International Journal for Innovative Engineering and Management Research
PEER REVIEWED OPEN ACCESS INTERNATIONAL JOURNAL
www.ijiemr.org

AI and Human Rights A Rights-Based Analysis of Ethical Governance

Mr. K Uttham Varma | Mr. Sharath Sanjeev | Mr. P Jashwanth | Mr. B Kiran

BS-CS, Vishwa Vishwani Institute of Systems and Management, Hyderabad, Telangana, India
Email: kuthamvarma@gmail.com | Mobile: 9989357808

Abstract

Artificial Intelligence (AI) is transforming governance, healthcare, education, employment, and social development. Despite its potential to promote progress, its growing influence raises ethical concerns regarding human dignity, equality, privacy, and fundamental freedoms. This article examines the relationship between AI ethics and human rights, focusing on algorithmic bias, systemic discrimination, digital surveillance, privacy violations, misinformation, employment displacement, and unequal technological access. It analyses accountability gaps in automated decision-making, transparency in algorithmic processes, and the protection of vulnerable and marginalised communities. Furthermore, it explores the responsibilities of governments, technology developers, and institutions in establishing effective regulatory frameworks. Particular emphasis is placed on human oversight, data protection, rights-based impact assessments, constitutional values, and inclusive governance. The article argues that technological innovation must be guided by ethical principles and human rights obligations. Transparent decision-making, enforceable legal safeguards, and participatory governance are essential to prevent discrimination and ensure accountability. A human-centred approach to AI governance can reconcile technological advancement with social justice, democratic values, and individual autonomy. Ultimately, AI should serve humanity by strengthening fundamental rights rather than undermining them.

Keywords: Artificial Intelligence, Human Rights, AI Ethics, Algorithmic Bias, Data Privacy, Ethical Governance, Accountability, Social Justice

1. Introduction

Artificial intelligence is increasingly used to classify information, generate content, recommend services, assess risk and support decisions in public and private institutions. These capabilities can improve access to information, assist medical diagnosis, support translation, identify patterns in complex data and make some services more efficient. They can also affect decisions about a person's education, employment, credit, healthcare, safety, speech and access to public benefits.

AI therefore raises questions beyond technical performance: who benefits, who bears the risk, what rights may be affected and how a person can challenge a decision.

Human rights provide a practical framework for answering these questions. Rights to equality, privacy, freedom of expression, work, education, health, association and an effective remedy are relevant throughout an AI system's lifecycle. A system may be statistically accurate on average yet perform poorly for a particular group. A decision may be efficient yet impossible to explain or contest. Data collection may be lawful in form yet excessive in scope or difficult to refuse. Evaluating AI through human rights helps identify such effects before they become routine features of institutional decision-making.

These questions matter as AI moves from optional tools into everyday services and public infrastructure. Automated decisions may be repeated at scale, while affected people often have little visibility into the data, criteria or institutional responsibility behind them. Risks are not inevitable properties of technology. They depend on choices about purpose, training data, design, deployment, oversight and remedies. Governance should protect people while permitting beneficial innovation under clear safeguards.

This article examines the ethical and human-rights implications of AI, with particular attention to India. It discusses discrimination, privacy and surveillance, expression and information integrity, due process, employment, access and inclusion. It then considers international principles and Indian constitutional and policy contexts before proposing a rights-based governance framework. Its central argument is that AI should be assessed by its impact on people's rights and agency, not solely by accuracy, speed or commercial value.

2. Objectives, Research Questions and Method

2.1 Objectives

The article has three objectives: to identify principal human-rights risks associated with AI-enabled decision-making; to examine the duties and responsibilities of governments, technology providers and deploying institutions; and to develop practical recommendations for transparent, accountable and inclusive AI governance relevant to India.

