International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

Algorithmic Discrimination and Human Rights in India: From Constitutional Equality to Responsible AI Governance

Ankul Prajapati
Law
National law Institute University, Bhopal

Shubham Sharma
Law
National law Institute University, Bhopal

Adarsh Sahu
Law
National law Institute University, Bhopal

Ashmit Chauhan
Law
National law Institute University, Bhopal

Abstract

Artificial intelligence is increasingly being used to make or support decisions that determine access to employment, credit, insurance, healthcare, education, public benefits, policing and other socially significant opportunities. The promise of automated decision making rests partly on the assumption that computational systems are more neutral than human decision makers. That assumption becomes difficult to sustain when the data used to train a system reflects historical inequality, when apparently neutral variables operate as proxies for protected characteristics, when datasets inadequately represent disadvantaged groups, or when automated predictions influence the very conduct that later appears to confirm them. Algorithmic discrimination therefore raises a question that is simultaneously technological, constitutional and human rights based. This paper examines that problem through the framework of equality, dignity, privacy, due process and effective remedy, with particular attention to Articles 14, 15, 16 and 21 of the Constitution of India. It analyses the Digital Personal Data Protection Act 2023, the Digital Personal Data Protection Rules 2025 and the India AI Governance Guidelines 2025, while comparing these developments with the European Union Artificial Intelligence Act, international human rights standards and selected regulatory approaches in the United States, the United Kingdom and Canada. The paper argues that the most important legal shift is from asking whether an algorithm was intentionally discriminatory to asking whether its design, data, deployment and institutional use produce unjustified disadvantage and whether adequate safeguards exist. It proposes a layered Indian governance model combining constitutional review, targeted statutory duties, impact assessments, meaningful explanation, human review, periodic fairness testing, audit powers, record keeping and effective remedies. The objective is not to prevent the use of artificial intelligence, but to ensure that technological delegation does not become a means of avoiding legal responsibility.

Keywords: artificial intelligence, algorithmic discrimination, human rights, constitutional equality, Article 14, substantive equality, indirect discrimination, privacy, explainability, responsible AI governance

1. INTRODUCTION

Artificial intelligence has moved from being a specialised technological field to becoming part of ordinary institutional decision making. Employers use automated tools to screen applications and rank candidates. Banks and financial technology companies use statistical models to evaluate creditworthiness and detect fraud. Hospitals increasingly use predictive systems to assist diagnosis and allocation of resources. Police forces have experimented with facial recognition, predictive policing and risk assessment tools. Public authorities use automated systems for welfare

administration, identity verification and the distribution of public services. Online platforms use recommendation, ranking and advertising systems to determine what information individuals see and which opportunities are presented to them. The growing use of such systems creates an important legal paradox. Artificial intelligence is frequently presented as a method of reducing human subjectivity. A machine, it is said, does not discriminate because it has no prejudice, emotion or personal motive. This assumption is legally and technologically unsound. An artificial intelligence system does not make decisions in a social vacuum. It is developed from data, selected according to institutional objectives, trained through particular methods, evaluated against chosen metrics and deployed within organisations that have their own incentives and histories. Every one of these stages can introduce or reproduce bias. The problem therefore cannot be reduced to a defective line of code. Historical inequality can become embedded in the data from which a model learns. A variable that appears neutral can operate as a proxy for race, caste, gender, disability or class. A system optimised for predictive accuracy may perform well on an aggregate population while producing substantially worse outcomes for a minority group. Human users may also give excessive weight to algorithmic recommendations because they perceive a computer generated result as objective. In some settings the resulting decision is then used as new data, creating a feedback loop that reinforces the original pattern.[1]

The human rights implications are substantial. The right to equality and non discrimination is implicated when an automated system produces unequal treatment or disproportionate disadvantage. Privacy is implicated when extensive personal information is collected, inferred and combined to predict behaviour or eligibility. Dignity and autonomy are implicated when individuals are classified without meaningful knowledge or the possibility of contesting the classification. Due process is implicated when an adverse decision cannot be understood or challenged. The right to an effective remedy is implicated when responsibility is divided among multiple technological and institutional actors. The central legal insight of this paper is that algorithmic discrimination should be treated as a problem of institutional responsibility rather than merely technical bias. The relevant question is not whether a computer intended to discriminate. The relevant questions are whether the system creates a prohibited or unjustified disparity, whether the use of the system can be justified, whether adequate safeguards were adopted, who controlled the relevant decision, whether the affected person was informed, and whether an effective remedy exists. This approach is particularly important in India because constitutional equality has increasingly moved away from a narrow focus on formal classification towards substantive equality and indirect discrimination.[2]

