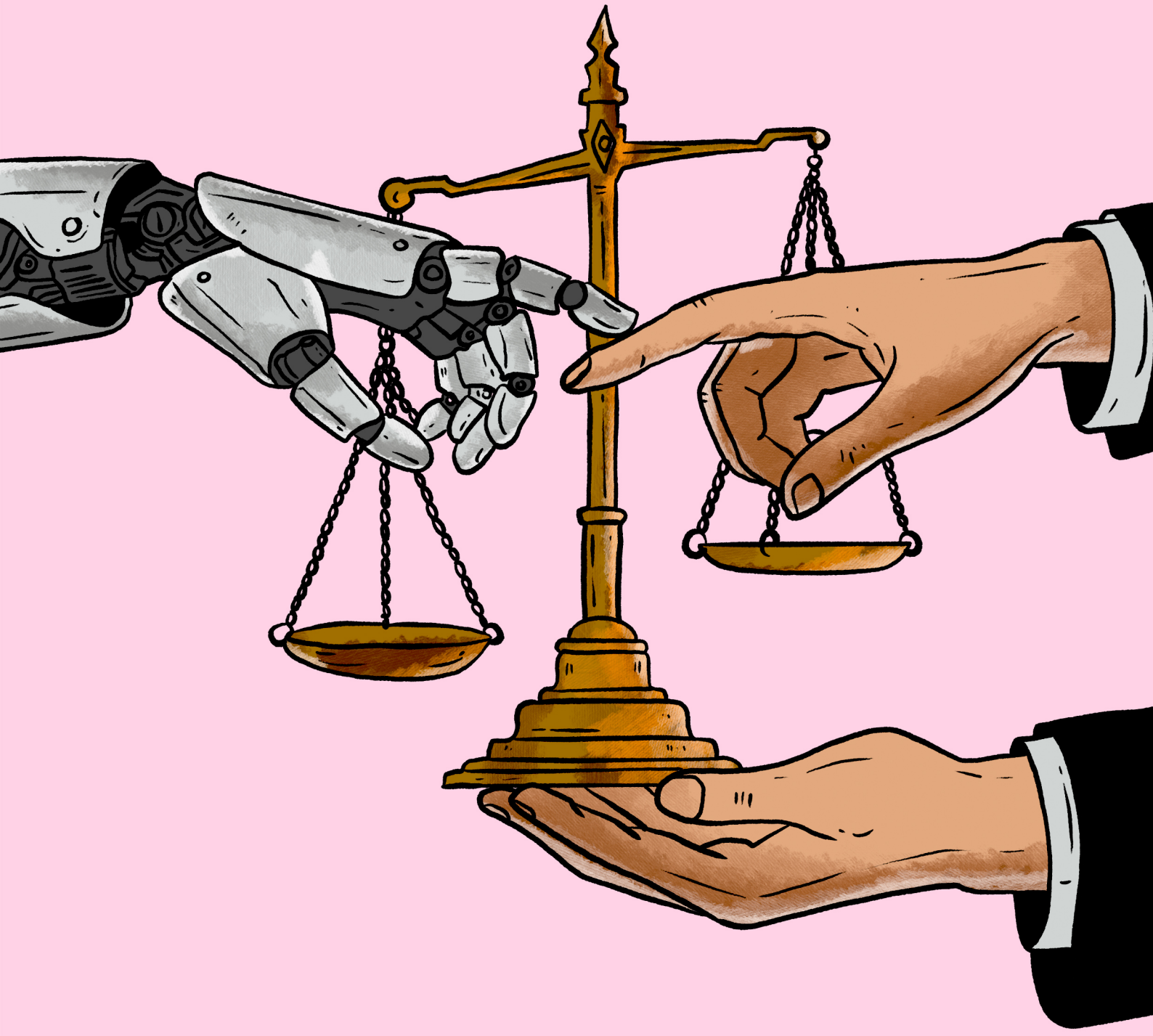
Civil Society Position Paper

Breaking Boundaries: Centring Human Rights in the National AI Strategy Towards #GoldenIndonesia2045



Authors

Parasurama Pamungkas
debby kristin
Siti Rochmah A. Desyana

Script Contributor

Aliansi Jurnalis Independen (AJI)
Center for Digital Society (CfDS)
Dewan Kesenian Jakarta (DKJ)
ECPAT
ELSAM
GIZ Asia
Humanis
ICT Watch
Lembaga dan Advokasi Independesi Peradilan (LeIP)
Perludem
Pusat Studi Hukum dan Kebijakan (PSHK)
PurpleCode Collective
SAFEnet
Tempo
Trend Asia
Wikimedia Foundation
Yayasan Tifa

Expert Reviewer

Prof. Dr. Sinta Dewi, S. H., LL. M.
(Faculty of Law - Padjajaran University)
Ir. Windy Gambetta, M. B. A.
(School of Electrical Engineering and Informatics - Bandung Institute of Technology)

Translated from Indonesian by

Marina Nasution

Layouter

Docallisme Studio

Published July, 2025 v

Executive Summary

The 2025–2029 National Medium-Term Development Plan (RPJMN) document identifies digital transformation as a key strategy to boost the efficacy of public services. Artificial Intelligence (AI) plays a crucial role in this public sector improvement process. To achieve the full potential of AI technology, the entire ecosystem that supports it must be supported by respect for human rights. The President-elect's Vision and Mission for 2024-2029, Asta Cita, point 4, implicitly explores this spirit. That point demonstrates the value of enhancing the quality of human resources by fostering a dedication to gender equality and the empowerment of women, youth, and persons with disabilities, as well as by enhancing science, technology, education, healthcare, and achievements in the sports sector. Therefore, it is crucial to structure AI governance in Indonesia based on human rights, including in the formulation of a national AI roadmap.

In 2025, the Ministry of Communication and Digital (KOMDIGI) formed a National Task Force to draft the Artificial Intelligence Roadmap (AI), hereinafter referred to as AI Indonesia 2025. The goal of this AI roadmap is to establish strategic AI governance in accordance with the principles of the UNESCO AI RAM 2024. The National Task Force is divided into seven working groups (hereinafter referred to as 'Working Groups'), namely the Ethics Working Group, Policy Working Group, Infrastructure and Data Working Group, Use-case Working Group, Talent Working Group, Industrial Research and Innovation Working Group, and Investment Working Group. Experts, government officials, industry players, civil society organisations, and academics are among the seven working groups that will assess and incorporate initiatives that serve as Indonesia's modalities for developing an AI ecosystem framework. This project involves mapping strategic issues and priority programs to achieve the vision of #GoldenIndonesia2045. Civil society organisations (CSOs), such as EngageMedia, ICT Watch, and SAFEnet, are involved in four working groups: the Ethics Working Group, the Policy Working Group, the Talent Working Group, and the Infrastructure and Data Working Group.

