The Indian Journal for Research in Law and Management

Open Access Law Journal – Copyright © 2026

Editor-in-Chief – Dr. Muktai Deb Chavan; Publisher – Alden Vas; ISSN: 2583-9896

ALGORITHMIC BIAS IN DIGITAL SYSTEMS: A CYBER SECURITY PERSPECTIVE IN INDIA

SHORT ARTICLE

- ***Ishti Singhal***

ABSTRACT

Artificial Intelligence is now extensively used in modern security systems for fraud detection, biometric authentication, facial recognition, behavioural analytics, surveillance and cyber threat monitoring. These systems can be biased in these areas, leading to potential vulnerabilities, skewed evidence weight, and flawed automated decision-making. While there are some protections in Indian law, the existing legal framework does not provide a comprehensive framework for the regulation of algorithmic transparency, bias audits, algorithmic explainability, or forensic validation of outputs generated by an AI system. This paper examines the nature of algorithmic bias, its impact on cybersecurity and cyber forensics, the existing Indian legal provisions, and the change needed to minimize the discriminatory harms, and strengthen the integrity, accountability and security of digital systems in India.

1. INTRODUCTION

Algorithmic bias refers to systematic and unfair distortions in the design, training, deployment or output of automated systems. It may stem from historical data bias, the use of unsuitable proxies, lack of representation of some social groups, incorrect assumptions in the design of the model, or institutional processes that introduce existing discrimination into technical processes. Algorithmic bias has been studied from an ethical and anti-discrimination perspective, but it can also be viewed from a cybersecurity lens as the integrity, reliability, and robustness of the outputs of digital systems are important for cybersecurity.

The cyber security aspect is better understood if the automated systems are considered not as tools, but as operational infrastructure. Whether a user is authenticated, flagged, monitored, or

investigated, it is determined by spam filters, intrusion detection systems, fraud analytics, biometric verification systems, facial recognition tools, and risk-scoring software. When these systems don't function equally in various communities or situations, they leave gaps that criminals could exploit and mistakes that investigators could mistake for fact.[1]

This is especially pertinent to the case of India. NITI Aayog's Responsible AI initiative outlines the following as key principles of governance for responsible use of AI in India: safety and reliability, inclusivity and non-discrimination, equality, privacy and security, transparency, accountability, and protection of human values. Another aspect of its facial recognition use case paper is the focus on fundamental issues that plague AI-powered systems, such as inaccuracy caused by bias or underrepresentation, opacity, failures in accountability and security issues stemming from data breaches and unauthorized access. In a country with significant linguistic, regional, caste, religious, gender and socio-economic differences, these risks are heightened as a system based on too small or unrepresentative samples may not fairly extrapolate to the population it serves.[2]

In this paper, the author argues that algorithmic bias in India must be seen through the lens of equality, privacy, cyber security and forensics reliability as a whole. It first lays out the basics of algorithmic bias, followed by a discussion on the effects of bias on cyber security applications and forensic systems, a review of the current Indian legal landscape, an analysis of the gaps in the regulatory framework, and some suggestions for change that would make Indian law more sensitive to the technical and constitutional dangers of biased digital systems.

2. ALGORITHMIC BIAS FUNDAMENTALS

2.1 TYPES OF BIAS

Algorithmic bias is not a single error but a range of errors, which can occur at any stage of a system's life cycle. Historical bias can happen when real-world data used to train an algorithm already reflects social inequalities, like over-policing, discriminatory lending, or unequal access to digital services, and these patterns are internalized as norms. Representation bias occurs when some groups are underrepresented or excluded from the training set, leading to better results for dominant groups and poorer results for other groups.[3]

[1] Navin Kumar, *AI in Cyber Security: Innovations and Implications for a Safer Digital World*, 13 Computer Science and Information Technology 1-7 (2025).
[2] Yoshita Sood, *Addressing Algorithmic Bias in India: Ethical Implications and Pitfalls*, SSRN (2022).
[3] Neha Verma, *Algorithmic Bias and Discrimination Under India's Dpdpa: Evaluating the Adequacy of Legal Frameworks for Ai-Driven Decisionmaking*, 5 Indian Journal of Integrated Research in Law

Measurement bias occurs when variables used by a system are used as an imprecise proxy for the trait or behaviour being measured. A model could, for example, use spending patterns, location, language, or device type as indirect indicators of risk, trustworthiness, or suspicious activity, even if these factors are related to structural inequality, and not to actual threat. The deployment bias arises when a model is used in a context very different from the context in which it is trained, for example, when the system used for verification is in a city but the target system is in a rural area, or when a commercial image classifier is used in a criminal investigation workflow.[4]

Such bias forms are not only unfair, but also negatively impact system performance in predictable ways. A tool that systematically misinterprets some groups, environments or behaviour patterns has known vulnerabilities, and in cyber security these vulnerabilities are points of attack.

