing consensus in critical data studies that disaggregated data is essential to identifying and remedying these harms.[36] However, this approach also carries risks. Without safeguards, such data can be co-opted to justify new forms of profiling or surveillance. Scholars caution that disaggregation must not become a technocratic fix to a political problem.[37] It must be paired with community consent, legal protections, and participatory governance mechanisms that ensure such data serves the interests of the communities it describes.

The invisibilization of caste, religion, and class in data is not simply a technical oversight, it is a political act with real-world consequences. Systems built on such erasures deny people the ability to be seen, heard, or served by the technologies that increasingly govern their lives. Confronting this requires more than bias audits or data collection protocols, it demands a rethinking of what it means to design just technologies in a deeply unjust world.

7 Surveillance and Disproportionate Impacts

India has rapidly expanded its use of AI-powered surveillance in the name of administrative efficiency and national security. From facial recognition systems at protests and airports to predictive policing tools in urban centers and politically sensitive regions, AI surveillance is becoming a core apparatus of state power.[38] While marketed

[34] Reuters. (2020, November 10). *"Unfair surveillance'? Online exam software sparks global student revolt." Times of India*, reporting on Thomson Reuters Foundation coverage.

[35] Winner, L. (1980). Do artifacts have politics? *Daedalus*, 109(1), 121–136.

[36] Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021). *Re-imagining algorithmic fairness in India and beyond.* In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency* (FAccT '21) (pp. 315–328). Association for Computing Machinery

[37] Benjamin, R. (2019). *Race after technology*. Polity.

[38] Internet Freedom Foundation. (2024, January 16). *Resist Surveillance Tech, Reject Digi Yatra*. Internet Freedom Foundation.

as neutral and technocratic, these systems disproportionately affect Muslims, Dalits, Adivasis, and the urban poor—communities already over-policed and under-protected.[39]

Surveillance systems such as Digi Yatra, AFRS, and the Jarvis prison monitoring platform illustrate the state's growing investment in real-time biometric and behavioural tracking.[40] These tools are often deployed without public consultation, legal transparency, or democratic oversight. In practice, they convert social and spatial disadvantage into algorithmic suspicion. Muslim neighbourhoods become 'high-risk zones'; poor, informal workers become data points for risk scoring.

Predictive policing, in particular, reflects the dangers of algorithmic circularity. Historical crime data, often shaped by caste and communal biases, are fed into machine learning models that then 'predict' future risk in the same communities.[41] The result is not predictive justice but pre-emptive punishment. Innocent individuals are flagged based on where they live, how they look, or what language they speak.

This form of surveillance threatens not just informational privacy but behavioural and decisional privacy, i.e. the freedom to think, act, and move without being watched.[42] When protestors are identified and tracked using facial recognition, or when students are monitored during exams through AI-powered webcams, surveillance becomes a tool of discipline and deterrence.[43] The chilling effect is especially severe for historically marginalized groups, for whom even minor errors in identification can result in disproportionate harm, including arrest, loss of services, or reputational damage.

India's legal framework provides few protections against such overreach. There are no binding transparency norms, audit mandates, or meaningful redress mechanisms for individuals misidentified or wrongly profiled by AI tools. State agencies often invoke national security to avoid scrutiny, citing exemptions in the Digital Personal Data Protection Act (DPDPA), 2023. This creates a governance gap where AI surveillance grows unchecked, especially in spaces of political dissent or social vulnerability.

AI-powered surveillance is not only a question of technology—it is a question of power. In the absence of legal safeguards and public accountability, it becomes a tool of

[39] Singh, S., & Mohanty, R. (2023). *Impacts and ethics of using Artificial Intelligence (AI) by the Indian Police. Police Practice and Research*, 24(3), 102–116.

