



LAWFOYER INTERNATIONAL JOURNAL OF DOCTRINAL LEGAL RESEARCH

[ISSN: 2583-7753]

Volume 4 | Issue 3

2026

DOI: <https://doi.org/10.70183/lijdlr.2026.v04.313>

© 2026 LawFoyer International Journal of Doctrinal Legal Research

Follow this and additional research works at: www.lijdlr.com

Under the Platform of LawFoyer – www.lawfoyer.in

After careful consideration, the editorial board of LawFoyer International Journal of Doctrinal Legal Research has decided to publish this submission as part of the publication.

In case of any suggestions or complaints, kindly contact (info.lijdlr@gmail.com)

To submit your Manuscript for Publication in the LawFoyer International Journal of Doctrinal Legal Research, To submit your Manuscript [Click here](#)

BALANCING INNOVATION AND HUMAN RIGHTS: REASSESSING PRIVACY AND AUTONOMY IN AI GOVERNANCE

Gaurav Anand¹ & Harsh Raj²

I. ABSTRACT

*Artificial Intelligence is increasingly shaping governance, public administration, commercial decision-making, and everyday digital interactions. While AI offers significant benefits in innovation, efficiency, security, welfare delivery, and economic growth, its dependence on large-scale data extraction raises serious concerns for privacy, autonomy, dignity, and equality. This paper examines the central research problem of how AI systems, when deployed by the State and powerful private actors, may transform individuals into data subjects through profiling, behavioural prediction, biometric identification, surveillance, and opaque automated decision-making. The study adopts a rights-based and doctrinal legal approach to assess whether existing constitutional and statutory protections are adequate to respond to these risks in India. The paper analyses the constitutional foundations of privacy, dignity, autonomy, and equality under Articles 14, 19, and 21 of the Constitution of India, with particular emphasis on the Supreme Court's recognition of privacy in *K.S. Puttaswamy v. Union of India*. It further evaluates the Digital Personal Data Protection Act, 2023, and identifies its limitations in addressing algorithmic bias, inferential data, automated decision-making, and AI-specific accountability. Comparative insights are drawn from the European Union's General Data Protection Regulation, the EU AI Act, the OECD AI Principles, and UNESCO's Recommendation on the Ethics of Artificial Intelligence. The paper argues that India requires a comprehensive, human-rights-oriented AI governance framework. It recommends the establishment of an independent AI Ethics and Accountability Commission, mandatory AI Impact Assessments and Human Rights Impact Assessments for high-risk systems, stronger*

¹ B.A. LL.B. (Hons.), Fourth Year, Gitarattan International Business School, Guru Gobind Singh Indraprastha University (GGSIU), New Delhi (India). Email: gaurav150994@gmail.com

² B.A. LL.B. (Hons.), Final Year, Gitarattan International Business School, Guru Gobind Singh Indraprastha University (GGSIU), New Delhi (India). Email: legal.harshraj@gmail.com

transparency and contestability rights, restrictions on biometric surveillance, and meaningful remedies for persons affected by automated decisions.

II. KEYWORDS

Artificial Intelligence Governance; Right to Privacy; Autonomy; Digital Personal Data Protection Act; Article 21.

III. INTRODUCTION

A. Background: AI, Datafication, and Digital Governance

AI is playing an essential part in boosting economy, innovation, security and improving our daily lives. From personalized healthcare and smart urban infrastructure to autonomous transportation and environmental monitoring. However, the implications of AI on politics, security, and information must be thoroughly considered to ensure that AI benefits society and reduces risks. AI is developing so rapidly that it is challenging for policy-making processes to keep pace.³ As AI is increasingly used by governments and private entities to improve efficiency and provide tailored services, there is growing apprehension about the potential infringements on basic rights, such as privacy, self-determination, and fairness. The problem is not just a technical one; it is a constitutional one. How can progress be made without jeopardizing the dignity and liberties that form the foundation of democratic states?

B. Problem Statement: When Innovation Collides with Rights

AI is data-hungry and thrives on data that is behavioral, personal, or affinity-based often, precisely the kinds of data that people are most likely to consider private. Data that was generated in one context is fed into another, violating user expectations of the privacy of their communications and the purposes for which their information is to be used. This new era also leverages uniquely identifying data, often about devices

³ Shoshana Zuboff, *The Age of Surveillance Capitalism* 95-118 (2019)

and infrastructure, that can be indirectly identifying of individuals in unprecedentedly intrusive ways.⁴

The fundamental question here is how can we strike a balance between technological innovation and the safeguarding of human rights to ensure that AI serves as an enabler rather than a means of domination.

C. Objectives of the Study

The objectives of the study are to:

1. conduct a human rights assessment of AI with specific emphasis on the rights to privacy, autonomy, and dignity;
2. assess the adequacy of Indian and international laws applicable to data, AI, and automated decision-making systems;
3. study and articulate structural issues such as profiling, consent fatigue, and surveillance that undermine constitutional values;
4. develop a rights-based governance framework for AI that balances innovation and constitutionally mandated rights; and
5. propose specific policy and institutional strategies to enhance human-centred AI governance in India.⁵

D. Research Questions

The study is guided by the following questions:

1. How do AI technologies impact fundamental rights to privacy, autonomy, and equality?
2. What gaps exist in India's legal and institutional frameworks governing AI?
3. What global models offer guidance for balancing innovation with rights protection?
4. How can India design AI governance systems that respect dignity, freedom, and democratic values?

