JURNAL PARADIGMA HUKUM Pembangunan
ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

GENDERED AI GOVERNANCE: INDONESIA'S HUMAN SECURITY FRAMEWORK

Muhammad Yuga[1], Herdiansyah Hamzah[2]
[1]Universitas Mulawarman, my931953@gmail.com

ABSTRACT

The concept of human security, first articulated by Mahbub-ul-Haq in the United Nations Development Programme's Human Development Report in 1994, emphasizes freedom from fear and freedom from want as cardinal principles, positioning individuals as the primary referent in security discourse. As artificial intelligence systems rapidly proliferate across Indonesia, these foundational human security concerns have manifested in distinctly gendered dimensions that demand urgent regulatory attention. This article examines the critical intersection between gender, human security, and AI governance in Indonesia's evolving regulatory landscape. Through systematic analysis of algorithmic discrimination cases affecting Indonesian women, particularly in the gig economy and online gender-based violence contexts, this research establishes the imperative for integrating gendered human security frameworks into national AI regulation. The study employs doctrinal legal research methodology combined with intersectional feminist analysis to propose a three-pillar regulatory model encompassing gender-responsive Smart Mix Approach, mandatory Human Rights Due Diligence, and gender-sensitive Regulatory Sandbox mechanisms. Findings reveal significant gender bias in AI systems, with online gender-based violence cases surging by 80.8 percent in 2024. The article argues that constitutional obligations under Indonesia's 1945 Constitution necessitate explicit gender mainstreaming in AI governance to protect women's fundamental rights from algorithmic discrimination.

Keywords: human security, gender mainstreaming, artificial intelligence regulation, algorithmic discrimination

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

A. INTRODUCTION

The rapid advancement of artificial intelligence technologies across Indonesia presents both unprecedented opportunities and significant challenges for human security, particularly from a gendered perspective. The concept of human security, as originally conceived in the 1994 Human Development Report, fundamentally shifted the security paradigm from state-centric approaches to people-centered frameworks that prioritize individual wellbeing, dignity, and fundamental freedoms (UNDP, 125). This theoretical reorientation becomes especially salient when examining how emerging technologies such as artificial intelligence interact with existing structural inequalities and power asymmetries within Indonesian society.

Indonesia's digital transformation journey, accelerated by national development priorities outlined in the RPJMN 2025-2029 and the President-elect's Asta Cita vision, positions artificial intelligence as an integral component of public sector enhancement and economic modernization (Kementerian Komunikasi dan Digital, 15). However, this technological integration occurs within a context where gender disparities remain pronounced across economic, social, and political spheres. The intersection of AI systems with pre-existing gender inequalities creates unique vulnerabilities that threaten the human security of Indonesian women and gender minorities, manifesting through algorithmic discrimination in employment platforms, technology-facilitated gender-based violence, and exclusionary biometric systems that fail to recognize diverse gender identities and expressions.

Recent empirical evidence demonstrates the urgency of addressing these gendered dimensions of AI governance. Data from 2024 indicates an alarming 80.8 percent surge in online gender-based violence cases, many facilitated or exacerbated by generative AI technologies capable of producing non-consensual intimate imagery (SAFEnet, 42). Female workers in Indonesia's burgeoning gig economy face systematic disadvantages through algorithmic management systems that penalize them for biological and social conditions including menstruation, pregnancy, and caregiving responsibilities (Pertiwi, 3). Biometric verification systems employed in public service delivery and financial sector applications demonstrate consistent failures in recognizing female profiles, particularly affecting women who wear

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

religious head coverings or possess physical characteristics underrepresented in training datasets (Trisianto, 1).

These developments raise fundamental questions about the adequacy of Indonesia's current regulatory frameworks to address the gendered impacts of artificial intelligence. The 1945 Constitution of Indonesia establishes comprehensive guarantees for human rights and gender equality, particularly through Article 28G which affirms every person's right to protection of personal dignity and freedom from discriminatory treatment (Konstitusi Republik Indonesia, 28). The enactment of Law Number 27 of 2022 concerning Personal Data Protection represents a significant milestone in establishing individual data sovereignty and privacy protections (Undang-Undang Perlindungan Data Pribadi, 5). However, these existing legal instruments were conceived prior to the widespread deployment of AI systems and consequently lack specific provisions addressing algorithmic discrimination, automated decision-making processes, and the unique ways these technologies can perpetuate or amplify gender-based harms.

