



DOI: <https://doi.org/10.38035/sijal.v4i2>
<https://creativecommons.org/licenses/by/4.0/>

Received: 06-09-2026 | Revised: 26-09-2026 | Accepted: 01-10-2026

Reorienting Human Rights Protection in an Artificial Intelligence-Based Digital Ecosystem

Acep Wahyu¹, Joko Sriwidodo², Lukman Hakim³

¹Universitas Bhayangkara Jakarta Raya, Bekasi, Indonesia, acepw1978@gmail.com

²Universitas Bhayangkara Jakarta Raya, Bekasi, Indonesia, -

³Universitas Bhayangkara Jakarta Raya, Bekasi, Indonesia, -

Corresponding Author: acepw1978@gmail.com¹

Abstract: Artificial intelligence mediates access to employment, finance, health care, education, public services, and security, thereby transforming human rights risks from identifiable human acts into distributed sociotechnical processes. This article examines whether Indonesia's existing legal framework adequately protects rights against opaque, autonomous, and transnational artificial intelligence systems and formulates a rights-oriented governance model. Using normative legal research, it combines statutory, conceptual, and comparative approaches and analyses Indonesian constitutional guarantees, human rights law, personal data protection law, electronic information law, global ethical standards, and the European Union's risk-based regulation. The study finds that current Indonesian rules provide fragmented protection but do not establish an integrated allocation of responsibility across the artificial intelligence life cycle. Privacy violations, discriminatory outcomes, deficient explanations, and ineffective remedies may consequently arise without a clearly accountable actor. The article proposes the RIGHTS-AI model, comprising risk classification, impact assessment, governance accountability, human oversight, transparency and traceability, safeguards and security, accessible remedies, and independent supervision. Its novelty lies in connecting individual remedies with preventive duties imposed on developers, providers, deployers, and public authorities. This model enables innovation while keeping automated power subordinate to human dignity, equality, legal certainty, and democratic accountability within Indonesia's digital ecosystem and strengthens sustainable rights protection.

Keywords: algorithmic accountability, artificial intelligence, digital governance, human rights, personal data protection.

INTRODUCTION

Artificial intelligence (AI) has moved from specialised computation into the ordinary infrastructure of contemporary life. Machine-learning systems classify information, detect patterns, generate predictions, rank individuals, and produce content at a scale that cannot be

replicated by unaided human decision makers. Public authorities and private organisations employ these capabilities in health care, education, finance, recruitment, transportation, commerce, communication, border control, policing, and administrative services. Generative AI further expands the field by producing text, images, audio, software code, and synthetic representations that may influence knowledge, identity, and public discourse. Consequently, AI is no longer merely an instrument used within social institutions. It increasingly constitutes the environment through which individuals obtain services, become visible to institutions, and exercise legal rights (Calo, 2017; Russell & Norvig, 2020).

The societal benefits are substantial. Automated systems may accelerate diagnosis, detect fraud, improve the allocation of public resources, personalise education, expand accessibility, and reduce repetitive administrative work. Properly designed technologies can support public welfare and make previously scarce expertise more widely available. Yet the same capacity to infer, classify, and predict produces a new concentration of informational power. Organisations can know individuals through patterns that individuals neither provided directly nor understand. An apparently technical score may determine whether a person receives credit, employment, insurance, medical treatment, public assistance, or heightened security attention. The efficiency of a system therefore cannot be separated from the legal significance of the decisions that it shapes (Floridi et al., 2018; Zuboff, 2019).

Human rights law traditionally presumes that legally significant conduct can be attributed to an identifiable actor. A public officer issues a decision, a corporation processes data, or an individual commits a harmful act. AI complicates this architecture. A single output may depend on data collected by one entity, labelled by another, processed through a model developed elsewhere, delivered through a cloud platform, configured by an intermediary, and adopted by an institution in a different jurisdiction. No participant may possess complete knowledge of the system. When harm occurs, each actor may point to another component in the chain, while the affected person encounters a technically complex and institutionally fragmented process. The problem is not that algorithms literally become legal persons; it is that distributed automation may obscure the persons and institutions that remain responsible.

Opacity intensifies this difficulty. Some machine-learning systems generate results through relationships that are difficult to explain in ordinary legal language. Commercial secrecy, cybersecurity concerns, technical complexity, and organisational incentives may further restrict disclosure. An individual may know the outcome but not the decisive factors, data sources, confidence level, or possibility of error. Without meaningful information, rights to object, correct data, seek human review, and obtain an effective remedy become formal promises rather than practical guarantees. The "black box" is thus not exclusively a technical feature. It is also a legal and institutional condition produced by choices about documentation, procurement, access, audit, and responsibility (Pasquale, 2015; Wachter et al., 2017).

Bias demonstrates that computational neutrality cannot be presumed. Historical data may encode unequal access to employment, credit, housing, education, health care, or law enforcement. Proxy variables may reproduce distinctions associated with gender, disability, ethnicity, age, location, or socioeconomic status even when protected characteristics are removed. A model optimised for aggregate accuracy can impose systematically higher error rates on underrepresented groups. Discrimination may therefore result without an express intention to discriminate. It can be embedded in problem formulation, data selection, label

definition, threshold setting, deployment context, and institutional reliance on automated recommendations (Barocas & Selbst, 2016; Buolamwini & Gebru, 2018; Selbst et al., 2019).