The consensus among the CSOs involved in the development of this policy paper calls for and recommends several key issues, namely:

- Establishing meaningful participation from multiple stakeholders by supporting the growth of innovation while maintaining public interest protection in compliance with AI safeguards.
- Incorporating legally binding regulations and enhancing self-compliance commitments through voluntary self-assessments, and providing redressal channels in the event of negligence.
- Ensuring sovereignty in data governance by strengthening infrastructure and improving independent data management systems while ensuring that technological advancements do not negatively impact the public.

Table of contents

Executive Summary II

Discussion 1

1. Building a Conceptual Framework for AI 1

a. Stringing the Definition of AI 1

b. Identifying the Actors 1

c. Developing General Principles of AI Governance from a Local Perspective: Decolonising AI Ethics 3

d. Mapping Risks (Fact and Potential) 3

d.1 Risks of Colonialism and Racism 4

d.2 Risks to the Rights of Persons with Disabilities 5

d.3 Risks to Journalism 5

d.4 Risks to Women and Gender Minorities 5

d.5 Risks to Children 6

d.6 Risks to Freedom of Expression 6

d.7 Risks to Freedom of Assembly and Association 6

d.8 Other Risks 7

2. Integrating Human Rights-Based Approaches 7

a. Outlining the Smart Mix Approach in a Roadmap 7

b. Initiating Human Rights Due Diligence 8

c. Implementation 9

c.1 Developing a Self-Assessment Toolkit 9

c.2 Implementing a Regulatory Sandbox 9

3. Governance Framework and Civil Society Engagement 10

a. The Multi-stakeholder Approach in the AI Governance Cycle 10

b. Establish a Cross-Sector Collaboration Forum 10

4. Redress Mechanism 11

a. Building an Effective Redress Channel 11

b. Developing an Accountability Model for AI 11

Recommendations 14

1. Ethics 14

a. Integration of Basic Principles Reflecting the Indonesian Context in the Roadmap 14

b. A Multi-stakeholder Approach through the Establishment of Collaborative Forums 14

2. Policy 14

a. Independence from the State's Securitisation Agenda 14

b. Framing a Smart-Mix Approach in the Roadmap 15

c. Initiating Human Rights Due Diligence in AI Governance 15

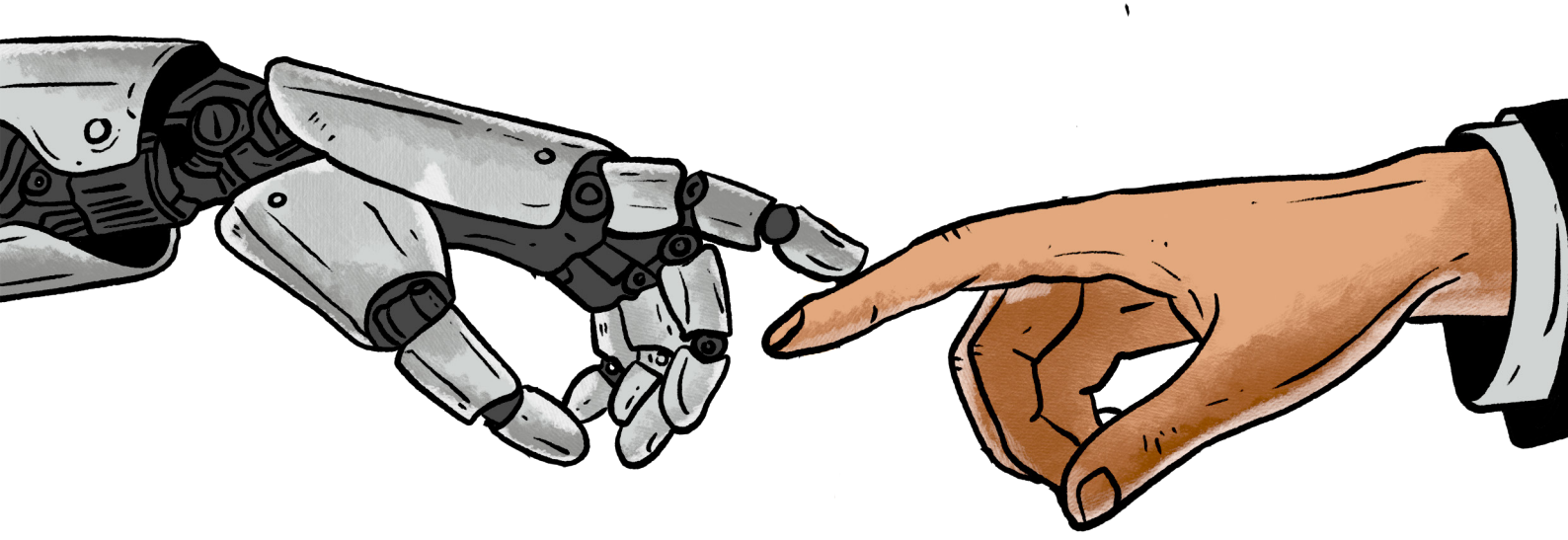
d. Recognising Diverse Risks in the Indonesian Context 16

e. Implementing a Data Protection Impact Assessment (DPIA) Mechanism in the Processing of Personal Data by AI Systems in the Public Sector 16

f. Building an Effective Redress Channel 16

g. Developing an AI Liability Model 16

h. Infrastructure and Data 16



Discussion

1. Building a Conceptual Framework for AI

a. Stringing the Definition of AI

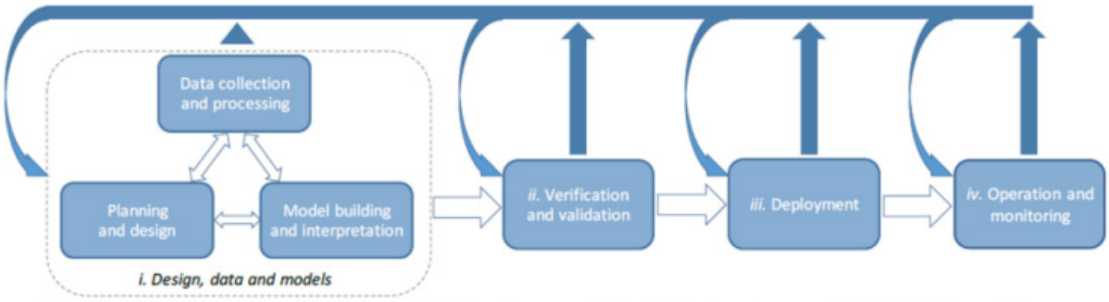
The definition of AI continues to evolve. What was once referred to as AI may not be considered so anymore. This shift is due to big tech companies continually advancing AI capabilities for their own interests, frequently distorting its boundaries and misinterpreting the term to encompass any type of algorithm or computer program.