The international community has increasingly recognized the necessity of human rights-centered approaches to AI governance. The United Nations Guiding Principles on Business and Human Rights provide foundational frameworks for corporate accountability in the technology sector (Ruggie, 18). UNESCO's Recommendation on the Ethics of Artificial Intelligence, adopted in 2021, establishes comprehensive principles including fairness, transparency, and non-discrimination (UNESCO, 7). The European Union's Artificial Intelligence Act, finalized in 2024, introduces risk-based regulatory classifications and mandatory human rights due diligence for high-risk AI applications (European Parliament, 12). These international developments provide instructive models while simultaneously highlighting the imperative for Indonesia to develop context-appropriate regulatory approaches that reflect local social structures, cultural values, and specific patterns of gender inequality.

This research addresses a critical gap in both academic scholarship and policy discourse by systematically examining how gendered human security principles should inform Indonesia's AI regulatory framework. The methodology employed combines doctrinal legal research examining Indonesia's constitutional provisions,

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

statutory frameworks, and regulatory instruments with intersectional feminist analysis that centers the experiences of women and gender minorities as primary subjects rather than peripheral considerations (Crenshaw, 140). This approach recognizes that gender operates not in isolation but intersects with other axes of marginalization including class, ethnicity, disability status, and geographic location. Empirical documentation draws from civil society monitoring reports, including the comprehensive position paper on AI governance compiled by Indonesian civil society organizations in 2025, which documents specific cases of algorithmic discrimination affecting vulnerable populations (Parasurama et al., 8).

The proposed three-pillar regulatory model responds directly to the identified patterns of gendered AI harms documented in Indonesia. The Smart Mix Approach pillar recognizes that effective AI governance requires both mandatory regulatory interventions by the state and voluntary compliance mechanisms by private sector actors (OHCHR, 21). The Human Rights Due Diligence pillar establishes obligations for AI developers and deployers to systematically identify, assess, prevent, and mitigate adverse human rights impacts throughout the AI lifecycle, with particular attention to gendered harms (Ebert et al., 25). The Regulatory Sandbox pillar creates controlled environments for testing AI applications while ensuring robust protections for potentially affected populations, enabling innovation while safeguarding human security (GIZ Asia, 33).

B. METHOD

This study adopts a doctrinal legal research methodology combined with an intersectional feminist analysis. It primarily relies on normative legal materials, including constitutional provisions, statutory regulations, and human rights instruments, to examine Indonesia's legal obligations in governing artificial intelligence. Through a statute and conceptual approach, the research analyzes the constitutional mandate under the 1945 Constitution concerning the protection of fundamental rights, particularly women's rights, in the context of algorithmic decision-making. The intersectional feminist framework is employed to critically assess structural gender bias in AI systems and to contextualize the increasing incidence of online gender-based violence. Based on this normative and critical

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

analysis, the study formulates a three-pillar regulatory model consisting of a gender-responsive Smart Mix Approach, mandatory Human Rights Due Diligence, and gender-sensitive Regulatory Sandbox mechanisms to strengthen gender mainstreaming in AI governance

C. DISCUSSION

C.1.Theoretical Foundations: Human Security and Gendered AI Governance

The human security paradigm represents a fundamental reconceptualization of security discourse that emerged in the post-Cold War era as scholars and policymakers recognized the limitations of traditional state-centric security frameworks. Mahbub-ul-Haq's formulation in the 1994 Human Development Report identified seven dimensions of human security: economic security, food security, health security, environmental security, personal security, community security, and political security (Haq, 23). This multidimensional approach shifts analytical focus from territorial integrity and military capabilities to the actual lived experiences of individuals and communities, emphasizing protection from chronic threats such as hunger, disease, and repression as equally significant to protection from sudden disruptions such as conflict or natural disasters.

The application of human security frameworks to artificial intelligence governance requires understanding how AI systems impact each of these security dimensions. Economic security becomes compromised when algorithmic management systems in gig economy platforms discriminate against women workers through rating mechanisms that fail to account for caregiving responsibilities or biological conditions (Adhari, 2). Health security faces threats when biometric medical systems trained on datasets lacking diverse representation produce inaccurate diagnoses or treatment recommendations for female patients. Personal security suffers erosion through technology-facilitated gender-based violence enabled by generative AI tools capable of producing non-consensual intimate imagery (Kristin et al., 18).

