IRIS
Information Resilience & Integrity Symposium

RESEARCH PAPER

The Algorithmic Bureaucrat: Navigating Regulatory Challenges of Generative AI in Indonesia's Public Sector

PANEL 2 | Surveillance and Privacy in Digital Development

Safer Internet Lab
Centre for Strategic and International Studies
Google
Universitas Gadjah Mada
Fakultas Ilmu Sosial dan Ilmu Politik
Center for Digital Society

Alia Yofira

Alia is a researcher at PurpleCode Collective, specializing in gender, human rights, and technology. She holds a Master of Laws (LLM) in Innovation, Technology, and the Law from the University of Edinburgh as a Chevening scholar. Her work spans legal and policy research, strategic advocacy, and stakeholder engagement on issues such as platform governance, data protection, mis/disinformation, technology-facilitated gender-based violence and AI governance. Alia collaborates with civil society, government, academia, and the tech industry to promote rights-based digital governance in Indonesia. She is passionate about building inclusive and accountable tech policy frameworks that center human rights and gender justice.

The Algorithmic Bureaucrat: Navigating Regulatory Challenges of Generative AI in Indonesia's Public Sector

Alia Yofira Karunian

This paper explores the policy and legal frameworks necessary to ensure the responsible use of generative artificial intelligence (GenAI) in Indonesia's public sector, amid a broader digital transformation aimed at enhancing governance efficiency. Employing a normative legal research methodology, the study examines current and emerging applications of GenAI across key public service sectors, with a focus on the health and migrant workers sector. In these areas, GenAI-powered chatbots are being piloted to assist with preliminary health assessments and provide real-time consular support to Indonesian nationals abroad. However, the deployment of GenAI raises significant legal and ethical challenges, including the legality of data scraping, risks of AI hallucinations, surveillance capitalism, linguistic inclusivity, and the effectiveness of human-computer interaction in emergency contexts. The paper argues that Indonesia's current reliance on soft law mechanisms, such as ethical guidelines, is insufficient to safeguard human rights in the age of AI. To ensure transparency, accountability, and compliance with legal standards, a transition toward binding regulatory frameworks is imperative. Special attention is given to the Personal Data Protection (PDP) Law, particularly regarding lawful data processing. To bridge existing governance gaps, the paper proposes incorporating community-led algorithmic auditing into the mandatory Data Protection Impact Assessment (DPIA) process, strengthening oversight of AI deployment in public services.

Keywords: Generative AI, Public Sector, Chatbot, UU PDP

Introduction

On one morning in Jakarta, I found myself seated in the crowded waiting area of a hospital. After more than two hours of waiting to register for a doctor's appointment, the exhaustion was palpable. Resting my head against the wall, I observed the cause of the delay unfold before my eyes. Several elderly patients ahead of me were struggling to complete the biometric verification process. "*Please use the hand sanitizer liquid to clean your fingers*," the registration staff repeatedly instructed, as the fingerprint reader failed time and again to recognize their prints. Ultimately, the staff abandoned the biometric scanner, resorting instead to photographing the patients to verify their identities. This scene raises a critical question: when studies have demonstrated a marked decline in fingerprint quality in individuals over 70 years of age (Beslay, Galbally, and Haraksim 2018), had the government adequately accounted for the unique needs of elderly patients when mandating biometric identity verification in hospitals?

This reflects a broader reality in Indonesia. The adoption of artificial intelligence (AI) in public services, including biometric systems, is part of a nationwide digital transformation effort aimed at enhancing governance efficiency in Indonesia. The Presidential Regulation No. 132 of 2022 concerning the Architecture of Electronic-Based Government Systems outlines strategic initiatives for integrating AI across various sectors, including the application of AI and Big Data in geospatial data management to support the One Data Indonesia framework, the deployment of AI in research and innovation services to improve administrative efficiency.

Against this backdrop, this paper seeks to examine a fundamental question: ***"****what policy frameworks are required to ensure the responsible use of AI, particularly generative AI (GenAI), in Indonesia's public sector?* In addressing this question, the paper will adopt a normative legal research methodology. The research will draw on literature reviews encompassing books, scholarly articles, news reports, press releases, and online resources retrieved via Google during the period of June to July 2025. This study will also undertake a comparative analysis of Indonesia's regulatory frameworks alongside those of the European Union (EU). Particular attention will be given to the EU's General Data Protection Regulation (GDPR) and Artificial Intelligence Act, and their parallels with Indonesia's Personal Data Protection Law and emerging AI governance ethical policies.

Accordingly, this paper is structured as follows: *First,* it provides an overview of GenAI use in the Indonesian public sector, with a focus on health and migrant worker sector. *Second,* it explores the opportunities and challenges arising from the deployment of GenAI in these sectors. *Third,* it analyzes the regulatory challenges in AI and data protection law, comparing the Indonesian and EU contexts. *Finally,* the conclusion presents the key findings and policy recommendations.

GenAI in Indonesian Public Sector: An Overview

The rapid advancement of Generative Artificial Intelligence (GenAI) has begun to reshape public sector operations worldwide, and Indonesia is no exception. This section provides a comprehensive overview of GenAI's relevance and emerging role within Indonesia's public sector. It begins with an introduction to the fundamental concepts and technologies that underpin GenAI, offering essential context for understanding its application in government functions. Following this, the discussion turns to how the Indonesian government is currently deploying or planning to deploy GenAI technologies across various sectors, highlighting key initiatives and areas of focus.