2.2 CYBER SECURITY LINKAGES

While traditionally cyber security focuses on confidentiality, integrity and availability, modern digital governance also demands trustworthiness of automated outputs. An algorithmic system that is biased becomes unreliable in terms of the integrity of the decisions it makes as similar inputs can produce dissimilar outputs based on who is being evaluated or what data is available. This makes it difficult to make decisions and to defend the actions taken. First link is the exploitability. Anomaly detection or authentication system that is biased may not detect malicious activity if the malicious activity conforms to a pattern that is not common in the system's training set. These blind spots can be learned by adversaries and inputs can be developed to bypass detection or cause misclassification.[5] The second linkage is to adversarial manipulation. Poisoning of training pipelines, datasets, updates, and labelling, combined with weak governance over these processes, can exacerbate the inherent bias and lead a model to even more unreliable results. Operational distortion is the third linkage. Automated triage systems are critical for security teams to prioritise incidents and allocate investigative resources. These systems are biased and can be over-observing some users while missing real attacks at other locations. To sum up, algorithmic bias is not a side effect of cyber security, but rather it directly impacts threat modelling, incident response, digital trust and resilience.

3. CYBER APPLICATIONS WITH BIAS

[4] Julia Angwin et al., 'Machine Bias', ProPublica, 23 May 2016.
[5] Kate Crawford, *Atlas of AI: Power, Politics, and the Planetary Costs of Artificial Intelligence* 19–46. (2021).

3.1 THREAT DETECTION SYSTEMS

AI-powered threat detection systems are a crucial component of cyber defence today. These systems can classify suspicious network traffic, evaluate transaction risk, detect malware, detect fraudulent activity, and analyse user behaviour patterns at scale. They are quick and cheap to produce, but can also replicate or amplify bias if trained on incomplete or biased data. Think about the use of behavioural analytics tools to detect unusual logins or risky transactions. The tool might classify rural, multilingual, or lower-bandwidth patterns as suspicious, even if they are legitimate, because it has been trained on patterns that are more typical of urban, English-speaking, well-banked users and using standard devices. This can result in false positives which can result in the denial of access to services, intrusive verification or freezes on accounts of already marginalised users.[6] Bias can also lead to false-negative results. A system that is trained on a small set of threat models could fail to identify malicious activity that is outside of the expectations of the dominant threat model. In the cybersecurity world, the inability to classify some patterns is not simply a technical limitation, it's an attack surface. The legal relevance is also very strong due to the fact that if these systems process personal data, organisations could have a responsibility under data protection and security law to have put in place the necessary technical and organisational measures and to have minimised the risks of automated processing.

3.2 FORENSIC TOOLS

These are even more pressing concerns in the context of forensic deployment of automated systems. The digital forensics industry is becoming increasingly automated in order to sort through evidence, identify persons of interest, cluster devices, detect anomalies, and analyse large numbers of images or communications. These tools can be helpful for efficiency, but can also introduce bias or obscurity into the evidence process.[7]

The most obvious example is facial recognition. The study by NITI Aayog on facial recognition points to design-related issues like inaccuracy due to bias or underrepresentation, lack of transparency in system design, accountability challenges, and security concerns like data breaches or unauthorized access. It also distinguishes between 1:1 authentication and 1:n identification applications, stating that live or large scale applications of identification are more

[6] Lilian Edwards & Michael Veale, *Slave to the Algorithm? Why a 'Right to an Explanation' Is Probably Not the Remedy You Are Looking For*, 16 Duke Law & Technology Review 18-84 (2017).

[7] Veena Kumari, *The Role of Artificial Intelligence in Digital Forensic Investigations: Legal and Ethical Challenges*, 11 INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT, (2026).

serious issues, as people may not know that their biometric information is being collected and used.

The use of facial recognition in law enforcement and public systems has already been a topic of debate in India due to potential for false or biased matching, which can lead to misdirected investigations, inappropriate surveillance, or wrongful suspicion. If an algorithmically generated lead directs the course of an investigation, then additional human review may not fully correct the initial distortion. From a cyber forensics' perspective, this compromises the integrity and reliability of digital investigations.