This development in AI technological capabilities has also influenced the process of defining AI from 2019 to 2025, as outlined by the OECD, the European Union (EU) AI Act, and Indonesia in its White Paper, "Indonesia's National AI Roadmap". According to the latest EU AI Act in June 2024, AI is defined as a machine-based system designed to operate with varying degrees of autonomy and capable of adapting after use for explicit or implicit purposes, inferring from inputs it receives, and producing outputs such as predictions, content, recommendations, or decisions that can affect the physical or virtual environment. Some consider this definition to have weaknesses, including the difficulty of distinguishing AI systems from other systems, the vagueness of each element's meaning, and the notion that AI serves as a means to achieve human goals rather than an intrinsic objective.

Karen Hao, a journalist and author of the book "Empire of AI", states that, in the broadest sense, AI refers to machines that can learn, reason, and act for themselves. By identifying patterns in vast amounts of training data, AI can be programmed to make decisions or generate responses when confronted with new situations, much like humans and animals can.

b. Identifying the Actors

The OECD defines AI actors as "those who have an active role in the life cycle of an AI system, including organisations and individuals that deploy and operate AI systems." The diagram below illustrates the life cycle of an AI system:



Source: OECD (2019) AI in Society: AI system lifecycle as defined and approved by the OECD Expert Group on AI (AIGO) in February 2019

dentifying key actors is crucial in determining the accountability of an AI system. The following is a description of the actors in each stage of the AI system life cycle and their responsibilities:

AI Development Stage
At this stage, actors must ensure that no hazards occur throughout AI development. Currently, companies constitute the vast majority of AI development actors, and they frequently control the three actors listed below for the AI systems they develop.

Actors:
a. **Developer**: the party that creates or expands (develops) AI systems. This could be a third party or an internal organisation.
b. **Data Provider**: the party that provides data sources for data-driven AI development (machine learning). This could be a third party or an internal organisation.
c. **Computing Power Provider**: the party that provides computing power, either on-premises (the company manages the entire infrastructure) or in the cloud (the third-party cloud computing service provider manages the infrastructure). This includes cloud computing power, which covers both virtual service providers (cloud computing) and hardware and software providers.

AI Deployment Stage
Actors in this stage are responsible for ensuring that the marketed AI system is adequately protected and complies with the legal and ethical requirements set by the buyer.

Actors:
a. **Implementer**: the party that makes the AI operational. This can be a developer or another authorised party, including an internal organisation.

b. **AI Service Provider**: the party that will run the AI service.

End User Stage
The government, private sector, and civil society are key actors that deploy AI to support their daily activities. High-risk AI users include government actors such as the military and law enforcement. End users are responsible for implementing AI systems in their operations, including ensuring the appropriateness of the system's output, as well as identifying any dual-use mechanisms that could endanger users.

Actors:
a. **User**: The end user who interacts with the AI system. For example, in a self-driving automobile, it is the driver; similarly, in a medical consultation, it is the patient.

b. **AI service provider**: This entity is not necessarily the same as the developer. The service provider is responsible for managing the AI system service, including providing a Service

Level Agreement (SLA). For example, if a ministry operates a chatbot service for the public, the service provider is either the ministry itself (e.g., the IT Division) or a third-party provider.

c. Developing General Principles of AI Governance from a Local Perspective: Decolonising AI Ethics

Policymakers and international organisations have developed principles and ethical practice guidelines related to the right to privacy as a form of human rights—generally focusing on information and data protection. Privacy is closely linked to the concepts of freedom, personal liberty, individuality, autonomy, and human dignity. Furthermore, the right to privacy can also be defined as a person's freedom to choose without coercion from others, including information often displayed through certain platforms.[1]

The principle of dignity is reflected in the second principle of Pancasila and Article 28G of the 1945 Constitution. When applied to the technological sphere, AI systems must respect the autonomy and dignity of every individual, especially in decision-making that impacts people's lives. To adopt a decolonial approach to AI, Indonesia must prioritise the most marginalised and vulnerable people—those who continue to suffer the brunt of the negative impacts of innovation and scientific progress—at the centre of technological design.[2] This obligation is also outlined in Law No. 27 of 2022 concerning Personal Data Protection (PDP Law), which guarantees the protection of personal data, which is considered a constitutional right of Indonesian citizens, and prohibits the illegal acquisition, collection, disclosure, and use of other people's personal data.

The state is obligated to recognise and treat citizens' data as a derivative of their privacy and autonomy rights. Therefore, the government must ensure the implementation of sovereignty in data governance based on human rights. With cross-border data flows—the exchange and processing of data between countries—being commonplace[3], and data becoming a traded commodity, the Indonesian government must be proactive in filling the gaps in regulations protecting data transfers across borders. Otherwise, the consequences could threaten the country's sovereignty.

d. Mapping Risks (Fact and Potential)

In 2024, the United Nations General Assembly (UNGA) issued several resolutions, one of which addressed the importance of harnessing the opportunities of safe, secure, and trustworthy AI for sustainable development. This resolution calls on member states and

---

[1]Bert-Jaap Koops et al., *"A Typology of Privacy,"* Penn Carey Law: Legal Scholarship Repository, 2017, https://scholarship.law.upenn.edu/cgi/viewcontent.cgi?article=1938&context=jil.

[2]Sábëlo Mhlambi and Simona Tiribelli, *Decolonizing AI Ethics: Relational Autonomy as a Means to Counter AI Harms*, 2023, https://link.springer.com/article/10.1007/s11245-022-09874-2.