⁴ Vidhi Centre for Legal Policy, *Regulating Automated Decision-Making in India* (2022)

⁵ OECD, *Principles on Artificial Intelligence* (2019)

E. Research Methodology

This study adopts a doctrinal and analytical legal research methodology to examine the impact of Artificial Intelligence on privacy, autonomy, dignity, and equality. The research is primarily based on an analysis of legal principles, constitutional provisions, statutory frameworks, judicial precedents, and international regulatory standards relevant to AI governance. The primary sources relied upon include the Constitution of India, particularly Articles 14, 19, and 21; the Digital Personal Data Protection Act, 2023; leading Indian constitutional decisions such as *K.S. Puttaswamy v. Union of India* and *E.P. Royappa v. State of Tamil Nadu*; the European Union's General Data Protection Regulation; and the European Union Artificial Intelligence Act.

The study also relies on secondary sources, including academic scholarship, journal articles, policy papers, expert reports, and institutional guidelines issued by bodies such as the OECD, UNESCO, NITI Aayog, the Vidhi Centre for Legal Policy, and civil society organisations. A comparative method is used to benchmark India's emerging AI governance framework against international approaches, particularly the EU's risk-based model and the ethical principles developed by the OECD and UNESCO. This methodology enables the paper to identify gaps in India's current legal framework and propose rights-based reforms for responsible AI governance.

F. Significance in Global and Indian Context

Internationally, a human rights approach to AI governance is considered the default normative response for democratic societies. In practice, countries including the US, Canada, Japan, and members of the European Union have incorporated such principles in their AI policies and created oversight mechanisms to ensure compliance.⁶ India is the world's largest digital democracy and thus faces special challenges at the intersection of AI and democratic practice.

Its context of large-scale biometric systems (Aadhaar), expanding digital public infrastructure, and growing use of algorithmic tools by the state make questions of AI and democratic governance particularly pressing. The Supreme Court's

⁶ European Parliament, *General Data Protection Regulation* (2016)

unprecedented recognition of privacy as a fundamental right in *K.S. Puttaswamy* (2017) creates a constitutional mandate for ensuring that AI deployment does not compromise dignity or autonomy.⁷ This paper is important because, although India now has a formal AI governance framework in the form of the India AI Governance Guidelines issued by MeitY under the IndiaAI Mission in November 2025, it still does not have a binding, AI-specific statute comparable to the European Union's AI Act.

The Guidelines mark an important policy development by articulating principles for safe, inclusive, transparent, accountable, and innovation-oriented AI adoption. However, their advisory and techno-legal character means that rights-protective legal and regulatory paradigms remain necessary before AI systems become deeply entrenched in governance.

G. Scope and Limitations

The research is premised on the structural dominance and use of AI by states and powerful private entities to influence people's lives and surveil them at a scale incompatible with liberal constitutionalism. Also assumed is that the state plays a pre-eminent role in ensuring realisation of civil and political rights and in protecting against human rights abuses by private actors. Where inferences are drawn from these broad conceptual anchors, they are unpacked with specific reference to Indian constitutional jurisprudence and/or global human rights trends.

IV. HUMAN RIGHTS IN THE AGE OF AI – CONCEPTUAL FOUNDATIONS

A. Introduction

Artificial Intelligence is transforming the relationship between citizens, the state, and digital applications. The systems are assuming additional supervision of the administrative state, the interfaces for democracies, and the coordinative capacities of social and economic activities. The resilience, privacy, autonomy, and human rights of users are progressively becoming dependent on the abilities and limitations of

⁷ *K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1

automated decision-making systems. Analyzing the implications requires a grounding in the theoretical foundations of the most directly affected human rights.

B. Understanding Human Rights: Privacy, Autonomy, Dignity, and Equality

Human rights are normative safeguards protecting individuals from arbitrary interference by the state or private actors. Four rights face particular exposure in AI-driven societies:

1. **Privacy:** Privacy protects individuals' ability to control their personal information and maintain a private sphere free from intrusion.⁸ Modern AI systems undermine privacy through pervasive data collection, behavioural prediction, and profiling.
2. **Autonomy:** Autonomy refers to individuals' ability to make free, informed choices about their lives. AI threatens autonomy through nudging, behavioural manipulation, and targeted persuasion.⁹
3. **Dignity:** Human dignity forms the basis of all fundamental rights. AI systems that categorise, rank, or score individuals risk reducing them to data points, thereby undermining dignity.
4. **Equality:** AI systems may discriminate based on caste, gender, religion, or socioeconomic characteristics if trained on biased datasets. Equality, guaranteed under Article 14 in India, demands that AI systems avoid unjustified differential treatment.¹⁰

These rights are interlinked: loss of privacy can undermine autonomy and dignity; biased algorithms threaten equality.

C. Digital-Era Expansion of the Article 21 Framework

The right to life and personal liberty is guaranteed by the Indian Constitution in Article 21. The apex court in its historic verdict in *K.S. Puttaswamy v. Union of India*

⁸ Samuel D. Warren & Louis Brandeis, *The Right to Privacy*, 4 Harv. L. Rev. 193 (1890)

⁹ Susser, Roessler & Nissenbaum, *Online Manipulation: Hidden Influences in a Digital World*, 28 Geo. L. Tech. Rev. 1 (2019)

¹⁰ *E.P. Royappa v. State of Tamil Nadu*, (1974) 4 SCC 3

declared that privacy is part of life and personal liberty under Article 21 and is the seed for the growth of human values.¹¹

In the digital era, this right extends to:

1. informational privacy,
2. data protection,
3. autonomy in digital decision-making,
4. informational self-determination,
5. protection against mass surveillance.