2.2 Research questions

The discussion is guided by three questions: (1) How can AI systems affect equality, privacy, liberty, expression, work and social rights? (2) What accountability gaps arise when consequential decisions are automated or supported by algorithmic tools? (3) Which legal, institutional and participatory safeguards can align AI deployment with human dignity and democratic values?

2.3 Method and scope

This is a conceptual and policy-oriented paper based on a focused review of human-rights instruments, ethical guidance, Indian constitutional and statutory materials, and relevant academic literature. It does not present original survey data, interviews or a technical audit of a particular AI system. Its examples explain plausible mechanisms of impact rather than claiming that a named institution has caused a specific violation. “AI system” broadly includes systems that infer, classify, recommend, predict or generate outputs that influence decisions or access to opportunities.

The analysis follows a lifecycle perspective: design and data collection, development and testing, deployment, monitoring and remedy. A system’s effects can change when it is used in a new setting, combined with other data or applied to a population unlike the one represented in its training and evaluation. A rights-based approach also requires attention to unequal social conditions; an identical system can impose different burdens on people with different levels of power, access or ability to appeal.

3. Human Rights as a Framework for AI Governance

Human rights are not a checklist applied only after a system causes harm. They provide standards for public action and business conduct throughout the system lifecycle. International instruments such as the Universal Declaration of Human Rights, the International Covenant on Civil and Political Rights and the International Covenant on Economic, Social and Cultural Rights recognise rights implicated by data-intensive and automated systems. These include equality before the law, privacy, freedom of opinion and expression, work, education, health and participation in public affairs.

The United Nations Guiding Principles on Business and Human Rights distinguish the State duty to protect human rights, the corporate responsibility to respect rights and the need for access to

remedy when harm occurs (United Nations Human Rights Council, 2011). For AI, governments should maintain effective safeguards; organisations should identify, prevent and mitigate adverse impacts connected to their systems; and affected people should be able to seek meaningful correction or remedy. Narrow legal compliance may not reveal broader or cumulative effects on rights.

UNESCO's Recommendation on the Ethics of Artificial Intelligence places human dignity and human rights at the centre of AI governance. It emphasises proportionality, privacy, fairness, transparency, accountability, human oversight, inclusion and ethical impact assessment (UNESCO, 2021). These principles distinguish technical functionality from social legitimacy, which also requires examining purpose, context, distribution of benefits and burdens, and availability of effective review.

3.1 Ethical principles and rights-based duties

Voluntary AI ethics principles such as fairness, accountability, transparency and safety can help organisations identify relevant values and improve design. Their practical effect, however, depends on whether institutions translate them into measurable duties, oversight and remedies. Human rights add established standards for public action and business conduct, and make clear that people affected by a system should be able to seek an effective remedy. These approaches are complementary: ethical principles can guide responsible design, while rights-based duties provide a firmer basis for accountability (United Nations Human Rights Council, 2011; UNESCO, 2021).

A rights-based approach does not make every technical or ethical question simple. Rights may interact, and transparency may need to be balanced with privacy, safety and security. The aim is to make decisions reviewable at the level appropriate to their impact. Where an AI-supported decision can materially affect liberty, livelihood, access to essential services or other important interests, the responsible institution should explain the system's role, preserve meaningful human review and provide a route to correct errors.

4. Key Human-Rights Challenges

4.1 Equality, bias and systemic discrimination

AI systems learn patterns from data and institutional practices that may reflect existing inequality. Historical records can encode unequal access to employment, credit, education or healthcare. When these patterns are used to predict future suitability or risk, a system may reproduce them under the appearance of objective measurement. Bias can also arise from incomplete data, poorly chosen labels, proxy variables, design assumptions or the way a system is used. Removing explicit categories such as gender or caste does not necessarily eliminate discrimination if other variables act as proxies.

The consequences depend on context. A hiring model may rank applicants using signals correlated with past exclusion. A risk assessment tool may direct greater scrutiny toward communities already subject to disproportionate monitoring. A language model may be less reliable for underrepresented languages or dialects. Aggregate accuracy can hide uneven error rates and unequal consequences. The relevant question is not only whether a system predicts correctly overall, but whether its use is fair and justified in light of the affected rights.

