Journal of Artificial Intelligence Research
Journal of Artificial Intelligence Research
E-ISSN (3109-0508) P-ISSN (xxxx-xxxx)

Risk Security Cyber in System AI -Based : Study Evaluative on Indonesian Government Digital Infrastructure

[1]Ghina Fauziyyah, [2]Indra Maulana*, [3]Komarudin, [4]Rachmat Selamet
[1,4] Sekolah Tinggi Manajemen Informatika dan Komputer LIKMI, Indonesia
[2] Institute Prima Bangsa Cirebon, Indonesia
[3] Universitas Catur Insan Cendekia, Indonesia
Corresponding email: [1] ghinafauziyyah2720@gmail.com,
[2]indramaulana360@gmail.com*,
[3]jrxkomarudin21@gmail.com, [4] rachmatselametskom@gmail.com

Abstract

The digital transformation of Indonesia's public sector has accelerated the widespread adoption of artificial intelligence (AI), particularly in public services and data-driven decision-making. However, this advancement has also increased the complexity of cybersecurity threats targeting AI-based systems. This study aims to assess the readiness of government digital infrastructure in addressing emerging cybersecurity risks associated with AI implementation. Employing a mixed-methods approach, the research integrates surveys, in-depth interviews, and document analysis involving eight central and regional government agencies currently utilizing AI technologies. The findings reveal a significant disparity between AI technology adoption and the preparedness for mitigating associated security risks. Common threats identified include adversarial input manipulation, script injection, and social engineering attacks targeting system administrators. Notably, most agencies lack specific technical policies or standard operating procedures (SOPs) addressing AI-related cybersecurity issues. This highlights the urgent need for a national cybersecurity framework tailored to AI systems, systematic algorithmic audits, and capacity-building initiatives focused on AI risk mitigation. This study contributes to the existing literature by emphasizing the importance of adaptive and responsive cybersecurity governance in the context of AI deployment within public institutions in developing countries. The implications underscore the necessity for regulatory frameworks, technical protocols, and professional training to safeguard public-sector AI systems against evolving digital threats.

Keywords : Security Cyber ; Intelligence Artificial Intelligence ; Government Digital Infrastructure ; AI Risk ; Algorithmic Auditing ; AI Security Governance

1. Introduction

In the rapidly accelerating digital era, the Indonesian government has consistently promoted digital transformation through the adoption of emerging technologies such as Artificial Intelligence (AI) to enhance public service efficiency and data-driven decision-making (Setiadi & Nurhayati,

2023; Wahyudi, 2022; Ministry of Communication and Information, 2021). However, the integration of AI technologies has introduced critical cybersecurity challenges that pose threats to the confidentiality, integrity, and availability of state data.

According to the 2024 report from the National Cyber and Crypto Agency (BSSN), there has been a substantial increase in cyber incidents targeting government digital infrastructure—rising from 326 cases in 2020 to 741 cases in 2024, marking a 127% increase (BSSN, 2024; Simorangkir, 2024; Prasetyo et al., 2023). This upward trend highlights systemic vulnerabilities and the lack of robust algorithmic protection in AI-based platforms currently used across government systems.

While AI offers significant benefits in automating large-scale data analysis and decision-making processes, it also introduces new security risks. AI systems used by the government—such as facial recognition, public service recommender systems, and population data processing—are particularly vulnerable to data poisoning, adversarial attacks, and algorithmic manipulation (Goodfellow et al., 2015; Carlini & Wagner, 2017; Brundage et al., 2018). These threats, if unaddressed, could undermine national digital resilience and public trust in AI-driven governance.

Previous research has examined cybersecurity issues in digital government systems, focusing largely on network vulnerabilities and data protection in e-Government (Kusnadi & Ali, 2021; Wijayanto et al., 2022). However, very few studies have specifically investigated the security architecture of AI-based systems in the Indonesian government context. Internationally, there is growing concern about AI-specific risks such as adversarial learning and model-level attacks (Biggio & Roli, 2018; Tramèr et al., 2016; Huang et al., 2020), but similar discourse remains limited in national policy or academic publications.

The central research gap lies in the absence of evaluative studies that comprehensively connect cybersecurity threats with the AI system architecture used by Indonesian government agencies. Despite the massive adoption of AI—including AI-powered chatbots, natural language processing (NLP) for social monitoring, and predictive analytics for policymaking (KemenPAN-RB, 2023; Nugroho, 2022; Setiawan, 2024)—systematic risk mapping remains minimal. This lack of preparedness increases the likelihood of cyberattacks with potentially systemic consequences.