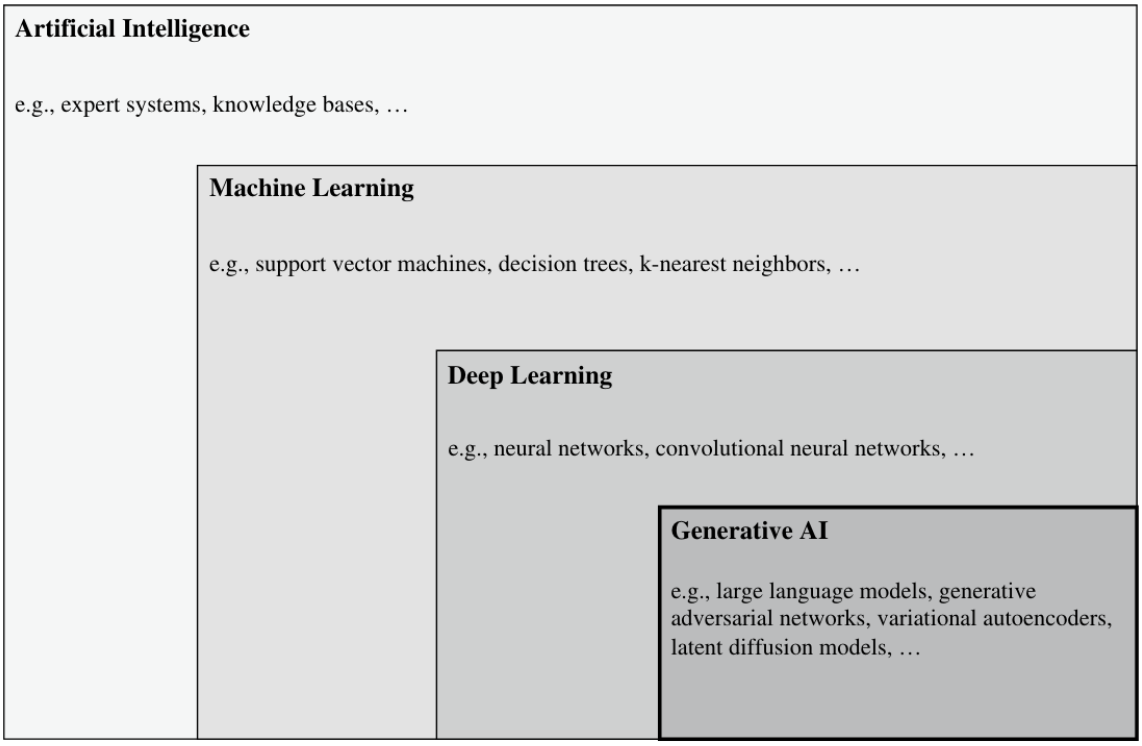
Introduction to Generative AI

Generative Artificial Intelligence, commonly referred to as "Generative AI" or "GenAI," refers to a class of AI systems, capable of "generating high-quality text, images, and other content based on the data they were trained on" ("What Is Generative AI?" 2023). Unlike traditional AI technologies, which primarily curate or aggregate existing digital content, Generative AI produces entirely new material in response to user prompts formulated in natural-language conversational interfaces (*Guidance for Generative AI in Education and Research* 2023, 11). This functionality is most commonly implemented in the form of chatbots, which are computer programs designed to simulate conversational exchanges with human users over digital platforms (Butterby and Lombard 2025, 1).

GenAI systems are accessed and operated by end users through application interfaces that facilitate prompting—a method of interaction whereby users provide instructions in natural language or other modalities to direct the system's outputs (Dang et al. 2022). The modality and content of a prompt directly influence the AI model's operation. For instance, a text-to-image system will interpret descriptive textual prompts to generate corresponding visual representations, whereas image-to-image systems use an initial image to guide the generation of further visual content (Banh and Strobel 2023).

The underlying technical foundations of GenAI are situated within the broader field of machine learning (ML), a subset of AI characterized by its ability to autonomously learn or improve performance through the analysis of large datasets without being explicitly programmed (Brynjolfsson and Mitchell 2017). Central to this improvement process is the Neural Networks (NN), interconnected layers of artificial neurons structured to mimic the human brain's processing mechanisms (Goodfellow, Bengio, and Courville 2016). These NN are implemented using deep learning (DL), an advanced subset ML that utilizes multilayered NN to identify complex patterns and correlations within data (Janiesch, Zschech, and Heinrich 2021). These models have unlocked a wide range of applications by enabling GenAI systems to autonomously produce creative and human-like content.

Figure 1. GenAI as the sub-field of AI systems



Source: Bahn and Strobel, 2023

Distinct types of NN architectures are applied within GenAI systems depending on the content modality. For example, text generation GenAI models typically rely on a class of NN known as general-purpose transformers, and in particular, Large Language Models (LLMs), which are capable of producing coherent and contextually appropriate textual content (*Guidance for Generative AI in Education and Research* 2023, 12). In contrast, audiovisual generation GenAI frequently employ Generative Adversarial Networks (GANs), sometimes combined with variational autoencoders (Nguyen et al. 2022).

The training of GenAI models often differs from that of traditional discriminative AI models. While ML-based discriminative AI models often employ supervised learning (Janiesch, Zschech, and Heinrich 2021), GenAI tends to employ semi-supervised learning approaches. This semi-supervised methodology combines a small volume of labeled data with a much larger corpus of unlabeled data, thereby enhancing the AI model's ability to learn from limited direct supervision while efficiently scaling to larger datasets (Banh and Strobel 2023). Previous study has shown that semi-supervised learning is prone to confirmation bias (Arazo et al. 2020), a tendency to overweight information that matches prior existing beliefs or choices.

GenAI in Indonesian Public Sector: Existing Application and Future Direction

This section examines the present and anticipated use of GenAI by the Indonesian government across key public service sectors. *First,* the discussion will address the application of GenAI in the Indonesian health sector, where GenAI-powered chatbots are being developed in a sandbox to assist in preliminary health assessments. *Second,* the analysis will turn to the migrant workers sector, where GenAI-powered chatbots are being utilised to provide real-time assistance and essential consular services to Indonesian nationals working abroad.

Health Sector

The Indonesian Ministry of Health (Kementerian Kesehatan, "Kemenkes") has articulated a strategic commitment to integrating AI into healthcare services. Further, at the 2025 APAC Health and Life Sciences Summit, Minister Sadikin stated that AI tools have the capacity to expedite diagnosis, minimise inpatient durations, and improve overall healthcare system quality and efficiency (*Antara News* 2025). In a June 2024 press release, the Minister of Health, Budi Gunadi Sadikin, described AI as a transformative tool capable of assisting medical professionals in interpreting complex biological data—such as genetic sequences, neuronal networks, and microbiome interactions—that exceed human cognitive capacity when analysed empirically (Kementerian Kesehatan 2024).

To operationalize this ambition, Kemenkes has entered into collaboration agreements with prominent technology entities. In July 2024, Kemenkes formalised a memorandum of understanding ("MoU") with Australian firm Harrison.ai to pilot AI-driven diagnostic applications across several state-run hospitals, including RSPON (stroke and neurological

imaging), Dharmais Cancer Hospital (oncological radiology and pathology), and Dr. M. Djamil Padang Hospital (tuberculosis radiology) (*Antara News* 2024b; *Antara News* 2024a).

Particularly on GenAI, Kemenkes and Google Cloud have announced an initiative to develop a controlled GenAI sandbox within Indonesia's "SATUSEHAT" digital health ecosystem (Kementerian Kesehatan 2024). This initiative aims to integrate AI-driven chatbot into SATUSEHAT platform, allowing Indonesian citizens to conduct preliminary self-assessment and obtain accurate pre-diagnoses prior to consulting a doctor or medical professional (*GovInsider* 2024). Indonesia's collaboration with global technology providers such as Google Cloud follows a wider regional trend, where similar sandbox environments for GenAI development have been established in partnership with governments across Southeast Asia, including Malaysia and Singapore (*TNGlobal* 2025; Hirdaramani 2023).

The integration of generative AI tools into the SATUSEHAT platform raises important concerns about data protection, particularly given the sensitivity of self-assessment and pre-diagnostic health information. While this initiative holds promise for improving access and efficiency in healthcare delivery, it also introduces potential risks related to surveillance and the misuse of personal health data. Kemenkes has stated that the GenAI sandbox will function within a "controlled" environment; however, details regarding data storage, access rights, and secondary use remain vague. Critical questions persist—such as whether user data will be anonymized, who will have access (including potential third-party developers), and how long this data will be retained.

Without a publicly available privacy policy, there is a risk of repeating past mistakes, such as those associated with the PeduliLindungi platform. A 2020 Citizen Lab report revealed that personal data collected through PeduliLindungi were shared with PT Telkom Indonesia without users' consent (Lin et al. 2020, 6). It remains unclear why this data was shared, how it was used, or whether it was repurposed for digital advertising. These unresolved issues highlight the urgent need for transparent regulatory frameworks and independent oversight to ensure that the use of AI in healthcare builds public trust rather than undermining individual privacy.