Feminist scholarship has long emphasized that security cannot be understood through gender-neutral analyses, as women and gender minorities experience distinct forms of insecurity shaped by patriarchal power structures, gendered divisions of labor, and systematic exclusion from decision-making processes (Mhlambi and

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

Tiribelli, 871). The integration of gender analysis into human security frameworks reveals how threats manifest differently across gender identities and how protective mechanisms must be designed with attention to these differences. In the context of artificial intelligence, gendered human security analysis illuminates three critical dimensions. First, AI systems frequently encode and perpetuate existing gender biases present in training data, development teams, and institutional contexts where they are deployed. Second, the impacts of AI-driven decisions fall disproportionately on women and gender minorities who occupy positions of economic precarity and social marginalization. Third, women remain significantly underrepresented in AI development, policy formulation, and governance institutions, resulting in frameworks that fail to anticipate or adequately address gendered harms (Noble, 45).

The concept of algorithmic discrimination has emerged as a central concern within technology studies scholarship examining how automated decision-making systems reproduce and amplify social inequalities. Algorithms operate through classification processes that sort individuals into categories based on available data attributes, historical patterns, and predetermined optimization objectives (Eubanks, 78). When training datasets reflect historical discrimination, when relevant variables correlate with protected characteristics such as gender, or when optimization metrics prioritize efficiency over equity, the resulting algorithmic systems will systematically disadvantage marginalized groups. Research by scholars including Safiya Noble, Virginia Eubanks, and Ruha Benjamin has documented extensive evidence of algorithmic discrimination across domains including employment screening, credit allocation, criminal justice risk assessment, and content moderation (Benjamin, 92).

In the Indonesian context, algorithmic discrimination manifests through multiple mechanisms that specifically disadvantage women. Platform-based gig economy applications including ride-hailing and delivery services employ algorithmic management systems that rate worker performance, allocate job opportunities, and determine compensation structures. Civil society research documented in the 2025 position paper reveals that these systems penalize female workers who maintain lower acceptance rates or shorter working hours due to caregiving obligations, safety concerns about working late hours, or biological conditions including menstruation and pregnancy (Nurbaiti, 1). The algorithmic systems treat these gendered realities as

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

individual performance deficiencies rather than structural conditions requiring accommodation, resulting in systematic income penalties for female workers relative to male counterparts.

C.2.Constitutional Framework and Legal Gaps

Indonesia's constitutional and legal architecture provides foundational principles that should govern artificial intelligence deployment and establish obligations for protecting citizens from algorithmic discrimination. The 1945 Constitution, particularly following the comprehensive amendments enacted between 1999 and 2002, establishes robust human rights protections and equality guarantees. Article 28G affirms that every person has the right to protection of self, family, honor, dignity, and property under their control, and the right to feel secure and protection from the threat of fear (Konstitusi Republik Indonesia, 28). This constitutional provision establishes individual dignity and security as fundamental rights that cannot be compromised by state or private actors, including through the deployment of algorithmic systems that subject individuals to discriminatory treatment or privacy violations.

Article 28I of the Constitution establishes that the rights to life, freedom from torture, freedom of thought and conscience, freedom of religion, freedom from enslavement, recognition as a person before the law, and freedom from retroactive prosecution are human rights that cannot be limited under any circumstances (Konstitusi Republik Indonesia, 28). Technology-facilitated gender-based violence through generative AI tools represents a clear violation of these non-derogable rights to dignity and freedom from torture. The constitutional principle of equality receives explicit articulation in Article 28I paragraph two, which states that every person has the right to be free from discriminatory treatment on any basis and is entitled to protection from such discriminatory treatment.

Law Number 27 of 2022 concerning Personal Data Protection represents Indonesia's primary statutory framework governing individual data rights and establishing obligations for entities that collect, process, or store personal information. This legislation recognizes personal data protection as a constitutional right and prohibits the acquisition, collection, disclosure, and use of others' personal data

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

without proper legal basis (Undang-Undang PDP, 7). For AI systems that necessarily process vast quantities of personal data during training and deployment phases, this legislation establishes clear compliance obligations. However, the Personal Data Protection Law was enacted prior to widespread recognition of the specific challenges posed by artificial intelligence, particularly concerning algorithmic decision-making, automated profiling, and the potential for AI systems to infer sensitive attributes including gender identity, sexual orientation, or health conditions from ostensibly non-sensitive data.