2. THE NATURE AND MECHANISMS OF ALGORITHMIC DISCRIMINATION

Algorithmic discrimination may be understood as unequal or unjustified treatment produced or materially influenced by an algorithmic system, where the outcome disadvantages an individual or group on a protected or socially significant ground. The concept covers more than direct use of a protected characteristic. A system can discriminate even when it never receives a field labelled caste, gender or race. Solon Barocas and Andrew Selbst demonstrated the importance of this point in their analysis of big data and disparate impact. Data driven systems can use apparently neutral variables that are correlated with protected characteristics. A residential address, educational institution, employment history, purchasing pattern or linguistic feature may function as a proxy for a protected or historically disadvantaged status. Removing the protected attribute from the dataset therefore does not necessarily remove discrimination. Algorithmic discrimination may occur at several stages. It may begin with the selection of the problem itself. An institution may decide to predict employee retention rather than evaluate workplace conditions, or to predict criminal risk rather than investigate individual circumstances. It may arise during data collection if the dataset excludes particular communities. It may arise during labelling if human annotators encode stereotypes into the training data. It may arise during model construction if the optimisation target ignores group disparities. It may arise during deployment if the system is used outside the environment in which it was tested. Finally, it may arise through human interaction with the system when decision makers blindly accept automated recommendations. Direct algorithmic discrimination is comparatively easy to

1 UNESCO, Recommendation on the Ethics of Artificial Intelligence (adopted 23 November 2021) <https://unesdoc.unesco.org/ark:/48223/pf0000381137> accessed 6 October 2026.
2Universal Declaration of Human Rights 1948, arts 1, 2 and 7; International Covenant on Civil and Political Rights 1966, arts 2 and 26.

understand. If a system explicitly assigns lower scores to candidates because they are women, or rejects an applicant because the applicant belongs to a particular caste, the use of the protected characteristic is apparent.[3]

Indirect discrimination is more difficult and more important. A facially neutral rule can place a protected group at a particular disadvantage even though the rule does not expressly refer to that group. The Indian Supreme Court's decision in Lt Col Nitisha v Union of India is especially relevant. The Court recognised indirect discrimination as part of substantive equality under Articles 14 and 15. It emphasised that formally neutral practices can perpetuate existing structures of disadvantage. This reasoning translates naturally to algorithmic systems. An automated hiring system that rewards uninterrupted employment history may appear neutral. Yet the criterion may disadvantage women who have taken career breaks because of unequal caregiving responsibilities. A credit model that rewards stable property ownership may appear neutral but may reproduce historical socioeconomic inequality. A predictive policing system trained on historical arrest data may treat neighbourhood arrest rates as evidence of future criminality even when the underlying arrest data reflects unequal policing. The significance of Nitisha is therefore greater than its immediate facts. It provides a constitutional vocabulary for analysing algorithmic discrimination without requiring proof that a developer intentionally encoded prejudice. If the effect of a formally neutral algorithm is to reproduce a structural disadvantage, the constitutional enquiry can focus on the relationship between the rule, its operation and the resulting inequality. Historical bias occurs when the data used by an artificial intelligence system reflects discriminatory social conditions. A model trained on past hiring decisions may learn that men were more frequently selected for senior technical roles. A system trained on historical lending decisions may learn that borrowers from particular communities received fewer loans. A predictive policing model may learn from historical arrest patterns that particular neighbourhoods are associated with crime.[4]

The resulting model may be statistically faithful to the past while being legally and socially unacceptable in the present. This reveals a basic limitation of the claim that artificial intelligence is objective. If the past is unequal, statistical reproduction of the past can reproduce inequality. The Amazon hiring example illustrates this problem. Amazon reportedly abandoned an experimental recruitment tool after discovering that it penalised resumes containing references associated with women. The system had learned from historical recruitment data generated in a male dominated technological environment. The algorithm did not need an explicit instruction to discriminate. The historical record supplied the discriminatory pattern. Sampling bias occurs when the training data does not adequately represent the population on which the system will operate. The problem is particularly visible in facial recognition. The Gender Shades study by Joy Buolamwini and Timnit Gebru demonstrated significant disparities in commercial gender classification systems, particularly for darker skinned women. The importance of this research lies not only in the error rates but in what the error distribution reveals about technological design. A system that performs extremely well for a majority group and poorly for a minority group cannot be described as equally reliable merely because its overall accuracy is high. In high consequence settings, the relevant question is whether errors are distributed in a way that produces unjustified harm. The same problem can arise in India through language, region, disability, caste and socioeconomic status. A model developed primarily from English language data may perform less effectively for speakers of Indian languages. A health model trained on private hospital data may not perform equally for rural populations. A hiring model developed using formal sector employment data may systematically undervalue informal work histories.[5]

Proxy discrimination is particularly significant because modern systems can infer characteristics that they were not explicitly given. A model may not receive caste as an input, but variables such as surname, school, locality, language, social network or educational history can correlate with caste. A model may not receive gender, but career history, job title, employment gap or online behaviour may correlate with gender. The legal significance of proxies is that removing protected attributes from the data is not enough. A system can become discriminatory through combinations of individually neutral variables. This makes traditional compliance strategies based on checking whether a prohibited field has been included inadequate. In India the problem is especially acute because social identities are embedded in geography, education, occupation, language and family networks. A technology that appears caste neutral at the surface level may nevertheless reproduce caste hierarchy through proxy variables. This is one reason why algorithmic auditing

3Solon Barocas and Andrew D Selbst (n 1).
4Jeffrey Dastin, 'Amazon Scraps Secret AI Recruiting Tool That Showed Bias Against Women' Reuters (10 October 2018).
5Anupam Chander, 'The Racist Algorithm?' (2017) 115 Michigan Law Review 1023.