Migrant Workers Sector

Indonesian female migrant workers (Perempuan Pekerja Migran Indonesia, or PPMI) continue to face substantial risks of violence abroad. In 2024, a total of 489 cases of violence against PPMI were recorded by the Indonesia National Commission on Violence Against Women (Komisi Nasional Anti Kekerasan Terhadap Perempuan or Komnas Perempuan), reflecting a

sharp increase of 90.27% from the previous year ("Organizing Data, Sharpening Direction: Reflections on Documentation and Trends in Violence Against Women Cases 2024" 2025, 72). The most frequently reported form of abuse was psychological violence (46.14%), followed by economic exploitation (37.22%), physical violence (12.93%), and sexual violence (3.71%) ("Organizing Data, Sharpening Direction: Reflections on Documentation and Trends in Violence Against Women Cases 2024" 2025, 75).

The victims primarily originated from East Java (Jawa Timur), West Java (Jawa Barat), West Nusa Tenggara (Nusa Tenggara Barat), and North Sumatra (Sumatera Utara), although in 302 cases the victim's province of origin remained unverified ("Organizing Data, Sharpening Direction: Reflections on Documentation and Trends in Violence Against Women Cases 2024" 2025, 73). Furthermore, 179 cases involved workers who migrated through non-procedural channels, leaving them outside the protection of formal government safeguards and further increasing their vulnerability to exploitation ("Organizing Data, Sharpening Direction: Reflections on Documentation and Trends in Violence Against Women Cases 2024" 2025, 75).

In response to this situation, on 21 April 2025, the Ministry of Foreign Affairs (Kementerian Luar Negeri or Kemenlu) in collaboration with UN Women have launched a Generative AI chatbot named SARI to assist Indonesian migrant workers abroad (*VOI* 2025a). SARI aims to serve as a first line of support outside traditional embassy or hotline services, facilitating accessible and dignified assistance for migrant workers across multiple time zones and legal jurisdictions (*VOI* 2025b). Importantly, the design and development of SARI have specifically taken into account the unique needs and circumstances of Indonesian female migrant workers, who comprise a significant portion of the overseas workforce. In addition to the Indonesian language, SARI has been programmed to understand and respond in select local languages (bahasa daerah), the selection of which was based on a demographic mapping of migrant worker populations originating from particular regions of Indonesia (*VOI* 2025a).

While SARI represents a promising shift toward more responsive and inclusive digital support, its long-term impact depends heavily on how it is governed and who gets to shape its evolution Although available information suggests that migrant workers were consulted during SARI's initial design, their continued involvement in the platform's evaluation and oversight remains crucial. Without sustained and meaningful participation, there is a risk that SARI could unintentionally reinforce existing power imbalances between state institutions, AI developers, and the very communities it is intended to serve. To avoid a top-down model of digital intervention, it is essential to embed mechanisms for ongoing consultation,

feedback, and accountability, ensuring that migrant workers have a meaningful role throughout the system's lifecycle. This approach is key to promoting democratic governance and ensuring that SARI remains responsive to the lived realities of its users.

Both Kemenlu and UN Women have asserted that SARI is designed to convey a sense of empathy beyond that of an ordinary machine, seeking to understand and reflect the emotional needs of its human users (*VOI* 2025a). Additionally. SARI aims to be the first line of responders for cases of violence against PPMI, complementing face-to-face consultation provided by Kemenlu ("Kementerian Luar Negeri Dan UN Women Memperkuat Pelindungan Perempuan Pekerja Migran Indonesia Melalui Inovasi Chatbot AI SARI" 2025). While SARI is presented as a supportive tool for female migrant workers, it is important to critically assess how AI-driven chatbots may also introduce new forms of dependency or disempowerment. A 2025 study found that AI-powered chatbots, especially those integrated into wellness apps, can create emotional bonds and dependencies similar to those formed in human relationships, potentially resulting in psychological risks such as ambiguous loss and unhealthy reliance (De Freitas and Cohen 2025). These risks underscore the importance of establishing safeguards and participatory governance mechanisms to ensure that tools like SARI genuinely enhance migrant worker agency, rather than constrain it.

GenAI in Indonesian Public Sector: Opportunities and Challenges

This section offers a critical analysis of the key challenges and concerns arising from the adoption of GenAI in the Indonesian public sector. Particular attention is given to the use of AI-driven chatbots, which are increasingly being integrated into public services in Indonesia, as elaborated in previous section.

In examining the broader implications of GenAI, this section addresses several pressing issues. These include *first,* the legality of data scraping practices, *second,* the risks associated with AI hallucinations (i.e., the generation of false or misleading information), and *third,* the potential for increased surveillance capitalism through AI-enabled systems. *Fourth,* this section explores concerns related to linguistic inclusivity, especially in a linguistically diverse nation such as Indonesia, and *lastly,* the effectiveness of human-computer communication in crisis or emergency contexts. Through this examination, the section seeks to provide a foundation for regulatory and policy considerations in the responsible deployment of GenAI technologies within the public sector.

The Legality of Data Scraping

One of the foremost privacy challenges posed by GenAI lies in the methods used to collect the data necessary to train such systems. Scholars note that GenAI models are frequently trained on vast datasets obtained either by scraping publicly available information from the internet or by repurposing user and customer data collected through various platforms (Solove 2024, 26). This unprecedented scale of data collection has raised serious concerns regarding individual privacy and data protection compliance. Unlike conventional AI systems that are typically designed for specific applications, GenAI models are created without a clearly defined context or purpose of use (Helberger and Diakopoulos 2023, 2). Their versatility and accessibility enable them to be deployed across numerous sectors and by a broad range of users, thereby amplifying both the scale and scope of personal data usage (Helberger and Diakopoulos 2023, 2). This dynamic and open-ended nature of GenAI makes it exceptionally difficult to assess and limit the contexts in which personal data, scraped or otherwise obtained, might be used.

Moreover, the pervasive practice of data scraping in GenAI development creates a fundamental conflict with established data protection principles, most notably the principle of data minimization under the Indonesian Data Protection Law (PDP Law) and the EU General Data Protection Regulation (GDPR). Scholars have argued that the indiscriminate extraction of personal data from online sources during model training renders compliance with purpose limitation and data minimization requirements practically unattainable (Solove 2024, 25). Because GenAI systems are designed for broad, undefined applications, it becomes nearly impossible to specify and limit the purposes for which scraped data will be processed. This legal tension underlines the need for clearer regulatory frameworks and more rigorous data governance practices in the development and deployment of GenAI technologies. Without such measures, the training processes of GenAI systems risk perpetuating large-scale privacy infringements on an unprecedented scale.

