THE ILLUSION OF NEUTRALITY: HOW 'OBJECTIVE' ALGORITHMS PERPETUATE DISCRIMINATION

Dishi, BBA LLB, Symbiosis Law School, Pune

Introduction

In 1936, the mathematician Alan Turing specified algorithms as determinate sequences of computational steps, mechanical procedures that progress toward logical conclusions with mathematical exactitude, insulated from subjective human judgment. This foundational characterization positioned algorithms as instruments of pure rationality, detached from the social contexts in which they would later operate.[1] For decades, Western technological discourse sustained this view: algorithms were cast as objective arbiters capable of transcending biases embedded in human decision-making systems. When automated risk assessment entered criminal justice in the 1990s, when financial institutions mechanized credit-worthiness determinations in the 2000s,[2] and when law enforcement agencies implemented predictive policing architectures in the 2010s, the underlying premise remained ever constant to imply that algorithmic systems would achieve fairness through their non-human nature.[3]

India's relationship with this technological narrative has been peripheral and central at the same time. While the Global North was busy building out algorithmic infrastructure, India became its key engineering workforce, with the IT services sector alone employing more than five million professionals by 2020.[4] But this technical capability has not been matched by a similar critical inquiry into algorithmic neutrality in Indian contexts. When Aadhaar's biometric authentication system started deciding who got food subsidy, when automated loan appraisal platforms began considering creditworthiness for first-time borrowers from marginalized communities, and when algorithmic hiring tools entered the recruitment process of Indian

[1] McGillivray, B., Alex, B., Ames, S., Armstrong, G., Beavan, D., Ciula, A., Colavizza, G., Cummings, J., De Roure, D., Farquhar, A. and Hengchen, S., 2020. The challenges and prospects of the intersection of humanities and data science: A white paper from The Alan Turing Institute.

[2] OMBUNA, J.G., 2018. *EVALUATION OF FACTORS INFLUENCING SMALL AND MICRO ENTERPRISES'CREDIT WORTHINESS BY COMMERCIAL BANKS IN NAKURU TOWN, KENYA* (Doctoral dissertation, KABARAK UNIVERSITY).

[3] McKay, C., 2020. Predicting risk in criminal procedure: actuarial tools, algorithms, AI and judicial decision-making. *Current Issues in Criminal Justice*, *32*(1), pp.22-39.

[4] Judy, R.W. and d'Amico, C., 1997. *Workforce 2020: Work and workers in the 21st century*. Hudson Institute, Herman Kahn Center, PO Box 26-919, Indianapolis, IN 46226; tele.

corporations, notions of objectivity were imported lock, stock, and barrel without interrogation of their founding premises.

This essay argues that discrimination in algorithms is sustained not through technical failure but through structural design-through the ossification of historical inequities within training data, through the deployment of socioeconomic proxies that reproduce caste and class hierarchies, and through operation within legal and technical opacity that precludes accountability. The Indian context, analysed later in this essay, manifests how algorithmic bias intersects with pre-existing systems of social stratification to magnify rather than mitigate discrimination.

The Global Architecture of Algorithmic Bias

The problem begins with a seemingly innocuous premise: machine learning algorithms identify patterns in past data to predict future behaviour. But the past is not an objective accounting of events; it is a record of who was in power, whose lives were valued, and whose were ignored. When ProPublica's 2016 investigation proved that COMPAS, a widely used criminal risk assessment tool, labelled Black defendants as high-risk at nearly twice the rate of white defendants who had committed similar crimes, its defenders said that the algorithm was simply reflecting statistical reality.[5] And they were right, but not in the way they meant it.[6] The algorithm did reflect reality, but that reality is the reality of a criminal justice system that has surveilled, arrested, and incarcerated Black Americans at disproportionate rates for decades.[7]

This embodies what has been recognized as the fundamental paradox of data-driven discrimination: algorithms achieve accuracy by learning from the past, yet the past is saturated with the very biases that technology aspires to transcend. Amazon's automated recruiting tool, trained on a decade of hiring data from a predominantly male technical workforce, learned to systematically downgrade résumés containing the term "women's" - for example, in "women's chess club captain".[8] The algorithm was not malfunctioning; it was operating precisely as

[5] Rodà, A., 2025. The COMPAS case: an educational journey for explaining fairness in AI-based applications.
[6] Washington, A.L., 2018. How to argue with an algorithm: Lessons from the COMPAS-ProPublica debate. *Colo. Tech. LJ*, *17*, p.131.
[7] Engel, C., Linhardt, L. and Schubert, M., 2025. Code is law: how COMPAS affects the way the judiciary handles the risk of recidivism: C. Engel et al. *Artificial Intelligence and Law*, *33*(2), pp.383-404.
[8] Andrews, L. and Bucher, H., 2022. Automating discrimination: AI hiring practices and gender inequality. *Cardozo L. Rev.*, *44*, p.145.

designed, optimizing for patterns that replicated historical gender exclusion.