[3]Singapore and New Zealand have already implemented this under their respective trade agreements. See Mhlambi and Tiribelli, *Decolonizing AI Ethics*, 871.

other stakeholders—including the private sector—to avoid or discontinue the use of AI systems that do not comply with international human rights law or pose an undue risk to the enjoyment of human rights, particularly for those in vulnerable situations.[4]

AI-related risks must be addressed, mitigated, and managed continuously on, from planning to operation. Two examples: first, ethical factors must be constantly considered and periodically evaluated from the time a system is intended through its operation. Fairness can be implicit in other features or attributes, not merely gender, ethnicity, and so on. For example, unconscious prejudice between specific jobs and certain genders will persist even if a recommendation system is developed. Bias checking is essential before, during, and after processing.

Second, regular evaluations of AI systems are required. While studies have proved their accuracy, this does not guarantee their correctness or validity. This risk is vital to consider. This consideration is especially relevant when using GenAI, as a normally functioning machine may suddenly display a fatal error. GenAI is only safe if the AI's errors are harmless and the user can understand whether the output is correct or incorrect.

d.1 Risks of Colonialism and Racism

No AI system is entirely objective; every system reflects the values and assumptions of its creators, as well as the biases embedded in the data used to train it. Currently, the majority of existing AI systems are created by Western companies and trained using data derived from environments, languages, physical characteristics, or nations centred around the United States and Europe.[5] These technologies could potentially ignore or narrow down non-Western identities and cultural perspectives[6], reinforcing stereotypes and leading to widespread misrepresentation of marginalised communities.[7] The UN special rapporteur has highlighted that AI systems pose a particular danger to human rights because their use has been shown to reinforce existing inequalities and penalise those already systematically oppressed, such as migrants and certain racialised groups.[8]

Such an effect is evident, for example, in the phenomenon of LLMs (e.g., ChatGPT), which tend to answer questions from a Western perspective because the majority of the learning materials are sourced from English-language sources, so that judgments of right and wrong are heavily influenced by this perspective. This also creates inequities in the quality

[4]United Nation General Assembly, *Seizing the Opportunities of Safe, Secure and Trustworthy Artificial Intelligence Systems for Sustainable Development*, UN Doc. A/RES/78/265 (2024), https://undocs.org/A/RES/78/265.

[5]"*Shifting Beyond the Global North: The Data Driving AI Is Not Fit for Purpose – The OSCE's Role in Promoting Human-Centric Outreach and Strong Community Frameworks Beyond Its Network*," IFIMES, n.d., https://www.ifimes.org/en/researches/shifting-beyond-the-global-north-the-data-driving-ai-is-not-fit-for-purpose-the-osces-role-in-promoting-human-centric-outreach-and-strong-community-frameworks-beyond-its-network/5510.

[6]B. Jones, E. Luger, and R. Jones, *Generative AI & Journalism: A Rapid Risk-Based Review*, 2023, https://www.pure.ed.ac.uk/ws/portalfiles/portal/372212564/GenAI_Journalism_Rapid_Risk_Review_June_2023_BJ_RJ_EL.pdf.

[7]*Ibid*.

[8]European Digital Rights (EDRi), *Structural Racism, Digital Rights and Technology* (2020), https://edri.org/wp-content/uploads/2020/09/Structural-Racism-Digital-Rights-and-Technology_Final.pdf.

of service, where users who do not speak English tend to receive lower-quality results than those who speak English.[9] This pattern is also evident in biometric systems trained with data predominantly from specific groups or training datasets that already contain biased data, leading to algorithmic bias and discriminatory outputs.[10] Case studies related to these risks can be found in Appendix Tables for Case Examples AI.1 and AI.2.

d.2 Risks to the Rights of Persons with Disabilities

The UN Special Rapporteur identified one impact of the implementation of AI technology on the fulfilment of the rights of persons with disabilities[11] as the use of AI systems by the private sector in the selection and recruitment process. Algorithms tend to screen candidates based on historical patterns and internal company data, defining the "ideal profile" of employees in narrow and normative terms, thus risking reinforcing the systemic exclusion of already under-represented groups—including persons with disabilities. This scenario occurred in Indonesia in December 2024, as shown in Appendix Table for Case Examples AI.3.

d.3 Risks to Journalism[12]

AI can provide false or misleading information, which is difficult to detect and could potentially be included in content published by journalists. This is exacerbated by the fact that most AI systems lack the capability or are not designed to indicate the origin of the information they use clearly. These systems are designed to generate content that sounds convincing and engaging to humans, but often fail to follow standard citation rules or allow users to trace ideas back to their sources. An incident related to this risk occurred recently during the #SaveRajaAmpat campaign (see Appendix Table, AI Case Example.4).

d.4 Risks to Women and Gender Minorities

AI systems, particularly those using biometric inference, pose significant risks to gender minority groups by reinforcing binary conceptions of identity and perpetuating structural discrimination due, among other things, to the unequal availability of gender minority profile data[13] (Appendix Table, AI Case Example.5). This applies to women, as is the case with female economic workers in Indonesia, who are frequently subjected to suspension

---

[9] *"Multilingual AIs Are Better at Responding to Queries in English,"* New Scientist, accessed July 2025, https://www.newscientist.com/article/2387574-multilingual-ais-are-better-at-responding-to-queries-in-english/.

[10] Ashwini K.P., *Contemporary Forms of Racism, Racial Discrimination, Xenophobia and Related Intolerance,* Report of the Special Rapporteur on Contemporary Forms of Racism to the UN Human Rights Council, A/HRC/56/68, June 3, 2024, https://docs.un.org/en/A/HRC/56/68.

[11] United Nations Special Rapporteur on the Rights of Persons with Disabilities, *Report on AI and the Rights of Persons with Disabilities,* A/HRC/49/52, December 28, 2021, https://undocs.org/A/HRC/49/52.

[12] B. Jones, E. Luger, and R. Jones, *Generative AI & Journalism: A Rapid Risk-Based Review,* 2023, https://www.pure.ed.ac.uk/ws/portalfiles/portal/372212564/GenAI_Journalism_Rapid_Risk_Review_June_2023_BJ_RJ_EL.pdf.