Despite the legal provisions outlined in Indonesia's PDP Law, current enforcement mechanisms remain limited, most notably due to the absence of an independent Data Protection Authority (DPA). DPAs play a critical role in mediating power imbalances between individuals and data controllers, ensuring accountability in cases of privacy violations such as unauthorized data scraping. Without such an institution, protections against extractive data practices, particularly in the context of GenAI training, are significantly weakened. Furthermore, the lack of meaningful community participation in decisions surrounding data use exacerbates the risk of exploitation. A more democratic approach would involve seeking

explicit consent from individuals whose data may be used in AI training, accompanied by robust investments in public AI literacy. This would help ensure that people not only understand their right to opt in or out of data collection but also recognize the potential implications of contributing their data to models that may be used for certain purposes. Such participatory frameworks are essential to aligning AI development with human rights standards.

AI Hallucinations: Errors and Harms

GenAI chatbots, particularly those based on large language models (LLMs), present notable risks of generating inaccurate, misleading, or entirely fabricated information, a phenomenon commonly referred to as "AI hallucination." This issue arises when generative AI tools perceive non-existent patterns or references, thereby producing outputs that are factually incorrect or nonsensical ("What Are AI Hallucinations?" 2023). These errors stem from the underlying architecture of AI systems, which rely on probabilistic pattern recognition drawn from training data rather than verifiable knowledge. Scholars have cautioned that AI models may fabricate academic papers or misattribute authorship merely because the outputs resemble plausible texts (Wolff, Lehr, and Yoo 2023, 6). This concern is particularly acute in high-stakes domains such as healthcare, where inaccurate responses can have severe consequences. For instance, Google's Gemini chatbot was found to generate fictitious medical journal articles in response to queries about thoracic outlet surgery, raising alarms about the reliability of AI-generated health information (Franklin 2024).

In addition to factual inaccuracies, AI hallucinations can result in substantial reputational harm to individuals. There have been documented instances of AI chatbots falsely accusing people of serious misconduct. In one case, a German journalist discovered that Microsoft's AI tool, Copilot, had wrongly described him as a child molester who had confessed to the crime, an entirely fictitious and defamatory assertion (*ABC News* 2024). Similarly, ChatGPT incorrectly described an Australian mayor to be involved in a foreign bribery incident in the early 2000s (*ABC News* 2023). In another case, a US law professor was also falsely accused by ChatGPT of sexually assaulting students on a trip that he "never took" while working at a school he "never taught at" (Verma and Oremus 2023). This shows that instances of AI hallucinations have been occurring widely across multiple countries.

The risks posed by AI hallucinations are especially pronounced for vulnerable populations in development contexts, who may lack the resources, digital literacy, or access to alternative information sources needed to verify AI-generated content. For example, users relying on AI

chatbots for legal or medical guidance abroad may act on inaccurate information, potentially resulting in harmful outcomes such as misdiagnosis, exploitation, or legal jeopardy. These risks are compounded when users place high levels of trust in the perceived authority of AI tools. A digital literacy survey by the Ministry of Communication and Digital Affairs revealed that digital literacy levels in Indonesia are highest among privileged, well-educated individuals living in urban and western regions, while men consistently outperformed women across all pillars of assessment (*Status Literasi Digital di Indonesia 2022*). These findings underscore how socio-economic, regional, and gender disparities shape unequal access to digital skills, further exacerbating the vulnerability of certain groups to AI hallucinations. Addressing these risks requires differentiated protections based on user context alongside targeted, gender-sensitive, and region-specific interventions to strengthen digital literacy and build resilience to AI hallucinations.

Governance mechanisms that enable community oversight, such as participatory model evaluation, feedback loops, and partnerships with CSOs, are also critical to ensuring accountability and contextual relevance. These approaches can help democratize the assessment of AI accuracy and promote equitable safeguards, particularly for those least able to challenge or interpret erroneous outputs on their own. Creating independent regulatory bodies dedicated to overseeing AI governance can significantly strengthen community oversight approaches by providing neutral oversight, balancing power dynamics, and ensuring that the perspectives of all stakeholders, particularly marginalized or vulnerable community, are meaningfully included in the oversight process (Cheong 2024). These bodies can also facilitate transparency, enforce accountability, and serve as a formal channel for community input in shaping AI policies and practices.

Language Inclusivity

A significant challenge in the development and deployment of GenAI, particularly in natural language processing (NLP) applications such as text generation, lies in the issue of language inclusivity. Despite the global reach of AI technologies, the vast majority of web-based training data is concentrated in a small number of high-resource languages, including English (Hamill-Stewart 2024). This disproportionate representation leaves speakers of low-resource languages (languages with limited digital presence and fewer linguistic datasets) largely excluded from the technological advancements of GenAI. Consequently, these communities face a heightened risk of receiving inaccurate, incomplete, or misleading AI-generated content, further exacerbating existing digital and informational inequalities.

Empirical studies have confirmed these concerns, particularly with respect to large language models (LLMs) that underlie many GenAI systems. Researchers have found that LLMs, while demonstrating high levels of linguistic fidelity and accuracy in high-resource languages such as English, frequently struggle with low-resource languages. Dr. Arthur Tang of RMIT Vietnam, for instance, has observed that LLMs produce notable errors in linguistic accuracy and informational relevance when used in Vietnamese, a low-resource language (Hoang 2024). A 2024 study exposed that an AI chatbot responding to a health-related query in Vietnamese incorrectly provided advice about Parkinson's disease when asked about heart problems, highlighting the critical risks of deploying GenAI chatbots in healthcare contexts where linguistic nuances are essential (Tang et al. 2024). These findings illustrate a broader truth about GenAI: the quality and inclusivity of its outputs are fundamentally limited by the scope and diversity of the training datasets on which these systems rely (Wolff, Lehr, and Yoo 2023, 6).

In the Indonesian context, efforts to address linguistic inclusivity in GenAI applications have begun to emerge. Indonesia, as the country with the second-highest number of living languages (742 in total) faces unique challenges and responsibilities in this regard (Yasyi 2022). A notable initiative is the development of the Sari chatbot, a collaborative project between Kemenlu and UN Women, which is designed to support Indonesian female migrant workers using both Bahasa Indonesia and several local languages (Bahasa Daerah). This approach reflects a commendable effort to make AI-driven chatbot more linguistically accessible. Nevertheless, similar considerations must guide the future deployment of GenAI chatbots in the Indonesian healthcare sector. In particular, during the government's ongoing GenAI sandbox initiatives, Kemenkes should prioritize the evaluation of language-related error rates, especially in chatbot applications, where miscommunication can have serious consequences for patient care.

Genuine language inclusivity in AI must extend beyond technical solutions to address deeper questions of cultural autonomy, community participation, and data sovereignty. A compelling example comes from New Zealand's *Papa Reo* project, which blends Indigenous knowledge with modern data science to support smaller language communities (Jones et al. 2023). *Papa Reo* empowers communities to collect their own data and build local expertise, ensuring the tools serve their needs. A core principle of *Papa Reo* is data sovereignty, that communities should retain full control over their data ("Papa Reo," n.d.). Te Hiku Media, one of the organizations behind the project, launched the *Kōrero Māori* campaign to collect speech data through a transparent, ethical, and community-driven process (Raj 2024). By gaining the

support of elders and using a *Kaitiakitanga* license, restricting data use solely for the benefit of the Māori people, they gathered over 300 hours of labeled speech from more than 2,500 volunteers in just 10 days (Raj 2024). This model demonstrates how inclusive, community-led AI development can preserve cultural identity and strengthen democratic participation. It offers a valuable framework for Indonesia as it navigates multilingual AI development.