The civil society position paper compiled in 2025 identifies significant gaps in Indonesia's existing legal frameworks when applied to artificial intelligence governance (Parasurama et al., 22). First, definitional challenges arise as AI systems encompass diverse technologies ranging from simple rule-based algorithms to complex neural networks with emergent capabilities. The lack of clear legal definitions creates uncertainty about which technologies fall within regulatory scope and what obligations apply. Second, the multi-actor nature of AI development and deployment complicates accountability assignment. A single AI system may involve data providers, algorithm developers, computational infrastructure operators, implementation entities, and service providers, each operating under different legal jurisdictions and contractual arrangements.

Third, the opacity of many AI systems, particularly those employing deep learning techniques, creates challenges for transparency and explainability. When individuals suffer adverse outcomes from algorithmic decisions, they often cannot obtain meaningful explanations of why the system reached its determination or what factors influenced the outcome (Burrell, 1). This opacity undermines procedural fairness and prevents effective challenges to discriminatory decisions. Fourth, the cross-border nature of AI development and deployment creates jurisdictional challenges. Many AI systems used in Indonesia are developed by multinational corporations operating from foreign jurisdictions, trained on datasets collected globally, and deployed through cloud infrastructure located outside Indonesian territory.

C.3.Documented Cases of Gendered Algorithmic Harm

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

The civil society position paper provides systematic documentation of algorithmic discrimination cases affecting Indonesian women across multiple domains. In the gig economy context, female ride-hailing and delivery drivers face systematic disadvantages through algorithmic management systems that fail to account for gendered realities (Pertiwi, 3). Research reveals that approximately eighty percent of ride-hailing application users are female passengers, yet female drivers constitute a small minority of the driver workforce. This gender imbalance creates specific vulnerabilities as male passengers frequently cancel rides upon discovering their assigned driver is female, often without providing reasons for cancellation. The algorithmic rating systems employed by these platforms treat such cancellations as negative indicators of driver performance, lowering the female driver's overall rating and consequently reducing her priority in future ride allocations.

These discriminatory algorithmic practices raise substantial questions of civil liability under Indonesian private law. The systematic disadvantaging of female drivers through biased algorithmic systems potentially constitutes a violation of fundamental principles governing contractual equality and fair dealing in commercial relationships. Under the general provisions of civil law governing obligations and torts, parties engaged in commercial transactions bear duties to refrain from conduct that causes unjustified harm to others. When platform companies deploy algorithmic management systems that systematically disadvantage female drivers based on gender-related factors beyond their control, such conduct may satisfy the elements of tortious liability requiring compensation for resulting economic losses. Female drivers who experience diminished income opportunities due to algorithmic bias possess potential grounds for civil claims seeking damages for the economic harm suffered, calculated based on demonstrable income reduction directly attributable to the discriminatory algorithmic treatment. The causal connection between algorithmic downgrading and reduced earning capacity establishes the foundation for claims of wrongful conduct resulting in compensable injury.

The financial services sector demonstrates another domain where algorithmic systems produce gendered exclusions with significant civil law implications. A documented case from December 2024 involved an individual with retinal detachment, a medical condition requiring surgical insertion of specialized silicone oil to stabilize

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

retinal positioning (Parasurama et al., 17). When this individual attempted to transfer banking services to a new mobile device, the bank's biometric verification system could not detect their iris pattern due to the presence of the medical device. This case demonstrates how artificial intelligence systems designed with narrow assumptions about normal human physiology exclude individuals with disabilities and medical conditions, creating barriers to accessing essential financial services.

The exclusionary effect of such algorithmic systems raises critical questions regarding the scope of duties owed by financial institutions to their customers under private law. Banking relationships create contractual obligations requiring financial institutions to provide reasonable access to account services for all customers, including those with disabilities or medical conditions. When biometric verification systems exclude customers based on physical characteristics resulting from medical treatment, the institution's failure to provide alternative verification methods may constitute a breach of contractual duties of care and good faith performance. The affected individual suffers both economic harm through inability to access their own financial resources and non-economic harm through the frustration and distress caused by unreasonable barriers to essential services. These harms give rise to potential claims for breach of contract and negligent conduct causing injury to another party. The financial institution's deployment of verification systems without adequate accommodation for persons with disabilities reflects a failure to exercise reasonable care in selecting and implementing customer service technologies, establishing grounds for liability based on negligent infliction of harm.