[40] Abhijit Ahaskar (2019). *Uttar Pradesh prisons turn to AI-based video surveillance to monitor inmates.* https://www.livemint.com/technology/tech-news/uttar-pradesh-prisons-turn-to-ai-based-video-surveillance-to-monitor-inmates-11573196335267.html *[accessed June 12th, 2024]*

[41] Ramachandran Murugesan (2021). *Predictive policing in India: Deterring crime or discriminating minorities?.* https://blogs.lse.ac.uk/humanrights/2021/04/16/predictive-policing-in-india-deterring-crime-or-discriminating-minorities/ *[accessed June 12th, 2024]*

[42] Solove, D. (2008). *Understanding privacy*. Harvard University Press.

[43] India Today. (2024). UPSC to deploy AI for exam surveillance.

structural domination. Any ethical AI policy must begin by asking not what can be surveilled, but who is being watched—and why.

8 Regulatory and Legal Gaps

Despite the rapid adoption of AI across state institutions in India, the country's legal and regulatory framework remains profoundly underdeveloped. The Digital Personal Data Protection (DPDP) Act, 2023, India's first comprehensive data protection law, offers limited safeguards against AI-driven discrimination and surveillance.[44] Its focus on consent and individual control over personal data does not address deeper structural harms like algorithmic bias, profiling, or exclusion.

Crucially, the Act grants sweeping exemptions to state actors in matters concerning sovereignty, public order, and national security, effectively insulating state-led AI surveillance from accountability.[45] There are no legal obligations for government agencies to disclose their use of AI systems, conduct impact assessments, or allow public auditing of algorithms.[46] This is particularly concerning given the growing evidence that AI systems deployed in India often reproduce social hierarchies and target already marginalized communities.

Furthermore, the DPDP Act does not mandate disaggregated data collection across caste, religion, gender, or class, nor does it require public agencies to publish impact assessments based on these variables. As a result, algorithmic harms to specific groups remain legally invisible, and therefore unaddressed.[47] There are also no remedies for individuals adversely affected by AI decisions, such as those misidentified by facial recognition or denied services due to algorithmic scoring.

In contrast, international frameworks such as the European Union's General Data Protection Regulation (GDPR) and the EU AI Act provide more robust protections. These include rights to explanation, obligations for transparency in automated decision-making, and mandatory human rights impact assessments for high-risk AI systems.[48] Similarly,

[44] *Digital Personal Data Protection Act*, 2023

[45] Krishna Preetham Kanthi. (2024). *Privacy, Surveillance, and State Interest: Appraising the DPDP Act through a Constitutional Perspective. Beyond Encryption: Tech & Data Protection,* https://www.ijlt.in/post/privacy-surveillance-and-state-interest-appraising-the-dpdp-act-through-a-constitutional-perspect *[accessed June 12th, 2024]*

[46] Internet Freedom Foundation. (2023). *DPDP Act analysis: Surveillance and public accountability*.

[47] Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021). *Re-imagining algorithmic fairness in India and beyond.* In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)*. arXiv:2101.09995v2

[48] *Regulation (EU) 2024/1689*

UNESCO's *Recommendation on the Ethics of Artificial Intelligence* (2021) emphasizes fairness, inclusivity, and the right to participate in decisions about AI systems that affect communities.[49]

India's **techno-legal discourse**—that is, the body of policy documents, official strategies, and legal debates surrounding emerging technologies—**remains innovation-driven but accountability-poor**. Government strategies such as *NITI Aayog's National Strategy for Artificial Intelligence* (2018) and *Digital India* emphasize economic growth and 'AI for All,' but devote little attention to human rights, transparency, or oversight mechanisms.[50] Scholars and policy analysts have similarly noted that India's regulatory imagination privileges technological innovation over ethical and legal accountability. The absence of a dedicated AI law, the lack of an independent oversight body, and the government's discretionary power to bypass privacy protections create a governance vacuum. Civil society actors have consistently demanded stronger legal frameworks that address the specific risks posed by AI, including casteist profiling, communal surveillance, and the erasure of minority voices from digital systems.[51] This discourse comprises both *state-led policy narratives*—framing AI and digital governance primarily as engines of national innovation—and *civil society critiques* highlighting the absence of enforceable accountability norms. The tension between these two positions defines India's techno-legal trajectory today.