AI technologies can infringe these expanded dimensions by creating persistent digital footprints, enabling profiling, and influencing individual choices without consent.

D. Theories of Autonomy: Informational, Decisional, and Cognitive Autonomy

Autonomy is not a singular concept; in the context of AI, it fractures into three distinct but overlapping dimensions:

1. **Informational Autonomy:** The definition of data autonomy is when you can decide what to do with your own data. AI-powered data ecosystems put this in jeopardy because they gather vast quantities of it without proper consent.
2. **Decisional Autonomy:** Being free to make decisions without being forced by external forces. AI-powered manipulation like target advertisement or political micro-targeting can lead to skewed decisions.¹²
3. **Cognitive Autonomy:** Having the freedom to think and form your opinions. Suggestion algorithms and echo chambers may influence your perceptions and limit the amount of information you are exposed to. AI can also pose a risk to independence through this indirect behavioral manipulation that users are unaware of.

¹¹ *K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1

¹² Zarsky, Privacy and Manipulation in the Digital Age, 20 J. Eth. & Info. Tech. 5 (2018)

E. Surveillance Capitalism and Behavioural Prediction

Shoshana Zuboff describes the modern data economy as “surveillance capitalism,” where human experience becomes raw material for commercial prediction and control.¹³ AI excels at extracting behavioural patterns, predicting future actions, and influencing preferences. This raises human rights concerns:

1. behavioural data is collected without genuine consent;
2. individuals become subjects of constant monitoring;
3. AI models are used to predict vulnerabilities for commercial or political manipulation.

Such practices undermine the freedom of thought and expression, both of which form part of constitutional and international human rights protections.

F. AI, Profiling, and the Threat of Manipulation

AI systems often depend on profiling, which is the classification of people into groups based on behavioral or demographic characteristics. Profiling can be used to:

1. deny services or opportunities,
2. create differential pricing,
3. identify “risky” individuals,
4. nudge users towards certain behaviours,
5. manipulate political opinions.¹⁴

All these issues raise questions about the existence of necessary conditions for rational agency, which are not solely dependent on internal cognitive processes but also on external conditions of agency. Given this, some scholars believe that external conditions of agency matter for AI in architecture, and elsewhere.

¹³ Shoshana Zuboff, *The Age of Surveillance Capitalism* 93–118 (2019)

¹⁴ Wachter, Normative Challenges of Algorithmic Profiling, 24 *Bus. & Info. Sys. Eng'g* 105 (2020)

G. Conclusion

Over the past few decades, human rights have adapted to the constraints and possibilities of emerging technologies. They have since been interpreted in ways that often succeed in placing limits on the power of new technologies modifying the expectations of States, corporations, and other entities that wield, develop, or promulgate such technologies. In the age of AI, the time has come to move from abstract allegations of rights violations to specific, fact-based arguments concerning how AI interferes with recognized human rights.

V. PRIVACY AND AUTONOMY RISKS IN AI SYSTEMS

A. Introduction

Artificial intelligence works by identifying patterns in data and drawing predictions or classifications based on those patterns. For all the promise AI holds for governance and business, these capabilities also pose deep risks to privacy, autonomy, dignity, and equality, risks arising from mass data extraction, behavioral monitoring, biometric identification, inscrutable decision-making, and ambient surveillance systems. This chapter identifies human rights directly threatened by AI systems and assesses the structural risks that render individuals, and especially minoritized groups, susceptible to algorithmic harms.

B. The Expanding Data Ecosystem: Personal, Sensitive, Biometric, and Inferred Data

AI systems thrive on volume, variety, and granularity of data. In the digital economy, data extends beyond personal identifiers to include:

1. **Personal Data:** Names, addresses, device identifiers, browsing history.
2. **Sensitive Personal Data:** Health status, caste, religion, financial data, sexual orientation.
3. **Biometric Data:** Fingerprints, iris scans, facial features, gait patterns.

4. **Inferred Data:** Predictions about behaviour, mental states, interests, or political preferences generated from seemingly innocuous data.¹⁵

The most concerning type is inferred information, as people cannot manage how such inferences are made or used the system can infer depression, pregnancy, political orientation, or credit risk, for example, with a high degree of certainty. This puts individuals at risk of discrimination, exploitation, and loss of self-determination.

C. Algorithmic Profiling and Behavioural Analytics

Creating patterns or classifying groups of people based on common traits or behaviours is known as profiling. AI does this by analyzing data to identify characteristics that people have in common.¹⁶

There are several risks associated with profiling:

1. discriminatory decisions in, for example, lending, hiring, and insurance;
2. personalised political targeting that aims to exploit psychological vulnerability;
3. individuals being excluded from opportunities based on the statistical behavior of their “lookalikes” (i.e. similar people who have gone on to behave in a certain manner in the past);
4. the widespread use of predictive and prescriptive analytics feeding ever more data into an ever-expanding maw of surveillance, since extensive data collection is, for the most part, necessary for effective profiling.

D. Surveillance, Mass Monitoring, and the New Intrusion

Surveillance systems powered by AI, such as facial recognition, gait analysis, and predictive policing, allow continuous monitoring of both physical and cyber spaces. In India, facial recognition technology (FRT) has been used in airports, railway stations, protests, and police work with no strong legal protections.¹⁷

Risks include:

¹⁵ Hildebrandt, *Privacy as Protection of the Incomputable Self*, 6 *Ethics & Info Tech.* 143 (2013)

¹⁶ Solon Barocas & Andrew Selbst, *Big Data’s Disparate Impact*, 104 *Calif. L. Rev.* 671 (2016)

¹⁷ Internet Freedom Foundation, *Panoptic Report on Facial Recognition in India* (2021)

1. continuous and invisible tracking of individuals;
2. mapping social networks without consent;
3. targeting of minority groups through over-policing;
4. chilling effects on speech, protest, and dissent.

