



Bhagwat Mani Tripathi

Dept of Philosophy,

University of Delhi

H112, Karampura, Delhi

(+91)6393202624

[tripathipratyush01@philosophy.du.ac.in](mailto:tripathipratyush01@philosophy.du.ac.in)

Article ID No. 2801

DOI No. <https://doi.org/10.5281/zenodo.21214664>

## Digital Untouchability: Caste Bias in the Age of Artificial Intelligence

### Abstract:

*Artificial Intelligence (AI) systems dominate access to employment, credit, healthcare, and public services across India and the questions of the algorithmic fairness that fights with one of the globally most entrenched social hierarchy: the caste system. This study examines how caste-based discrimination is reproduced and amplified within Artificial Intelligence Systems installed across Indian digital infrastructure. Drawing analysis of critical algorithmic studies, post-colonial theories and case analysis; anyone can argue that training data has sourced from deep rooted historical social structures encoded in caste biasness and homogenous upper caste composition of India shapes evolution priorities.*

*In this paper, the researchers have tried to analyze three domains- automated hiring platforms, facial recognition systems and credit scoring algorithms; where proxy variables like name, dialectic, geography, and educational institutions act as surrogate for caste identity and generate discriminatory results without any explicit caste markers. The paper further critiques the insufficiency of western fairness theories which fail to account for the intersectional and legally protected & preserved caste in Indian constitutional contexts. We conclude this paper by proposing a theoretical framework for caste-conscious algorithm auditing, goading collaborations between AI researcher, Dalits rights organization and policy makers to be accountable mechanism with perspective of India's unique socio-technical landscapes.*

**Keywords:** Caste discrimination, Algorithmic bias, Proxy variables, Algorithmic auditing and Postcolonial AI

### 1. Introduction:

The global expansion of AI systems has been accompanied by emerging scholarly and activist attention to the algorithmic biases, fairness, and accountability. However, the most of this discourse has left in western epistemologies and contexts; addressing categories of discrimination identified within liberal democratic frameworks mainly in the form of race, gender, and disability. The AI governance of India a critical gap: about complete absence of caste from algorithmic fairness discussion. Despite the fact, the caste discrimination being recognized constitutionally, legally prohibited, and materially consequential for Schedule Caste (SC) and Schedule Tribe (ST) citizens along with other marginalized and vulnerable communities. As AI



systems adjudicate access to the financial resources, economic opportunity, healthcare, and state & central services. A biased hiring algorithm affects thousands of job seekers. The discriminatory facial recognition systems enable surveillance and exclusion at mass scale. A flawed credit scoring algorithm denies capital to the particular community. Yet, besides these systematic dangers; the sophisticated question of caste-based algorithmic discrimination still largely unexamined in the context of Indian AI policy, research, and corporate practices. This absence is not only reflected any accident but it depicts *algorithmic erasure*<sup>1</sup>. This erasure is particularly insidious because it works through technical languages and neutral processes and produces social hierarchies that have been generated within the algorithmic system. When the hiring through these algorithmic systems eliminates candidates from a particular geographical area or educational background, it is engaging in the caste discrimination, when facial recognition system not able to recognize dark-skinned-face- it harms Dalits and Adivasis community; and when a credit algorithm denies loan to borrower without formal employment histories, they exclude the working poor.

This paper progresses as follows: section 2 reflects essential context on caste in India, constitutional recognition and relationship to data & technology; section 3 investigates three critical domains where discrimination of algorithm operates; section 4 provides critique to western fairness frameworks and their insufficiency for context of India; section 5 proposes a caste-conscious algorithmic auditing framework. We conclude this paper by arguing that post-colonial AI approaches should be the perspective of historically marginalized communities and caste and also provide perspective for non-western structure.

## 2. Literature Review:

### 2.1 Caste as a Social Organization Systems:

Caste system in India is an ancient and hierarchical system of social organization that progressively shapes material life, opportunity, and dignity across contemporary India. The philosophical concepts of ritual purity and pollution; caste gives birth to endogamous groups (Jatis) that organized into the wider perspective of *Varna*<sup>2</sup> that have historically determined

---

<sup>1</sup> the systematic exclusion of particular caste from the technical discussion; even caste remains materially embedded in the data training and design choices that constitutes these systems.( Bjorn Aslak Juliussen, Jon Petter Rui, Dag Johansen, Algorithms that forget: Machine unlearning and the right to erasure, Volume 51,2023, <https://doi.org/10.1016/j.clsr.2023.105885> )

<sup>2</sup> refers to a fourfold social hierarchy described in Brahmanical texts, hymn, called the *Purushasukta*, portrays the Brahmin (priest), the Kshatriya (noble), the Vaishya (commoner), and the Shudra (servant) formed in the cosmogonic sacrifice from the mouth, arms, thighs, and feet of the Lord Vishnu as stated in the Rig Ved. birth, <https://www.britannica.com/topic/varna-Hinduism>



occupational status (Ambedkar, 1944<sup>3</sup>; Dirks, 2002<sup>4</sup>). Therefore, caste is not only cultural and religious context but also it operates as a well-developed system of exploitation, legal discrimination, and epistemological coordination.