The Indonesian constitutional framework recognises dignity, equality before the law, legal certainty, personal security, communication, information, and protection of the person, family, honour, and property. Law Number 39 of 1999 concerning Human Rights provides a general rights framework, while Law Number 27 of 2022 concerning Personal Data Protection regulates the processing of personal data and the obligations of controllers and processors. Law Number 1 of 2024, which amended the Electronic Information and Transactions Law, addresses electronic systems and unlawful digital conduct. Minister of Communication and Information Circular Letter Number 9 of 2023 provides ethical guidance for AI. Together these instruments establish important normative foundations, but none offers a comprehensive legal regime for AI systems that differentiates risk, allocates life-cycle duties, guarantees explanations, or coordinates independent supervision.

The regulatory gap is increasingly significant because AI applications do not remain confined to consumer convenience. Facial recognition, predictive analytics, biometric verification, automated eligibility assessments, and generative systems may affect rights at scale. Personal Data Protection Law can govern much of the data layer, yet not every harmful inference is easily characterised as unlawful data processing. Electronic information law may address particular misuse but does not supply a complete framework for algorithmic fairness or high-risk automated decisions. Ethical guidance is valuable as a source of principles, but a circular letter cannot by itself establish the enforceable institutional architecture required for serious rights impacts. Fragmented regulation risks producing fragmented accountability.

International developments offer relevant, although not mechanically transferable, lessons. The United Nations Educational, Scientific and Cultural Organization's Recommendation on the Ethics of Artificial Intelligence places human dignity, human rights, proportionality, fairness, transparency, accountability, and human oversight at the centre of AI governance (UNESCO, 2021). The European Union's Artificial Intelligence Act adopts differentiated obligations based on risk, prohibits selected practices, and imposes governance requirements on high-risk systems (European Parliament & Council of the European Union, 2024). These instruments demonstrate a movement from voluntary ethics toward operational duties, documentation, risk management, conformity assessment, monitoring, and enforcement. Their value for Indonesia lies in the governing logic rather than in direct institutional transplantation.

Existing scholarship has extensively examined privacy, algorithmic discrimination, opacity, and accountability. It has shown how automated scoring can undermine due process, how data-intensive systems reproduce inequality, and how abstract notions of fairness may fail in concrete institutional settings (Citron & Pasquale, 2014; Eubanks, 2018; Mittelstadt et al., 2016; Noble, 2018). However, analysis in the Indonesian context often remains divided between data protection, electronic transactions, AI ethics, and sector-specific governance. Less attention has been paid to integrating constitutional rights, legal-system capacity, risk-based regulation, and responsibility across the entire AI life cycle. This creates a research gap between recognising principles and designing an enforceable model capable of translating them into institutional duties.

This article addresses two questions. First, why are existing Indonesian human rights protections insufficient for harms generated by autonomous, opaque, and transnational AI systems? Second, what legal and institutional model can combine preventive governance with accessible remedies while preserving socially beneficial innovation? The research aims to evaluate the existing framework and formulate a reorientation model that joins substantive rights, institutional responsibility, technological design, and public accountability. It uses the rule of law and human rights protection as its normative foundation, Friedman's legal-system theory to analyse substance, structure, and culture, and Radbruch's triad of justice, legal certainty, and utility to evaluate regulatory balance (Friedman, 1975; Radbruch, 2006).

The central argument is that protection must shift from a predominantly reactive and individual model to a preventive, systemic, and life-cycle model. Remedies after harm remain indispensable, but they cannot correct structural risks on their own. Human rights safeguards must operate before deployment through risk classification, rights-impact assessment, lawful and representative data governance, security testing, documentation, and institutional authorisation. They must continue during operation through monitoring, explanation, contestability, human review, and incident reporting, and after harm through correction, compensation, sanctions, and systemic remediation. The proposed RIGHTS-AI model gives this reorientation an operational structure and constitutes the principal novelty of the article.

METHODS

This study uses normative legal research to evaluate the coherence and adequacy of norms governing human rights in AI-based digital environments. Normative inquiry is appropriate because the research questions concern the allocation of legal duties, the interpretation of rights, and the design of a governance model rather than the measurement of individual attitudes or system performance. Primary legal materials include the 1945 Constitution of the Republic of Indonesia, Law Number 39 of 1999 concerning Human Rights, Law Number 27 of 2022 concerning Personal Data Protection, Law Number 1 of 2024 concerning the Second Amendment to the Electronic Information and Transactions Law, and Ministerial Circular Letter Number 9 of 2023 concerning AI Ethics. International and comparative materials include UNESCO's 2021 Recommendation and Regulation (EU) 2024/1689, the Artificial Intelligence Act.

The research combines statutory, conceptual, and comparative approaches. The statutory approach identifies rights, duties, enforcement mechanisms, and normative gaps within Indonesian law. The conceptual approach examines dignity, privacy, equality, due process, transparency, accountability, legal certainty, justice, utility, and human control. The comparative approach uses international instruments to identify regulatory principles capable of functional adaptation to Indonesia. Comparison is limited to problem-solving concepts, particularly risk classification, impact assessment, documentation, supervision, and remedies; it does not assume that foreign institutional arrangements can be copied without regard to Indonesia's constitutional and administrative context.

Legal materials were collected through document-based research and analysed qualitatively. The analysis proceeded in four stages: identifying AI-related rights risks; mapping those risks against existing Indonesian norms; evaluating gaps through legal-system theory and the values of justice, certainty, and utility; and synthesising a rights-oriented model.

Doctrinal interpretation was combined with sociotechnical reasoning because algorithmic harm emerges from interaction among law, data, technical design, organisational practice, and social context. The resulting model is normative and prescriptive. It proposes principles and institutional mechanisms that require further empirical testing, sectoral specification, and legislative deliberation.