Chatbot and Crisis Communication

Chatbots present notable opportunities in the provision of information and support, particularly in sensitive contexts. Prior research has shown that individuals are often more comfortable discussing personal or sensitive matters with chatbots, perceiving them as non-judgmental entities in contrast to human interlocutors (Pickard, Roster, and Chen 2016). Survivors, in particular, may disclose intimate details of their experiences to chatbots that they would otherwise withhold from human listeners (Park and Lee 2020). However, such disclosure requires the assurance of privacy and anonymity, as individuals are significantly more likely to engage with sensitive questions under conditions that safeguard their confidentiality (Joinson and Paine 2012).

Nevertheless, the sensitive nature of the data collected by chatbots gives rise to material concerns regarding privacy and security. Scholars have cautioned that over-privileged chatbots, those granted extensive access to personal data or excessive operational capabilities, pose substantial risks in terms of unauthorized data access, system vulnerabilities, and potential breaches (Edu et al. 2022, 6). Further concerns have been documented in the crisis management context, where individuals expressed doubts over chatbots' resilience against hacking, system failures, and the reliability of the content provided (Xiao and Yu 2025, 12). Beyond these technical vulnerabilities, users also face challenges regarding the credibility and objectivity of information disseminated by chatbots; some participants found chatbots less reliable than conventional search engines in this regard (Maeng and Lee 2022, 16).

Moreover, socio-economic disparities also exacerbate access barriers. Minoritized communities, including neurodivergent individuals, deaf persons, older adults, and those with limited literacy, often struggle to utilize chatbot (Butterby and Lombard 2025, 2). In the specific case of the SARI chatbot, such barriers may prevent the most vulnerable groups, non-procedural PPMI with low literacy, from accessing critical support during emergencies. Importantly, chatbots must be programmed to assess users' safety needs promptly; in situations of imminent danger, automated tools are not suitable substitutes for emergency

services, and clear directives to contact such services must be in place (Butterby and Lombard 2025, 3).

Given these limitations, chatbot deployment must be carefully developed to complement, rather than replace, human services. While this is already the case for SARI chatbot, to maintain their relevance, safety, and effectiveness, chatbot systems require ongoing review, quality assurance, and evaluation (Butterby and Lombard 2025, 5).

GenAI in the Indonesian Public Sector: Regulatory Challenges

This section seeks to critically examine the regulatory challenges associated with the use of GenAI in the Indonesian public sector. The analysis focuses on two key areas of legislation that are particularly relevant to GenAI governance in Indonesia: the existing policy and regulatory framework on artificial intelligence and data protection.

Artificial Intelligence

The rapid adoption of AI technologies in government activities in Indonesia has not yet been matched by a comprehensive regulatory framework to safeguard against ethical risks and misuse. At present, Indonesia adopts a soft law approach to AI governance, relying primarily on non-binding ethical guidance. One of the key instruments in this regard is the Circular Letter of the Ministry of Communication and Informatics No. 9 of 2023 concerning Artificial Intelligence Ethics (the "AI Ethics Circular Letter"). This Circular seeks to provide ethical guidance for business actors and Electronic System Operators (ESOs) across both private and public sectors. It is designed to support these entities in developing internal policies and conducting activities involving AI-driven consultation, analysis, and programming. In addition, the Circular sets out a range of ethical values and guiding principles intended to ensure that the use of AI remains aligned with applicable laws and regulations. The key ethical values outlined in the Circular are as follows:

Table 1. The AI ethics circular letter

Ethical Value	Explanation
Inclusivity	AI development shall serve the collective interest, upholding the principles of equality, fairness, and peace.

Ethical Value	Explanation
Humanity	AI systems shall respect human rights, foster social cohesion, and honor the diversity of beliefs and viewpoints.
Safety	AI shall safeguard user safety and the security of personal data, ensuring privacy and preventing harm to any party.
Accessibility	AI technology should be accessible to all groups without discrimination, in compliance with ethical standards.
Transparency	Data use in AI systems must be transparent so authorized users can understand how data is managed and prevent misuse.
Credibility and Accountability	AI-generated decisions and information must be reliable and accountable to the public.
Personal Data Protection	AI systems must protect personal data in accordance with the law.
Sustainable Development and the Environment	AI should be developed with consideration for its impact on people, the environment, and other living beings, supporting sustainability and social welfare.
Intellectual Property	The development and use of AI must comply with intellectual property rights protection laws.

Source: The Circular Letter on AI Ethics, 2023

The primary limitation of the AI Ethics Circular Letter lies in its legal nature. The Circular Letter functions solely as a set of guidelines and does not impose binding legal obligations on actors within the AI ecosystem. While the issuance of this Circular represents a positive initial step by the Indonesian government in addressing AI governance, the rapid expansion of AI use in the public sector highlights the need for a more comprehensive regulatory framework. Mere ethical guidance is insufficient to address the complex risks and potential harms associated with AI deployment.

Understanding the limitations of soft law approaches like the Circular Letter requires situating them within a broader spectrum of regulatory tools. AI governance requires a mix of approaches, each suited to different levels of risk and the maturity of a country's AI

ecosystem (Appaya and Ng 2024, 8). Soft law, like the AI Ethics Circular Letter, supports innovation but offers limited protection due to its non-binding nature. Industry self-governance, such as technical standards, can influence company practices but lacks enforceability and is inadequate for high-risk sectors. Regulatory sandboxes allow safe experimentation but are resource-intensive and hard to scale. Hard law provides clear, enforceable rules essential for high-stakes applications, though it risks lagging behind technological change.

Hard law, by contrast, establishes binding rules necessary for regulating high-stakes applications, though it may struggle to keep pace with technological advances. Given Indonesia's current institutional capacity and regulatory gaps, hard law is essential to set baseline safeguards, while soft law, industry standards, and sandboxes can complement it by supporting flexible, adaptive oversight. As such, Indonesia should begin transitioning from reliance on non-binding ethical guidance toward a more enforceable, comprehensive regulatory framework that ensures responsible and accountable AI deployment across sectors.

In developing such a hard law approach, Indonesia can draw important lessons from other jurisdictions that have begun regulating GenAI. GenAI presents unique challenges for regulatory frameworks, especially those built on a risk-based approach. As discussed in Section B, GenAI systems, characterized by the lack of a predefined purpose and the broad scale of their deployment, complicate the classification of such systems into traditional risk categories, such as those used in the EU AI Act. This raises questions about the feasibility of determining whether GenAI falls within the "high-risk" or "non-high-risk" categories (Helberger and Diakopoulos 2023, 2).

Scholars argue that the current risk-based approach is ill-suited to address the evolving nature of GenAI. The lack of a predefined purpose, the scale of usage, and the extraction of large-scale training data introduce new and complex risks that existing frameworks do not fully anticipate (Helberger and Diakopoulos 2023, 6). Accordingly, some scholars propose that GenAI should be treated as a distinct general risk category, warranting regulatory consideration beyond the existing frameworks for application-specific AI systems (Helberger and Diakopoulos 2023, 6). These discussions offer valuable insights for Indonesian policymakers. They highlight the importance of reassessing how GenAI fits into Indonesia's evolving AI governance landscape.