Regulation cannot merely be reactive or sectoral. It must be proactive, intersectional, and rooted in constitutional values of equality, justice, and fraternity. Without this, AI will continue to operate in India as a class-blind, caste-silent, and surveillance-heavy apparatus of governance.

9 Toward an Inclusive AI Policy

An equitable AI governance framework in India must begin with the recognition that neutrality is not justice. The country's socio-technical systems operate in the shadow of caste, communalism, and economic inequality. Without explicit safeguards, AI technologies will continue to reinforce these asymmetries under the guise of modernization.

[49] UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*.

[50] NITI Aayog. (2018). *National Strategy for Artificial Intelligence: #AIforAll.* Government of India. Available at: https://www.niti.gov.in

[51] Gupta, M. (2025). *Regulating Artificial Intelligence in India: A Legal Imperative for Ethical Accountability and Responsible Innovation.* Lawful Legal.; *AI Regulation in India: Between Innovation and Accountability.* (2024). The Policy POV.; Agarwal, A., & Nene, M. J. (2025). *Incorporating AI Incident Reporting into Telecommunications Law and Policy: Insights from India.* arXiv preprint.

The first step toward an inclusive framework is the mandatory collection and use of disaggregated data. AI systems must be able to reflect how their outcomes affect people differently based on caste, religion, gender, and economic status.[52]However, this disaggregation must not be technocratically imposed. It must be co-designed with the communities it aims to represent, governed by data sovereignty principles, and accompanied by ethical safeguards to prevent misuse in surveillance or profiling.[53]

Second, AI systems used in public governance should be subject to mandatory algorithmic audits, especially for high-risk applications like policing, welfare, education, and credit. These audits must include fairness assessments, not just accuracy checks.[54] Audit bodies should be independent, publicly funded, and include diverse representation from civil society, academia, and impacted communities.

Third, India must establish legal protections against algorithmic discrimination, modelled on both international best practices and its own constitutional guarantees under Articles 14, 15, and 21. These protections should include the right to explanation, the right to opt-out of automated decision-making, and remedies for algorithmic harms.[55]

Fourth, a class-aware AI policy must be participatory. This means involving marginalized communities not only as subjects of impact assessments, but as co-creators in design, deployment, and oversight. Grassroots organizations, public interest technologists, and community media must be empowered to critique, shape, and challenge AI systems.[56]

Finally, AI education and policy discourse must move beyond elite institutions and urban centers. Public education campaigns on algorithmic rights, data justice, and digital harm are essential to counter the opacity that currently shields AI systems from scrutiny.[57] Without a broad democratic base, technological governance risks becoming a tool for elite consolidation.

In a class-based democracy, technological design must be accountable to those it most affects. Equity is not an afterthought; it is the measure of legitimacy. India's constitutional values of justice, equality, and fraternity must be at the centre of any AI governance

[52] Kitchin, R. (2014). *The data revolution: Big data, open data, data infrastructures and their consequences*. Sage.

[53] Couldry, N., & Mejias, U. A. (2019). *The costs of connection*. Stanford University Press.

[54] Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. *Proceedings of the AAAI/ACM Conference on AI Ethics and Society*.

[55] European Commission. (2021). *Proposal for a Regulation on Artificial Intelligence*.

[56] D'Ignazio, C., & Klein, L. F. (2020). *Data feminism*. MIT Press.

[57] Zuboff, S. (2019). *The age of surveillance capitalism*. PublicAffairs.

model. Without this, AI will not be a tool for liberation, but a new instrument of exclusion in digital form.

10 Conclusion

As India deepens its investment in Artificial Intelligence across governance, welfare, and security, it must confront a difficult truth: AI systems, if left unchecked, will not disrupt social hierarchies, they will entrench them. Designed and deployed within a structurally unequal society, these technologies do not merely reflect injustice; they encode, amplify, and automate it.