Fairness assessment should begin before deployment and continue as data and use conditions change. Institutions should identify relevant groups, test for disparate outcomes, document limitations and provide processes for challenging errors. Where sensitive data are used to detect discrimination, safeguards are needed to prevent the assessment from becoming another source of surveillance. People affected by a system should help define the harms to be measured and the remedies that are practical. (Noble, 2018).

4.2 Privacy, data protection and digital surveillance

AI can infer sensitive attributes from data that appear ordinary when viewed separately. Location histories, browsing patterns, voice, images, purchasing behaviour and social connections may be combined to infer interests, health status, identity or vulnerability. Risks include collection beyond necessity, reuse for new purposes, insecure retention, opaque sharing and profiling without meaningful understanding or choice. People may be unable to participate in essential services without disclosing data, weakening the practical value of consent.

Surveillance concerns extend beyond secrecy. Persistent monitoring can chill speech, association and movement, particularly when people do not know what is recorded, how long it is retained or how it may affect future treatment. Both state and private systems can create these risks. Privacy protections should include clear legal authority, necessity and proportionality, purpose limitation, data minimisation, security, retention controls and independent oversight. Privacy by design treats these protections as engineering and institutional requirements rather than user preferences alone. The commercial use of behavioural data also raises questions about prediction and influence (Zuboff, 2019).

The Supreme Court of India has recognised privacy as a constitutionally protected right connected with liberty and dignity, and has discussed legality, legitimate aim and proportionality in assessing restrictions (Justice K. S. Puttaswamy (Retd.) v. Union of India, 2017). India's Digital Personal Data Protection Act, 2023 establishes a statutory framework for digital personal data, with commencement phased by government notification. As of October 2026, the commencement schedule should be checked provision by provision; the Act's principal processing obligations are scheduled to take effect in 2027 under the November 2025 notification (Government of India, 2023, 2025). Ethical privacy safeguards remain relevant throughout implementation.

4.3 Freedom of expression, information and democratic participation

Automated content ranking, moderation and account restrictions can affect which information is visible and whose speech reaches an audience. Moderation systems may struggle with satire, quotation, political dissent, documentation of violence or expression in languages with fewer resources. Over-removal can restrict lawful expression and access to information; under-enforcement can leave people exposed to threats, incitement or targeted harassment. Effective protection from abuse must coexist with safeguards against unjustified restriction.

AI-generated or manipulated content creates additional challenges for information integrity. Synthetic images, audio and video can be used for legitimate creative work, but can also impersonate individuals or mislead audiences. Automated systems may help identify known forms of manipulation, yet labels can be mistaken and provenance signals incomplete. Responses should be transparent, proportionate and open to appeal. Restrictions on speech should rest on clear rules, legitimate aims and due process, rather than opaque assessments that people cannot contest.

Democratic participation depends on access to diverse and reliable information. Personalised systems can make relevant material easier to find, but may also narrow exposure or amplify content according to criteria users cannot see. Evidence about recommendation effects must be assessed carefully; exposure is not persuasion, and platform algorithms are only one part of a larger information ecosystem. Independent research access and meaningful system disclosure can support public scrutiny without indiscriminate publication of source code or personal data.

4.4 Due process, accountability and access to remedy

When an AI-supported decision affects rights or opportunities, accountability can become fragmented among the developer, vendor, deploying institution and decision-maker. A person may receive a denial or restriction without knowing which data were used, which criteria mattered or who can correct the result. This opacity creates a practical barrier to due process and remedy. The organisation that relies on a system should remain responsible for decisions made with its assistance; automated outputs cannot excuse a failure of human responsibility.