Biometric verification systems have also demonstrated consistent failures in recognizing individuals who wear religious head coverings, creating a distinct category of civil law concerns. Indonesia's Muslim-majority population includes millions of women who observe religious requirements to cover their hair and sometimes faces through wearing various forms of modest dress (Trisianto, 1). When government agencies and private sector entities deploy facial recognition systems for identity verification without ensuring these systems can accommodate covered faces, they effectively force women to choose between religious observance and access to employment, education, or services. The July 2021 case involving civil service examination candidates in Kediri exemplifies this problem, as prospective

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

government employees experienced repeated verification failures and some ultimately removed their head coverings to satisfy the algorithmic requirements.

This situation presents violations of personal rights protected under civil law principles governing individual dignity and freedom from discrimination. The forced removal of religious head coverings to satisfy algorithmic requirements constitutes interference with fundamental personal liberties protected through civil remedies. Women who experience denial of examination opportunities or employment access due to their religious dress suffer concrete harms in the form of lost career opportunities and emotional distress from violations of their deeply held religious convictions. These harms establish grounds for civil claims seeking compensatory relief for both economic losses stemming from denied opportunities and non-economic damages for the violation of personal dignity and religious freedom. The measure of damages in such cases encompasses calculation of lost earning potential from denied employment, as well as compensation for the psychological harm and social stigmatization experienced through forced compromise of religious identity. The entities deploying incompatible verification systems bear responsibility for ensuring their technological choices do not create discriminatory barriers, and their failure to do so establishes liability for resulting harms.

The emergence of generative artificial intelligence has created unprecedented capabilities for technology-facilitated gender-based violence through the production of non-consensual intimate imagery, raising novel questions of civil liability. The April 2025 case at Udayana University in Bali, where a student allegedly collected thousands of photographs of female classmates and used generative artificial intelligence tools to create explicit imagery, demonstrates the severity of this threat (Kid.kdf, 1). The affected women experienced profound violations of dignity, privacy, and sexual autonomy. Many victims reported psychological trauma, social stigmatization, and fear of continued exploitation through potential wider distribution of the fabricated images.

This conduct constitutes multiple violations of personal rights that establish comprehensive civil liability for the perpetrator and potentially for technology providers who facilitate such misuse. The unauthorized creation of non-consensual intimate images, regardless of whether such images depict actual events or artificially

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

generated scenarios, violates protected interests in personal privacy, dignitary integrity, and control over one's own image and reputation. These violations satisfy the elements required for civil claims based on wrongful conduct causing injury to another party. The victims possess standing to pursue multiple forms of civil relief addressing both the immediate harm and ongoing risks. Injunctive relief preventing further distribution of the images serves to mitigate continuing violation of privacy interests. Compensatory damages address the substantial psychological trauma documented through victim testimony and clinical assessment, as well as reputational harm affecting social relationships, educational opportunities, and future employment prospects. The particularly egregious nature of deliberately creating and distributing fabricated intimate imagery supports claims for enhanced damages reflecting the severity of the dignitary violation and the calculated nature of the harmful conduct.

Beyond the direct perpetrator's liability, this category of cases raises important questions regarding potential secondary liability of technology platform providers. When companies develop and deploy generative artificial intelligence systems capable of producing realistic intimate imagery, they assume responsibilities to implement safeguards preventing misuse for harmful purposes. Platforms that fail to incorporate adequate protections against non-consensual image generation or that do not provide effective mechanisms for victims to request removal of such content may bear liability based on their enabling role in facilitating the underlying harm. The scope of such secondary liability depends on factors including the foreseeability of misuse, the adequacy of implemented safeguards, and the responsiveness to reports of harmful content. As generative artificial intelligence capabilities become more sophisticated and accessible, the duty of care required of technology providers necessarily expands to encompass reasonable measures preventing foreseeable harms to third parties.

C.4.Three-Pillar Regulatory Framework

This article proposes a three-pillar regulatory framework specifically designed to integrate gendered human security principles into Indonesia's AI governance system while respecting the country's institutional capacities, developmental priorities, and constitutional values. The framework consists of a gender-responsive Smart Mix

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

Approach combining mandatory regulation with voluntary industry initiatives, mandatory Human Rights Due Diligence requirements for AI developers and deployers, and gender-sensitive Regulatory Sandbox mechanisms enabling innovation while protecting vulnerable populations.