in India must be sensitive to local patterns of social stratification rather than importing fairness metrics developed for a different social environment. Algorithmic discrimination can become self reinforcing. Suppose a predictive policing system identifies a neighbourhood as high risk. Police resources are then concentrated in that area. Increased police presence produces more recorded incidents and arrests. Those records are fed back into the system, which then interprets the higher recorded crime rate as evidence that the neighbourhood was correctly identified as high risk. The algorithm has therefore become part of the causal process generating the data on which it relies. The model is no longer simply predicting an independent reality. It is helping to create the reality that later appears to validate the model. This problem has direct implications for equality and liberty. It also complicates judicial review because the institution may claim that the algorithm is accurate while ignoring the fact that the algorithm helped produce the very evidence used to establish its accuracy. A human decision maker may discriminate. A machine may produce discriminatory results. A third possibility is that the human decision maker may become more willing to accept discrimination because it has been expressed as a numerical score. This phenomenon is sometimes called automation bias. It is particularly dangerous in policing, healthcare, credit and employment because a person may treat an automated recommendation as more objective than a human judgement. Human oversight therefore cannot mean merely placing a person somewhere at the end of the process. Meaningful oversight requires the person to understand the system sufficiently, have access to relevant information, possess authority to depart from the recommendation and actually exercise independent judgement.[67]

3. ALGORITHMIC DISCRIMINATION, HUMAN RIGHTS AND CONSTITUTIONAL EQUALITY

The Universal Declaration of Human Rights establishes equality in dignity and rights and equal protection of the law. The International Covenant on Civil and Political Rights prohibits discrimination and guarantees equal protection. The International Covenant on Economic, Social and Cultural Rights similarly requires rights to be exercised without discrimination. These guarantees are technologically neutral. They do not cease to apply because a decision is produced by software. The Committee on the Elimination of Racial Discrimination has specifically recognised the risk that artificial intelligence and algorithmic profiling can reproduce and reinforce racial bias. It has recommended legal, administrative and technical safeguards, regular monitoring, independent auditing and remedies where algorithmic profiling creates unacceptable human rights risks. The UNESCO Recommendation on the Ethics of Artificial Intelligence takes a similarly broad approach. It identifies human dignity, human rights, fairness, non discrimination, transparency and human oversight as central principles across the entire AI lifecycle. Its importance lies in recognising that ethical risk begins before deployment and continues through monitoring, evaluation and termination. Discrimination is not only a problem of unequal distribution. It can also undermine dignity by reducing individuals to statistical categories. A person rejected for employment, denied credit or treated as a security risk may be unable to understand what aspect of their life produced the result. The dignity dimension becomes stronger when the system infers sensitive characteristics without the person's knowledge. The individual may become the object of prediction rather than an agent capable of understanding and contesting the decision. This concern connects algorithmic governance with the constitutional recognition of privacy as a component of dignity. In Justice K S Puttaswamy v Union of India, the Supreme Court recognised privacy as a fundamental right under Article 21 and connected privacy with dignity and autonomy. Algorithmic systems intensify the importance of this principle because modern models can generate new information from existing data.[8]

The legal concern is therefore not limited to unauthorised disclosure. Even lawfully collected information may become intrusive when aggregated, inferred and used for a new purpose. A person's location history, purchasing behaviour, education, employment history and online activity may individually appear harmless but collectively become a basis for predicting creditworthiness, employability, health risk or political preference. The relationship between privacy and discrimination is also reciprocal. Extensive data collection can facilitate more accurate personalisation, but it can also enable more precise exclusion. At the same time, removing sensitive attributes to protect privacy can make it harder to

6 National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework 1.0 (2023) <https://www.nist.gov/itl/ai-risk-management-framework> accessed 6 October 2026.
7Solon Barocas and Andrew D Selbst (n 1).
8Universal Declaration of Human Rights 1948, arts 1, 2 and 7; International Covenant on Civil and Political Rights 1966, arts 2 and 26; Committee on the Elimination of Racial Discrimination, General Recommendation No 36 on Preventing and Combating Racial Profiling by Law Enforcement Officials (2020).

audit whether a model discriminates. Effective governance must therefore permit controlled access to sensitive data for legitimate fairness testing while preventing unnecessary use of that data in decision making. Where an algorithmic decision produces a serious legal or practical consequence, the affected person must have a meaningful opportunity to understand and challenge the decision. A formal appeal is inadequate if the institution cannot identify the basis of the decision. The European Union Artificial Intelligence Act provides a useful example. Article 86 creates a right to obtain clear and meaningful explanations of the role of a high risk AI system and the main elements of the decision where the statutory conditions are met. The United States approach to credit decisions provides another example. The Consumer Financial Protection Bureau has stated that creditors cannot avoid adverse action explanation requirements merely because they use complex algorithms. These developments demonstrate an important principle: technological complexity cannot become a defence against legal accountability. Article 14 is the principal constitutional foundation for addressing algorithmic discrimination in India. Its guarantee of equality before the law and equal protection of the laws is not limited to traditional classifications. Indian constitutional doctrine has developed towards a substantive understanding of equality that examines whether legal and institutional practices perpetuate disadvantage.[9]