AI-enabled surveillance transforms privacy from a right into a privilege that can be withdrawn based on opaque technological judgments.

E. Consent Fatigue and the Collapse of the Notice-and-Consent Model

Traditional data protection frameworks rely on the model of “notice and consent,” where users consent to data collection after reading legal notices. However, AI’s complexity and the scale of data processing make meaningful consent impossible.

Problems with Consent:

1. users lack technical understanding to evaluate risks;
2. consent is bundled into lengthy privacy policies;
3. refusal often denies essential services, especially in welfare contexts;
4. power imbalance between data subjects and data controllers.¹⁸

The Digital Personal Data Protection Act, 2023, now operationalised through the Digital Personal Data Protection Rules, 2025, gives statutory effect to India’s consent-and-notice framework for digital personal data. The Rules strengthen the operational content of consent by requiring clearer notices, consent management, data principal rights, breach-related obligations, and specific safeguards for children’s data. However, the framework remains substantially dependent on the notice-and-consent model.

In AI-driven environments, where data processing is complex, continuous, and often inferential, individuals may still be unable to meaningfully understand or control downstream uses of their data. Further, while the Act and Rules regulate digital personal data processing, they do not comprehensively govern AI systems

¹⁸ Nissenbaum, A Contextual Approach to Privacy, 79 Wash. L. Rev. 119 (2004)

themselves, including automated profiling, algorithmic decision-making, model opacity, or discriminatory outcomes arising from biased datasets.

F. Biometric and Facial Recognition Technologies

Biometric systems have greater dangers as biometric clues are permanent, unique, and cannot be replaced by others. If leaked, biometric data cannot be modified as passwords. AI systems that apply facial recognition have shown inaccuracies based on race and gender, with higher rates of incorrect identification for women and minority populations.¹⁹

Risks in India:

1. linking biometric databases with welfare schemes (Aadhaar),
2. police use of FRT without legal regulation,
3. use of facial analytics to identify protestors or dissidents,
4. potential creation of 360-degree surveillance systems through CCTV-AI integration.²⁰

These technologies are endangering our physical autonomy and deepening social inequities. They turn our bodies into a key that can be spied on, replicated, and used without constraint.

G. Chilling Effect on Free Speech and Informational Self-Determination

Automated content moderation, AI-based risk assessment tools, and other types of surveillance technology contribute to what's known as the "panopticon effect" where people change their actions because they know they're being watched. This impacts:

1. freedom of speech under Article 19(1)(a);
2. freedom of movement;
3. academic and political expression;

¹⁹ Joy Buolamwini & Timnit Gebru, *Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification*, 81 *Proc. Mach. Learning Res.* 1 (2018)

²⁰ IFF, *Facial Recognition Technology in India – State of Surveillance* (2022)

4. participation in protests and democratic activities.²¹

Predictive policing tools can result in the categorization of certain individuals or areas as “high risk,” prompting undue scrutiny and self-censoring behavior. Consequently, the chilling effect violates informational self-determination, which is the freedom to develop one’s public persona without unreasonable pressure or monitoring.

H. Case Studies Demonstrating Privacy and Autonomy Risks

1. **Cambridge Analytica (Global):** The scandal exposed how psychological profiling of millions of Facebook users enabled political manipulation during elections.²² This demonstrated how AI-driven microtargeting could distort democratic autonomy.
2. **Aadhaar and Welfare Delivery (India):** Algorithmic matching errors led to wrongful welfare exclusion in several states, illustrating how automated systems can undermine individual dignity and bodily autonomy.²³
3. **Facial Recognition in Policing (India):** FRT systems deployed by Delhi Police and other agencies have shown high error rates and lack statutory oversight. Their use during protests raises concerns about chilling dissent and targeting minority groups.²⁴

These examples reveal how AI can compromise both privacy and autonomy when deployed without adequate safeguards.

I. Conclusion

AI systems present grave threats to privacy and autonomy as they rely on massive data analytics, profiling, and ubiquitous surveillance. The invisible, continuous, and pre-emptive nature of such technologies not only heightens privacy and autonomy risks but also erodes the very conceptual underpinnings of these human rights. With India increasingly moving toward AI-driven governance structures, these risks

²¹ Jonathan Penney, Chilling Effects: Online Surveillance and Self-Censorship, 7 Berkeley Tech. L.J. 1 (2017)

²² Cadwalladr & Graham-Harrison, Revealed: Cambridge Analytica and Facebook Data Scandal, The Guardian (2018)

²³ Reetika Khera, Aadhaar Failures and Exclusion, 52 Econ. & Pol. Wkly. 5 (2017)

²⁴ IFF, *Delhi Police’s Use of FRT* (2020)

demand a rights-based risk preventive and protective regulatory response. The next chapter will evaluate the efficacy of current Indian laws and global responses such as the GDPR and the EU AI Act, at protecting privacy and autonomy vis-a-vis AI.