4. INDIAN LEGAL FRAMEWORK

4.1 CONSTITUTIONAL PRINCIPLES

The Constitution of India furnishes the strongest normative foundation for regulating detrimental algorithmic bias. Article 14 guarantees equality in the eyes of the law and protects against arbitrary action, Article 15 forbids discrimination on certain grounds. Automated tools, which produce systematically unequal outcomes without adequate safeguards, when implemented by a state agency, can be challenged as arbitrary, discriminatory, or disproportionate.[8]

There is also constitutional privacy doctrine. The paper on FRT by NITI Aayog clearly ties facial recognition risks to privacy, informational autonomy, anonymity, freedom of movement and the need for legal safeguards like necessity and proportionality. These principles suggest that highly intrusive AI systems, especially those used for surveillance or biometric identification, must be based on a clear legal foundation, have a legitimate purpose, and be subject to appropriate safeguards commensurate with the potential for harm.[9]

However, constitutional litigation is not the answer, as it is often reactive and specific. The values need to be operationalized in the technical sense and this needs to be done through a more robust statutory and regulatory framework.

4.2 PROVISIONS OF THE IT ACT, 2000

The Information Technology Act, 2000 is India's main cyber law. It covers issues of unauthorised access, computer offences, intermediary liability and some aspects of data-security responsibilities. While the statute was not designed for AI, it is still relevant as many

[8] Sandra Fredman, *Discrimination Law* 176–210 (3rd ed. 2022).

[9] Cathy O'Neil, *Weapons of Math Destruction: How Big Data Increases Inequality and Threatens Democracy* 3-20 (2016).

AI-based systems are used in biased ways across digital platforms, surveillance systems and security operations, which are all covered by cyber law.

One of the key ideas is reasonable security practice. In the current discussion on AI governance in India, it is observed that the IT framework and related security regulations remain applicable to data processing in AI systems until new regulations are fully implemented. An automated system that is based on low quality data, has no protection for manipulation, and has predictable discriminatory failures can be claimed not to be in reasonable security practice, especially if the data being processed is sensitive or personal. But the IT Act does not define the concept of high-risk AI systems, mandate fairness testing, impose algorithmic accountability or provide for specific remedies for unfair automated outputs. So, the statute is helpful but incomplete.[10]

4.3 DPDP ACT, 2023 ANALYSIS

The Digital Personal Data Protection Act, 2023 is highly pertinent since numerous algorithmic systems rely on the collection and processing of digital personal data. The commentary to the Act suggests that the processing of personal data using AI, such as creating and training AI models, may fall under the scope of the Act's definition of processing. The Act sets obligations for data fiduciaries to take suitable technical and organizational measures, establish reasonable safeguards and manage personal data responsibly.[11]

One of the developments is the growing demand on SDF to comply with more stringent obligations in regard to algorithmic systems. There are some indications that such organizations might need to conduct regular algorithmic audits, data protection impact assessments, and due diligence to ensure that the algorithmic software does not present risks to the data principals. This is important because it starts to connect data governance and algorithmic accountability.

However, the DPDP framework has some drawbacks. It does not currently create a general right to be free from algorithmic discrimination, and it doesn't go far enough in regulating explainability, human review, or contestability of automated decisions. This leaves the Act as a means of regulating AI that relies on personal data but not as a solution to bias in digital systems.

4.4 POLICY GUIDANCE

[10] Siva Vignesh S.K.V & Nagarjun D.N, *Legal Challenges of Artificial Intelligence in India's Cyber Law Framework: Examining Data Privacy and Algorithmic Accountability via a Comparative Global Perspective*, 6 International Journal For Multidisciplinary Research (2024).

[11] *AI and Privacy – Navigating the DPDP Act in the Age of Emerging Tech*, (Jan. 12, 2026), https://www.thelegaljournalontechnology.com/post/ai-and-privacy-navigating-the-dpdp-act-in-the-age-of-emerging-tech.

In the absence of hard law, India's soft law offers valuable direction. The principles of accountability, inclusivity, non-discrimination, privacy and security, transparency and safety are identified as the key elements of AI governance in NITI Aayog's Responsible AI approach documents. The same guidelines warn against the use of black-box algorithms in critical public services and urge to make clear the disclosures on training data, testing procedures, and steps taken to reduce spurious correlations and bias.[12]

The techno-legal framework by the Office of the Principal Scientific Adviser (PAS) also reflects a shift towards a more integrated approach to AI governance, drawing on existing laws like the IT Act and the DPDP Act, while recognizing the need for more formalized regulations. The documents are important because they show that the issue of bias is a concern in Indian governance discourse, albeit one that is not yet actionable.