A rights-respecting process should provide timely notice, an intelligible explanation appropriate to context, access to relevant records where appropriate, and a way to request human review. Reviewers need authority to change the outcome rather than merely confirm the machine's recommendation. Institutions should keep records sufficient to investigate complaints, monitor disparate impacts and show that safeguards operate. Remedies may include correcting data, reconsidering a decision, restoring access, compensation where legally available, or changing the system to prevent recurrence. (Pasquale, 2015).

4.5 Work, livelihoods and social and economic rights

AI can automate tasks, alter skill requirements and reorganise work. It may reduce repetitive or hazardous tasks and create new forms of employment, while also displacing some workers, intensifying monitoring or shifting risk to workers with limited bargaining power. Effects vary by sector, region, education and access to retraining. Broad claims that AI will either eliminate work or benefit all workers overlook this diversity.

A rights-based response considers working conditions as well as job numbers. Algorithmic management may allocate shifts, measure performance or recommend discipline through metrics workers cannot inspect. Workers need clear information about automated monitoring and

evaluation, access to challenge inaccurate records and protection against discriminatory or arbitrary decisions. Governments and employers should assess transitions, provide meaningful training and social protection, and consult workers before deploying systems that substantially alter working conditions.

4.6 Health, education and access to essential services

AI tools in healthcare, education and public services may help allocate attention and identify needs, but errors can affect fundamental interests. A clinical decision-support system may perform unevenly for populations missing from its data. An educational tool may misclassify a learner or narrow opportunities based on past performance. A benefits-screening system may delay support because of incomplete records or an unexplained risk score. People affected may have little ability to choose another provider or refuse automated processing.

Institutions should validate systems for the population and setting in which they will be used, preserve qualified human decision-making, communicate uncertainty and create routes for correction. High-impact decisions about healthcare, education, eligibility or liberty should not be delegated to an automated output without appropriate review and accountability. Public procurement should require evidence of safety, accessibility, data protection and ongoing performance rather than relying solely on a vendor's claims.

4.7 Unequal access and the digital divide

The benefits of AI depend on access to devices, connectivity, accessible interfaces, relevant language support and digital skills. Unequal access can exclude people from new services or make them dependent on intermediaries. Systems designed around dominant languages, urban users or standardised communication may not serve rural communities, people with disabilities, older adults or linguistic minorities equally. Inclusion means more than nominal access: systems should be usable, affordable, appropriate and open to feedback from affected users.

Digital inclusion also requires caution about data extraction. Communities should not have to surrender extensive personal data to receive basic services. Public institutions should offer non-digital or assisted routes where feasible, ensure accessibility and assess whether automated service delivery creates new barriers. Participation in design and evaluation can reveal practical obstacles that technical teams may not anticipate.

5. India: Constitutional Values and Governance Context

India's constitutional framework offers important reference points for evaluating AI. Article 14 protects equality before the law and equal protection; Article 19 protects specified freedoms, subject to constitutionally permitted restrictions; and Article 21 protects life and personal liberty. The Supreme Court's privacy jurisprudence links privacy to liberty, dignity and autonomy. These rights are relevant when public or private systems process personal information, classify people, allocate opportunities or restrict expression. Their application depends on the facts and applicable law, and this article does not offer case-specific legal advice.

India's social and linguistic diversity requires context-sensitive assessment. Data and evaluation sets may not represent all languages, dialects, regions, genders, castes, disabilities or economic circumstances. An error that appears minor in a general test may have serious consequences for a person seeking healthcare, public benefits or employment. Human-rights impact assessments should identify affected groups and consider direct and indirect discrimination, while using privacy safeguards when gathering information needed to test equality.

The Digital Personal Data Protection Act, 2023 and the Information Technology Act and related intermediary rules form part of the evolving policy environment. The DPDP Act's phased commencement means institutions should avoid presenting all statutory duties as already operative; current commencement notifications and rules should be consulted. Beyond data protection, AI governance intersects with constitutional freedoms, sectoral regulation and administrative accountability. Implementation should avoid fragmented oversight that leaves affected people unsure where to complain or seek redress.