An algorithmic decision made by a public authority can therefore be challenged where the decision is arbitrary, irrational, disproportionate or discriminatory. The fact that a public authority relied on a vendor's software should not alter the constitutional obligation. The State cannot outsource its constitutional duties to a private technological system. This becomes particularly important where automated systems are used for welfare eligibility, policing, immigration, public employment or access to government services. In these contexts the Constitution governs not only the final decision but the institutional process through which the decision is reached. Article 15 prohibits discrimination on specified grounds and permits affirmative measures for disadvantaged groups. Article 16 guarantees equality of opportunity in public employment and permits constitutionally recognised forms of affirmative action. Algorithmic systems used in recruitment therefore raise two separate issues. First, the system must not unlawfully disadvantage candidates because of protected characteristics. Second, in appropriate public sector contexts, the system must not erase legally required affirmative action by treating historical inequality as irrelevant noise. This point is important for India because a simplistic definition of algorithmic fairness could conflict with substantive equality. Treating every group identically is not always constitutionally sufficient. A system can be formally neutral while reproducing a historically unequal starting position. The constitutional concept of equality may require the opposite: attention to structural disadvantage and, where legally authorised, measures that counteract it. Lt Col Nitisha v Union of India is particularly relevant because the Supreme Court recognised indirect discrimination and connected it to substantive equality. The judgment rejected the assumption that discriminatory intent is always necessary. It examined how apparently neutral criteria could interact with structural conditions to disadvantage women. Algorithmic systems create precisely this type of legal problem. A developer may not intend to discriminate. The employer may not intend to discriminate. Yet the system can apply a neutral criterion that disproportionately disadvantages a protected group because the criterion is built upon unequal historical conditions. The decision therefore supports a shift in legal analysis from intention to impact, justification and structural context. At the same time, Nitisha does not solve every problem. Indian indirect discrimination doctrine remains under development, particularly regarding proof of causation, appropriate comparators and justification. Those uncertainties become more difficult when the causal chain includes data, model design, vendor decisions and institutional deployment. Article 21 provides the constitutional foundation for challenging opaque and intrusive algorithmic systems that affect personal autonomy and dignity. Puttaswamy established privacy as a fundamental right and rejected a narrow understanding of privacy as merely secrecy.[10]

4. INDIA'S DATA PROTECTION AND AI GOVERNANCE FRAMEWORK

The Digital Personal Data Protection Act 2023 is important to algorithmic governance because many artificial intelligence systems depend on personal data. The Act regulates the processing of digital personal data and establishes duties for data fiduciaries, rights for data principals and institutional arrangements for enforcement. However, the Act should not be described as an artificial intelligence discrimination law. It does not create a comprehensive right against algorithmic discrimination comparable to a dedicated equality statute. Its primary concern is the processing of digital

9Justice K S Puttaswamy (Retd) v Union of India (2017) 10 SCC 1.
10Lilian Edwards and Michael Veale, 'Slave to the Algorithm? Why a Right to an Explanation Matters' (2017) 16 Duke Law and Technology Review 18.

personal data. This distinction matters. A model can discriminate without using sensitive personal data. Conversely, a model may process personal data lawfully and still produce an unfair outcome. Data protection and equality regulation therefore overlap but cannot substitute for one another. The legal landscape changed substantially with the notification of the Digital Personal Data Protection Rules 2025 and the phased commencement of the Act in November 2025. The commencement framework is staggered, with some institutional and definitional provisions taking effect immediately and core compliance provisions being brought into force later. The Rules likewise contain staged commencement dates. This phased approach has two consequences for algorithmic governance. First, organisations must distinguish between the law as enacted and the provisions that are presently operative. Second, India is moving towards a more operational data governance system even without creating a general AI statute. For algorithmic discrimination, the most relevant principles include purpose limitation, notice, security, rights of data principals and accountability of data fiduciaries. Yet these principles still do not answer the central equality question: what happens when the processing is lawful but the resulting decision is discriminatory? An individual may exercise a right to access or correct personal data and still be unable to challenge an algorithmic model. The data may be accurate. The discriminatory outcome may arise from the model's statistical relationships rather than from an incorrect data point.[11]

Suppose a bank accurately records an applicant's address, occupation, education and income. If the model uses these variables to produce a lower credit score for applicants from a historically disadvantaged locality, correcting the underlying data will not solve the problem. The issue is the relationship between variables and outcome. A comprehensive governance framework must therefore distinguish three questions: whether the data was lawfully processed, whether the data is accurate, and whether the resulting decision is fair and lawful. These are separate questions. India's AI Governance Guidelines 2025 represent a significant development because they move beyond general statements about responsible AI and establish a proposed governance architecture. The Guidelines are based on seven principles: trust is the foundation, people first, innovation over restraint, fairness and equity, accountability, understandable by design, and safety, resilience and sustainability. The Guidelines deliberately reject a one size fits all regulatory model. They favour a risk based, technology neutral and sector oriented approach. Existing regulators are expected to play a central role, while proposed institutions such as the AI Governance Group, the Technology and Policy Expert Committee and the AI Safety Institute provide coordination and technical capacity. This model is sensible in one respect. Artificial intelligence is used differently in banking, healthcare, employment, policing and entertainment. A single regulator may lack the expertise to supervise all of these areas. Yet the approach also creates a risk of fragmented accountability. A person affected by an algorithm may not know whether to approach a sector regulator, data protection authority, consumer forum, equality body or court. The Guidelines expressly state that AI systems should be designed and tested to ensure fair and unbiased outcomes and should not discriminate against marginalised communities. This is important because it recognises discrimination as a governance risk rather than merely an ethical concern.[12]