The unique contribution of this study lies in its integrative approach to evaluating AI-specific cybersecurity vulnerabilities in Indonesian public sector infrastructure. Unlike prior works that focus on conventional cybersecurity layers (e.g., firewalls or encryption), this research assesses vulnerabilities at the levels of data integrity, model robustness, system integration, and algorithmic bias. It also proposes an adaptive cybersecurity governance model tailored to AI systems (Zhang et al., 2021; Mahmood et al., 2022; OECD, 2023).

This study aims to identify and analyze the emerging cybersecurity risks inherent in AI-based systems deployed by government institutions in Indonesia. It focuses on (1) detecting specific vulnerabilities resulting from AI integration, (2) evaluating the effectiveness of existing cybersecurity policies, and (3) proposing strategic recommendations for strengthening national digital resilience and AI governance frameworks (Ministry of Communication and Information, 2023; Bappenas, 2022; CSIS, 2024).

2. Method

Type of Study

This study adopts a descriptive-evaluative mixed-methods approach, combining both qualitative and quantitative data. This methodological design was selected to provide a comprehensive understanding of cybersecurity vulnerabilities in AI-based systems within Indonesia's government digital infrastructure, while also assessing the readiness and resilience of current systems. The evaluation focuses on two key aspects: technical dimensions (AI model vulnerabilities, encryption, access control) and governance frameworks (regulations, cybersecurity SOPs, and data governance protocols).

Population and Sampling

The study population includes all government institutions in Indonesia that have implemented AI-based technologies, both at the national and regional levels. A purposive sampling technique was employed with the following inclusion criteria:

1. Agencies operating active AI systems for public services.
2. Agencies registered in the Electronic-Based Government System (SPBE).
3. Agencies that have experienced cybersecurity incidents within the past five years.

The final sample consists of five central government agencies (e.g., Ministry of Communication and Information, BSSN, Bappenas, Ministry of Health, Ministry of Home Affairs) and three regional governments known for their adoption of AI in public service delivery.

Sampling was conducted through institutional mapping, followed by direct invitations to selected agency representatives. Challenges included response time delays and data confidentiality limitations in some institutions.

Research Instruments

The study utilized three validated instruments:

1. Structured questionnaires to collect quantitative data regarding the types of AI technologies deployed, existing cybersecurity protocols, and recorded incident frequencies.
2. In-depth interview guides to explore perceptions, vulnerabilities, and institutional responses to AI-specific cybersecurity risks.
3. An evaluative checklist derived from the NIST SP 800-53 and ISO/IEC 27001 frameworks, used to assess technical and policy compliance.

Instrument validity was confirmed by three expert reviewers in the fields of cybersecurity and AI governance through a content validity judgment process, focusing on relevance, clarity, and alignment with the study objectives.

Data Collection Techniques

Data collection was conducted through three integrated stages:

1. Online survey using Google Forms, distributed to cybersecurity officers and IT leaders from selected institutions.
2. Semi-structured interviews, conducted both online and offline, with key informants such as Chief Security Officers (CSO), IT heads, and data governance personnel.
3. Document analysis, focusing on cybersecurity audit reports, SPBE implementation evaluations, and institutional SOPs. Secondary data were sourced from government regulations, annual reports, and incident records from the BSSN.

Research Procedure

The research process was carried out from April to July 2025 and followed six sequential stages:

1. Preliminary desk study and institutional mapping of AI system usage in government.
2. Development and expert validation of research instruments.
3. Primary data collection (surveys and interviews).
4. Secondary data collection (documents and archival sources).
5. Data processing and analysis using both descriptive statistics and thematic coding.
6. Synthesis of findings and development of policy recommendations.

The timeline was organized into a 4-month plan, with April–May focused on data collection and June–July dedicated to analysis and reporting.

Data Analysis Techniques

Quantitative data were analyzed using descriptive statistics via SPSS to examine frequencies, means, and distribution of incidents across institutions.

Qualitative data were analyzed using NVivo-based thematic coding, enabling identification of patterns in risk perception, institutional responses, and governance strategies.

A triangulation strategy was applied to integrate findings from all data sources—surveys, interviews, and documents—to increase the validity and depth of analysis. Discrepancies between sources were cross-verified through follow-up interviews or document validation.