RESULTS AND DISCUSSION

Artificial Intelligence and the Transformation of Human Rights Protection

AI transforms human rights protection by changing the location, speed, and visibility of legally significant power. In conventional administration, a person can generally identify the institution making a decision and the rule used to justify it. Automated systems divide that process among datasets, models, interfaces, organisational protocols, and human operators. The final decision may still be formally signed by a person, but the practical range of available choices can be shaped by a system that classifies an applicant as low priority, high risk, or unsuitable. Power is exercised not only through a final legal act but through the informational architecture that determines which facts become relevant and how a person is represented.

This transformation changes the meaning of vulnerability. Individuals are not merely data subjects whose information is collected; they become objects of inference. AI can estimate preferences, health conditions, financial reliability, political orientation, emotional state, or future behaviour from correlations among seemingly ordinary data. Inferred information may be more consequential than the data originally disclosed, yet individuals may have little awareness of its creation. The law must therefore address not only collection and storage but also the production and use of probabilistic knowledge. Privacy concerns the integrity of social context and the ability to maintain appropriate boundaries, not only secrecy (Nissenbaum, 2010; Solove, 2008).

Automated prediction also creates temporal harm. A person may be treated today according to an estimated future risk rather than proven conduct. Predictive policing, fraud detection, child-welfare analytics, credit scoring, and employee monitoring can shift institutional action from responding to events toward managing anticipated behaviour. Such practices may be legitimate in carefully bounded settings, but they carry a danger that probability becomes a substitute for evidence. When an adverse decision is based on group correlations, the individual may have no meaningful opportunity to demonstrate that the prediction does not apply. Due process consequently requires more than formal notice; it requires intelligible reasons and a genuine possibility of contestation (Citron & Pasquale, 2014).

Generative AI introduces additional rights concerns. Synthetic media can facilitate creativity and accessibility, but it can also imitate voices, fabricate images, generate defamatory content, automate deception, and overwhelm information environments. The harm is not limited to false statements about a particular person. Large-scale synthetic content can reduce trust in authentic evidence and create a "liar's dividend," allowing genuine material to be dismissed as artificial. Public discourse becomes vulnerable when the cost of producing plausible manipulation falls dramatically. Legal responses must distinguish harmful deployment from legitimate expression and must avoid using misinformation concerns as a pretext for disproportionate restrictions on speech.

AI systems can produce cumulative harm even when individual outputs appear minor. A slightly biased ranking repeated across millions of decisions may shape labour markets, access to education, or public-resource distribution. Conventional litigation focuses on a particular claimant and a particular loss, making dispersed statistical harm difficult to prove. An affected individual may lack access to comparative data needed to establish discrimination. Regulators, auditors, civil-society organisations, and researchers therefore require lawful access to aggregate evidence. Collective and systemic mechanisms complement individual remedies by identifying patterns that no single claimant can demonstrate.

The state has multiple roles in this environment. It is a regulator responsible for protecting rights, a purchaser of AI systems, a deployer in public administration, and a holder of extensive datasets. These roles may conflict. An institution seeking efficiency can become dependent on a vendor whose model is difficult to audit. Procurement secrecy may shield public decision-making tools from scrutiny. Effective human rights protection requires the state to impose at least the same degree of transparency and accountability on its own use of AI as it requires from private actors. Public authority cannot be delegated to an algorithm in a manner that weakens constitutional responsibility.

Corporations also exercise quasi-public power when their platforms structure access to information, employment, commerce, or communication. Human rights law has historically focused on the vertical relationship between state and individual, while business obligations often operate through regulation and due diligence. AI makes this division increasingly porous. A platform's ranking or moderation system may significantly affect expression and livelihood, while a private scoring model may determine access to essential services. The state's duty to protect therefore includes establishing enforceable corporate obligations, and corporate responsibility requires human rights due diligence across technological supply chains.

The concept of meaningful human control responds to these transformations but should not become a symbolic requirement. Placing a person nominally "in the loop" does not guarantee protection if the person lacks time, expertise, authority, or access to contrary evidence. Automation bias may lead reviewers to accept outputs because the system appears objective. Human oversight is meaningful only when reviewers understand the system's limits, can depart from recommendations, record reasons, obtain adequate information, and are institutionally protected from pressure to follow automated results. Control must be designed as a capability, not merely a procedural label.

These developments demonstrate why traditional ex post protection is insufficient. A lawsuit after denial of employment cannot easily reverse widespread discriminatory training practices. Compensation following a data breach does not remove compromised biometric identifiers. A correction after unlawful surveillance cannot fully restore autonomy or trust. Preventive duties are therefore essential. Risk assessment, data governance, testing, and documentation should operate before deployment, while continuous monitoring should detect changes in performance and context. Human rights protection must become part of technological governance rather than an external response activated only after visible damage.

Human Rights Risks and the Fragmentation of Indonesian Law

The 1945 Constitution supplies the highest normative foundation for AI governance. Constitutional guarantees of dignity, equal treatment, legal certainty, personal security,

communication, and information apply regardless of whether interference is executed directly by a human official or mediated by technology. The technologically neutral character of constitutional rights is an advantage because it prevents new tools from escaping basic norms. Nevertheless, constitutional generality must be translated into specific duties. A person affected by automated classification needs to know who must provide notice, what explanation is required, when human review is mandatory, and which institution can order correction or suspension.

Law Number 39 of 1999 reinforces protection of dignity, equality, privacy, and freedom from discrimination. It also supports the principle that public and private power should be accountable when it impairs rights. However, the statute predates contemporary machine learning and does not address distributed algorithmic decision making. Its principles remain applicable, but enforcement may be difficult when claimants cannot identify the relevant system or obtain technical evidence. Human rights norms must therefore be connected to procedural rights concerning information, traceability, audit, and access to remedies.