The Guidelines also recognise that additional obligations may be necessary for sensitive applications or vulnerable groups. This supports a proportionate model in which not every algorithm requires the same degree of scrutiny. However, principles alone cannot create an enforceable right. If fairness remains voluntary, a person harmed by an algorithm may still have no direct legal remedy. The long term challenge is therefore to determine which principles should remain guidance and which should become enforceable duties. The Guidelines call for explanations and disclosures that help users and regulators understand how AI systems work, what their outputs mean and what outcomes are intended. This is preferable to a simplistic demand for complete disclosure of source code. Source code transparency is not always necessary to provide meaningful accountability. A person affected by a credit decision does not necessarily need thousands of lines of code. The person needs to know the significant factors that affected the outcome, whether the system was used, who made the final decision and how the decision can be challenged. This distinction between technical transparency and decision transparency should inform Indian regulation. Trade secrets and cybersecurity interests can justify limits on disclosure of technical information, but they should not justify secrecy about the existence of an automated decision or the principal reasons for an adverse outcome.[13]

11Digital Personal Data Protection Act 2023, ss 4, 6, 7, 8 and 10.
12 Ministry of Electronics and Information Technology, India AI Governance Guidelines (5 November 2025) <https://www.meity.gov.in/static/uploads/2025/11/5376af07c07efc41e0e0c8d6bfc2e4c5.pdf> accessed 6 October 2026.
13Information Technology Rules 2021, as amended in 2026; Information Technology Act 2000.

5. INTERNATIONAL DEVELOPMENTS AND LESSONS FOR INDIA

UNESCO's Recommendation on the Ethics of Artificial Intelligence places human dignity and human rights at the centre of AI governance. It recognises fairness and non discrimination, transparency, human oversight and accountability across the AI lifecycle. The Recommendation is not a treaty and does not create directly enforceable obligations. Its significance is normative. It demonstrates an emerging international consensus that AI governance cannot be separated from human rights law. The European Union Artificial Intelligence Act is the most developed horizontal AI regulatory framework currently in operation. It adopts a risk based structure and distinguishes prohibited practices, high risk systems and transparency obligations. The Act prohibits certain practices, including harmful manipulation, exploitation of vulnerabilities, certain forms of social scoring and certain forms of predictive criminal risk assessment based solely on profiling. High risk systems include specified uses in employment, access to essential services, creditworthiness, law enforcement, migration, education and justice. The Act is particularly relevant to discrimination because it imposes obligations concerning risk management, data governance, technical documentation, record keeping, human oversight and monitoring for high risk systems. Article 27 also requires certain deployers to conduct fundamental rights impact assessments before deploying high risk systems. Article 86 creates a right to obtain clear and meaningful explanations of the role of certain high risk systems in individual decisions. The United States has taken a more fragmented approach. Rather than relying on one comprehensive federal AI statute, existing anti discrimination, consumer protection and sectoral laws have been applied to algorithmic systems. The Consumer Financial Protection Bureau has made clear that lenders using complex algorithms remain subject to adverse action explanation requirements. The Equal Employment Opportunity Commission has recognised the civil rights implications of automated employment systems. The Federal Trade Commission has also acted against harmful deployment of facial recognition technology.[14]

The American approach demonstrates that existing law can sometimes regulate AI without an AI specific statute. Its weakness is fragmentation. Rights and remedies depend heavily on the sector and the applicable federal or state law. The English Court of Appeal decision in R v Chief Constable of South Wales Police, commonly known as Bridges, provides a useful example of judicial scrutiny of facial recognition. The court held that the police use of live automated facial recognition was not in accordance with the law for the purposes of Article 8 of the European Convention on Human Rights and found non compliance with the public sector equality duty. The case is important because the court did not treat the technology as inherently unlawful. Instead, it examined whether the legal framework governing its use was sufficiently precise and whether the public authority had properly addressed equality obligations. The lesson for India is that judicial review can regulate AI through existing legal principles if courts are willing to examine the entire decision making process rather than only the final output. In Ewert v Canada, the Supreme Court of Canada considered psychological and actuarial assessment tools used for Indigenous offenders. The Court emphasised the obligation of correctional authorities to ensure that information used in decision making is accurate and reliable for Indigenous offenders. Ewert is highly relevant to algorithmic discrimination because it demonstrates that an institution cannot assume that a statistical tool is reliable merely because it has been validated generally. Validation must be relevant to the population on whom the tool is used. This principle should be adopted in India. An AI system used on Indian populations cannot be assumed to be fair merely because it performed well in another country or on a different demographic dataset.[15]

6. INDIAN CASE STUDIES, HIGH RISK USES AND THE NEED FOR EFFECTIVE REMEDIES

The COMPAS controversy remains one of the most influential examples of algorithmic risk assessment. ProPublica reported racial disparities in risk classifications. The Wisconsin Supreme Court considered the use of COMPAS in State v Loomis and permitted consideration of the assessment subject to limitations. The deeper lesson is not simply that COMPAS was biased. It is that criminal justice decisions are uniquely sensitive because liberty is at stake. An individual should not be deprived of liberty through an opaque statistical prediction that cannot be meaningfully challenged. The EU AI Act's restrictions on certain predictive criminal risk assessments reflect the seriousness of this risk. India should be cautious before allowing predictive policing or automated criminal risk scores to influence arrest, bail, sentencing or