This paper has argued that the apparent neutrality of AI masks profound representational and regulatory failures. In a context marked by caste stratification, religious marginalization, and economic exclusion, the absence of disaggregated data, legal safeguards, and participatory governance leaves vulnerable communities disproportionately exposed to algorithmic harm.

An equitable AI future in India requires more than technical correction. It demands a structural reckoning. Regulation must be rights-based. Data must be collected with care, consent, and justice in mind. And communities most affected must not be relegated to footnotes, they must be cantered as architects of the systems that govern them.

In a constitutional democracy that promises justice, equality, and dignity for all, technology must be held to those same standards. AI must be accountable not only to efficiency metrics, but to the people, and especially to those whom history has taught to expect neither fairness nor visibility from the state. Only then can Artificial Intelligence become a tool for social transformation, rather than digital domination.

References

Abhijit Ahaskar (2019). Uttar Pradesh prisons turn to AI-based video surveillance to monitor inmates. https://www.livemint.com/technology/tech-news/uttar-pradesh-prisons-turn-to-ai-based-video-surveillance-to-monitor-inmates-11573196335267.html [accessed June 12th, 2024]

Agarwal, A., & Nene, M. J. (2025). *Incorporating AI Incident Reporting into Telecommunications Law and Policy: Insights from India.* arXiv preprint.

AI Regulation in India: Between Innovation and Accountability. (2024). The Policy POV.

Barocas, S., Hardt, M., & Narayanan, A. (2019). Fairness and machine learning. fairmlbook.org.

Benjamin, R. (2019). Race after technology: Abolitionist tools for the new Jim code. Polity.

Chahal, V, Hooda, S. (2024). Auditing AI: What is it and why does it matter for India? Observer Research Foundation.

Couldry, N., & Mejias, U. A. (2019). The costs of connection: How data is colonizing human life and appropriating it for capitalism. Stanford University Press.

Crawford, K. (2021). Atlas of AI: Power, politics, and the planetary costs of artificial intelligence. Yale University Press.

D’Ignazio, C., & Klein, L. F. (2020). Data feminism. MIT Press.

Digital Personal Data Protection Act, 2023

Eubanks, V. (2018). Automating inequality: How high-tech tools profile, police, and punish the poor. St. Martin's Press.

European Commission. (2021). Proposal for a Regulation on Artificial Intelligence.

Fricker, M. (2007). Epistemic injustice: Power and the ethics of knowing. Oxford University Press.

Ghosh, B., & Arora, S. (2019). Smart as democratically transformative? An analysis of ‘Smart City’ sociotechnical imaginary in India. IDS/Steps Centre Working Paper 109.

Gupta, M. (2025). *Regulating Artificial Intelligence in India: A Legal Imperative for Ethical Accountability and Responsible Innovation.* Lawful Legal.

India Today. (2024). UPSC to deploy AI for exam surveillance.

Internet Freedom Foundation. (2023). DPDP Act analysis: Surveillance and public accountability.

Internet Freedom Foundation. (2024, January 16). Resist Surveillance Tech, Reject Digi Yatra. Internet Freedom Foundation.

Jaffrelot, C. (2021). Modi's India: Hindu nationalism and the rise of ethnic democracy. Princeton University Press.

Jain, G., & Parsheera, S. (2021). Cinderella's shoe won't fit Soundarya: An audit of facial processing tools on Indian faces. arXiv. https://doi.org/10.48550/arXiv.2112.09326

Jamil, G. (2017). Muslim Women Speak: Of Dreams and Shackles.

Jasanoff, S., & Kim, S.-H. (2009). Containing the atom: Sociotechnical imaginaries and nuclear power in the United States and South Korea. Minerva, 47(2), 119–146. https://doi.org/10.1007/s11024-009-9124-4

Khurana, L, et. al. (2025). *Fintech And Financial Inclusion In India: A Data-Driven Analysis Of Digital Payments And Banking Access.* Journal of Informatics Education and Research.

Kitchin, R. (2014). The data revolution: Big data, open data, data infrastructures and their consequences. Sage.