Law Number 27 of 2022 is the most directly relevant statutory instrument because AI development and deployment frequently involve personal data. Its principles of lawful processing, purpose limitation, accuracy, security, transparency, and accountability provide a foundation for data-driven governance. Rights of data subjects can support access, correction, objection, and protection against harmful processing. Yet AI exposes several boundaries. Models can generate inferences that were not explicitly collected, combine data from multiple contexts, or operate on anonymised datasets that may later become re-identifiable. A complete AI regime must clarify how data-protection principles apply to inferred data, model outputs, group privacy, and decisions with significant legal or comparable effects.

Sensitive data create heightened risks. Biometric, health, genetic, financial, criminal, and children's data may expose individuals to irreversible or discriminatory consequences. Facial templates cannot be replaced like passwords. Health predictions can influence employment or insurance. Children's behavioural data can shape educational opportunities before they are capable of understanding long-term implications. Risk should therefore be assessed by purpose, context, scale, reversibility, and effect, not solely by data category. A system processing ordinary location or purchasing data may infer sensitive characteristics and become functionally high risk.

Discrimination presents a doctrinal and evidentiary challenge. Indonesian constitutional and human rights principles prohibit discriminatory treatment, but algorithmic discrimination may be indirect, statistical, and distributed. A rule may appear neutral while producing disproportionate adverse effects. Establishing causation can require access to model documentation, testing results, error rates, and comparative outcomes that remain under organisational control. The burden of proof should not rest entirely on individuals who lack this information. For high-risk systems, providers and deployers should have affirmative duties to test for disparate impact, document mitigation, and disclose relevant evidence to supervisors and affected persons.

The Electronic Information and Transactions framework addresses electronic systems, prohibited digital acts, and institutional responsibilities, and its 2024 amendment is relevant to synthetic media and automated misuse. Nonetheless, a prohibition-oriented approach does not resolve all governance problems. Harm can occur through lawful systems used in unreasonable

ways, such as excessive workplace monitoring, opaque credit scoring, or automated service exclusion. The issue may not be illegal content but the design and institutional adoption of a decision process. AI governance therefore requires prospective duties governing development and use, not only sanctions for specified electronic misconduct.

Ministerial Circular Letter Number 9 of 2023 is an important recognition that AI should be guided by inclusiveness, humanity, security, accessibility, transparency, credibility and accountability, personal data protection, sustainable development, and intellectual property. Its principles align with international ethics frameworks. However, the instrument's soft-law character limits enforceability, institutional coordination, and remedial consequences. Ethical principles can guide interpretation and organisational conduct, but rights protection cannot depend solely on voluntary compliance where systems influence employment, credit, health, education, security, or public benefits.

Sectoral regulation may fill some gaps but can also produce inconsistency. Health, finance, education, labour, and law enforcement have different risk profiles and institutional mandates. Sector-specific rules are necessary because an explanation suitable for product recommendation is inadequate for medical diagnosis or criminal justice. Yet basic rights safeguards should remain common across sectors. Without a horizontal framework, similar risks may receive different protection depending on the administering agency, and systems spanning several sectors may exploit jurisdictional boundaries. A general AI statute or binding regulatory framework should establish minimum principles, while sectoral regulators specify stricter requirements.

Transnational operations create additional difficulties. A model may be trained abroad, accessed through a foreign platform, hosted in another jurisdiction, and used to decide matters concerning Indonesian residents. Contractual chains and data transfers complicate supervision. Domestic law should define when obligations apply based on effects within Indonesia, the location of affected persons, and the offering or use of AI services in the Indonesian market. Local representation, regulatory cooperation, evidence preservation, and cross-border enforcement mechanisms are necessary for high-risk foreign providers. Otherwise, national rights become dependent on the accessibility of actors beyond domestic jurisdiction (Kuner, 2013).

The principal deficiency is therefore not a total absence of law but the absence of integration. Constitutional, human rights, data-protection, electronic-system, consumer, administrative, and sectoral norms each illuminate part of the problem. Their institutions, procedures, and sanctions do not yet form a coherent AI accountability chain. Individuals may be referred among controllers, sectoral agencies, system operators, and vendors. Fragmentation increases transaction costs, delays remedies, and enables responsibility shifting. The legal task is to connect existing protections through common definitions, life-cycle duties, coordinated supervision, and accessible complaint pathways.

Accountability Gaps in Autonomous, Opaque, and Transnational Systems

Accountability requires more than identifying someone who can be blamed after harm. It involves answerability, traceability, reviewability, and enforceability. An accountable institution must explain what it did, preserve evidence of how a system operated, allow independent evaluation, correct errors, and accept legal consequences. AI weakens these

conditions when organisations do not document model versions, data provenance, performance limitations, human interventions, or changes after deployment. The first response to opacity should thus be mandatory record keeping. If a system cannot be reconstructed sufficiently for lawful review, it should not be used for high-impact decisions.

Responsibility should follow functional control and capacity to prevent harm. Developers influence architecture, training, and testing; providers determine documentation and system access; deployers select contexts, thresholds, and operational procedures; data controllers determine purposes and lawful bases; public authorities and corporate leaders approve institutional reliance. These roles may overlap, but overlap should not become an excuse for absence of responsibility. Legislation can assign differentiated duties and joint responsibility where several actors materially contribute to a risk. Contracts may allocate costs internally but should not eliminate statutory protection for affected persons.