The Smart Mix Approach, derived from the United Nations Guiding Principles on Business and Human Rights, recognizes that effective governance requires both binding legal obligations enforced by the state and voluntary compliance mechanisms adopted by private sector actors (OHCHR, 21). This approach acknowledges that rapid technological change often outpaces legislative processes, creating governance gaps where neither pure command-and-control regulation nor complete industry self-regulation proves adequate. For Indonesia's AI governance, the mandatory regulatory component should establish clear prohibitions on AI applications that pose unacceptable risks to fundamental rights. Certain applications should face outright prohibition including social credit scoring systems that evaluate citizens' trustworthiness based on behavior or personal characteristics, real-time biometric identification in public spaces for mass surveillance purposes, and AI systems specifically designed to manipulate behavior in ways that cause physical or psychological harm.

Beyond absolute prohibitions, high-risk AI systems should face mandatory compliance requirements including pre-deployment human rights impact assessments with specific attention to gender-disaggregated analysis (Lindblad Kernell and Veiberg, 17). Organizations deploying AI for employment decisions, credit allocation, access to government services, or law enforcement applications must systematically assess potential impacts on women and gender minorities, document mitigation measures, and establish ongoing monitoring mechanisms. Transparency obligations represent another essential mandatory component. Organizations deploying consequential AI systems must disclose to affected individuals that automated decision-making is being employed, provide meaningful information about the logic involved, and explain the significance and envisaged consequences of such processing.

Human rights due diligence represents a systematic process through which organizations identify, assess, prevent, mitigate, and account for adverse human rights

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

impacts of their operations, products, and services (Ebert et al., 21). This concept, established through the UN Guiding Principles on Business and Human Rights and increasingly translated into mandatory legal requirements across multiple jurisdictions, applies with particular importance to AI systems given their potential for widespread and systematic rights impacts. For Indonesia, mandatory HRDD requirements for AI systems align with constitutional obligations under the 1945 Constitution while operationalizing these abstract principles through concrete procedural requirements.

The HRDD process for AI systems should encompass the complete lifecycle from initial development through ongoing deployment. During the development phase, organizations must conduct preliminary impact assessments examining potential rights implications before systems are built. Dataset assessment represents a critical HRDD component given that algorithmic discrimination frequently originates in biased training data (Noble, 67). Organizations must document the sources, composition, and characteristics of datasets used for AI training, conducting statistical analysis to identify potential biases. Algorithm design and testing phases must incorporate gender analysis and diverse participation. Development teams should include women and gender minorities in substantive roles rather than token representation, ensuring diverse perspectives inform design decisions.

Regulatory sandboxes represent controlled environments where innovative technologies can be tested under regulatory supervision with appropriate safeguards for potentially affected populations (GIZ Asia, 33). This governance instrument has gained prominence internationally as regulators recognize the challenge of developing appropriate rules for rapidly evolving technologies before understanding how they function in practice. For Indonesia's AI governance framework, regulatory sandbox mechanisms offer particular value given the country's dual objectives of fostering domestic AI innovation while ensuring robust rights protections.

Gender-sensitive sandbox design requires several specific features distinguishing these mechanisms from generic innovation testing environments. First, sandbox participation criteria should explicitly prioritize applications addressing gender inequality or improving outcomes for women and gender minorities. Second, sandbox safeguards must include mandatory gender impact monitoring throughout testing periods. Participating organizations should be required to collect

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

gender-disaggregated data on system usage, outcomes, and user experiences, analyzing whether applications produce different results or effects across gender identities (European Parliament, 45). Third, sandbox governance structures should include meaningful participation from women and gender minorities as decision-makers rather than merely as test subjects.

D. CONCLUSION

The integration of gendered human security principles into Indonesia's artificial intelligence governance framework represents not merely a technical regulatory challenge but a fundamental constitutional imperative grounded in the nation's commitment to human dignity, equality, and non-discrimination. The empirical evidence documented throughout this research demonstrates that algorithmic discrimination against Indonesian women constitutes a present reality rather than a hypothetical future concern, with documented cases spanning employment platforms, financial services, biometric verification systems, and technology-facilitated gender-based violence. The 80.8 percent surge in online gender-based violence cases during 2024, the systematic disadvantaging of female gig economy workers through algorithmic management systems that penalize caregiving responsibilities and biological conditions, and the repeated failures of biometric verification technologies to recognize women wearing religious head coverings collectively establish the urgency of regulatory intervention.

Indonesia's existing constitutional and legal frameworks provide foundational principles for addressing these challenges, particularly through Article 28G and Article 28I of the 1945 Constitution establishing rights to dignity, security, and freedom from discriminatory treatment. However, current statutory instruments including the Personal Data Protection Law were conceived prior to widespread AI deployment and consequently lack specific provisions addressing algorithmic discrimination, automated decision-making transparency, and the unique mechanisms through which AI systems perpetuate gender-based harms. This regulatory gap creates accountability deficits where affected individuals lack effective remedies, organizations deploying discriminatory systems face minimal consequences, and structural patterns of algorithmic discrimination persist unchecked.