The Constitution of India (1950)<sup>5</sup> acknowledged the caste discrimination as a cornerstone injustice that need *affirmative actions*<sup>6</sup>. The Article 15<sup>7</sup> and Article 17<sup>8</sup> of Indian Constitution explicitly prohibited discrimination based on caste. The Article 16<sup>9</sup> and Article 46<sup>10</sup> mandate reservation for SCs and STs- which is a legal framework that has no resemblance in western democracies (Thoart, 2009)<sup>11</sup>. This perspective of the Indian Constitution generates a unique category of discrimination that demands specific policy and legal demand. However, India's digital infra expands this constitutional commitment to caste justice that faces challenges from automated algorithmic systems.

## 2.2 Caste, Data, and Technology Sector:

The practices and composition of the technology sector in India shape how caste dynamism is embedded in AI-systems. India's global dominance in IT and software engineering; the sector is dominated by upper-caste professionals. Research (Thorat and Negi, 2022)<sup>12</sup> finds that among top tech companies in India, upper caste comprises 45%-60% while SCs comprise 8%-10% which is below constitution reservation 15% for federal employment. This skewed composition in tech companies portrays both educational marginalization and discriminatory practices.

The upper-caste composition matters for the evolution and development of AI. The individual designing and displaying algorithms are disproportionate from historically dominated castes carrying many biases. The data that has been used to train these algorithms is referenced from the historical records that reflect centuries old caste discrimination and their exclusion. When

<sup>3</sup> Ambedkar, B.R. (1944). Annihilation of Caste. [Annihilation of caste Dr.B.r.ambedkar.pdf](#)

<sup>4</sup> Dirks, N.B. (2001). Castes of Mind: Colonialism and the Making of Modern India. Princeton University Press.

<sup>5</sup> The constitution of India (1950), <https://www.legislative.gov.in/constitution-of-india>

<sup>6</sup> affirmative action in India is a constitutional designed policy to address historical injustices against Scheduled Castes (SC), Scheduled Tribes (ST), and Other Backward Classes (OBC).

<sup>7</sup> stated that Prohibition of discrimination on grounds of religion, race, caste, sex or place of birth, The constitution of India (1950), <https://www.legislative.gov.in/constitution-of-india>

<sup>8</sup> abolition of untouchability in all forms , <https://www.legislative.gov.in/constitution-of-india>

<sup>9</sup> equality of opportunity in matters of public employment birth, <https://www.legislative.gov.in/constitution-of-india>

<sup>10</sup> upliftment of educational and economic interests of Scheduled Castes, Scheduled Tribes and other weaker sections, <https://www.legislative.gov.in/constitution-of-india>

<sup>11</sup> Thorat, S. (2009). Caste, Category and Exclusion in Modern India." Economic and Political Weekly, 44(45), 51-59

<sup>12</sup> Thorat, S., & Negi, R. (2022). Caste Discrimination and Occupational Choice: Empirical Evidence from Indian Engineering Firms.



these data are processed into a machine learning model- the algorithms amplify the historical patterns of discrimination.

### 3. Emergence of Problem:

#### 3.1 Automated Hiring Platforms:

The recruitment algorithm represents a crucial picture where caste-based discrimination becomes technologically mediated. The recruitment platforms like LinkedIn filters thousands of applicants for a particular job. A study conducted by Narayan and colleagues found that resumes associated with Schedule Caste & Schedule Tribe received lower interviews than identical names of upper caste provided skills and experiences are the same.

What makes these algorithmic discriminations unique; is its function through the proxy variables. The algorithmic recruitment trained on historical hiring associated between particular names and educational institutions. The educational institutions serve particular proxy prestigious institutions like IITs and Delhi University enrolled disproportionately upper-caste students both due to caste-based access and reservation policies. Algorithms trained on historical patterns thus make sure that graduating from IITs have increased employment probability. Similarly, geographical coefficients work as caste proxy. There are certain regions in India that have different geographical compositions; the coastal area of Maharashtra and Karnataka have the higher constituents of upper-caste, while the agricultural region of Uttar Pradesh and Bihar have larger SC/ST population. When these recruitment algorithms use geographical variables they create geography-based caste discrimination. A candidate from the marginalized area of UP and Bihar faces lower chances of employment not because of their skills and capability but the algorithm trained a unique association between rural origin and lower employment outcomes. Eventually, dialectic and language serve an additional proxy. The accent carried by caste associations in India reflects both educational patterns and cultural variables. The recruitment algorithms incorporate these linguistic patterns and these systems may penalize candidates whose linguistic patterns are edged by regional, educational and specific caste context.