Autonomy is often overstated in public debate. Current systems do not become morally independent actors merely because outputs are generated without immediate human instruction. Their goals, data, interfaces, and deployment conditions are established by organisations. The relevant legal question is therefore not whether the machine is guilty but whether human institutions designed reasonable safeguards around automated operations. Claims of technological autonomy must not become a liability shield. An organisation choosing to rely on an AI system assumes a duty to understand its appropriate use, foreseeable limitations, and rights impacts.

Opacity must also be differentiated. Source-code disclosure is not always necessary or sufficient. Affected persons generally need meaningful reasons: the fact that AI was used, the purpose, principal factors relevant to the outcome, applicable criteria, data categories, review procedures, and avenues for challenge. Regulators and qualified auditors require deeper access to technical documentation, training and validation information, logs, and testing results. Courts may require protected disclosure under procedural safeguards. A layered transparency model can protect legitimate intellectual property and security interests while preventing secrecy from defeating accountability.

Explainability should be connected to decision significance. Low-risk applications may require only basic notice. High-risk decisions affecting liberty, livelihood, essential services, or legal status require specific, timely, and comprehensible explanations. Counterfactual information what material change could have produced a different result can assist contestation, although it does not reveal every form of structural bias (Wachter et al., 2017). Explanations must not be reduced to generic statements drafted to satisfy compliance. Their adequacy depends on whether the recipient can understand the basis of the decision and effectively respond.

Audit provides an important accountability tool but is not self-sufficient. Technical audits can test accuracy, robustness, privacy, security, and differential performance. Legal audits can assess lawful basis, proportionality, non-discrimination, procedural fairness, and remedial mechanisms. Sociotechnical audits examine the organisational context in which outputs are interpreted and acted upon. Independence, access, competence, and public reporting determine whether audits are credible. A market for superficial certification could legitimise harmful systems. High-risk audit standards should therefore be established or recognised by public authorities and supplemented by powers to investigate incidents.

Procurement is a critical but frequently overlooked site of accountability. Public bodies and regulated firms can require vendors to provide documentation, audit access, incident support, data-governance assurances, and termination rights. Contracts should prohibit clauses that prevent legally required explanations or regulatory examination. Procurement officers need multidisciplinary expertise to assess not only cost and performance but also rights risks. When the state purchases opaque technology without adequate safeguards, it effectively privatises part of public decision making while retaining constitutional responsibility for the result.

Effective remedies must be accessible before irreversible harm. Individuals should be able to obtain notice, correction, suspension of an adverse outcome, and human review without first mastering complex technical evidence. Complaint mechanisms should be free or affordable, time-bound, accessible to persons with disabilities, and available in understandable language. Where system-wide risk is plausible, regulators should be empowered to order preservation of evidence, independent testing, temporary suspension, or recall. Collective actions and representation by authorised organisations may be necessary because individual claims cannot always expose statistical discrimination.

Friedman's legal-system theory clarifies why legal substance alone is insufficient. Substantive rules can define duties, but legal structure determines whether institutions have jurisdiction, expertise, independence, and resources to enforce them. Legal culture influences whether organisations view compliance as a meaningful responsibility or a documentation exercise. Indonesia's AI framework must therefore develop all three dimensions. It needs binding norms, coordinated and technically capable institutions, and a culture of rights-by-design among developers, public officials, business leaders, professionals, and users (Friedman, 1975).

Radbruch's values illuminate the regulatory balance. Legal certainty requires clear classifications, duties, and consequences. Justice requires protection against discrimination, surveillance, arbitrary exclusion, and denial of remedy. Utility requires that regulation preserve beneficial innovation and proportionality. A governance regime focused only on certainty may become rigid; one focused only on utility may externalise harm; one invoking justice without operational standards may remain indeterminate. Risk-based and rights-oriented regulation can integrate the three values by increasing obligations as foreseeable harm becomes more serious (Radbruch, 2006).

Comparative Lessons for Indonesian Artificial Intelligence Governance

UNESCO's Recommendation on the Ethics of Artificial Intelligence provides a globally endorsed normative framework that explicitly grounds AI governance in human rights and dignity. It promotes proportionality, safety, fairness, sustainability, privacy, transparency, responsibility, human oversight, and public awareness. Its significance lies in treating ethics as an institutional undertaking rather than a list of aspirations. Ethical impact assessment, governance capacity, education, stakeholder participation, and monitoring translate values into processes. For Indonesia, the Recommendation can guide interpretation of constitutional duties and help align domestic policy with internationally recognised principles without displacing national legal authority (UNESCO, 2021).

The European Union AI Act offers a different but complementary lesson. It classifies certain practices as unacceptable, regulates high-risk systems through mandatory requirements, and provides transparency obligations for specified applications. Its life-cycle orientation addresses risk management, data governance, technical documentation, record keeping, accuracy, robustness, cybersecurity, human oversight, post-market monitoring, and incident reporting. The Act does not eliminate all uncertainty or rights concerns, and its implementation is institutionally demanding. Nevertheless, it demonstrates how general values can be converted into differentiated and enforceable responsibilities (European Parliament & Council of the European Union, 2024).

Indonesia should avoid copying the EU scheme verbatim. Market structure, administrative capacity, technological development, public-service priorities, and constitutional institutions differ. Compliance costs that large multinational firms can absorb may disproportionately burden local start-ups, universities, and small enterprises. A proportional framework should distinguish experimentation from deployment, research from commercial use, and low-impact tools from systems affecting fundamental interests. Regulatory sandboxes can support responsible innovation, but participation should not suspend core rights or conceal experiments involving vulnerable groups.