VI. LEGAL AND REGULATORY LANDSCAPE – INDIA AND THE WORLD

A. Introduction

You cannot understand AI governance without considering the legal and policy guidelines that govern data, privacy, automated decision-making, and human rights. India has a nascent legal landscape, but global jurisdictions have put in place more comprehensive structures. This chapter reviews India's constitutional and statutory guardrails, identifies deficiencies in the existing legal and policy environment, and measures them against international standards such as the EU GDPR, EU AI Act, OECD Principles, and UNESCO recommendations to glean pertinent takeaways for enhancing India's legal framework.

B. Indian Constitutional Protections: Articles 14, 19, and 21

The Indian Constitution provides a strong normative foundation for rights-based AI regulation.

1. **Article 14 – Equality and Non-Arbitrariness:** Article 14 says that all states should protect against discrimination that has no objective and reasonable justification. Issues like algorithmic bias, decision-making being hidden from the public, and unequal treatment based on incorrect or incomplete data are breaking both the obligations to be fair and ensure equal treatment.²⁵
2. **Article 19 – Freedom of Expression and Movement:** Surveillance, facial recognition, and algorithmic content moderation powered by artificial intelligence have the potential to chill speech, limit civic dissent, and impact

²⁵ *E.P. Royappa v. State of Tamil Nadu*, (1974) 4 SCC 3

democratic engagement. Such restrictions must pass the Article 19(2) proportionality test.

3. **Article 21 – Privacy, Dignity, and Autonomy:** In *Puttaswamy*, the Supreme Court held that privacy is a fundamental right intrinsic to dignity, autonomy, and personal liberty.²⁶

C. The Digital Personal Data Protection Act, 2023 (DPDPA)

The Digital Personal Data Protection Act, 2023, enacted as Act No. 22 of 2023 and operationalised through the Digital Personal Data Protection Rules, 2025, constitutes India's first comprehensive statutory framework for digital personal data protection. The Rules give practical effect to the Act by detailing notice and consent requirements, data principal rights, obligations of data fiduciaries, breach-reporting mechanisms, protections relating to children's personal data, and the constitution and functioning of the Data Protection Board of India. They also introduce a phased compliance framework, allowing regulated entities time to align their systems and practices with the new regime.

For AI governance, the DPDP Rules, 2025 are significant because they strengthen procedural safeguards around personal data processing, particularly through clearer consent standards, accountability duties, security safeguards, and institutional enforcement through the Data Protection Board. Nevertheless, their regulatory focus remains data protection rather than comprehensive AI governance. They do not directly regulate automated profiling, algorithmic explainability, bias testing, human oversight, model-risk classification, or contestability of AI-generated decisions.

Consequently, while the DPDPA and the DPDP Rules may reduce certain privacy risks associated with excessive or unlawful data collection, they do not fully address the broader constitutional concerns raised by AI systems, especially discriminatory outcomes, inferential profiling, opaque automated decision-making, and autonomy-related harms. India therefore still requires AI-specific legal safeguards alongside its data protection framework.

²⁶ *K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1

D. Global Frameworks: Learning from International Best Practices

1. European Union General Data Protection Regulation (GDPR)

The GDPR is the world's most influential data protection law. Key rights relevant to AI include:

- the right to access and correct data;
- the right to restrict processing;
- the right not to be subject to solely automated decisions producing legal effects;
- the right to explanation (interpreted through Recital 71).²⁷

These rights give people control over algorithmic decision-making, something that is currently not protected in India.

2. European Union AI Act (2024)

The EU AI Act is the world's first comprehensive AI regulation. It adopts a risk-based framework, classifying AI as:

- unacceptable risk (prohibited),
- high risk (strict compliance requirements),
- limited risk, and
- minimal risk.²⁸

High-risk AI systems such as those used in policing, healthcare, and public administration must comply with transparency, human oversight, accuracy testing, and robustness requirements. This approach provides a regulatory blueprint India can adapt while preserving constitutional mandates.

3. OECD Principles on Artificial Intelligence (2019)

²⁷ European Parliament, *General Data Protection Regulation*, art. 22 & Recital 71 (2016)

²⁸ Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), OJ L, 2024/1689, 12.7.2024.

The OECD Principles emphasise:

- human-centred values,
- transparency,
- robustness and safety,
- accountability mechanisms.²⁹

These global norms guide democratic AI governance and have been adopted by leading economies worldwide.

4. UNESCO Recommendation on the Ethics of AI (2021)

UNESCO's ethical guidelines call for:

- protecting human rights,
- banning mass surveillance,
- preventing bias in datasets,
- ensuring diversity and inclusivity in AI design.³⁰

The Recommendation is particularly useful for developing countries designing governance systems from the ground up.

E. India AI Governance Guidelines, 2025

In November 2025, the Ministry of Electronics and Information Technology, under the IndiaAI Mission, released the India AI Governance Guidelines as India's first comprehensive government-issued framework for AI governance. The Guidelines adopt a flexible, pro-innovation, techno-legal approach and are structured around seven guiding principles or "sutras": Trust is the Foundation, People First, Innovation over Restraint, Fairness & Equity, Accountability, Understandable by Design, and Safety, Resilience & Sustainability. They seek to promote responsible AI adoption

²⁹ OECD, *Principles on Artificial Intelligence* (2019)

³⁰ UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (2021)

across sectors while addressing risks to individuals and society through proportionate and adaptive governance mechanisms.

Unlike the European Union AI Act, which creates a binding, risk-tiered legal regime for unacceptable-risk, high-risk, limited-risk, and minimal-risk AI systems, the Indian Guidelines are primarily advisory and principle-based. Their strength lies in recognising the need for fairness, transparency, accountability, safety, and human-centricity in AI development and deployment. They also reflect India's preference for sectoral regulation, innovation sandboxes, capacity building, and the use of existing legal frameworks rather than immediate enactment of a comprehensive AI statute.