[13] Daniel Leufer, "Computers Are Binary, People Are Not: How AI Systems Undermine LGBTQ Identity," *Access Now*. https://www.accessnow.org/how-ai-systems-undermine-lgbtq-identity/.

due, among other things, to the unequal capacity of biometric technology and application algorithms to recognise female worker profiles (see Appendix Table, AI Case Example 6). Furthermore, generative AI has the potential to be a tool for combating Technology-Facilitated Gender-Based Violence (TGBV) because it can produce visuals that resemble real individuals through a simple prompting process (see Appendix Table of AI Case Examples.7). Consequently, victims experience restricted access to essential services, vulnerability due to stigmatisation and objectification, and isolation.

d.5 Risks to Children

The use of generative AI often involves the collection and analysis of data about children, including personal information and behavioural patterns, which has the potential to violate children's privacy if misused[14], as in the case of child photo misuse in Appendix Table of Case Examples AI.8. Furthermore, the impact of prolonged interaction with generative AI systems on children's mental health is noteworthy, as children have found it difficult to distinguish the boundaries between interactions with humans and interactions with AI, as seen in cases of teen suicide involving generative AI (see Appendix Table, AI Case Examples AI.9).

d.6 Risks to Freedom of Expression

In a 2018 report, the UN Special Rapporteur highlighted the significant impact of AI use on freedom of expression and the dynamics of public information. One example is the potential for generative AI tools to generate misleading or hateful content, which not only endangers vulnerable and marginalised groups but can also result in serious legal consequences in jurisdictions with strict regulations on hate speech and the spread of misinformation in cybercrime[15] (see Appendix Table of Case Examples AI.10).

d.7 Risks to Freedom of Assembly and Association

The UN Special Rapporteur (2019) presented findings on various countries' strategies to restrict access to digital technology and implement intensive mass monitoring and surveillance practices, both in digital and physical[16] settings (see Appendix Table Case Examples AI.11). These actions are considered to have a significant impact on reducing civil liberties, particularly in hindering the exercise of the rights to free and peaceful assembly, discussion, and association, guaranteed under the international human rights framework.

---

[14]"*How Is Artificial Intelligence Reshaping Early Childhood Development?*" UNICEF, October 2024, https://www.unicef.org/media/163786/file/2024-10_Blog%20ECD%20and%20AI_cw_zj_am.pdf.pdf.

[15]United Nations Special Rapporteur on the promotion and protection of the right to freedom of opinion and expression, *Report on Artificial Intelligence Technologies and Implications for Freedom of Expression and the Information Environment*, A/73/348, August 29, 2018, https://www.ohchr.org/en/calls-for-input/report-artificial-intelligence-technologies-and-implications-freedom-expression-and.

[16]United Nations Special Rapporteur on the rights to freedom of peaceful assembly and of association, *Report on Opportunities and Challenges for the Rights to Freedom of Peaceful Assembly and of Association in the Digital Age,* A/HRC/41/41, May 17, 2019. https://ap.ohchr.org/documents/dpage_e.aspx?si=A/HRC/41/41.

d.8 Other Risks

AI also has an impact on copyright, media, and the creative ecosystem. Generative AI, for example, can generate text and visual data because it adapts various data used to train it. This training data is taken from creative workers, often without compensation or even consent from those who own the copyright to the creative product[17], as was the case with several Indonesian artists and creative workers listed in Appendix Table Case Examples AI.12. Furthermore, predictive AI systems have the potential to stifle the growth of MSMEs and gig workers, especially those who rely on e-commerce and e-hailing systems. The priority system based on non-transparent and uncontested partner assessments has a direct impact on their well-being, including lower daily incomes and longer working hours[18] (see Appendix Table Case Example AI.13). This practice has a broad and systematic impact on the economic rights of individuals from many segments of society.

2. Integrating Human Rights-Based Approaches

This section presents several recommendations and a conceptual framework surrounding the smart mix approach within the United Nations Guiding Principles on Business and Human Rights (UNGPs), human rights due diligence initiatives, and voluntary assessment mechanisms.

a. Outlining the Smart Mix Approach in a Roadmap

State-based regulations and voluntary initiatives from companies developing and using AI are integral to technology governance, including AI. These two regulatory bases can interact and complement one another, and must be tailored to address the various risks in each region, taking into account their context and locality.

The B-Tech Project outlines how the UNGPs can serve as a roadmap to bridge the governance gap in the technology era, emphasising the state's responsibility to adopt a balanced mix of voluntary and mandatory measures that require technology companies to respect human rights. Conceptually, the smart mix encompasses the following:

a. National and international measures include policies and regulations at both the national and global levels.
b. Mandatory and voluntary instruments combine legally obligatory regulations with

[17]K. Lin et al., "Governance of Generative AI in Creative Work: Consent, Credit, Compensation, and Beyond," ACM Conference on Human Factors in Computing Systems (CHI 2025), Yokohama, May 2025, https://www.researchgate.net/publication/388231892_Governance_of_Generative_AI_in_Creative_Work_Consent_Credit_Compensation_and_Beyond.

[18]Kunying Xin, et al., "Moral Hazard in Ride Hailing Services: Provide Disincentives from Ratings System Failures," *Lecture Notes in Education Psychology and Public Media* 105, no. 1 (July 2025): 118-131, https://www.researchgate.net/publication/393613868_Moral_Hazard_in_Ride_Hailing_Services_Provide_Disincentives_from_Ratings_System_Failures.

voluntary initiatives such as industry standards or company commitments.

c. Multi-stakeholder involvement includes affected communities, civil society, the state, the private sector, and academia.

b. Initiating Human Rights Due Diligence

Due diligence is a structured procedure carried out by business actors to detect and assess potential and actual impacts on individual rights. This mechanism is designed to ensure that business practices do not violate human rights, either directly or indirectly.[19] This procedure includes prevention and mitigation measures, continuous tracking of their implementation, and accountability through transparent reporting and communication of potential and actual impacts.[20]

A wave of responsible business regulation is impacting the global market with the emergence of numerous mandatory human rights due diligence (mHRDD) regimes, both enacted and still under formulation in various jurisdictions. This trend reflects an increasing drive from business actors and investors, as well as the active support of civil society organisations, to build an effective and impactful mHRDD legislative framework that guarantees human rights protection in business practices across sectors.[21]

One of the challenges in conducting human rights impact assessments is the inherent uncertainty in projecting the use of new technologies and their potential human rights impacts, particularly when such technologies lack empirical precedent or established ethical frameworks to draw on.[22] Furthermore, various social, cultural, and economic determinants impact humanity's multidimensional reality, including race, gender, religion, social stratification, level of well-being, sexual orientation, gender identity, citizenship status, physical condition and disability, and age. Therefore, the assumption of uniformity among victims of human rights violations must be critically examined. The analysis of the impacts of services, products, and business activities must take into account many perspectives and uncover frequently hidden forms of structural discrimination.[23]

---

[19]Isabel Ebert, 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), 21.