Data Protection

This sub-section critically examines the regulatory challenges posed by generative AI (GenAI) in the context of Indonesia's Personal Data Protection Law (Law No. 27 of 2022, or "PDP Law"), with particular focus on two key data protection issues: the absence of a clear legal basis for personal data processing in AI training and the protection of data subject rights. Scholars have noted that the use of personal data in GenAI training processes creates significant implications for various data subject rights. Given the scope of these issues and limitations of space, this paper will focus specifically on the right to transparency. Transparency is a cornerstone of data protection laws ensuring that individuals are informed about how their personal data is collected, used, and processed.

Legal Basis for Data Processing

PDP Law closely mirrors the European Union's General Data Protection Regulation ("GDPR"), particularly regarding the legal bases for personal data processing. Both frameworks stipulate that any processing of personal data must be based on one of six lawful basis: consent, contractual obligation, legal obligations, protection of vital interests, performance of a public interest, or legitimate interests. However, scholars have noted that several of these bases are difficult to apply in the context of GenAI training. Specifically, it is hard to justify the use of personal data for training AI models as necessary to fulfill a legal obligation, protect an individual's vital interests, or perform a task in the public interest (Wolff, Lehr, and Yoo 2023, 15). This disconnect highlights the challenges of applying traditional data protection frameworks to the complexities of AI development.

Moreover, even the legitimate interests basis, which often serves as a flexible justification for data processing, faces significant limitations when applied to GenAI. Scholars argue that because there is typically no direct relationship between AI developers and the data subjects whose personal data is scraped from public sources, and because these individuals do not reasonably expect their data to be used for AI training, the legitimate interests basis is largely inapplicable (Wolff, Lehr, and Yoo 2023, 16). This leaves only two viable legal grounds: consent or contractual obligations. Yet, both pose practical hurdles, as GDPR and the PDP Law require that consent be specific, affirmative, and informed, often through an explicit opt-in, and that data controllers disclose their processing activities and limit data collection to what is strictly necessary (Wolff, Lehr, and Yoo 2023, 16). These requirements present a significant compliance responsibility for GenAI developers relying on publicly available

personal data for training purposes. Therefore, it is important for GenAI developers to meet one or more legal basis for data processing.

Right to Transparency

GenAI systems, which are typically trained on vast amounts of data, introduce significant challenges to meeting transparency obligations under PDP Law. Scholars have noted that, the nature of AI systems requires the large-scale processing of complex datasets, making it difficult for data controllers to comply with transparency requirements (Kesa and Kerikmäe 2020, 7). Indonesia's Personal Data Protection (PDP) Law in Article 16(2) mandates that personal data be processed in a limited, specific, and transparent manner. Further, Article 79 of the draft Government Regulation implementing the PDP Law clarifies that transparency means ensuring that data subjects understand both that their data is being processed and how such processing occurs.

However, the technical complexity of machine learning models limits the extent to which meaningful transparency can be achieved in practice. The UK Information Commissioner's Office (ICO) has emphasized that the complexity of AI systems renders the processing opaque to the very individuals whose personal data is being used ("Big Data, Artificial Intelligence, Machine Learning and Data Protection" 2017, 27). Moreover, the ICO has observed that providing clear explanations of the innerworkings complex analytics in privacy notices is particularly difficult, as such processes are often too technical to be meaningfully explained in lay terms ("Big Data, Artificial Intelligence, Machine Learning and Data Protection" 2017, 63). These difficulties raise compliance challenges for AI developers seeking to balance technical opacity with legal transparency obligations.

To address this transparency gap, scholars and regulators have advocated for the adoption of algorithmic auditing. The ICO, for example, recommends algorithmic audits as a practical method to improve transparency in explaining how machine learning systems generate outputs (Butterworth 2018). Algorithmic auditability requires that AI developers design systems in such a way that independent parties can assess and verify how particular outcomes are produced (Diakopoulos and Friedlerarchive 2016). This approach not only enables greater transparency into the decision-making processes of AI systems but also allows developers to trace results back to specific aspects of the training data or model architecture, thereby facilitating error detection and accountability (Kesa and Kerikmäe 2020, 11).

To fully harness the potential of algorithmic auditing as a tool for transparency and accountability, it is crucial to design audit processes that meaningfully include community participation. While audits are often conducted by well-resourced entities, communities most affected by AI systems frequently lack the institutional access to engage in meaningful oversight. To address this gap, audit capacity should be distributed more equitably through participatory mechanisms. These could include community-led audit ("An Introduction to Community-Led AI Audits" 2025), co-created auditing criteria, and public reporting platforms, all of which help ensure that AI systems are assessed through the perspectives and lived experiences of the communities most affected by them.

Conclusion

Indonesia's current reliance on soft law approaches, such as ethical guidelines, is insufficient to address the human rights risks posed by GenAI. While these frameworks lay important groundwork, they lack legal enforceability. As GenAI adoption expands in the public sector, a transition to binding legal frameworks becomes necessary. Hard law instruments would provide the clarity and accountability needed to align AI development with human rights standards. Furthermore, GenAI presents distinct challenges that do not fit neatly within conventional risk-based regulatory models. GenAI systems create risks that traditional AI classifications may not adequately capture. This reality highlights the urgent need to reassess Indonesia's AI governance landscape and ensure that emerging risks are properly addressed.

A critical legal concern relates to the basis for processing personal data. GenAI developers often train models using publicly available personal data, while Indonesia's PDP Law requires that all data processing be grounded in a lawful basis. This presents significant compliance challenges. Therefore, the Indonesian Data Protection Supervisory Body should issue clear guidance to AI developers on how to establish a legal basis for data processing activities in the context of AI training. Transparency also remains a significant challenge. GenAI systems process vast amounts of data, complicating compliance with transparency obligations under the PDP Law. To bridge this gap, the government should consider adopting algorithmic auditing practices. Integrating algorithmic audits as part of the mandatory Data Protection Impact Assessment (DPIA) under PDP Law would enhance transparency and accountability in AI deployment.

To ensure that regulatory development and implementation are both inclusive and democratically accountable, multistakeholder and community participation must be central to processes such as clarifying legal bases for AI training data and designing algorithmic

audits. Algorithmic auditing capacity should be more equitably distributed by enabling affected communities, who often lack access, to participate meaningfully through community-led audits, co-developed criteria, and public reporting mechanisms grounded in lived experience. Meaningful language inclusivity in AI requires more than technical and legal solutions, it must empower communities to maintain cultural autonomy and data sovereignty, as demonstrated by New Zealand's Papa Reo initiative, where Indigenous groups govern their own language data through ethical, community-led frameworks rooted in Indigenous values. Such models of community-driven AI development and oversight not only preserve local identity but also offer a path toward democratic AI governance rooted in public accountability.

In summary, responsible GenAI use in Indonesia's public sector requires a stronger regulatory framework, inclusive governance mechanisms, and transparency mechanisms to protect human rights while fostering technological progress.