14UNESCO (n 1).
15European Commission, 'AI Act' <https://digital-strategy.ec.europa.eu/en/policies/regulatory-framework-ai> accessed 6 October 2026.

surveillance decisions. The Amazon example demonstrates historical bias in employment. The system learned from past recruitment patterns and penalised terms associated with women. Amazon abandoned the tool after the problem became apparent. The case also reveals a regulatory weakness. Voluntary abandonment after internal testing is preferable to continued discrimination, but it does not provide a general institutional mechanism ensuring that other employers conduct comparable testing before deployment. Indian employment law should therefore require stronger procedural safeguards for automated recruitment, particularly where an employer uses AI to reject candidates at scale. The Facebook housing advertising controversy illustrates a different form of discrimination. The issue was not merely the content of an advertisement. It was the algorithmic delivery system that determined who would see particular opportunities. The United States Department of Justice later brought an action concerning Meta's housing advertising system, alleging that its algorithms relied in part on protected characteristics and contributed to discriminatory distribution. The settlement required changes to the system.[16]

The Indian lesson is important because digital advertising affects employment, housing, credit, insurance and political participation. A platform may claim that it does not discriminate because advertisers select broad categories, while the delivery algorithm independently optimises who receives the advertisement. Regulation must therefore examine both user instructions and platform optimisation. The Gender Shades research, Bridges and the FTC's action involving Rite Aid together demonstrate the risks of biometric systems. Rite Aid was accused of deploying facial recognition without adequate safeguards and of experiencing higher false positives in communities with larger Black and Asian populations. Biometric errors can have severe consequences. A false recommendation in entertainment may be inconvenient. A false identification in policing may lead to detention, search or violence. India therefore requires a higher threshold for biometric deployment by public authorities. Systems should undergo independent testing on relevant Indian demographic groups before operational use, with continuing monitoring after deployment. The study by Obermeyer and others concerning a widely used health management algorithm showed how an apparently race neutral healthcare system could produce racially unequal outcomes because healthcare expenditure was used as a proxy for healthcare need. Lower expenditure among Black patients, reflecting unequal access and historical disadvantage, caused the model to underestimate their health needs. This example is important for India because healthcare data can encode socioeconomic inequality. A model trained on treatment patterns may learn that a group receives less treatment and interpret this as evidence that the group requires less care. Algorithmic fairness in healthcare therefore cannot be achieved merely by removing caste, religion or gender from the dataset. The model must be evaluated against substantive outcomes.[17]

Caste presents a distinctive challenge. Unlike race in many Western datasets, caste is deeply embedded in Indian social, educational, occupational and geographic structures. A model may infer caste from surnames, locality, school, occupation, language or social connections even if caste is never directly collected. This creates a paradox. Excluding caste from a dataset can make auditing difficult, while using caste directly may create privacy and discrimination concerns. The appropriate response is controlled and lawful use of sensitive attributes for auditing rather than their unrestricted use for decision making. The constitutional commitment to substantive equality also means that fairness cannot always mean identical treatment. In contexts where affirmative action is legally required, an algorithm that ignores reservation requirements is not neutral. It may be discriminatory precisely because it refuses to account for legally recognised disadvantage. Disability raises a separate issue because fairness often requires accommodation rather than equal treatment in the narrow sense. The Supreme Court in Vikash Kumar v Union Public Service Commission emphasised reasonable accommodation and the need to understand disability in an individualised and intersectional manner. An AI hiring or examination system that evaluates candidates only according to standard behavioural or productivity patterns may disadvantage persons with disabilities. A facial recognition system may perform poorly on persons with certain disabilities. A speech recognition model may misinterpret speech differences. The legal principle should therefore be that accessibility is part of fairness. A system is not fair merely because its error rate is similar across groups if the system is inaccessible to persons with disabilities. India's linguistic diversity creates another form of algorithmic inequality. Natural language systems trained primarily on English or a small number of high resource languages can perform poorly for speakers of other Indian languages and dialects.[18]

16State v Loomis 881 NW 2d 749 (Wis 2016).
17Jeffrey Dastin (n 3).
18Federal Trade Commission, Federal Trade Commission v Rite Aid Corporation (2023).

This may affect access to public services, education, healthcare and legal information. If an automated grievance system fails to recognise a citizen's language accurately, the harm may appear technological but its effect is unequal access to rights. Responsible AI governance in India must therefore treat linguistic inclusion as an equality issue, not merely as a market opportunity. Discrimination rarely occurs along one axis. A rural woman with a disability from a disadvantaged caste may experience an algorithmic system differently from an urban woman without a disability. A fairness test that checks only gender may show acceptable performance while masking severe disadvantage at the intersection of gender, caste, disability and geography. Indian constitutional equality is increasingly attentive to structural and intersectional disadvantage. AI auditing should follow the same direction. Organisations should examine relevant subgroups where the context makes intersectional harm reasonably foreseeable. Algorithmic impact assessment should become a central governance mechanism for high consequence AI systems. An assessment should occur before deployment and continue throughout the system lifecycle. A meaningful assessment should identify the purpose of the system, the legal basis for its use, the groups likely to be affected, the data sources, known limitations, potential discriminatory effects, privacy risks, accessibility concerns, human oversight mechanisms, security risks and available remedies. The assessment should also ask a prior question that is often neglected: should the automated system be used at all? Not every decision should be automated. In criminal justice, immigration, welfare termination, employment dismissal and medical triage, the potential cost of error may be too high for certain forms of automation. The European Union's fundamental rights impact assessment model provides a useful reference point. UNESCO's ethical impact assessment approach and NIST's risk management framework offer additional methodological resources. India can adapt these models without copying them mechanically.[19]