Risk classification should be based on function and impact rather than on technology labels. The same underlying model may be low risk when suggesting entertainment and high risk when screening job applicants or supporting criminal investigation. Relevant criteria include the importance of the affected interest, scale of deployment, vulnerability of affected groups, degree of autonomy, opacity, reversibility, data sensitivity, and availability of human alternatives. Categories should be periodically reviewed because uses and social consequences evolve. A static list alone will become obsolete.

Some practices may warrant prohibition where their purpose or effect is incompatible with dignity and democratic society. Possible examples include manipulative systems exploiting children's or vulnerable persons' susceptibility, indiscriminate biometric surveillance without strict legal authority, and social scoring that produces disproportionate exclusion across unrelated contexts. Prohibition should be defined carefully to protect legitimate research, accessibility, and narrowly authorised public interests. Exceptions must be necessary, proportionate, time-limited, independently supervised, and subject to judicial review.

High-risk systems require prior assessment and continuing oversight. An AI human rights impact assessment should identify affected rights, stakeholders, foreseeable direct and indirect harms, discriminatory effects, alternatives, mitigation measures, residual risks, and remedies. Assessment should begin before procurement or development decisions become irreversible. It should be updated when the model, data, purpose, scale, or context materially changes. Public-sector assessments should generally be disclosed in accessible form, with limited redactions for lawful security or confidentiality concerns.

Institutional design should build on existing Indonesian authorities while preventing gaps. Data-protection supervision, sectoral regulators, the National Human Rights Commission, consumer-protection bodies, administrative oversight, courts, and cybersecurity institutions each hold relevant mandates. A central coordinating mechanism can establish standards, maintain a registry of high-risk public systems, coordinate investigations, and

resolve jurisdictional overlap. Sectoral expertise should remain with relevant regulators, while human rights and data-protection bodies ensure that economic or security objectives do not displace fundamental safeguards.

Meaningful participation is another comparative lesson. Individuals and communities affected by AI should not be treated merely as sources of data or recipients of completed policy. Workers, disability organisations, children's advocates, minority communities, consumer groups, technical experts, and civil society can identify contextual harms missed by developers. Participation is especially important when systems are deployed in public services or high-impact sectors. Consultation must occur early enough to influence design and should be supported by accessible information rather than technical disclosure without explanation.

International interoperability is necessary because AI supply chains are transnational. Indonesia benefits from standards that allow evidence, audits, and risk documentation to be recognised across borders while preserving domestic rights. Cooperation with foreign regulators can support incident response and enforcement against providers without local technical infrastructure. Interoperability should not mean adopting the weakest common denominator. Domestic constitutional guarantees and social conditions may justify stronger protections, particularly for public-sector surveillance, biometric processing, and decisions involving vulnerable populations.

Comparative experience ultimately indicates that ethical principles, risk-based obligations, and institutional enforcement must operate together. Ethics without enforceability can become reputational language; rules without technical standards can be difficult to apply; enforcement without rights-based direction can prioritise formal compliance. Indonesia's opportunity is to create a framework that is technologically realistic, administratively feasible, constitutionally grounded, and responsive to local inequality. This requires an integrated model rather than isolated amendments to existing statutes.

The RIGHTS-AI Model for Preventive and Systemic Human Rights Protection

This article proposes RIGHTS-AI as an integrated model for reorienting Indonesian human rights protection. The acronym identifies eight mutually supporting components: **R**isk classification; **r**ights **I**mpact assessment; **G**overnance accountability; meaningful **H**uman oversight; **T**ransparency and traceability; **S**afeguards and security; **A**ccessible remedies; and **I**ndependent supervision. The model links legal substance, institutional structure, and compliance culture throughout the AI life cycle. It is not intended as a purely technical checklist. Each component expresses a legal duty derived from dignity, equality, privacy, due process, and accountable government.

Risk classification is the gateway. Before development, procurement, or deployment, an organisation should determine the foreseeable rights impact of the intended use. Low-risk systems may remain subject to general data, consumer, security, and transparency duties. Limited-risk systems may require notice and user control. High-risk systems affecting employment, education, health, credit, essential services, biometric identification, law enforcement, migration, or public benefits require enhanced obligations. Unacceptable-risk practices should be prohibited. Classification must be documented, reviewable, and revisited when a system's function or context changes.

Rights impact assessment converts classification into contextual analysis. The assessment should identify purposes, legal bases, affected groups, data flows, model capabilities, expected benefits, reasonably foreseeable misuse, direct and cumulative harms, differential impacts, alternatives, and mitigation. It should evaluate necessity and proportionality: whether AI is needed, whether a less intrusive method could achieve the purpose, and whether residual risk is justified by a legitimate objective. The process should include relevant stakeholders and should not be conducted solely by the team promoting deployment.

Governance accountability allocates responsibility to identifiable persons and institutions. Every high-risk system should have an accountable deploying institution, a designated senior officer, documented provider responsibilities, and clear procedures for escalation and incident response. Developers, providers, importers, distributors, controllers, processors, and deployers should bear duties corresponding to their functional control. An organisation should not be able to escape liability through outsourcing. Contracts must support audit, evidence preservation, correction, and regulatory access, while joint responsibility should apply when harm results from interconnected failures.

Meaningful human oversight preserves human agency and constitutional responsibility. High-impact decisions should not be based solely on automated output where rights, liberty, livelihood, or access to essential services are materially affected. Reviewers must be trained, informed of system limitations, authorised to override recommendations, and given sufficient time and evidence. Institutions should monitor override rates and patterns of automatic acceptance to detect automation bias. In security or law-enforcement contexts, oversight must be especially strict because error can lead to surveillance, coercion, or deprivation of liberty.