[20]Emil Lindblad Kernell and Cathrine Bloch Veiberg, *Guidance on Human Rights Impact Assessment of Digital Activities: Introduction* (Copenhagen: The Danish Institute for Human Rights, 2020), 17.

[21]United Nations, Mandate of the Working Group on the issue of human rights and transnational corporations and other business enterprises, October 22, 2020, https://www.ohchr.org/sites/default/files/Documents/Issues/Business/RecommendationsLegislativeProposal.pdf.

[22]Ashley Nancy Reynold, "Human Rights Due Diligence within the Tech Sector: Developments and Challenges," *Business & Human Rights Resource Centre*, https://www.business-humanrights.org/en/blog/human-rights-due-diligence-within-the-tech-sector-developments-and-challenges/.

[23]Bonita Meyersfeld, "The Rights of Women and Girls in HRIA: The Importance of Gendered Impact Assessment," in *Handbook on Human Rights Impact Assessment*, ed. Nora Götzman (Cheltenham: Edward Elgar Publishing, 2019), 154.

c. Implementation

Human rights due diligence in the development and use of AI can leverage Indonesia's existing modalities in business and human rights initiatives. The National Action Plan for Human Rights (RAN HAM) and the National Strategy for Business and Human Rights (Stranas BHAM) can serve as a general framework for implementing human rights due diligence. In the context of AI, the obligation to implement human rights due diligence should consider the scale and revenue of developers and users. In addition to the criteria for entities subject to due diligence obligations, Indonesia must also consider the engagement period before such obligations are fully implemented. This period is useful for preparing the due diligence ecosystem and providing space for AI development companies to implement improvements.

c.1 Developing a Self-Assessment Toolkit

After establishing a safety net through state regulations, businesses' next responsibility is to complement it by ensuring that respect for human rights is embedded in voluntary initiatives. A self-regulatory approach has several advantages that can support the efficiency and effectiveness of policy implementation. One of these is a relatively high level of compliance because self-regulatory mechanisms are generally based on contractual agreements between stakeholders.

Additionally, this model reduces regulatory costs for the state as the relevant entities independently carry out the monitoring and enforcement processes. In practice, online platforms' application of a rating model is a self-regulatory enforcement that allows consumers to submit feedback on services, pushing platform providers to make necessary improvements proactively. Its primary goal is to establish standards or assess business practices to ensure they align with sustainability targets, particularly those related to human rights protection and reducing social and environmental impacts.[24]

This self-assessment model would be useful as a voluntary mechanism prior to human rights due diligence becoming mandatory. This mechanism could also cover less strictly regulated AI systems, particularly those not classified as high-risk and therefore not subject to third-party conformity assessments.

c.2 Implementing a Regulatory Sandbox

An AI regulatory sandbox, also known as an AI regulatory testbed, serves as a learning tool for developing a more appropriate regulatory framework because it is context-sensitive, stable, and encourages local innovation. The best time to implement it is before a country finalises its AI-specific regulatory framework. This tool offers mutual benefits: regulators

[24]OECD, *The Role of Sustainability Initiatives in Mandatory Due Diligence: Background Note on Regulatory Developments Concerning Due Diligence for Responsible Business Conduct* (2022), https://mneguidelines.oecd.org/the-role-of-sustainability-initiatives-in-mandatory-due-diligencenote-for-policy-makers.pdf.

gain early visibility and practical insights into real-world applications and their associated risks, while innovators, particularly startups and local developers, benefit from guidance, support, and regulatory clarity throughout the product development process.

They view regulatory sandboxes as a means to identify implementation issues, particularly for local developers, and make necessary changes before broader legislation takes effect.

This approach is based on research involving more than 60 AI regulatory sandbox initiatives across Asia. The European Union, through its comprehensive AI Act (EU AI Act), mandates the establishment of regulatory sandboxes by August 2026, as the EU views them as a mechanism to identify implementation challenges, particularly for local developers, and to make necessary adjustments before broader regulations are implemented.

3. Governance Framework and Civil Society Engagement

The implementation of a multi-stakeholder policymaking process has been in place since the dawn of internet governance. This approach prioritises collaboration among stakeholders, including government, the private sector, civil society, the technical community, and academia. Transparency, inclusivity, and collaborative decision-making are the key principles of this model, which guarantee the inclusion and consideration of all voices in policy formulation.

a. The Multi-stakeholder Approach in the AI Governance Cycle

Given the complexity, broad scope, and potential impacts of AI technology, both governments and businesses are required to adopt a participatory, inclusive, and transparent approach to the procurement and implementation of AI systems. This approach requires the active involvement of a wide range of stakeholders, including independent human rights experts, civil society organisations, the academic community, the labour force, and the individuals and groups directly affected.

Governments and corporations need to do the following:[25]

a. Publicly announce any projects involving the use of AI tools, promoting transparency and enabling public debate.
b. Before deploying the tool, engage relevant stakeholders and actors early in the AI product lifecycle.
c. Establish a multi-stakeholder forum with adequate authority.

[25] Access Now, *Submission to the United Nations Working Group on Business and Human Rights to Inform UN Human Rights Council 59th Session Report on "The Use of Artificial Intelligence and the UN Guiding Principles on Business and Human Rights (UNGPs)"*, January 15, 2025, https://www.ohchr.org/sites/default/files/documents/issues/business/workinggroupbusiness/wg-business-cfis/2025/subm-use-artificial-intelligence-cso-access-now.pdf.

b. Establish a Cross-Sector Collaboration Forum

The AI roadmap in Indonesia needs to mandate the establishment of a multi-stakeholder forum with the following capabilities:

a. Connected Recovery Access: This forum serves as a referral centre for victims or affected parties to obtain holistic recovery services (psychosocial, legal, economic, etc.), and includes a referral point for several sectoral recovery channels.
b. Risk Documentation and Mitigation: With systematic incident documentation, the forum serves as a critical database for trend analysis, risk pattern identification, and evidence-based policy development.
c. Cross-Sector Consultation: It serves as a space for dialogue between the government, academics, civil society organisations, affected groups, and the private sector, resulting in more participatory and inclusive policies.