Bibliography

Legislations

Law No. 27 of 2022 concerning Personal Data Protection

The Presidential Regulation No. 132 of 2022 concerning the Architecture of Electronic-Based Government Systems

The Circular Letter of the Ministry of Communication and Informatics No. 9 of 2023 concerning Artificial Intelligence Ethics

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation)

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

Books

Goodfellow, Ian, Yoshua Bengio, and Aaron Courville. 2016. *Deep Learning*. Adaptive Computation and Machine Learning. The MIT Press.

Joinson, Adam N., and Carina B. Paine. 2012. "Self-Disclosure, Privacy and the Internet." In *The Oxford Handbook of Internet Psychology*, edited by Adam N. Joinson, Katelyn Y. A. McKenna, Tom Postmes, and Ulf-Dietrich Reips. Oxford Handbooks Online. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780199561803.013.0016.

Journal Article and Conference Papers

Arazo, Eric, Diego Ortego, Paul Albert, Noel E. O'Connor, and Kevin McGuinness. 2020. "Pseudo-Labeling and Confirmation Bias in Deep Semi-Supervised Learning." arXiv:1908.02983. Preprint, arXiv, June 29. https://doi.org/10.48550/arXiv.1908.02983.

Banh, Leonardo, and Gero Strobel. 2023. "Generative Artificial Intelligence." *Electronic Markets* 33 (1): 63. https://doi.org/10.1007/s12525-023-00680-1.

Brynjolfsson, Erik, and Tom Mitchell. 2017. "What Can Machine Learning Do? Workforce Implications." *Science* 358 (6370): 1530–34. https://doi.org/10.1126/science.aap8062.

Butterby, Kate, and Nancy Lombard. 2025. "Developing a Chatbot to Support Victim-Survivors Who Are Subjected to Domestic Abuse: Considerations and Ethical Dilemmas." *Journal of Gender-Based Violence* 9 (1): 153–61. https://doi.org/10.1332/23986808y2024d000000038.

Butterworth, Michael. 2018. "The ICO and Artificial Intelligence: The Role of Fairness in the GDPR Framework." *Computer Law & Security Review* 34 (2): 257–68. https://doi.org/10.1016/j.clsr.2018.01.004.

Cheong, Ben Chester. 2024. "Transparency and Accountability in AI Systems: Safeguarding Wellbeing in the Age of Algorithmic Decision-Making." *Frontiers in Human Dynamics* 6 (July). https://doi.org/10.3389/fhumd.2024.1421273.

Dang, Hai, Lukas Mecke, Florian Lehmann, Sven Goller, and Daniel Buschek. 2022. "How to Prompt? Opportunities and Challenges of Zero- and Few-Shot Learning for Human-AI Interaction in Creative Applications of Generative Models." Version 1. Preprint, arXiv. https://doi.org/10.48550/ARXIV.2209.01390.

De Freitas, Julian, and I. Glenn Cohen. 2025. "Unregulated Emotional Risks of AI Wellness Apps." *Nature Machine Intelligence* 7 (6): 813–15. https://doi.org/10.1038/s42256-025-01051-5.

Edu, Jide, Cliona Mulligan, Fabio Pierazzi, Jason Polakis, Guillermo Suarez-Tangil, and Jose Such. 2022. "Exploring the Security and Privacy Risks of Chatbots in Messaging Services." *Proceedings of the 22nd ACM Internet Measurement Conference*, October 25, 581–88. https://doi.org/10.1145/3517745.3561433.

Franklin, Gary M. 2024. "Google's New AI Chatbot Produces Fake Health-Related Evidence-Then Self-Corrects." *PLOS Digital Health* 3 (9): e0000619. https://doi.org/10.1371/journal.pdig.0000619.

Helberger, Natali, and Nicholas Diakopoulos. 2023. *ChatGPT and the AI Act*. February 16. https://doi.org/10.14763/2023.1.1682.

Janiesch, Christian, Patrick Zschech, and Kai Heinrich. 2021. "Machine Learning and Deep Learning." *Electronic Markets* 31 (3): 685–95. https://doi.org/10.1007/s12525-021-00475-2.

Jones, Peter-Lucas, Keoni Mahelona, Suzanne Duncan, and Gianna Leoni. 2023. "Kia Tangata Whenua: Artificial Intelligence That Grows from the Land and People." *Ethical Space: International Journal of Communication Ethics* 2023 (2/3). https://doi.org/10.21428/0af3f4c0.9092b177.

Kesa, Aleksandr, and Tanel Kerikmäe. 2020. "Artificial Intelligence and the GDPR: Inevitable Nemeses?" *TalTech Journal of European Studies* 10 (3): 68–90. https://doi.org/10.1515/bjes-2020-0022.

Maeng, Wookjae, and Joonhwan Lee. 2022. "Designing and Evaluating a Chatbot for Survivors of Image-Based Sexual Abuse." *CHI Conference on Human Factors in Computing Systems*, April 29, 1–21. https://doi.org/10.1145/3491102.3517629.

Nguyen, Thanh Thi, Quoc Viet Hung Nguyen, Dung Tien Nguyen, et al. 2022. "Deep Learning for Deepfakes Creation and Detection: A Survey." *Computer Vision and Image Understanding* 223 (October): 103525. https://doi.org/10.1016/j.cviu.2022.103525.

Park, Hyanghee, and Joonhwan Lee. 2020. "Can a Conversational Agent Lower Sexual Violence Victims' Burden of Self-Disclosure?" *Extended Abstracts of the 2020 CHI Conference on Human Factors in Computing Systems*, April 25, 1–8. https://doi.org/10.1145/3334480.3383050.

Pickard, Matthew D., Catherine A. Roster, and Yixing Chen. 2016. "Revealing Sensitive Information in Personal Interviews: Is Self-Disclosure Easier with Humans or Avatars and under What Conditions?" *Computers in Human Behavior* 65 (December): 23–30. https://doi.org/10.1016/j.chb.2016.08.004.

Solove, Daniel J. 2024. "Artificial Intelligence and Privacy." SSRN Scholarly Paper 4713111. Social Science Research Network, February 1. https://doi.org/10.2139/ssrn.4713111.

Tang, Arthur, Neo Tung, Huy Quang Nguyen, et al. 2024. "Health Information for All: Do Large Language Models Bridge or Widen the Digital Divide?" *BMJ*, October 11, e080208. https://doi.org/10.1136/bmj-2024-080208.

Wolff, Josephine, William Lehr, and Christopher S. Yoo. 2023. "Lessons from GDPR for AI Policymaking." SSRN Scholarly Paper 4528698. Social Science Research Network, August 1. https://doi.org/10.2139/ssrn.4528698.

Xiao, Yi, and Shubin Yu. 2025. "Can ChatGPT Replace Humans in Crisis Communication? The Effects of AI-Mediated Crisis Communication on Stakeholder Satisfaction and Responsibility Attribution." *International Journal of Information Management* 80 (February): 102835. https://doi.org/10.1016/j.ijinfomgt.2024.102835.

Official Reports

Appaya, Sharmista, and Jeremy Ng. 2024. *Global Trends in AI Governance: Evolving Country Approaches*. The World Bank. https://documents1.worldbank.org/curated/en/099120224205026271/pdf/P1786161ad76ca0ae1ba3b1558ca4ff88ba.pdf.