However, from a rights-based perspective, the Guidelines leave important gaps. They do not create enforceable individual rights against algorithmic harms, mandate AI Impact Assessments for all high-risk systems, impose binding transparency or explainability obligations, or establish a statutory right to contest automated decisions. Therefore, while the Guidelines significantly update India's AI governance landscape, they do not eliminate the need for a dedicated legal framework addressing profiling, biometric surveillance, automated decision-making, discrimination, and remedies for AI-related rights violations.

F. Conclusion

India has witnessed the increased deployment of AI systems that affect various human rights and lack effective regulation to monitor, prevent or address these impacts. These deployments and cognitive biases have disproportionate and unique impacts on historically marginalized populations. The intersecting nature of these biases could exacerbate social discrimination, undermine the value of equal citizenship, and undermine social relations among diverse communities in India.

VII. BALANCING INNOVATION WITH HUMAN RIGHTS TOWARD A RESPONSIBLE AI ARCHITECTURE

A. Introduction

Artificial Intelligence is expected to spur economic growth, increase administrative efficiency, and foster public-sector innovation in fields as diverse as health,

transportation, the environment, criminal justice, welfare, and the economy. However, it must be ensured that AI is not deployed in ways that undermine constitutional rights or subvert democratic values. To realize this, India needs a comprehensive strategy for fostering the development and deployment of rights-protective AI innovation.

B. Principles for Human-Centric Innovation

Human-centric innovation positions human rights, dignity, and welfare at the core of AI development and deployment. Key principles include:

1. **Primacy of Human Dignity:** AI must never undermine intrinsic human worth. Systems that score, rank, or categorise individuals without transparency threaten dignity and must be restricted.³¹
2. **Autonomy and Informational Self-Determination:** AI systems should support individuals' ability to make informed choices. Practices involving covert manipulation, dark patterns, or microtargeting compromise autonomy.
3. **Fairness and Non-Discrimination:** AI models must be tested for disparate impacts on caste, gender, religion, socioeconomic status, or disability.³²
4. **Accountability and Redress:** Stakeholders including developers, vendors, and state agencies must be accountable for harms caused by AI.

A rights-based AI governance model ensures that innovation enhances social equity rather than exacerbating structural inequalities.

C. Privacy-by-Design and Autonomy-by-Design

Privacy-by-design embeds privacy into the architecture of AI systems rather than relying on post-hoc controls.

Core components:

1. data minimisation,

³¹ Floridi et al., AI4People – Ethical Framework for a Good AI Society, 28 Minds & Machines 689 (2018)

³² Barocas & Selbst, Big Data's Disparate Impact, 104 Calif. L. Rev. 671 (2016)

2. purpose limitation,
3. decentralised data storage,
4. user control mechanisms, and
5. algorithmic transparency.

Autonomy-by-design extends these principles by ensuring that AI systems do not manipulate, coerce, or unduly influence individuals. Interfaces must avoid dark patterns, ensure informed consent, and provide exit options.³³

D. Oversight Institutions for AI and Data Protection

India must establish an Independent AI Ethics & Accountability Commission with statutory authority.

Mandate should include:

1. certification of high-risk AI systems,
2. mandatory audits of public-sector AI,
3. monitoring algorithmic impacts on rights,
4. adjudicating complaints and providing redress,
5. issuing binding codes of conduct and guidelines.³⁴

Composition: Experts from law, data science, ethics, social sciences, cybersecurity, and public policy.

Enforcement Powers: The Commission should have the power to halt harmful AI deployments, impose penalties, and mandate corrections. Strong oversight institutions prevent unregulated, opaque, and discriminatory AI deployment.

E. AI Impact Assessments and Human Rights Assessments

AI Impact Assessments (AIAs) must be made mandatory for all high-risk AI systems.

Pre-Deployment AIAs should evaluate:

³³ Cavoukian, Privacy by Design: The 7 Foundational Principles (2011)

³⁴ Vidhi Centre for Legal Policy, *A Framework for India's AI Regulator* (2023)

1. potential discriminatory impacts,
2. privacy risks,
3. proportionality of data collection,
4. accuracy and error rates,
5. vulnerability impacts on marginalised groups.³⁵

Post-Deployment AIAs: Regular reviews ensure that models do not drift into biased or harmful behaviour.

Human Rights Impact Assessments (HRIAs): HRIAs evaluate whether AI systems comply with constitutional rights, including dignity, equality, and autonomy.

F. Corrective Mechanisms, Remedies, and Accountability

A responsible AI architecture must provide meaningful redress for individuals affected by automated decisions.

Recommended mechanisms:

1. right to explanation,
2. right to human review,
3. right to contest automated decisions,
4. error correction mechanisms,
5. compensation for harm caused by AI systems.³⁶

Vendor accountability must be clearly defined. Private firms supplying AI tools to the government should share liability for algorithmic harms.

G. Conclusion

Ensuring innovations are assessed, managed, deployed, and governed responsibly is no less important. While innovation is vital for sustained growth, human and institutional capabilities are necessary to recognize the ethical implications, assess the

³⁵ Reisman et al., *Algorithmic Impact Assessments: A Practical Framework*, AI Now Institute (2018)

³⁶ OECD, *Responsible AI Principles* (2019)

risks, and take timely action. The lack of a governance mechanism fosters existing inequities, perpetuates unfair discrimination, and exacerbates various subjugations within society.