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

The three-pillar regulatory framework proposed in this article offers a comprehensive yet implementation-feasible approach to integrating gendered human security into Indonesia's AI governance. The gender-responsive Smart Mix Approach recognizes that effective regulation requires both mandatory state intervention for high-risk applications and voluntary industry leadership for broader responsible AI development. Mandatory prohibitions on inherently rights-violating applications combined with binding transparency and impact assessment requirements for high-risk systems establish baseline protections, while voluntary codes of conduct and industry commitments create space for proactive responsibility and adaptive learning as technologies evolve.

The mandatory Human Rights Due Diligence pillar operationalizes constitutional obligations through concrete procedural requirements applicable throughout the AI lifecycle from initial development through ongoing deployment. By requiring systematic identification and mitigation of gendered harms, mandating dataset diversity assessment, ensuring algorithmic design incorporates intersectional analysis, and establishing ongoing monitoring mechanisms, HRDD transforms abstract equality principles into actionable organizational practices. The phased implementation approach with an initial engagement period allows organizations to develop necessary capabilities while ensuring meaningful compliance rather than superficial checkbox exercises.

The gender-sensitive Regulatory Sandbox pillar balances Indonesia's legitimate innovation and economic development objectives with robust rights protections by creating controlled testing environments under regulatory supervision. By prioritizing applications addressing gender inequality, mandating gender-disaggregated impact monitoring, ensuring meaningful participation from women and gender minorities in governance structures, and establishing clear exit criteria requiring demonstrated equity before graduation to unrestricted deployment, sandboxes enable beneficial innovation while preventing premature scaling of inadequately tested systems.

Implementation of this regulatory framework requires coordinated action across multiple institutional actors. The Ministry of Communication and Digital Technology, working in coordination with the Ministry of Law and Human Rights, the Ministry of Women's Empowerment and Child Protection, and the National Human Rights

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

Commission, should establish an interagency coordination mechanism specifically addressing AI and human rights (Parasurama et al., 56). Civil society organizations including SAFEnet, ICT Watch, EngageMedia, and women's rights groups must participate meaningfully in policy formulation, implementation oversight, and grievance mechanism design, leveraging their expertise in documenting technology-related rights violations and understanding affected communities' lived experiences.

The National Task Force established in 2025 to develop Indonesia's national AI roadmap provides a critical policy window for integrating these recommendations into foundational governance frameworks. The participation of civil society organizations in various working groups demonstrates existing momentum toward rights-based approaches, yet sustained advocacy supported by rigorous evidence remains essential to ensure gender considerations receive substantive attention rather than superficial acknowledgment. International cooperation and technical assistance from organizations including UNESCO, UNDP, and bilateral development partners can support capacity building for regulatory agencies, provide comparative learning opportunities from other jurisdictions' implementation experiences, and facilitate access to expertise in human rights impact assessment methodologies.

Several areas require continued research and policy attention beyond the scope of this article. First, the environmental impacts of AI development including energy consumption by data centers and carbon emissions from large-scale model training intersect with gender and climate justice considerations, as women in developing nations disproportionately bear climate change burdens (Hao, 275). Second, the labor implications of AI automation extending beyond gig economy platforms to encompass broader employment sectors require gender-disaggregated analysis of displacement risks and benefit distribution. Third, the governance challenges posed by cross-border data flows and extraterritorial AI deployment necessitate regional cooperation frameworks within ASEAN to establish common standards while respecting national sovereignty.

Indonesia's approach to AI governance carries significance extending beyond national borders, as the country's experience will inform policy development across Southeast Asia and the Global South more broadly. By centering gendered human

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

security principles, prioritizing the experiences of marginalized populations, and refusing to accept that algorithmic discrimination represents an inevitable technological externality, Indonesia can demonstrate that innovation and human rights represent complementary rather than competing objectives. The constitutional commitments embodied in the 1945 Constitution, the vibrant civil society sector documenting and challenging technology-related rights violations, and the policy opportunity presented by current AI roadmap development collectively position Indonesia to develop governance frameworks that protect human dignity in the algorithmic age while fostering beneficial technological advancement. The challenge facing Indonesian policymakers is not whether to regulate AI but how to do so effectively, and the answer must place gendered human security at the foundation.