#### 3.2 Facial Recognition System:

This system has evolved as a critical infra for identity verification, surveillance, and security purposes across both state and private institutions in India. *The Aadhar*<sup>13</sup> programme maintain the digital and biometric data of 1.3 billion Indians which interface with facial recognition system for authentication. But the problem is well documented- the facial recognition system has disparity across skin tone, with darker skinned higher error rate (Bualamwini and Gebru, 2018).

---

<sup>13</sup>Aadhar is the world's largest biometric identity system which is maintained by Unique Identification Authority of India; there are approximately 134 crores live Aadhar holders. (Press Information Bureau posted on 18 March, 2026) , <https://www.pib.gov.in/PressReleasePage.aspx?PRID=2241778&reg=3&lang=2>



This bias is often discussed in term of 'skin tone' in Indian context; skin tone carries caste hierarchies are historically articulated through racial ideologies with belief that darker skin with lower caste and *ritual pollution*<sup>14</sup> (Narayan, 2019)<sup>15</sup>.

Indian facial recognition systems are often trained on data that has been dominated by upper-caste faces which reflect both tech sector caste composition and deliberate choices about face constitutes representation data. When this system fails to recognize the darker-skin, these systems exclude Dalits, Adivasis, and other vulnerable & marginalized communities from digital authentication.

### 3.3 Credit Scoring Algorithms:

Financial inclusion insists on a critical context for algorithmic discrimination. These algorithms determine access to microfinance, small loans, and formal credits for millions of Indians to escape from the trap of poverty. But many SC/ST and other vulnerable communities historically engaged in occupations like leatherwork, scavenging, and agricultural labor, when algorithms use occupational classification as a factor of risk and sideline those occupations which resemble the historically marginalized community. Educational qualification acts as another proxy. The credit rating algorithm many times uses education as a credit indicator. Since such communities' face education barriers rooted in both poverty and social discrimination. The lower educational qualification acts as higher credit risk. Let us understand this with an analogy: a Dalit woman, who is not able to complete her secondary education due to economic & social conditions, the algorithm encodes structural inequality as individual risk. There is also a crucial factor in the form of residential pattern or ownership that acts as a caste proxy. The historical discrimination confined SC/ST to limited asset ownership and when algorithms assess creditworthiness based on ownership as collateral, they are disadvantaged because they have limited assets.

### 4. Research Gap:

Despite scholarly attention to the biasness of algorithms in India. The existing research is largely derived from the western framework that focuses on gender, race, and disability. The very limited work on caste and technology addressed digital access and caste identity in social media context but did not systematically observe how automated decisions making encoded caste discrimination. There is no established framework for caste consciousness algorithmic auditing. Indian AI policy corporate AI ethics guidelines similarly and fail to acknowledge caste as subject of high risk. This paper addresses this gap in documentary evidence of caste-based algorithmic

---

<sup>14</sup> Religious or spiritual contamination caused by actions and life events like death or menstruation.

<sup>15</sup> Narayan, B. (2019). *Khaki and the Saffron: Caste, Class and Communalism*. Aleph Book Company.



discrimination and also by proposing a contextually grounded framework for caste conscious algorithmic auditing.

### 5. Rational of Study:

The rationale of this study is three dimensional: constitutional, empirical, and epistemic.

Constitutionally, India made a legal commitment to eliminate caste discrimination through Articles like 15, 16, 17, and 46. The development of AI-bridged decision making across the domains like employment, credit, and health depicts direct threat and portrays proxy discrimination. The absence of AI governance is not merely an academic flaw but failure of constitutional duty. Empirically, the scale of harm to SC and ST communities revolving around 25% (census of India,2011). As the algorithmic system increasingly mediated the availability of opportunities that earlier guaranteed through human processes. The need for legal & technical accountability becomes urgent. Epistemically, this study will examine Care & AI not only in a specific Indian context but also advances the argument for culturally situated approaches to AI-Ethics and governance.

### 6. Methodology:

This study adopts a qualitative, interdisciplinary methodology drawing on legal analysis, and critical data study. This approach combines analysis of the Indian constitution and legislative framework which review literature on caste demography in the tech sector, algorithmic bias, facial recognition, and credit scoring.