Transparency and traceability operate at different levels. Individuals should receive clear notice when AI materially influences a decision, along with the system's purpose, major factors, data categories, and procedures for review. Regulators and auditors need technical documentation, validation results, risk assessments, model and data provenance, logs, incident records, and information about material changes. Public authorities should maintain a publicly accessible registry of high-risk AI systems, subject to narrowly tailored security exceptions. Traceability ensures that responsibility can be reconstructed even when systems evolve.

Safeguards and security encompass privacy by design, data minimisation, quality controls, bias testing, robustness, cybersecurity, accessibility, and continuous monitoring. Training and validation data should be relevant and sufficiently representative for the intended context. Performance should be evaluated across groups likely to experience different outcomes. Security measures should address model manipulation, data poisoning, unauthorised extraction, and adversarial use. Organisations must monitor concept drift and changing social conditions because a model that performed acceptably at deployment can become inaccurate or discriminatory over time.

Accessible remedies make rights practical. Affected individuals should have a single, understandable entry point for complaints even when several organisations participate in the system. Remedies should include information, correction, reconsideration by a competent person, suspension of an adverse decision where necessary, deletion or restriction of unlawful data, compensation, and institutional remediation. Time limits should reflect the urgency of the

interest affected. Vulnerable persons should receive assistance, and authorised organisations should be able to represent collective interests when systemic harm is alleged.

Independent supervision completes the model. Internal compliance is necessary but cannot substitute for external authority. Supervisors need legal powers, technical competence, adequate resources, independence, and access to documentation. They should coordinate across data protection, human rights, consumer protection, cybersecurity, and sectoral regulation. Powers should include inspection, audit, corrective orders, temporary suspension, administrative sanctions, referral for civil or criminal proceedings, and publication of significant findings. Judicial review remains essential to protect both affected persons and regulated entities.

The components operate as a cycle rather than a sequence completed once. Risk classification informs assessment; assessment shapes safeguards and human oversight; governance determines who implements them; transparency and traceability enable supervision; monitoring identifies incidents; remedies correct individual and systemic harm; supervisory findings update standards and future classifications. This feedback structure makes the model adaptive. AI governance must learn from failures because models, markets, and social contexts evolve more quickly than conventional legislative amendment.

The model also distinguishes organisational accountability from individual blame. Developers and officials may act in good faith while institutional systems still produce harm. A mature regime should encourage reporting, correction, and learning while imposing stronger consequences for concealment, reckless deployment, repeated non-compliance, or prohibited practices. Proportionate enforcement protects innovation by creating predictable expectations. It also prevents responsible organisations from being disadvantaged by competitors that externalise rights risks.

Implementation can proceed incrementally. The government can first require a registry and impact assessment for high-risk public-sector systems, integrate AI clauses into procurement rules, and issue coordinated standards under existing statutory mandates. Sectoral pilots can develop expertise in finance, health, education, labour, and public services. Legislative reform can then establish horizontal definitions, risk categories, common rights, role-based duties, supervisory coordination, and sanctions. Capacity building should include judges, regulators, procurement officials, developers, lawyers, auditors, and civil society.

Small and medium-sized organisations require support to meet obligations. Standardised documentation, open testing tools, regulatory guidance, shared audit facilities, and proportionate fees can reduce compliance burdens. Universities and public research institutions can contribute independent expertise. Sandboxes should require rights assessments, informed safeguards, incident reporting, and exit criteria. They should not become deregulated spaces. Innovation is sustainable when trust and accountability are built into development rather than added after public controversy.

Public literacy is part of legal culture. Individuals should understand that automated outputs can be contested and that AI is not inherently objective. Professionals using AI need discipline-specific competence: clinicians must evaluate diagnostic support, educators must understand learning analytics, and officials must recognise when automated recommendations implicate administrative-law duties. Legal education should integrate technology, data, ethics,

and human rights. Rights-oriented governance depends on institutional judgment as well as formal rules.

The novelty of RIGHTS-AI lies in joining two forms of protection often treated separately. Individual protection supplies notice, explanation, objection, review, and remedy. Systemic protection supplies risk classification, assessment, design safeguards, documentation, audit, and supervision. Either approach alone is incomplete. Individual remedies cannot reveal every structural pattern, while systemic certification can overlook the lived consequences of a particular decision. The model places the individual and the system in a reciprocal accountability structure.

The model also provides a principled balance among certainty, justice, and utility. Clear categories and role-based duties enhance legal certainty. Impact assessment, non-discrimination testing, human control, and remedies advance justice. Proportional obligations, phased implementation, and support for responsible experimentation preserve utility. Innovation is not conceptualised as freedom from regulation but as the development of socially valuable technology under conditions of trust. This approach positions law as a means of directing digital transformation toward constitutional purposes rather than merely reacting to technological harm (Pound, 1907).

The research finding is therefore not simply that Indonesia needs more AI rules. The deeper requirement is a change in the object and timing of human rights protection. Law must govern the sociotechnical system, not only the final actor or harmful output, and must intervene across the life cycle before risks become irreversible. RIGHTS-AI supplies the connective architecture that existing fragmented norms lack. It translates constitutional values into preventive duties, links private innovation with public accountability, and makes remedy part of design rather than an afterthought.

CONCLUSIONS AND RECOMMENDATIONS

Conclusion

AI changes human rights risk by distributing decision-making power across data, models, platforms, organisations, and jurisdictions. Its use can improve services and social welfare, but it can also enable intrusive inference, algorithmic discrimination, opaque exclusion, synthetic manipulation, and accountability gaps. Indonesian constitutional, human rights, personal data protection, and electronic information rules provide essential safeguards, while national AI ethics guidance expresses relevant principles. Their weakness is fragmentation: they do not yet establish common risk categories, role-based duties, life-cycle documentation, meaningful explanation, coordinated supervision, or remedies adapted to automated decisions.