But the philosophy of "garbage in, garbage out" with respect to training datasets fundamentally misapprehends the problem.[9] The issue is not only that training data contain obvious errors or anomalies that could be mitigated through better data hygiene. Rather, as legal scholar Deborah Hellman has argued, discrimination is embedded in ostensibly correct data that reflect genuinely disparate life circumstances produced by structural inequality.[10] When an algorithm infers that applicants from certain zip codes exhibit higher loan default rates, it learns a real correlation and equates it with causation-a causal narrative generated through decades of redlining, employment discrimination, and wealth extraction from marginalized communities. The data are not dirty per se; the society that created them is.[11] But these global patterns of algorithmic bias do not simply replicate themselves uniformly across contexts. Instead, they interact with local systems of stratification in ways that magnify and transform discrimination. In India, where the deployment of algorithms has accelerated without regulatory infrastructure to match, these dynamics materialize through caste, class, linguistic and regional hierarchies that long predate digital technology yet find novel expression through it.[12]

Algorithmic Discrimination in the Indian Context

Consider a rural woman in Jharkhand who comes to a ration shop to collect her household's subsidized grain. She presses her finger onto the biometric scanner-the same finger that has been used for years to transplant rice seedlings, to collect firewood, and to harvest mahua flowers-but it flashes red: authentication failed. A second try yields the same result. The shopkeeper is unconvinced; no Aadhaar verification means no rations. She walks away with unrealized entitlement, not because she is not entitled to it, but because her fingertips have been worn smooth by manual labour, illegible to a system that implicitly defaults to sedentary, indoor work as normative.

This is not a hypothetical scenario. Surveys led by the Jean Drèze–linked research collective

[9] Zhu, Z., 2023. *Embracing Data-Centric AI: Practical and Provable Solutions to Weakly Supervised Data* (Doctoral dissertation, University of California, Santa Cruz).
[10] Hellman, D., 2020. Measuring algorithmic fairness. *Virginia Law Review*, *106*(4), pp.811-866.
[11] Johnson, I., McMahon, C., Schöning, J. and Hecht, B., 2017, May. The effect of population and" structural" biases on social media-based algorithms: A case study in geolocation inference across the urban-rural spectrum. In *Proceedings of the 2017 CHI conference on Human Factors in Computing Systems* (pp. 1167-1178).
[12] Biju, P.R. and Gayathri, O., 2025. Indic approach to ethical AI in automated decision making system: implications for social, cultural, and linguistic diversity in native population. *AI & SOCIETY*, pp.1-26.

report hundreds of such authentication failures across Jharkhand throughout 2017–2018, with the exclusion rates significantly higher among elderly agricultural workers, construction laborers, and beedi rollers.[13] Most of these occupations are held by Scheduled Castes, Scheduled Tribes, and Other Backward Classes in large proportions. The Aadhaar system represents what sociologist Ruha Benjamin has called "the New Jim Code": discrimination functioning through technological mechanisms rather than overtly exclusionary rules, so that it would appear neutral and yet spawn systematically unequal outcomes.[14] The algorithm does not have to name caste as such; by privileging biometric features representative of non-manual work, it effectively keeps out those whose economic fortunes have left their signature marks on their bodies.

Discriminatory structures extend beyond biometric identification to financial access. Over the last five years, digital lending platforms have mushroomed in India, promising credit to the "unbanked" through algorithmic assessments that bypass traditional banking infrastructure.[15] Yet such systems frequently rely on proxy data-smartphone specifications, application use, social network analyses, and even linguistic patterns in text messages-that strongly correlate with caste and class position. Applicants who communicate mainly in a regional language rather than English, whose contacts include few smartphone users, and whose device is of lower-cost specification tend to receive systematically worse credit terms or outright rejection. Such algorithmic credit scoring reproduces exactly those exclusions it claims to transcend, excluding Dalit and Adivasi borrowers who cannot boast the same digital footprint as upper-caste urban privilege.[16] A similar pattern emerges in employment. Automated resume screening systems employed by Indian corporations sift candidates by educational pedigree, for instance, IITs, NITs, certain private institutions; English-language proficiency; urban residential addresses; and prior employment at recognizable firms. Each of these criteria may seem neutral on their face. Collectively, however, they serve as caste and class filters, favoring applicants whose families could afford elite education, whose homes provided an environment to learn English, and whose social networks supplied corporate connections. A skilled engineer from a