The analysis is guided by a post-colonial AI framework that centered the lived experience of the historically marginalized communities. The empirical claim is drawn from secondary sources including peer-review research, government data, and documented studies of caste-based discrimination in the Indian labour market.

### 7. Findings:

In this section we will propose an auditing framework finding that investigates these inadequacies and this algorithmic principle consists of five compartments-

#### 7.1 Caste Aware Dalit Audit:

The process of auditing should examine whether proxy variables, trained data and model outputs generate caste-based discriminations. This needs- First, *proxy variables analysis*; the process of audit where variables serve as caste proxies like name, location, education etc. It maps how these variables co-exist with historical caste discrimination and second, *contamination*



of *historical data*; it interrogates whether training data sourced from historically structured caste discrimination like police arrest data, credit histories etc.

### 7.2 Community Based Assessment:

It proposes that caste-conscious auditing should not be purely a technical audit but should involve affected, marginalized, and vulnerable communities. It audits the mechanism as follows: First, *participatory auditing*; it consists of SC/ST communities and Dalit right organizations & examining algorithmic systems and its impact. Second, *accountability mechanism*; it insists that vulnerable communities should have their voices in shaping what algorithmic justice means in their context.

### 7.3 Intersectionality Analysis:

The discrimination based on caste intersects with gender, religion, and other dimensions needed for marginalization, it can be understood as follows: First, *gender caste intersectionality*; Dalit and vulnerable women face collective discrimination from gender as well as caste hierarchy and henceforth, algorithmic auditing is substituted to their intersectionality. Second, *religious caste intersectionality*; in the process of hiring and credit scoring, the discrimination based on caste and religion are often coupled together.

### 7.4 Algorithmic Transparency on Caste:

The organizations and institutions that are deeply influenced by algorithms should publicly reveal as follows- first, *training data composition*; it insists how representatives are training data along the caste group? Second, *proxy variables*; it will reflect which variables function as caste proxies? Third, remediation measures, what steps should be taken to mitigate the caste-based discrimination?

### 7.5 Multi-Stakeholder Governance:

The efficient auditing needs coordination among AI researchers who develop technical methods to identify caste variables proxies and disparities impact in the context of caste. The Dalit rights organizations should also reframe what algorithmic justice means from the glasses of marginalized and vulnerable communities. The policymakers should also create a legal framework for caste impact assessment.

### 7.6 Impacts for Postcolonial AI:

This investigation finds an emerging post-colonial AI scholarship by- first, *centering non-western fairness approaches*; despite applying western fairness principle universally, the post-colonial AI



should evolve approaches that have specific social structure and legal frameworks. Second, *interrogating AI colonialism*, as machine systems expand globally there is always the risk of AI colonialism that imposes western rationality and governance to non-western contexts. And third, *addressing structural inequality*; other than treating inequality as a problem that needs fixing, post-colonial AI should investigate how algorithmic perpetuation rooted in colonialism, caste, and power asymmetries.

## 8. Conclusion:

AI systems have dominated access to the basic life requisite across India. However, despite the constitutional right for caste discrimination these algorithmic systems amplify caste-based discrimination through proxy variables and data biases. To address these crises, there is a need to move beyond purely technical approaches to algorithmic fairness. It requires to consider caste as a fundamental category that needs attention in algorithmic design, deployment as well as auditing. Most prominently, it requires recognition that algorithms are not neutral but a social technology that codes and amplifies the existing power structures.

## References:

- Ambedkar, B.R. (1944). Annihilation of Caste. [Annihilation of caste Dr.B.r.ambedkar.pdf](#)
- Buolamwini, J., & Gebru, T. (2018). Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification. Conference on Fairness, Accountability and Transparency.
- Census of India. (2011). Scheduled Caste and Scheduled Tribe Population. Office of the Registrar General and Census Commissioner.
- Bjorn Aslak Juliussen, Jon Petter Rui, Dag Johansen, Algorithms that forget: Machine unlearning and the right to erasure, Volume 51, 2023, <https://doi.org/10.1016/j.clsr.2023.105885> )
- Dirks, N.B. (2001). Castes of Mind: Colonialism and the Making of Modern India. Princeton University Press.
- Narayan, B. (2019). Khaki and the Saffron: Caste, Class and Communalism. Aleph Book Company.
- Thorat, S. (2009). Caste, Category and Exclusion in Modern India." Economic and Political Weekly, 44(45), 51-59.
- Thorat, S., & Negi, R. (2022). Caste Discrimination and Occupational Choice: Empirical Evidence from Indian Engineering Firms.
- Zuboff, S. (2019). The Age of Surveillance Capitalism. PublicAffairs.
- The constitution of India (1950), <https://www.legislative.gov.in/constitution-of-india>