The debate about explainability often becomes technically confused. A legal right to explanation does not necessarily require disclosure of the complete source code or model weights. What matters is meaningful information about the decision. A useful explanation should ordinarily identify that an automated system was used, identify the institution responsible for the decision, explain the principal factors that materially influenced the outcome, disclose important limitations and provide a route for contesting the result. Human oversight is frequently proposed as a solution to algorithmic discrimination. But a human rubber stamp is not meaningful oversight. The human reviewer must have authority to question the system, access to relevant evidence, sufficient training, adequate time and an institutional incentive to exercise independent judgement. If employees are penalised for departing from automated recommendations, human oversight becomes illusory. Institutions should preserve records showing which version of an AI system was used, what data was supplied, what output was produced, who reviewed the output and what final decision was taken. Without records, a person challenging an automated decision faces an almost impossible evidentiary burden. AI systems involve multiple actors. A developer may build the model. A data provider may supply training data. A vendor may customise the system. An employer or bank may deploy it. A public authority may use the resulting output. The law should therefore allocate responsibility according to control and function. The entity that chooses the purpose and deploys the system should ordinarily bear responsibility for lawful use. Developers should have duties concerning known design risks. Vendors should not be permitted to contract away all responsibility for predictable discriminatory outcomes. One of the biggest problems in algorithmic discrimination litigation is information asymmetry. The affected individual normally does not have access to the model, training data, audit reports or internal testing. A fair legal framework should therefore permit courts and regulators to require production of relevant technical records. Where a claimant establishes a substantial disparity and shows that an automated system materially influenced the decision, the burden should shift sufficiently to require the institution to demonstrate lawful purpose, reasonable safeguards and an adequate justification. This does not mean that every statistical disparity should create automatic liability. Disparities can arise by chance or from legitimate factors. The point is procedural: the party controlling the system is better placed to explain its operation. Remedies should include more than monetary compensation. Depending on the context, an affected person should be able to obtain reconsideration by a human decision maker, correction of inaccurate data, deletion of unlawfully processed data, suspension of the relevant system, independent audit, restoration of access to a service, compensation and appropriate declaratory relief. For public authorities, constitutional remedies under Articles

19Joy Buolamwini and Timnit Gebru (n 4).

32 and 226 remain important. Courts should also be willing to issue structural directions where systemic algorithmic practices affect large groups.[2021]

7. TOWARDS A RIGHTS BASED GOVERNANCE MODEL FOR INDIA

India does not need to choose between complete deregulation and a single massive AI statute. A layered model is more appropriate. First, constitutional standards should govern public sector use. Any AI system affecting fundamental rights should satisfy legality, legitimate purpose, proportionality, non discrimination, procedural fairness and effective remedy. Second, Parliament should introduce targeted statutory duties for high consequence private sector uses. Employment, credit, insurance, healthcare, education and essential services are obvious candidates. Third, sector regulators should issue detailed rules. The RBI should address credit and financial AI. SEBI should address algorithmic systems in securities markets. IRDAI should address insurance. Health authorities should address medical AI. Labour authorities should address automated employment decisions. Fourth, high consequence AI systems should undergo documented impact assessments before deployment. Assessments should be updated when the system is materially changed or deployed in a new population. Fifth, affected individuals should receive notice where AI materially influences a decision about them. They should have access to a meaningful explanation and a human review mechanism. Sixth, organisations should conduct periodic fairness testing. Testing should examine not only aggregate accuracy but group specific error rates and, where relevant, intersectional outcomes. Seventh, regulators should have powers to inspect technical records and commission independent audits. Confidential business information can be protected through controlled disclosure rather than complete secrecy. Eighth, India should develop a public incident and harm reporting mechanism. Regulators need empirical information about which systems cause which harms. Without incident data, risk based governance becomes speculative. Ninth, procurement rules should become an important regulatory instrument. Government departments should not purchase high consequence AI systems without contractual requirements concerning audit access, documentation, bias testing, data governance, security and remedies.[22]

Tenth, regulatory sandboxes should be used for controlled experimentation rather than as exemptions from rights. A sandbox should not become a space in which vulnerable individuals are exposed to systems without adequate protection. The India AI Governance Guidelines provide a pragmatic foundation, but their voluntary and principle based character creates a central weakness. A principle stating that AI should be fair does not by itself give an individual a cause of action. The Guidelines themselves recognise that existing law should be used to address AI harms and that additional obligations may be introduced where necessary. This creates a sensible evolutionary model. However, evolution should not become indefinite delay. India also faces an institutional capacity problem. Effective AI regulation requires technical expertise, legal expertise and domain expertise. A regulator may understand discrimination law but not machine learning. A technology agency may understand models but not constitutional rights. The proposed whole of government approach is therefore valuable, but it must be accompanied by independent expertise and public accountability. There is also a danger in relying too heavily on voluntary commitments. Responsible companies may undertake audits while irresponsible companies do not. Voluntary standards can create useful baselines but should not replace enforceable minimum protections in high consequence contexts. The most defensible Indian model is therefore one in which regulation remains proportionate. Low risk systems can largely be governed through existing law and voluntary standards. High consequence systems should face mandatory duties. Prohibited applications should be identified where the risk cannot be adequately mitigated. This approach would align the India AI Governance Guidelines with constitutional principles while avoiding unnecessary regulatory burdens.[23]