4. Redress Mechanism[26]

a. Building an Effective Redress Channel

The Indonesian AI roadmap has yet to establish an effective redress mechanisml. Meanwhile, this procedural mechanism is inherent in AI governance and a logical consequence of recognising that AI technology has human rights impacts at both the micro and macro scales. The absence of a redress mechanism reflects how states and companies continue to pay insufficient attention to the perspectives and needs of actual and potential users, as well as redress mechanisms for human rights violations related to their activities, in both the design and operation of these mechanisms.[27] States play a critical role in ensuring and protecting human rights. However, the way technology companies implement their corporate responsibility to respect human rights can also significantly influence the performance of the remedy ecosystem in practice.[28]

At the same time, technology companies can complement aspects of the state-based regulatory ecosystem through self-regulation for direct redress, as an integral part of the smartmix approach described above. AI companies, both developers and deployers, should establish and participate in effective operational-level grievance mechanisms for individuals and communities who may be negatively impacted by their activities.[29]

---

[26] Also called recovery mechanisms, it focuses on how the system as a whole operates to provide recovery to affected people rather than on individual roles, attributes, and mechanisms.

[27] United Nations Human Rights Office of the High Commissioner, *Access to Remedy and the Technology Sector: Understanding the Perspectives and Needs of Affected People and Groups*, A B-Tech: Foundational Paper (2020), 1.

[28] United Nations Human Rights Office of the High Commissioner, *Access to Remedy and the Technology Sector: A "Remedy Ecosystem" Approach*, A B-Tech Foundational Paper (2020), 1.

[29] *Ibid,* 5-6.

b. Developing an Accountability Model for AI

States are responsible for ensuring that everyone within their jurisdiction is protected from human rights violations enshrined in the International Convention on Human Rights. To fulfil this responsibility, states must create and implement national policies that guarantee equal access to rights for all, such as providing a swift, appropriate, and adequate legal remedy mechanism to accommodate those whose rights have been violated and require justice.

In complex AI systems, disruptive factors for traditional accountability schemes include, first, the involvement of multiple actors in component procurement, dataset collection, and algorithm programming, which creates challenges in determining accountability and is coupled with low transparency within these systems; second, the interaction between AI elements and other components of smart device technology creates uncertainty about which system is ultimately responsible for the incident in question. The 2019 case of algorithmic discrimination against women by Apple Card exemplifies this (see Appendix Table of AI Case Examples.14). The government requires a new approach to resolve these conflicts, one that reduces the burden of proof for consumers and takes into account the harms caused by AI system products. In the European Union, for example, the revised Product Liability Directive (PLD) lowers the burden of proof by defining AI as a "product" and expanding product liability for AI systems, allowing victims to sue for damages caused by opaque and autonomous AI systems.[30]

Indonesia's Consumer Protection Law has detailed the rights and obligations of users and business actors and stipulates that when a user experiences a loss, the victim and their heirs can negotiate the form and amount of compensation with the service provider.[31] In the context of technology, this mechanism protects victims of AI incidents and has the potential to improve the quality of AI as a usable product. In addition to the Consumer Protection Law, digital technology governance in Indonesia is reflected in the Regulation of the Minister of Communication and Information Technology (Permenkominfo) on Private and Public Electronic Transactions (PSE) and Government Regulation No. 71 of 2019 concerning the Implementation of Electronic Systems and Transactions (PP PSTE), which adopts the definition of "business actor" from the Consumer Protection Law in the context of regulating the obligations of electronic system providers.[32] Therefore, the obligations of

[30]European Parliament, *New Product Liability Directive*, September 2024. The PLD itself has received criticism from local developers regarding its stance on strict liability, which does not accommodate the possibility of misuse of AI by its users. Further information is available in EuroCommerce, "Adapting product liability rules to the digital age and Artificial Intelligence," *EuroCommerce Position Paper*, June 2022, 2.

[31]Undang-undang Republik Indonesia Nomor 8 Tahun 1999 tentang Perlindungan Konsumen [Indonesian Consumer Protection Law or PK Law], Pasal 19–27 (1999). Law 8/1999 on Consumer Protection (PK). Articles 19 to 27 of the PK Law emphasise that business actors are responsible for providing compensation for damage, pollution, and consumer losses resulting from consuming goods produced or traded. This responsibility also includes ensuring that similar losses will not recur.

[32]Business actors in the Consumer Protection Law and the Government Regulation Number 71 of 2019 concerning the Implementation of Electronic Systems and Transactions (PP PSTE) are every individual or business entity, whether in the form of a legal entity or not a legal entity, which is established and domiciled or carries out activities within the legal territory of the Republic of Indonesia, either alone or together through an agreement to carry out business activities in

business actors stipulated in the above regulations apply to parties conducting business activities through electronic systems.

Challenges of the Consumer Protection Law:

1. The traditional transaction model between producers and consumers (end users) continues to influence this law.
2. The concept of business actors and consumers in this law does not yet accommodate the complexity of new technology life cycles, which involve multiple actors.

Opportunities in the Consumer Protection Law:

1. Introducing a presumption of liability[33] in the AI life cycle provides an alternative to ensuring responsible AI.
2. A proof model that shifts the burden of proof to business actors would solve the problem of opacity, which often presents a significant obstacle for victims in providing evidence.
3. Redress channels that represent communal losses will be relevant in the recognition of relational autonomy.

various economic fields.

[33]This principle of presumptive liability holds the defendant responsible for the losses incurred and can only be exempt from compensation if they can prove they were not negligent or at fault.

Recommendations

To ensure AI aligns with efforts towards #GoldenIndonesia2045 goals and supports digital downstreaming, sovereignty in data governance is essential, encompassing infrastructure, software, and data management. The development of a national AI roadmap must also make human rights the foundation of its governance. This list of recommendations addresses three issue groups:

1. Ethics

a. Integration of Basic Principles Reflecting the Indonesian Context in the Roadmap

To decolonise the AI ethical framework in Indonesia, the white paper must establish additional principles that are integral to the AI governance framework, specifically social justice, collective responsibility, restorative justice, and effective reparation.

b. A Multi-stakeholder Approach through the Establishment of Collaborative Forums

The white paper should develop a multi-stakeholder framework to encompass the policymaking process, AI system assessment, and digital literacy. Mechanisms must be put in place to involve relevant actors at the beginning of the AI product lifecycle, prioritising inclusivity by providing a dedicated space for affected groups, such as the disability community, women, gender minorities, and children, to contribute meaningfully throughout all stages of the AI development and usage process. Furthermore, the government must establish a multi-stakeholder collaborative forum that serves as a referral mechanism for affected individuals or groups to access available redressal pathways across various sectors. This forum plays a strategic role in documenting various forms of risk within the national context and has the potential to serve as a consultative platform that bridges communication between the government and other stakeholders. This step is crucial for accountability and transparent assessment of the potential risks and impacts of AI systems as well as ensuring that redress referral mechanisms align with human rights standards.