Regulation must also protect against misuse of AI by public authorities. Systems for surveillance, policing, identity verification or public communication can affect privacy, equality and liberty. Any state use should have a clear legal basis, a legitimate and specific purpose, necessity and proportionality, safeguards against misuse, independent review and effective remedies. Public authorities should be transparent about significant AI uses, subject to narrowly tailored exceptions that protect genuine security interests without eliminating accountability.

6. Governance Responsibilities and Policy Recommendations

6.1 Human-rights impact assessment across the lifecycle

Governments, public institutions and technology providers should conduct rights-based impact assessments before high-impact systems are procured or deployed and repeat them after major changes. Assessments should identify affected rights and groups, examine foreseeable direct and indirect harms, evaluate alternatives and specify mitigation, monitoring and remedy. They should be multidisciplinary and include affected communities, with public summaries where disclosure does not create a concrete privacy or security risk. Assessment should influence whether and how a system is used, not simply document a decision already made.

6.2 Transparency and explainability suited to the decision

People should be informed when AI materially supports decisions about them, particularly in essential services and other high-impact contexts. An explanation should identify the system's role, principal factors relevant to the outcome, important limitations and available review routes. Regulators and independent researchers may require deeper technical documentation, data access or audit logs. Transparency should be proportionate: useful disclosure must support accountability while protecting personal data, security and legitimate confidential information.

6.3 Human oversight and meaningful appeal

Human review is necessary when a decision can significantly affect rights, livelihood, safety or access to essential services. Reviewers should have adequate expertise, access to relevant context and authority to override an automated recommendation. Complaint channels should be accessible, timely and available in relevant languages. Institutions should monitor whether appeals produce corrections and address patterns of error rather than treating each complaint as isolated.

6.4 Fairness, accessibility and inclusive participation

Systems should be tested across relevant populations and contexts before and after deployment. Testing should consider differences in error rates, severity of consequences and whether a person can realistically seek correction. Design teams and evaluators should include expertise in law, social science, accessibility, language and the affected sector. Consultation should be early enough to change design choices and should not substitute for legal duties or independent oversight.

6.5 Privacy, data governance and security

Data collection should be limited to legitimate purposes and necessary inputs, with appropriate retention limits, security controls and restrictions on secondary use. Organisations should assess privacy risks at the design stage, protect sensitive data and ensure that people can exercise applicable rights without excessive burden. Public-sector data linkage and biometric or behavioural monitoring require particular scrutiny. Security measures should address unauthorised access, data leakage and misuse of generated outputs.

6.6 Work transition and social protection

Employers and governments should assess how automation affects jobs, work quality, surveillance and bargaining power. Workers and their representatives should receive information and opportunities to participate before major changes. Training and transition programmes should be accessible and linked to actual employment pathways. Automated management should not remove a worker's ability to understand, question and correct consequential evaluations.

6.7 Independent oversight and access to remedy

Regulators need clear mandates, technical and human-rights expertise, adequate resources and independence from the organisations they oversee. Audits should be conducted by qualified bodies with access sufficient to evaluate claims and safeguards. People affected by AI-related harm should have access to judicial and non-judicial remedies. Organisations should maintain incident reporting, complaint handling and remediation processes, and publish aggregate information about significant impacts and corrective action.

6.8 International cooperation and public literacy

AI systems and data flows often cross borders, while human-rights protections and enforcement capacity vary. Governments should cooperate on common standards for assessment, transparency, safety and remedies while preserving the ability to adopt stronger domestic safeguards. Public education should build understanding of AI, data rights and automated decision-making in accessible languages and formats. Literacy supports informed participation, but responsibility for system design and rights protection remains with governments and deploying organisations.