Algorithmic discrimination is not a hypothetical future problem. It is a present legal problem produced by the interaction between data, statistical models, institutional incentives and existing social inequalities. The central mistake is to assume that automation removes discrimination. In many circumstances automation changes its form. Human prejudice can become historical data. Social inequality can become a predictive variable. Institutional choices can become optimisation

20Committee on the Rights of Persons with Disabilities, General Comment No 6 on Equality and Non Discrimination (2018); Vikash Kumar (n 6).
21Lt Col Nitisha (n 2); Justice K S Puttaswamy (Retd) (n 3).
22National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework 1.0 (2023) and Artificial Intelligence Risk Management Framework Generative Artificial Intelligence Profile (2024).
23Regulation (EU) 2024/1689 (Artificial Intelligence Act), art 27.

objectives. Discriminatory outcomes can become feedback data. Human rights law provides a principled response because its central concern is not the identity of the decision maker but the protection of human dignity, equality, autonomy and effective remedy. A decision does not become legitimate because it was produced by a machine. Indian constitutional law is capable of responding to this challenge. Article 14 provides a foundation for substantive equality. Article 15 addresses discrimination and permits measures designed to address disadvantage. Article 16 protects equality of opportunity in public employment. Article 21 protects dignity and privacy. The jurisprudence in Puttaswamy, Anuj Garg, Nitisha and Vikash Kumar demonstrates that Indian constitutional doctrine can move beyond formal neutrality and examine structural disadvantage, indirect discrimination and reasonable accommodation. The principal gap lies in the private sector and in the absence of clear procedural rules for high consequence automated decisions. The Digital Personal Data Protection framework is important but cannot substitute for anti discrimination regulation. The India AI Governance Guidelines provide a valuable governance architecture, but their principles will require targeted legal implementation if they are to produce enforceable rights.[24]

India should therefore develop a layered framework based on risk, rights and responsibility. High consequence AI should be subject to prior impact assessment, meaningful transparency, independent testing, human oversight, record keeping, accessibility and effective remedies. Responsibility should be allocated across the AI value chain according to control and foreseeable risk. Courts and regulators should have access to technical evidence. Sensitive attributes should be used carefully for auditing where necessary to identify discrimination, while being protected from misuse. The strongest lesson from comparative law is that technology neutral legal principles can regulate artificial intelligence effectively when institutions are willing to apply them seriously. Bridges demonstrates the value of judicial review. Ewert demonstrates the importance of population specific validation. The CFPB's approach demonstrates that technological complexity does not excuse failure to explain adverse decisions. The EU AI Act demonstrates the usefulness of risk based regulation. India's own constitutional jurisprudence demonstrates that formal neutrality is not sufficient where a rule reproduces structural inequality. Artificial intelligence should therefore be treated neither as inherently discriminatory nor as inherently objective. It is a sociotechnical instrument whose legal legitimacy depends on how it is designed, trained, deployed and governed. The appropriate legal question is not whether an algorithm is biased in the abstract. It is whether the system, in its real institutional context, respects the rights of the people over whom it exercises influence. The future of AI governance in India should be measured by this standard. Innovation is valuable only when its benefits are broadly accessible and its costs are not systematically imposed on those who are already disadvantaged. The constitutional promise of equality cannot be delegated to an algorithm. If artificial intelligence is to become part of public and private decision making, accountability must remain human, responsibility must remain identifiable and rights must remain enforceable.[25]

The effectiveness of this framework will ultimately depend on whether institutions treat fairness and accountability as continuing legal obligations rather than as technical exercises completed before deployment. An impact assessment that is prepared only to satisfy a procurement requirement, an explanation that merely repeats the output of a model, or a human reviewer who lacks authority to disagree with an automated recommendation would provide the appearance of protection without its substance. Indian law should therefore emphasise institutional responsibility at every stage of the lifecycle. Developers should document foreseeable limitations and testing conditions. Deployers should establish procedures for review, escalation and correction. Regulators should have sufficient technical capacity to examine evidence rather than accepting assurances of accuracy. Courts should be willing to examine the practical operation of automated systems when fundamental rights are affected, while remaining attentive to legitimate confidentiality and intellectual property concerns. Most importantly, individuals must remain visible within the legal process. The fact that an adverse decision has passed through a statistical model cannot diminish the person's entitlement to equality, dignity, privacy and an effective remedy. A rights based approach to artificial intelligence is consequently not an attempt to place technology outside the law. It is an insistence that technological innovation operate within the same constitutional conditions that govern every other exercise of institutional power.[26]

24Lt Col Nitisha (n 2); Anuj Garg (n 2); Justice K S Puttaswamy (Retd) (n 3); Vikash Kumar (n 6).
25 OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449 (2019), updated 2024 <https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449> accessed 6 October 2026.