VIII. WAY FORWARD & CONCLUSION

A. Introduction

AI presents technical and legal challenges that have no easy solutions. Instead, India must adapt to continued uncertainty and rapid change, guided by core constitutional principles. The chapter concludes with a roadmap for robust, accountable, and rights-based AI governance in India, identifying what specific measures are needed, who must act, and when they must do it.

B. Guiding Principles for India's Future AI Policy

India's AI future must be grounded in constitutional values. The following principles should guide all regulatory efforts:

1. **Rights-Centric Governance:** All AI systems used by the government must adhere to the basic principles of equality, freedom of speech and life. To achieve this, discrimination must be removed, justice must be ensured, and human control must be maintained.³⁷
2. **Human Dignity and Autonomy:** We must not allow AI to dehumanize us into mere pieces of data. Protection by governance is needed to safeguard our right to make choices and decisions as well as our independence in our thoughts, beliefs, and ideas, especially from influences like manipulation, behavioral nudges, and unclear profiling technologies.
3. **Transparency and Contestability:** Citizens must be informed when AI influences decisions and must have accessible avenues to challenge these decisions.

³⁷ OECD, *Principles on Artificial Intelligence* (2019)

4. **Proportionality:** State use of AI, especially in surveillance and policing, must meet strict proportionality standards established by the Supreme Court in *Puttaswamy*.³⁸

These principles anchor AI governance within constitutional morality and democratic norms.

C. Strengthening Autonomy, Privacy, and Dignity in Governance

To build a rights-protective ecosystem, India must adopt structural safeguards:

1. **Mandatory Disclosure of AI Use:** Government agencies should publish transparency reports detailing where AI is used, its purpose, datasets, and safeguards.³⁹
2. **Right to Explanation and Human Review:** Individuals must have the right to know why an automated decision was made and the right to request human review.
3. **Data Minimisation and Purpose Limitation:** AI deployments should collect only the data strictly necessary for specific tasks.
4. **Avoiding Manipulative and Coercive AI Practices:** AI systems used for political microtargeting, behavioural manipulation, or exploitative nudging must be prohibited or strictly regulated.⁴⁰

These measures ensure that AI enhances, rather than suppresses, privacy and autonomy.

D. Building Institutional Capacity and Algorithmic Literacy

India requires strong institutional capacity to govern AI effectively.

1. **Judicial Capacity Building:** Courts must be trained to understand algorithmic logic, bias testing, explainability challenges, and rights-based AI assessment.⁴¹

³⁸ *K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1

³⁹ Access Now, *Algorithmic Transparency for the Public* (2021)

⁴⁰ Susser, Roessler & Nissenbaum, *Online Manipulation*, 28 Geo. L. Tech. Rev. 1 (2019)

⁴¹ Vidhi Centre for Legal Policy, *AI and the Indian Judiciary: Capacity Building Report* (2023)

2. **Regulatory Expertise:** Regulators must be equipped with interdisciplinary teams including lawyers, technologists, ethicists, and sociologists.
3. **Administrative Competence:** Public officials must be trained to critically evaluate algorithmic outputs and resist automation bias.
4. **Public Awareness and Literacy:** Algorithmic literacy campaigns can empower citizens to recognise when AI affects them, understand their rights, and seek redress.⁴²

Capacity building ensures that institutions can meaningfully oversee AI rather than defer blindly to technological systems.

E. Citizen Participation in AI Governance

Democratic legitimacy demands that citizens be active participants in the design of AI governance frameworks.

Mechanisms include:

1. open public consultations,
2. citizen assemblies on digital rights,
3. multi-stakeholder forums involving civil society, industry, and academia,
4. grievance redressal portals for AI-related harms.⁴³

Citizen participation enhances inclusivity, exposes blind spots in policy design, and builds public trust in AI.

F. Strategic Policy Recommendations

Based on global best practices and India's constitutional imperatives, the following policy recommendations are essential:

1. **Enact a Comprehensive AI Regulation Law:** India should build upon the India AI Governance Guidelines, 2025 by converting core rights-protective principles into enforceable statutory obligations. A comprehensive AI

⁴² UNESCO, *AI Ethics and Public Awareness Guidance* (2021)

⁴³ OECD, *Public Consultation Guidelines for AI Governance* (2022)

regulation law should define high-risk AI systems, mandate human oversight, require AI Impact Assessments and Human Rights Impact Assessments, provide transparency and contestability rights, and create remedies for persons affected by automated decisions. The Guidelines provide an important policy foundation, but binding legislation is necessary to ensure that fairness, accountability, transparency, and safety are not left solely to voluntary compliance.

2. **Establish an Independent AI Ethics & Accountability Commission:** It should conduct audits, certify AI systems, and investigate harms.⁴⁴
3. **Mandatory AI Impact Assessments (AIAs):** High-risk AI systems must undergo pre- and post-deployment AIAs, including discrimination testing.
4. **Limitations on Biometric and Facial Recognition Technologies:** Use must be subject to judicial authorization, strict proportionality, and independent oversight.
5. **Strengthen the Digital Personal Data Protection Act:** Amendments should address profiling, inferential data, automated decision-making, and meaningful consent.
6. **Promote Ethical AI Innovation:** Through regulatory sandboxes, startup incentives, open-source AI frameworks, and industry ethical guidelines.

These reforms will help India build a robust, future-proof AI governance structure.