Human rights protection must consequently be reoriented from a predominantly reactive model toward preventive and systemic governance without abandoning individual remedies. The proposed RIGHTS-AI model integrates risk classification, rights impact assessment, governance accountability, human oversight, transparency and traceability, safeguards and security, accessible remedies, and independent supervision. Its life-cycle structure places duties on developers, providers, deployers, data actors, and public authorities according to their capacity to prevent and correct harm. By connecting systemic safeguards with contestable individual decisions, the model aligns legal certainty, justice, and utility and keeps technological power subordinate to dignity and democratic responsibility.

Limitations

This research is normative and relies on legal materials, policy instruments, and interdisciplinary scholarship. It does not empirically measure the operation of particular AI systems or the institutional capacity of every Indonesian regulator. Comparative analysis focuses on functional lessons from international instruments and the European Union and does not examine all jurisdictions. The proposed model therefore requires sector-specific validation, stakeholder consultation, and testing against concrete applications. Rapid technological and regulatory development may also alter particular risks and implementation options.

Recommendations

Indonesia should develop a binding horizontal AI governance framework grounded in constitutional rights and coordinated with personal data protection, electronic-system, consumer, administrative, and sectoral law. Initial priorities should include a public registry of high-risk government AI, mandatory rights impact assessment, rights-preserving procurement clauses, layered transparency, independent audit, meaningful human review, and a unified complaint pathway. Sectoral regulators should issue context-specific standards while a coordinating mechanism ensures consistency. Future studies should empirically examine AI use in public services, finance, employment, health, education, and security; assess the capacity of supervisory institutions; and test RIGHTS-AI through participatory case studies involving affected communities.

Author Contributions

The author was responsible for conceptualisation, legal analysis, comparative analysis, model development, drafting, revision, and approval of the final manuscript.

REFERENCES

- Barocas, S., & Selbst, A. D. (2016). Big data's disparate impact. *California Law Review*, 104(3), 671-732. <https://doi.org/10.15779/Z38BG31>
- Buolamwini, J., & Gebru, T. (2018). Gender shades: Intersectional accuracy disparities in commercial gender classification. *Proceedings of Machine Learning Research*, 81, 77-91. <https://proceedings.mlr.press/v81/buolamwini18a.html>
- Calo, R. (2017). Artificial intelligence policy: A primer and roadmap. *UC Davis Law Review*, 51(2), 399-435.
- Citron, D. K., & Pasquale, F. (2014). The scored society: Due process for automated predictions. *Washington Law Review*, 89(1), 1-33.
- Eubanks, V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. St. Martin's Press.
- European Parliament & Council of the European Union. (2024). Regulation (EU) 2024/1689 laying down harmonised rules on artificial intelligence. *Official Journal of the European Union*. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., Luetge, C., Madelin, R., Pagallo, U., Rossi, F., Schafer, B., Valcke, P., & Vayena, E. (2018). An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and Machines*, 28(4), 689-707. <https://doi.org/10.1007/s11023-018-9482-5>
- Friedman, L. M. (1975). *The legal system: A social science perspective*. Russell Sage Foundation.

- Kuner, C. (2013). *Transborder data flows and data privacy law*. Oxford University Press.
- Law Number 1 of 2024 concerning the Second Amendment to Law Number 11 of 2008 concerning Electronic Information and Transactions (Indonesia).
- Law Number 27 of 2022 concerning Personal Data Protection (Indonesia).
- Law Number 39 of 1999 concerning Human Rights (Indonesia).
- Minister of Communication and Information of the Republic of Indonesia. (2023). *Circular Letter Number 9 of 2023 concerning artificial intelligence ethics*.
- Mittelstadt, B. D., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The ethics of algorithms: Mapping the debate. *Big Data & Society*, 3(2), 1-21. <https://doi.org/10.1177/2053951716679679>
- Nissenbaum, H. (2010). *Privacy in context: Technology, policy, and the integrity of social life*. Stanford University Press.
- 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.
- Pound, R. (1907). The need of a sociological jurisprudence. *Green Bag*, 19(10), 607-615.
- Radbruch, G. (2006). Statutory lawlessness and supra-statutory law (B. L. Paulson & S. L. Paulson, Trans.). *Oxford Journal of Legal Studies*, 26(1), 1-11. <https://doi.org/10.1093/ojls/gqi041>
- Republic of Indonesia. (1945). *The 1945 Constitution of the Republic of Indonesia*.
- Russell, S., & Norvig, P. (2020). *Artificial intelligence: A modern approach* (4th ed.). Pearson.
- Selbst, A. D., Boyd, D., Friedler, S. A., Venkatasubramanian, S., & Vertesi, J. (2019). Fairness and abstraction in sociotechnical systems. In *Proceedings of the Conference on Fairness, Accountability, and Transparency* (pp. 59-68). Association for Computing Machinery. <https://doi.org/10.1145/3287560.3287598>
- Solove, D. J. (2008). *Understanding privacy*. Harvard University Press.
- UNESCO. (2021). *Recommendation on the ethics of artificial intelligence*. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- Wachter, S., Mittelstadt, B., & Russell, C. (2017). Counterfactual explanations without opening the black box: Automated decisions and the GDPR. *Harvard Journal of Law & Technology*, 31(2), 841-887.
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. PublicAffairs.