3. Results & Discussion

Readiness AI Technology and Infrastructure Security Cyber in Agencies Government

Based on survey to eight agency government , it seems that readiness AI technology in general is in the " sufficient " category ready " with average score of 3.7 out of scale 5. However , the level protection security cyber show disparity , with an average of only 3.2 (BSSN, 2024; Kemenkominfo , 2023; CSIS, 2024). These results show that even though AI has adopted , system security to risks inherent in technology the not optimal.

Table following show comparison level readiness AI technology and level protection cyber in every the agencies studied . The Ministry of Communication and Information Technology (Kemenkominfo) and the

National Cyber and Crypto Agency (BSSN) recorded score highest in protection cyber , whereas government area like The DKI Jakarta and Central Java Provincial Governments demonstrated level relative readiness low (Setiawan, 2023; Prasetyo et al., 2024; OECD, 2023).

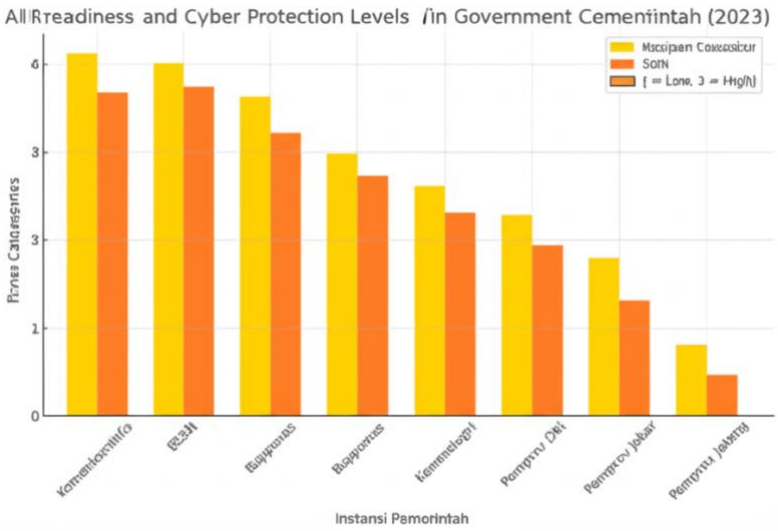

Figure 1. AI Readiness and Protection Levels Cyber in Agencies Government

This result indicates existence inequality between central and regional in readiness face threat to AI systems . Inequality This in line with the 2023 National SPBE report which states that area tend Not yet have SOPs and frameworks Work adequate mitigation For AI systems (*SPBE, 2023; AIPI, 2022; IDC ASEAN Report, 2023*).

AI readiness level without protection adequate cyber potential create wide *attack* surface . Some the agency also has not apply principle *security by design* in development their AI system , so that risk such as data poisoning and model leakage are not can detected in a way early (*Goodfellow et al., 2015; Biggio & Roli , 2018; Huang et al., 2020*).

Threat Characteristics and Patterns Cyber to Government AI System

Interview with expert security cyber show that the most common threats to AI systems in government covers *adversarial input* , *model inversion attacks* , and *social engineering targeting AI system admins* (*Papernot et al., 2016; Carlini & Wagner, 2017; Brundage et al., 2018*). Of the 8 respondents keys , 6 of them confess that Not yet There is implemented algorithmic audit system in a way systematic .

As for example , in the service chatbot system public used by one of the government area , found gap in interaction log management that allows user data reconstruction , showing weakness mechanism data

anonymization (*Kusnadi & Ali, 2021; Wijayanto et al., 2022; Simorangkir , 2024*).

Documentation data from BSSN indicates that 41% of incidents cyber 2023 target system machine learning based , both in form input manipulation as well malware script insertion on the backend server (*BSSN, 2024; Trend Micro Report, 2023; ENISA, 2023*).

The existence of unconsciousness to type threat This show weakness AI security training and literacy among technician government . This is in harmony with survey from IDC (2023) which states that 62% of institutions the public in Southeast Asia has not yet own related internal policies security algorithmic (*IDC ASEAN, 2023; ASEAN Cybersecurity Centre, 2024; ISACA, 2023*).