---

[13] Menon, S., 2017. Aadhaar-based Biometric Authentication for PDS and Food Security: observations on implementation in Jharkhand's Ranchi District. *Indian Journal of Human Development*, *11*(3), pp.387-401.
[14] Benjamin, R., 2023. Race after technology. In *Social Theory Re-Wired* (pp. 405-415). Routledge.
[15] Mobius, M., Casanova, L., Pandit, S. and Ninia, J., 2025. *The Digital Currency Revolution: Central Bank Digital Currencies, Crypto, and the Future of Global Finance*. Springer Nature.
[16] Dharmaraj, M.G., 2024, October. Surveillance, Complicity, and Abolitionist Refusal in India and Beyond. In *AI for People, Democratizing AI: Second EAI International Conference, CAIP 2023, Bologna, Italy, November 24-26, 2023, Proceedings* (Vol. 591, p. 46). Springer Nature.

government college in a tier-three city, possessing technical skills comparable to that of an IIT graduate, can be systematically weeded out before human evaluation. In this context, the algorithm is not merely gatekeeping through neutrality. It gatekeeps via strategically deployed neutrality to automate the social reproduction of privilege at scales and speeds beyond the reach of individual human biases.

The Proxy Problem: Circumventing Anti-Discrimination Frameworks

The discriminatory mechanisms described earlier all share a common structural feature: they work through variables that do not have an explicit connection with the protected categories but are strongly correlated with them. This is the proxy problem-continuing to be algorithmic discrimination's most sophisticated evasion of legal constraint. Traditional anti-discrimination frameworks-designed to combat explicit exclusions such as "No Dalits," "Brahmins Only," or "Upper Castes Preferred"-prove strikingly ineffective against systems that achieve identical exclusionary outcomes through facially neutral criteria.[17]

The mechanism operates across jurisdictions. In the United States, ZIP code functions as a powerful proxy for race, encoding decades of residential segregation. Credit history proxies for class position, penalizing those whose poverty hindered access to formal financial systems. These correlations are not accidental; they are the residue of historical discrimination rendered into algorithmic input.[18] The algorithm need never directly perceive race or class; it achieves discrimination through their shadow variables.

India's proxy landscape maps onto its distinctive hierarchies. Pincode operates analogously to the American ZIP code but encodes both caste concentration and class stratification. Certain pincodes in cities such as Chennai, Nagpur, or Patna contain higher proportions of Scheduled Caste residents, a spatial distribution produced by decades of residential segregation and economic marginalization.[19] When algorithms incorporate pincode as a risk factor, they effectively discriminate on caste without explicit mention. Educational institutions function analogously as proxies for caste; the Indian Institute of Technology system, while formally

[17] Sundaram, D., 2023. " If you like Hindutva... you might also like...": How Facebook's Recommendation Model Reinforces Hindu Nationalism's Casteism. *American Religion*, *5*(1), pp.118-129.
[18] Arora, P., 2019. Politics of algorithms, Indian citizenship, and the colonial legacy. *Global digital cultures: Perspectives from south Asia*, pp.37-52.
[19] Gagandeep, Kaur, J., Mathur, S., Kaur, S., Nayyar, A., Singh, S.P. and Mathur, S., 2024. Evaluating and mitigating gender bias in machine learning based resume filtering. *Multimedia Tools and Applications*, *83*(9), pp.26599-26619.

open through a competitive examination, draws disproportionately from upper-caste families able to afford intensive coaching. An algorithm that privileges IIT credentials thus advantages caste position. Language proficiency operates similarly: English fluency correlates strongly with upper-caste status, given that English-medium education remains financially inaccessible to most Dalit and Adivasi families.[20] Even patterns of smartphone ownership, device cost, application usage, and digital literacy function as class and caste proxies within algorithmic credit assessment.

The Indian constitutional framework, while proscribing caste discrimination under Essays 15 and 17, drafted these provisions for an analog world of overt exclusions. They have no purchase against algorithmic systems that discriminate through correlation rather than classification. In contrast to the European Union's General Data Protection Regulation, which ensures transparency and establishes rights to explanation for automated decision-making, or even the United Kingdom's Equality Act, which recognizes indirect discrimination through facially neutral criteria, Indian law does not provide any equivalent safeguards.[21] The Information Technology Act of 2000-which, in any event, acquired much of its current form and substance only in 2008-is both literal and substantive generations too late for the algorithmic deployment it would seek to regulate. At the time of going to press, the Personal Data Protection Bill, which was extensively debated but not yet enacted, has at best token provisions related to algorithmic accountability. This regulatory vacuum allows proxy discrimination to thrive, legally uncontestable despite producing outcomes indistinguishable from explicit caste bias.[22]