Beslay, L, J Galbally, and R Haraksim. 2018. *Automatic Fingerprint Recognition: From Children to Elderly*. JRC Technical Reports EUR 29265 EN. The European Commission.

Big Data, Artificial Intelligence, Machine Learning and Data Protection. 2017. Information Commissioner's Office (ICO). https://ico.org.uk/media2/migrated/2013559/big-data-ai-ml-and-data-protection.pdf.

Guidance for Generative AI in Education and Research. 2023. UNESCO. https://doi.org/10.54675/ewzm9535.

Lin, Pellaeon, Jeffrey Knockel, Irene Poetranto, Stephanie Tran, Justin Lau, and Adam Senft. 2020. *Unmasked II: An Analysis of Indonesia and the Philippines' Government-Launched COVID-19 Apps*. 136. The Citizen Lab. https://tspace.library.utoronto.ca/bitstream/1807/106207/1/Report%23136--unmaskedii.pdf.

"Organizing Data, Sharpening Direction: Reflections on Documentation and Trends in Violence Against Women Cases 2024." 2025. The National Commission on Violence Against Women (Komnas Perempuan), March.

Status Literasi Digital Di Indonesia 2022. 2022. Ministry of Communication and Informatics & Katadata Insight Center. https://web.kominfo.go.id/sites/default/files/ReportSurveiStatusLiterasiDigitalIndonesia2022.pdf.

News Articles

ABC News. 2023. "Victorian Mayor May Sue OpenAI after ChatGPT 'accuses' Him in Bribery Case." April 6. https://www.abc.net.au/news/2023-04-06/hepburn-mayor-flags-legal-action-over-false-chatgpt-claims/102195610.

ABC News. 2024. "When AI Wrongly Described This Man as a Criminal, It Thrust Him into Uncharted Legal Territory." November 3. https://www.abc.net.au/news/2024-11-04/ai-artificial-intelligence-hallucinations-defamation-chatgpt/104518612.

Antara News. 2024a. "Kemenkes-Australia kerja sama terapkan AI tingkatkan layanan kesehatan." Indonesia. July 9. https://www.antaranews.com/berita/4189470/kemenkes-australia-kerja-sama-terapkan-ai-tingkatkan-layanan-kesehatan.

Antara News. 2024b. "Kemenkes wujudkan efisiensi pelayanan kesehatan Indonesia dengan AI." Indonesia. July 10. https://www.antaranews.com/berita/4191693/kemenkes-wujudkan-efisiensi-pelayanan-kesehatan-indonesia-dengan-ai.

Antara News. 2025. "Menkes: Adopsi AI dalam inovasi medis bantu bangun sistem kesehatan." Indonesia. June 3. https://www.antaranews.com/berita/4876353/menkes-adopsi-ai-dalam-inovasi-medis-bantu-bangun-sistem-kesehatan.

Diakopoulos, Nicholas, and Friedlerarchive. 2016. "How to Hold Algorithms Accountable." *MIT Technology Review*. https://www.technologyreview.com/2016/11/17/155957/how-to-hold-algorithms-accountable/.

GovInsider. 2024. "Kemenkes Gandeng Google Cloud Kembangkan Inovasi AI Generatif Sektor Kesehatan." May. https://govinsider.asia/indo-en/article/kemenkes-gandeng-google-cloud-kembangkan-inovasi-ai-generatif-sektor-kesehatan.

Hirdaramani, Yogesh. 2023. "100 Solutions in 100 Days: Singapore Government Jumpstarts Generative AI Capabilities with Sandboxes, Workshops." *GovInsider*. https://govinsider.asia/intl-en/article/100-solutions-in-100-days-singapore-government-jumpstarts-generative-ai-capabilities-with-sandboxes-workshops.

Raj, Aaron. 2024. "Preserving Indigenous Languages with AI." *TechWire Asia*, January 23. https://techwireasia.com/2024/01/preserving-indigenous-languages-with-ai/.

TNGlobal. 2025. "Selangor State Government Launches AI Initiative with Google Cloud." March 24. https://technode.global/2025/03/24/selangor-state-government-launches-ai-initiative-with-google-cloud/.

Verma, Pranshu, and Will Oremus. 2023. "ChatGPT Invented a Sexual Harassment Scandal and Named a Real Law Prof as the Accused." *The Washington Post*, April 5. https://www.washingtonpost.com/technology/2023/04/05/chatgpt-lies/.

VOI. 2025a. "Kemlu RI dan UN Women Luncurkan Chatbot AI SARI yang Humanis, Dilengkapi Berbagai Bahasa Daerah." https://voi.id/berita/477218/kemlu-ri-dan-un-women-luncurkan-chatbot-ai-sari-yang-humanis-dilengkapi-berbagai-bahasa-daerah.

VOI. 2025b. "Kemlu Siapkan Chatbot AI SARI yang Humanis untuk Pelayanan dan Pelindungan WNI." https://voi.id/berita/459935/kemlu-siapkan-chatbot-ai-sari-yang-humanis-untuk-pelayanan-dan-pelindungan-wni.

Other Sources

"An Introduction to Community-Led AI Audits." 2025. *Civic Tech Field Guide*, April 30. https://civictech.guide/audit-intro/.

Hamill-Stewart, Chris. 2024. "The 'Missed Opportunity' with AI's Linguistic Diversity Gap." World Economic Forum, September 27. https://www.weforum.org/stories/2024/09/ai-linguistic-diversity-gap-missed-opportunity/.

Hoang, Ngoc. 2024. "Research Reveals AI's Double-Edged Impact on Health Information." https://www.rmit.edu.vn/news/all-news/2024/oct/research-reveals-ais-double-edged-impact-on-health-information.

IBM Research. 2023. "What Is Generative AI?" https://research.ibm.com/blog/what-is-generative-AI.

Kementerian Kesehatan. 2024. "Teknologi AI Peluang Besar Tingkatkan Layanan Kesehatan." June 3. https://kemkes.go.id/id/teknologi-ai-peluang-besar-tingkatkan-layanan-kesehatan.

"Kementerian Luar Negeri Dan UN Women Memperkuat Pelindungan Perempuan Pekerja Migran Indonesia Melalui Inovasi Chatbot AI SARI." 2025. https://kemlu.go.id/berita/kementerian-luar-negeri-dan-un-women-memperkuat-

pelindungan-perempuan-pekerja-migran-indonesia-melalui-inovasi-chatbot-ai-sari?type=publication.

"Papa Reo." n.d. Accessed July 28, 2025. https://papareo.nz.

"What Are AI Hallucinations?" 2023. September 1. https://www.ibm.com/think/topics/ai-hallucinations.

Yasyi, Dini. 2022. "Indonesia Ranks #2 on Countries with the Most Languages. Here's the Complete List List." Seasia.Co. https://seasia.co/2022/01/25/indonesia-ranks-2-on-countries-with-the-most-languages-here-s-the-complete-list-list.

INFORMATION RESILIENCE & INTEGRITY SYMPOSIUM

Generative AI and Information Resilience in the Asia-Pacific: Actions and Adaptations

Faculty of Social and Political Sciences
Universitas Gadjah Mada

21 August 2025

saferinternetlab.org/iris @saferinetlab @cfds_ugm iris@saferinternetlab.org