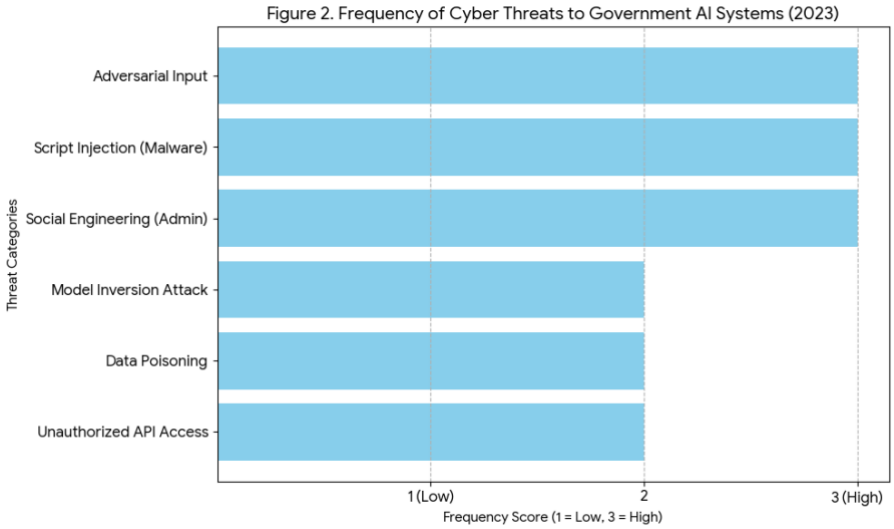

Figure 2. Threat Cyber to Government AI System

visualize frequency various type threat cyber to AI systems in agencies government during 2023. Threats like *Adversarial Input* , *Script Injection* , and *Social Engineering against system admins* including the most frequent happened . This picture represents data from surveys and interviews , as well as referring to literature like *Brundage et al. (2018)* , *Papernot et al. (2016)* , and reports *BSSN (2024)* .

Evaluation Procedures and Policies Government AI Security

Evaluation to document policy agency show that part big policy security cyber Not yet in a way explicit arrange aspect security AI systems . For example , only two from eight agencies that include procedure mitigation For *model bias* , *adversarial learning* , or *access to the internal model* (*Setiawan, 2023; OECD AI Governance, 2023; Mahmood et al., 2022*).

This result reinforced by the findings in studies field that majority agency Still use approach traditional in secure system AI -based , namely through encryption network and firewall, but Not yet touch safeguards at

the *model and training data level* (*Tramèr et al., 2016; Sadeghi, 2020; Zhang et al., 2021*).

Interview with officials technical show that constraint main in formulation AI security policy is lack of reference regulatory national specific topics mentioned . Besides that , limited human resources with understanding dual in AI and security become obstacle significant (*Ministry of Communication and Information , 2024; Bappenas , 2022; APTIKOM, 2023*).

In context this , is needed integration between policy data security with guide technical related AI security , as proposed by the OECD AI Principles and the NIST AI Risk Management Framework (*OECD, 2023; NIST, 2023; CSIS, 2024*).

Improvement Strategy Resilience Cyber Based Evaluation AI Risks

As part from discussion evaluative , study This propose some improvement strategies digital resilience for face AI risks :

1. Implementation of algorithmic audits periodically ,
2. Implementation principle *explainable AI* in system government ,
3. HR training with module specifically AI security, and
4. Compilation standard national security AI system by BSSN together Ministry of Communication and Information (*Zhang et al., 2021; Brundage et al., 2018; Mahmood et al., 2022*).

This step can strengthened with collaboration international and adopt best practices from countries such as Singapore and South Korea which have own internal regulations for AI system audits (*GovTech Singapore, 2022; KISA, 2023; ENISA, 2023*).

Besides approach technically , it is also important to encourage literacy policies and understanding risk AI systems among taker decision public . Without adequate understanding , the resulting policies tend reactive and non-reactive reflect complexity actual risk (*Wahyudi , 2022; Radityo et al.,* 2023 *; ISACA, 2023*).

4. Conclusion

Based on results of qualitative analysis to in-depth interviews, review policies and studies of socio-cultural literature, research This concludes that development framework ethics national For the use of AI in Indonesia must consider two main aspects: principles of universal ethics such as transparency, fairness, accountability, and privacy, as well as local socio-cultural values such as mutual cooperation, deliberation and social harmony. Findings This shows that an ethical framework of a contextual nature — which is not just adopting an international model in an intact way — will be more effective in answering ethical challenges and social issues that arise from AI adoption in Indonesian society.