The Opacity Shield: Technical Complexity as Accountability Barrier

The proxy problem is compounded by what might be termed algorithmic inscrutability-the technical and legal barriers that prevent those harmed by discriminatory systems from understanding, challenging, or remedying that harm. Algorithms deployed in consequential domains-credit scoring, employment screening, risk assessment-typically operate as proprietary black boxes, their internal logic shielded by trade secret protections and technical

[20] Deshpande, K.V., Pan, S. and Foulds, J.R., 2020, July. Mitigating demographic bias in AI-based resume filtering. In *Adjunct publication of the 28th ACM conference on user modeling, adaptation and personalization* (pp. 268-275).
[21] Das, A.K., 2018. European Union's General Data ProtectionRegulation, 2018: A brief overview. *Annals of Library and Information Studies (ALIS)*, *65*(2), pp.139-140.
[22] Kaminski, M.E., 2018. Binary governance: Lessons from the GDPR's approach to algorithmic accountability. *S. Cal. L. Rev.*, *92*, p.1529.

complexity.[23]

All that the applicant may get at best is some vague explanation like "insufficient credit history" or "risk profile does not meet the criteria." The particular features that generated this decision, their relative weights, the training data that generated these weights, and the testing protocols utilized for bias detection-these remain closed. Firms claim legitimate interests in protecting proprietary techniques from rivals, a claim usually accepted by Indian courts. This creates a deep asymmetry: the algorithm has complete information about the applicant while the applicant knows nothing about the algorithm.[24]

Technical complexity magnifies opacity. Even when the algorithmic logic is in theory accessible, comprehension requires deep expertise in machine learning, statistics, and software engineering. To take a well-known example, a Dalit borrower rejected for credit by an algorithmic system suspects that it has discriminated against her on grounds of caste through proxy variables. But how would she contest this discrimination when it is nestled deep within technical processes which she cannot understand? Legality, designed to operate with adversarial contestation of observable decisions by identifiable actors, falters before challenges to opaque systems whose discriminatory effects arise from interactions among thousands of variables.[25]

The irony is deep. India positions itself as the backbone of the world's IT services, providing technical talent to technology firms across the globe. Indian engineers architect, develop, and maintain algorithmic systems deployed across sectors and continents. Yet this technical capacity has not translated into a regulatory infrastructure for algorithmic accountability within India itself. The judiciary possesses limited technical expertise to evaluate claims of algorithmic discrimination. Regulatory bodies lack resources and frameworks to audit deployed systems. Civil society organizations, while documenting exclusionary outcomes, cannot access the algorithmic architecture that produces those outcomes.[26] The result is discrimination that operates in plain sight yet remains effectively immune to legal challenge-

---

[23] Vedder, A. and Naudts, L., 2017. Accountability for the use of algorithms in a big data environment. *International Review of Law, Computers & Technology*, *31*(2), pp.206-224.
[24] Wadden, J.J., 2022. Defining the undefinable: the black box problem in healthcare artificial intelligence. *Journal of Medical Ethics*, *48*(10), pp.764-768.
[25] Kumar, V.A., 2022. Caste discrimination in agricultural credit. *Issue 4 Indian JL & Legal Rsch.*, *4*, p.1.
[26] Von Eschenbach, W.J., 2021. Transparency and the black box problem: Why we do not trust AI. *Philosophy & technology*, *34*(4), pp.1607-1622.

not because it is legal, but because it is illegible.

Conclusion

The myth of algorithmic neutrality abides not because it is true, but because it is useful. It allows governments to automate welfare distribution while disclaiming responsibility for resulting exclusions. It allows corporations to scale discriminatory hiring and lending practices while invoking mathematical objectivity as a shield.[27] It allows caste and class hierarchies to be reproduced through technical means that evade constitutional prohibition.

The evidence presented herein demonstrates that algorithms do not malfunction when they discriminate; they function precisely as systems trained on inequitable data, operating through carefully selected proxies, shielded by deliberate opacity that is inevitably maintained. India is at a crossroads. The speed at which algorithms are being deployed in governance, finance, jobs, and criminal justice far outpaces the construction of regulatory mechanisms that might constrain them. This is less a function of bad timing than of political economy: the same corporate and state actors who deploy these systems benefit from regulatory absence. Yet the technical wherewithal does exist within India to build accountable algorithmic infrastructure.

What remains absent is the political will to mandate transparency, establish meaningful audit mechanisms, extend anti-discrimination law to include proxy variables, and invest in judicial and regulatory capacity to adjudge algorithmic harms. The central question is not whether algorithmic systems can be made fair in principle. It is whether there exists collective commitment to demand fairness from systems designed, whether consciously or not, to perpetuate advantage. The answer will determine whether India's digital future amplifies historical injustice-or finally disrupts it.

[27] Phillips-Brown, M., 2023. Algorithmic neutrality. *arXiv preprint arXiv:2303.05103*.