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

REFERENCES

Adhari, Luthfi Maulana. "Jalan Terjal Ojol Perempuan, Bertaruh Pada Aspal dan Algoritma: Hasil Riset Konde (1)." Konde.co, April 30, 2025.

Benjamin, Ruha. *Race After Technology: Abolitionist Tools for the New Jim Code*. Cambridge: Polity Press, 2019.

Burrell, Jenna. "How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms." *Big Data & Society* 3, no. 1 (2016): 1-12.

Crenshaw, Kimberlé. "Mapping the Margins: Intersectionality, Identity Politics, and Violence Against Women of Color." *Stanford Law Review* 43, no. 6 (2021): 1241-1299.

Ebert, Isabel, Thorsten Busch, and Florian Wettstein. *Business and Human Rights in the Data Economy: A Mapping and Research Study*. Berlin: German Institute for Human Rights, 2020.

Eubanks, Virginia. *Automating Inequality: How High-Tech Tools Profile, Police, and Punish the Poor*. New York: St. Martin's Press, 2018.

European Parliament. *New Product Liability Directive - Q4 2020*. September 2024.

GIZ Asia. *Regulatory Sandboxes for AI: Lessons from Asia*. Bangkok: Deutsche Gesellschaft für Internationale Zusammenarbeit, 2024.

Hao, Karen. *Empire of AI*. New York: Penguin Press, 2025.

Haq, Mahbub-ul. *Reflections on Human Development*. New York: Oxford University Press, 2015.

Kementerian Komunikasi dan Digital. *RPJMN 2025-2029: Transformasi Digital Indonesia*. Jakarta: Kementerian PPN/Bappenas, 2024.

Kid.kdf. "Universitas Udayana Buka Suara Soal Diduga Mahasiswa Edit Foto Asusila." CNN Indonesia, April 25, 2025.

Konstitusi Republik Indonesia. *Undang-Undang Dasar Negara Republik Indonesia Tahun 1945*. Jakarta: Sekretariat Jenderal MPR RI, 2020.

Lindblad Kernell, Emil, and Cathrine Bloch Veiberg. *Guidance on Human Rights Impact Assessment of Digital Activities: Introduction*. Copenhagen: Danish Institute for Human Rights, 2020.

Mhlambi, Sabelo, and Simona Tiribelli. "Decolonizing AI Ethics: Relational Autonomy as a Means to Counter AI Harms." *AI and Ethics* 3 (2023): 869-882.

Noble, Safiya Umoja. *Algorithms of Oppression: How Search Engines Reinforce Racism*. New York: NYU Press, 2018.

Nurbaiti, Alya. "Hari-Hari Lady Ojol: Narik, Kerja Domestik, dan Sulit Sejahtera." Project Multatuli, July 20, 2023.

OHCHR. *Access to Remedy and the Technology Sector: A "Remedy Ecosystem" Approach, A B-Tech Foundational Paper*. Geneva: United Nations, 2020.

Parasurama, Pamungkas, debby kristin, and Siti Rochmah A. Desyana. *Kertas Posisi Masyarakat Sipil: Peta Jalan Kecerdasan Artifisial Indonesia*. Jakarta: SAFEnet, 2025.

Pertiwi, Salsabila Putri. "Jalan Terjal Ojol Perempuan, Bertaruh Pada Aspal dan Algoritma: Hasil Riset Konde (2)." Konde.co, May 1, 2025.

Ruggie, John Gerard. *Just Business: Multinational Corporations and Human Rights*. New York: W.W. Norton, 2013.

ISSN: 2528-7486/ 2654-9298 Vol. 11 No. 1, Februari 2026

SAFEnet. *Laporan Tahunan Kekerasan Berbasis Gender Online 2024*. Jakarta: Southeast Asia Freedom of Expression Network, 2024.

Trisianto, Hendra. "Gara-Gara Kamera Jahat, Hampir Saja Gagal Ikut Tes CPNS." Laman Pemerintah Kabupaten Probolinggo, September 27, 2021.

Undang-Undang Perlindungan Data Pribadi. *Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Perlindungan Data Pribadi*. Jakarta: Sekretariat Negara, 2022.

UNDP. *Human Development Report 1994*. New York: Oxford University Press, 2015.

UNESCO. *Recommendation on the Ethics of Artificial Intelligence*. Paris: United Nations Educational, Scientific and Cultural Organization, 2021.