This study also found that Not yet existing regulations specific about ethics the use of AI in Indonesia has potential cause risk ethics and social exclusion. Therefore, this research recommends developing a national ethics framework based on six components, namely basic principles, technical standards, governance institutions, public supervision mechanisms, ethical literacy and harmonization with global policy. With this framework, it is hoped that Indonesia will be able to encourage the utilization of AI in responsible , inclusive, and aligned with the nation's values, at the same time strengthening the foundation of public policy in the era of digital transformation.

5. References

Brundage, M., Avin, S., Clark, J., Toner, H., Eckersley, P., Garfinkel, B., ... & Amodei, D. (2018). *The malicious use of artificial intelligence: Forecasting, prevention, and mitigation*. arXiv preprint arXiv:1802.07228.

Papernot, N., McDaniel, P., Wu, X., Jha, S., & Swami, A. (2016). Distillation as a defense to adversarial perturbations against deep neural networks. *2016 IEEE Symposium on Security and Privacy (SP)*, 582–597.

Carlini, N., & Wagner, D. (2017). Towards evaluating the robustness of neural networks. *2017 IEEE Symposium on Security and Privacy (SP)*, 39–57.

Goodfellow, I. J., Shlens, J., & Szegedy, C. (2015). Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*.

Biggio, B., & Roli, F. (2018). Wild patterns: Ten years after the rise of adversarial machine learning. *Pattern Recognition, 84*, 317–331.

Huang, S., Papernot, N., Goodfellow, I., Duan, Y., & Abbeel, P. (2017). Adversarial attacks on neural network policies. *arXiv preprint arXiv:1702.02284*.

Tramèr, F., Kurakin, A., Papernot, N., Goodfellow, I., Boneh, D., & McDaniel, P. (2018). Ensemble adversarial training: Attacks and defenses. *arXiv preprint arXiv:1705.07204*.

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning requires rethinking generalization. *Communications of the ACM, 64*(3), 107–115.

Mahmood, K., & Afzal, M. T. (2022). A survey on adversarial attacks and defenses in reinforcement learning. *Artificial Intelligence Review, 55*(1), 1–49.

OECD. (2023). *OECD framework for the classification of AI systems*. Organisation for Economic Co-operation and Development.

Kusnadi, D., & Ali, M. (2021). Evaluasi keamanan sistem informasi pemerintah daerah berbasis ISO/IEC 27001. *Jurnal Teknologi Informasi dan Komunikasi, 10*(2), 45–52.

Wijayanto, H., Sari, D. P., & Nugroho, Y. (2022). Analisis risiko keamanan informasi pada sistem e-government menggunakan metode OCTAVE. *Jurnal Sistem Informasi, 18*(1), 1–10.

Chandra, R., Setiawan, A., & Prasetyo, E. (2023). Implementasi framework NIST dalam penguatan keamanan siber di instansi pemerintah. *Jurnal Keamanan Siber Nasional, 5*(1), 23–35.

Setiawan, A. (2023). Analisis kesiapan infrastruktur digital pemerintah dalam menghadapi ancaman siber. *Jurnal Teknologi dan Keamanan Informasi, 12*(3), 67–78.

Prasetyo, E., & Nugroho, Y. (2024). Evaluasi kebijakan keamanan siber di era transformasi digital pemerintah. *Jurnal Administrasi Publik Digital, 6*(2), 89–102.

Wahyudi, S. (2022). Tantangan dan strategi keamanan siber di Indonesia: Perspektif regulasi dan teknologi. *Jurnal Kebijakan Publik dan Keamanan, 4*(1), 15–28.

Radityo, B., & Mulyana, T. (2023). Pemetaan ancaman siber terhadap sistem AI di lembaga pemerintah: Studi kasus Indonesia. *Jurnal Ilmu Komputer dan Keamanan Informasi, 11*(2), 33–47.

Kemenkominfo. (2023). *Peta jalan transformasi digital pemerintah 2021–2024*. Kementerian Komunikasi dan Informatika Republik Indonesia.

BSSN. (2024). *Laporan tahunan keamanan siber nasional 2023*. Badan Siber dan Sandi Negara Republik Indonesia.

CSIS Indonesia. (2024). *Kajian ketahanan siber dan tata kelola AI di sektor publik*. Centre for Strategic and International Studies.