7. Discussion

A human-rights approach does not require rejecting AI or assuming every automated system is harmful. It asks whether the purpose is legitimate, the system is necessary and proportionate, its effects are fairly distributed, and affected people can understand and challenge consequential outcomes. This approach permits innovation while requiring stronger safeguards where decisions are difficult to reverse or where people depend on a service. It also draws attention to cumulative effects: multiple low-impact data uses can combine into intrusive profiling, and several automated steps can obscure responsibility even if no single step appears decisive.

The framework must address both public and private power. States have obligations to protect rights and may use AI in service delivery, administration or security. Businesses have responsibilities to respect rights and conduct due diligence, while users need effective remedies. Accountability should follow the institutions that choose, procure and deploy systems; the complexity of a model cannot transfer responsibility to the model itself. Independent scrutiny is necessary because organisations often possess far more information about a system than the people affected by it.

In India, implementation should be guided by constitutional values, social diversity and practical access to remedies. Rules should be clear enough to constrain arbitrary decisions, while sufficiently adaptable to respond to changing evidence. Human-rights impact assessments, independent audits and participatory processes can connect broad principles to concrete system design. Their value depends on whether findings lead to changes, whether affected communities can participate meaningfully and whether remedies are available when prevention fails.

8. Conclusion

AI can support progress in governance, healthcare, education, employment and social development, but its benefits do not automatically protect human rights. Algorithmic bias, surveillance, privacy violations, misinformation, employment disruption and unequal access can undermine dignity, equality, autonomy and democratic participation. Accountability gaps become especially serious when decisions are opaque, difficult to contest or embedded in essential services.

The response should be grounded in enforceable rights protections and practical institutional safeguards. Human-rights impact assessment, data protection, transparent decision-making,

independent oversight, meaningful human review, accessible remedies and inclusive participation should guide AI throughout its lifecycle. In India, these measures should be implemented consistently with constitutional values and sensitivity to linguistic and social diversity. AI should serve humanity by strengthening fundamental rights, not by making their protection more difficult to claim.

References

Constitution of India. (1950). Government of India.

Government of India. (2023). The Digital Personal Data Protection Act, 2023 (Act No. 22 of 2023). Ministry of Law and Justice. https://www.meity.gov.in/writereaddata/files/Digital%20Personal%20Data%20Protection%20Act%202023.pdf

Government of India. (2025). Notification on phased commencement of the Digital Personal Data Protection Act, 2023 (G.S.R. 843(E), 13 November 2025). Ministry of Electronics and Information Technology. https://www.meity.gov.in/static/uploads/2025/11/c56ceae6c383460ca69577428d36828b.pdf

International Covenant on Civil and Political Rights, 999 U.N.T.S. 171 (1966). https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-civil-and-political-rights

International Covenant on Economic, Social and Cultural Rights, 993 U.N.T.S. 3 (1966). https://www.ohchr.org/en/instruments-mechanisms/instruments/international-covenant-economic-social-and-cultural-rights

Justice K. S. Puttaswamy (Retd.) v. Union of India, (2017) 10 SCC 1.

United Nations. (1948). Universal Declaration of Human Rights. https://www.un.org/en/about-us/universal-declaration-of-human-rights

United Nations Human Rights Council. (2011). Guiding principles on business and human rights: Implementing the United Nations “Protect, Respect and Remedy” framework.

United Nations. https://www.ohchr.org/Documents/Publications/GuidingPrinciplesBusinessHR_EN.pdf

UNESCO. (2021). Recommendation on the ethics of artificial intelligence. https://www.unesco.org/en/legal-affairs/recommendation-ethics-artificial-intelligence

Noble, S. U. (2018). Algorithms of oppression: How search engines reinforce racism. New York University Press.

Pasquale, F. (2015). The black box society: The secret algorithms that control money and information. Harvard University Press.

Zuboff, S. (2019). The age of surveillance capitalism: The fight for a human future at the new frontier of power. PublicAffairs.