G. Conclusion

Artificial Intelligence has become a defining feature of modern governance, economic activity, and social life. However, its rapid deployment also raises serious constitutional and human-rights concerns, particularly in relation to privacy, autonomy, dignity, equality, and democratic participation. This paper has shown that AI systems affect fundamental rights by enabling large-scale data extraction, inferential profiling, biometric identification, behavioural prediction, surveillance, and opaque automated decision-making. These risks are especially serious where

⁴⁴ Vidhi Centre for Legal Policy, *A Framework for India's AI Regulator* (2023)

individuals are unable to understand, contest, or meaningfully refuse the use of AI systems that shape access to welfare, employment, credit, policing, public services, or civic participation.

The analysis further demonstrates that India's existing legal framework provides an important but incomplete foundation for AI governance. Articles 14, 19, and 21 of the Constitution, together with the Supreme Court's recognition of privacy in *K.S. Puttaswamy v. Union of India*, establish strong rights-based limits on arbitrary, disproportionate, and dignity-infringing uses of technology. The Digital Personal Data Protection Act, 2023 and the DPDP Rules, 2025 strengthen India's data protection regime, while the India AI Governance Guidelines, 2025 provide a formal policy framework for responsible AI. Nevertheless, these instruments do not yet create a binding, comprehensive AI-specific regulatory structure comparable to the EU AI Act, nor do they fully address automated profiling, algorithmic discrimination, explainability, human oversight, or remedies for AI-related harms.

The paper therefore argues that India must adopt a rights-centric AI governance model that balances innovation with constitutional accountability. Such a model should include an independent AI Ethics and Accountability Commission, mandatory AI Impact Assessments and Human Rights Impact Assessments for high-risk systems, stronger transparency and contestability rights, limits on biometric surveillance, meaningful human review, and effective remedies for affected individuals. Responsible AI governance should not be understood as a barrier to innovation; rather, it is a necessary condition for ensuring that technological progress remains compatible with dignity, autonomy, equality, and democratic constitutionalism.

IX. BIBLIOGRAPHY

A. Books

1. Shoshana Zuboff, *The Age of Surveillance Capitalism* (2019).

B. Journal Articles

1. Samuel D. Warren & Louis Brandeis, *The Right to Privacy*, 4 Harv. L. Rev. 193 (1890).

2. Susser, Roessler & Nissenbaum, Online Manipulation: Hidden Influences in a Digital World, 28 Geo. L. Tech. Rev. 1 (2019).
3. Solon Barocas & Andrew Selbst, Big Data's Disparate Impact, 104 Calif. L. Rev. 671 (2016).
4. Hildebrandt, Privacy as Protection of the Incomputable Self, 6 Ethics & Info. Tech. 143 (2013).
5. Joy Buolamwini & Timnit Gebru, Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification, 81 Proc. Mach. Learning Res. 1 (2018).
6. Jonathan Penney, Chilling Effects: Online Surveillance and Self-Censorship, 7 Berkeley Tech. L.J. 1 (2017).
7. Zarsky, Privacy and Manipulation in the Digital Age, 20 J. Ethics & Info. Tech. 5 (2018).
8. Wachter, Normative Challenges of Algorithmic Profiling, 24 Bus. & Info. Sys. Eng'g 105 (2020).

C. Reports, Policy Papers & Expert Guidelines

1. Vidhi Centre for Legal Policy, *Regulating Automated Decision-Making in India* (2022).
2. Vidhi Centre for Legal Policy, *Regulating AI in India* (2023).
3. Vidhi Centre for Legal Policy, *A Framework for India's AI Regulator* (2023).
4. Vidhi Centre for Legal Policy, *AI and the Indian Judiciary: Capacity Building Report* (2023).
5. OECD, *Principles on Artificial Intelligence* (2019).
6. OECD, *Public Consultation Guidelines for AI Governance* (2022).
7. Access Now, *Algorithmic Transparency for the Public* (2021).
8. UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (2021).
9. UNESCO, *AI Ethics and Public Awareness Guidance* (2021).

10. Internet Freedom Foundation, *Panoptic Report on Facial Recognition in India* (2021).
11. Internet Freedom Foundation, *Facial Recognition Technology in India – State of Surveillance* (2022).
12. Reisman et al., *Algorithmic Impact Assessments: A Practical Framework*, AI Now Institute (2018).
13. Gebru et al., *Datasheets for Datasets*, 178 Commun. ACM 1 (2021).
14. NITI Aayog, *Responsible AI for All* (2021).
15. Cadwalladr & Graham-Harrison, *Revealed: Facebook–Cambridge Analytica Data Scandal*, The Guardian (2018).
16. Reetika Khera, *Aadhaar Failures and Exclusion*, 52 Econ. & Pol. Wkly. 5 (2017).
17. Ministry of Electronics and Information Technology, Government of India, *India AI Governance Guidelines* (2025).

D. Statutes & Regulations

1. Digital Personal Data Protection Act, No. 22 of 2023 (India).
2. European Parliament, *General Data Protection Regulation* (GDPR), Regulation (EU) 2016/679.
3. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act), OJ L, 2024/1689, 12.7.2024.

E. Cases

1. *K.S. Puttaswamy v. Union of India*, (2017) 10 SCC 1 (India).
2. *E.P. Royappa v. State of Tamil Nadu*, (1974) 4 SCC 3 (India).

F. International Standards & Instruments

1. UNESCO, *Recommendation on the Ethics of Artificial Intelligence* (2021).
2. OECD, *Principles on Artificial Intelligence* (2019).