5. REGULATORY GAPS AND CASES

Currently, India has 5 major weaknesses. First, there is no specific AI statute that defines high-risk applications and prescribes specific duties for these. Second, transparency is still limited, and individuals who are affected by an automated system might not be aware when or how an automated system has affected a decision. Third, there are no clear forensic standards in India to validate the outputs generated by AI or AI-assisted before they affect any criminal or quasi-criminal proceeding.[13]

Fourth, there is a specific risk of bias in India because of entrenched social inequalities, for example, data could be based on caste, sex, religion, language, or socio-economic disadvantage. Algorithmic bias and discrimination in India is a crisis that is unfolding and can be seen as a threat to marginalised communities, according to a 2025 scholarly evaluation. Fifth, institutional oversight is piecemeal, with no one responsible for fairness testing, audit standards, incident reporting, and cross-sector enforcement for high-risk algorithmic systems.

Facial recognition is an example of the debates that highlight these gaps. The NITI Aayog study on facial recognition technology identifies issues of underrepresentation, opacity, liability, grievance redress, privacy and freedom of movement as concerns to the rights of citizens. If such systems are used for security or law-enforcement, the impact of bias is not just commercial, it's on liberty, dignity, and due process.[14]

[12] Anjana Karumathil, *Responsible AI Approach Document for India: A Critical Assessment*, Indiaai (2022).
[13] Dharish David, *Algorithmic Bias and Discrimination in India: A Looming Crisis*, 11 Sage Journals (2025).
[14] *Leveraging Responsible*, Vidhi Centre for Legal Policy https://vidhilegalpolicy.in/leveraging-responsible/.

6. RECOMMENDED REFORMS

6.1 STATUTORY AMENDMENTS

It is high time that India update regulations on cyber and data without waiting for harm to be done. The IT Act should clearly state that reasonable security practices for high-risk digital systems include dataset governance, bias testing, adversarial robustness checks, documentation of model limitations, and post deployment monitoring. This recognizes that insecure AI isn't just about hacking, but about unreliable and discriminatory automation.

The DPDP regime should be clarified to require special protection for important automated decision making. These protections should consist of notices to impacted individuals, a right to meaningful explanations, human review at request, and required fairness audits when automated systems have a material effect on access to services, identity verification or legal exposure.

6.2 ASSESSMENT REQUIREMENTS

Algorithmic Impact Assessments should be mandatory in India for high-risk applications like public surveillance, cyber threat intelligence, fraud detection, welfare screening, biometric verification, and law enforcement. These evaluations should include a consideration of the purpose, legal basis, data quality, representativeness, differential error rates, risk of adversarial manipulation, security measures, and redress mechanisms. It is important to have an ex ante assessment because it is easier to reduce bias before deployment than after widespread use.

6.3 FORENSIC STANDARDS

Explicit admissibility and validation rules should be in place for AI-assisted forensic tools. The deploying authority should keep audit logs, validation studies, testing records, model version history and known limitations and confidence levels documentation before using such a tool in an investigation or relying on it in legal proceedings. Access to and challenging of defence should also be acknowledged, as the algorithm's output should not be unreviewable black box evidence.[15]

6.4 OVERSIGHT MECHANISMS

The institutional reform is also needed in India. Technical standards should be set by a specialized AI oversight body, or a dedicated unit within an existing digital regulator, which would also be responsible for approving high-risk deployment procedures, investigating complaints and coordinating between sectors. These obligations should then be further

[15] Frank Pasquale, THE BLACK BOX SOCIETY: THE SECRET ALGORITHMS THAT CONTROL MONEY AND INFORMATION 8 (2015).

specified for the banking, health, law enforcement, education, airports, and critical infrastructure sectors, given that the potential for bias varies by sector.[16]

7. CONCLUSION

The issue of algorithmic bias in digital systems is not just an ethics problem; it is a fundamental problem for cyber security, forensic trustworthiness and constitutional governance in India. A biased system may be unable to grant access when it should, or unable to detect a real threat when one exists, or mislead automated risk assessment, or introduce opaque or unreliable output into a digital investigation.

India already has a partial legal basis and principles of equality in the constitution, privacy doctrine, IT Act, DPDP Act, and Responsible AI policy. But these are not yet a full suite of bias audits, transparency requirements, rights to human review, forensic validation standards, and specific oversight. The most practicable approach is to adopt a multi-layered reform strategy that strengthens the current cyber and data legislation, embed constitutional principles in technical regulation, and recognise algorithmic bias as a legal wrong and cyber vulnerability.

[16] Solon Barocas & Andrew D. Selbst, *Big Data's Disparate Impact*, 104 Cal. L. Rev. 671, 677 (2016).