2. Policy

a. Independence from the State's Securitisation Agenda

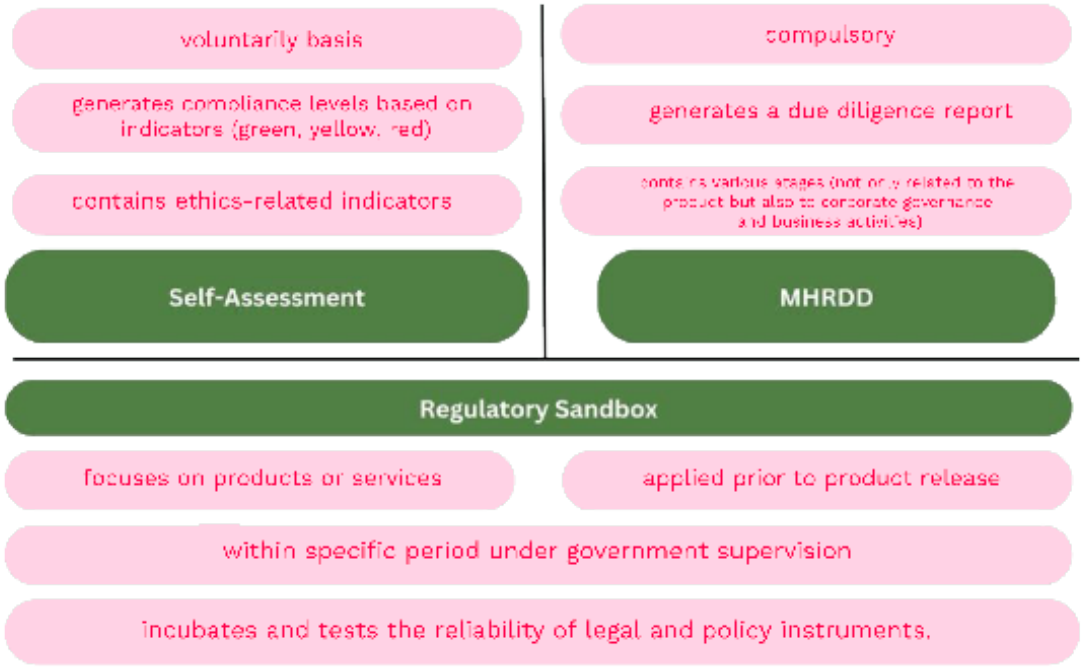
Policymaking in the AI sector must strive to be free from the influence of the state's securitisation agenda. Current dynamics in the European Union, the United States, and Israel demonstrate the rising tension between national security interests and the direction of AI industrial policy. These tensions have driven the expansion of the private sector—particularly technology companies and military entities—into state institutional roles, resulting in power imbalances among other stakeholders.

b. Framing a Smart-Mix Approach in the Roadmap

This approach can increase trust among investors and trading partners, thereby stimulating economic growth and supporting #GoldenIndonesia2045. This approach includes several parts, such as creating policies and rules at both the national and international levels; using both mandatory and voluntary instruments by mixing legally binding regulations with non-mandatory initiatives such as industry standards and company commitments and self-assessments; and involving multiple stakeholders, such as affected communities, civil society, government bodies, businesses, and academia, to ensure the governance's legitimacy and inclusivity.

c. Initiating Human Rights Due Diligence in AI Governance

Companies must include the active and meaningful participation of civil society actors, affected groups, and vulnerable communities at all stages of the due diligence process for digital technology advancements to reach their functional potential while minimising their negative impacts. This involvement aims to enhance the legitimacy of risk assessments and ensure that principles of fairness and inclusivity are reflected in the development and implementation of technology . In addition to the criteria for entities subject to due diligence obligations, Indonesia must consider the engagement period before completely implementing them. This period is useful for preparing the due diligence ecosystem and providing space for AI development companies to implement improvements



d. Recognising Diverse Risks in the Indonesian Context

The Indonesian AI white paper should acknowledge several AI risks that reflect inequalities in social structures. This recognition aims to clarify that the impacts of AI are not individual and demonstrate that technology does not operate in a vacuum but can reproduce and reinforce unequal power relations—asymmetrical and hierarchical—including colonial legacies that perpetuate systemic injustice and implicit structural inequalities. Some of these risks include colonialism and racism, as well as risks to persons with disabilities, journalistic work, women and gender minorities, children, freedom of expression, freedom of assembly and association, and others.

e. Implementing a Data Protection Impact Assessment (DPIA) Mechanism in the Processing of Personal Data by AI Systems in the Public Sector

To support the provision of AI systems for public services, the processing of personal data requires adopting systemic measures that balance the legal basis for data processing with the interests of state administration, which creates an unequal power relationship between personal data subjects (the public) and data controllers (the government). In this regard, the DPIA can address this imbalance through a series of risk identification, assessment, and mitigation processes in accordance with the PDP Law framework.

f. Building an Effective Redress Channel

The white paper should establish a discussion of effective redress as a logical consequence of recognising the diverse risks that may occur in the AI lifecycle. The government and AI development companies must understand and consider the interactions between various types of redress mechanisms when designing, operating, and utilising them. The white paper should outline a more coherent, less fragmented, and more easily navigable redress landscape, considering the contributions companies can make, both individually and collectively, to addressing gaps and enhancing the redress ecosystem's functionality .

g. Developing an AI Liability Model

The government must explore appropriate liability models by leveraging existing legal frameworks and implementing necessary reforms to ensure effective protection. In Indonesia, one modality is "presumption of liability," which requires the defendant to prove their innocence and obliges companies to ensure the safety of their products.

h. Infrastructure and Data

Both governments and companies must formulate and implement policies to mitigate environmental impacts, such as the enormous carbon emissions associated with the

development of large language models (LLMs) . The next step is to ensure that AI companies and data centre developments do not harm society. A typical data centre consumes a substantial amount of energy, accounting for about 2% of global electricity usage.[34] They must avoid using dirty energy sources like coal, as the entire lifecycle, from mining to power generation, has a significant impact on society. There needs to be a principle of fairness for society in the energy usage process at data centres—the backbone and future of AI.

[34]Karen Hao, *Empire of AI* (New York: Penguin Press, 2025), 275.