Krishna Preetham Kanthi. (2024). Privacy, Surveillance, and State Interest: Appraising the DPDP Act through a Constitutional Perspective. Beyond Encryption: Tech & Data Protection, https://www.ijlt.in/post/privacy-surveillance-and-state-interest-appraising-the-dpdp-act-through-a-constitutional-perspect [accessed June 12th, 2024]

Kumar, A. (2022). Ignoring caste and denying development. Data4SDGs.

Marda, V. (2020). Algorithmic accountability in India: A civil society perspective. Medianama.;

Marda, V. (2018). *Artificial Intelligence Policy in India: A Framework for Engaging the Limits of Data-Driven Decision-Making.* Philosophical Transactions A: Mathematical, Physical and Engineering Sciences

NITI Aayog. (2018). *National Strategy for Artificial Intelligence: #AIforAll.* Government of India. Available at: https://www.niti.gov.in

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. NYU Press.

O'Neil, C. (2016). Weapons of math destruction: How Big Data Increases Inequality and Threatens Democracy. Crown Publishing.

Obermeyer, Z., Powers, B., Vogeli, C., & Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464), 447-453. https://doi.org/10.1126/science.aax2342

Panigrahi, S. (2022). Marginalized Aadhaar: India’s Aadhaar biometric ID and mass surveillance. *ACM Interactions, 29*(2), 16–22.; Frontline. (2024, December 12). Mandatory Aadhaar authentication leads to exclusion of the marginalised from PDS.; The Hindu. (2017, February 18). Aadhaar no standout performer in welfare delivery.

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. Proceedings of the AAAI/ACM Conference on AI Ethics and Society.

Ramachandran Murugesan (2021). Predictive policing in India: Deterring crime or discriminating minorities?. https://blogs.lse.ac.uk/humanrights/2021/04/16/predictive-policing-in-india-deterring-crime-or-discriminating-minorities/ [accessed June 12th, 2024]

Raji, I. D., & Buolamwini, J. (2019). Actionable auditing: Investigating the impact of publicly naming biased performance results of commercial AI products. Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 429–435. https://doi.org/10.1145/3306618.3314244

Regulation (EU) 2024/1689

Reuters. (2020, November 10). ‘‘Unfair surveillance’? Online exam software sparks global student revolt.’ Times of India, reporting on Thomson Reuters Foundation coverage.

Rina Chandran**. (2023).** India’s scaling up of AI could reproduce casteist bias, discrimination against women and minorities. https://scroll.in/article/1055846/indias-scaling-up-of-ai-could-reproduce-casteist-bias-discrimination-against-women-and-minorities [accessed June 12th, 2024]

Sambasivan, N., Arnesen, E., Hutchinson, B., Doshi, T., & Prabhakaran, V. (2021). Re-imagining algorithmic fairness in India and beyond. In Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT ’21) (pp. 315–328). Association for Computing Machinery

Singh, S., & Mohanty, R. (2023). Impacts and ethics of using Artificial Intelligence (AI) by the Indian Police. Police Practice and Research, 24(3), 102–116.

Solove, D. (2008). Understanding privacy. Harvard University Press.

Teltumbde, A. (2018). Republic of caste: Thinking equality in the time of neoliberal Hindutva. Navayana.

Thorat, S., & Neuman, K. (2012). Blocked by caste: Economic discrimination in modern India. Oxford University Press.

UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence.

United Nations High Commissioner for Human Rights. (2021). The right to privacy in the digital age (A/HRC/48/31). https://www.ohchr.org/en/documents/thematic-reports/ahrc4831-right-privacy-digital-age-report-united-nations-high [accessed June 15th, 2024]

Vaidehi, R., Reddy, A. B., & Banerjee, S. (2021). Explaining caste-based digital divide in India. arXiv.

Winner, L. (1980). Do artifacts have politics? Daedalus, 109(1), 121–136.

Yengde, S. (2019). Caste matters. Viking.

Zuboff, S. (2019). The age of surveillance capitalism. PublicAffairs.