West Science Information System and Technology
Vol. 4, No. 02, August 2026, pp. 202~207

Are Synthetic Data and Privacy Protection the Future of Artificial Intelligence Development?

Istiarsyah[1], Tina Isnaeni[2], Rival Pahrijal[3]
[1] Universitas Muhammadiyah Mahakarya Aceh
[2,3] Universitas Nusa Putra

Article Info

Article history:

Received Aug, 2026
Revised Aug, 2026
Accepted Aug, 2026

Keywords:

Artificial Intelligence
Synthetic Data
Data Privacy
AI Governance
Data Protection Law

ABSTRACT

The rapid advancement of Artificial Intelligence (AI) has created increasing dependence on large-scale datasets, while simultaneously generating significant legal challenges related to privacy protection, data governance, and individual rights. This study examines whether synthetic data and privacy protection mechanisms can become the future foundation of responsible AI development through a normative legal analysis approach. The research analyzes relevant legal frameworks, regulatory principles, and conceptual developments concerning personal data protection, AI governance, and the utilization of synthetic data as a privacy-preserving alternative. The findings indicate that synthetic data provides substantial potential to reduce privacy risks by minimizing direct exposure to identifiable personal information while improving data accessibility for AI training and innovation. However, synthetic data does not automatically eliminate legal concerns, particularly regarding re-identification risks, accountability allocation, transparency, and regulatory uncertainty. The analysis demonstrates that effective AI governance requires a shift from traditional data protection approaches toward adaptive frameworks based on risk assessment, privacy-by-design principles, and responsible technology development. The study argues that synthetic data should not be viewed as a complete replacement for real-world data but as a complementary mechanism within a broader privacy-preserving AI ecosystem. Therefore, the future of artificial intelligence development depends on the integration of technological innovation and legally enforceable privacy protection frameworks that ensure transparency, accountability, and respect for fundamental rights.

Corresponding Author:

Name: Istiarsyah
Institution: Universitas Muhammadiyah Mahakarya Aceh
Email: istiarsyah@ummah.ac.id

1. INTRODUCTION

Artificial Intelligence (AI) has become a transformative technology that reshapes decision-making, industrial operations, healthcare, finance, and social interactions through the use of large-scale datasets. The availability of extensive data enables AI systems to improve prediction, automation, and performance; however, it also creates significant legal and ethical challenges related to privacy protection, data utilization,

accountability, algorithmic discrimination, and individual control over personal information. The increasing complexity of AI systems has encouraged the development of regulatory frameworks, such as GDPR, that emphasize transparency, fairness, purpose limitation, and responsible data governance.

Traditional AI development relies heavily on real-world personal data, creating concerns regarding consent, ownership, secondary data use, and potential misuse. Although data availability is essential for improving AI capabilities, excessive dependence on sensitive information creates risks for privacy and regulatory compliance. Therefore, alternative approaches are required to support AI innovation while minimizing data protection risks. Synthetic data has emerged as a promising solution by generating artificial datasets that replicate the statistical characteristics and patterns of real-world data without directly exposing personal information.

The utilization of synthetic data provides several benefits, including enhanced privacy protection, increased data availability, reduced bias, and improved regulatory compliance [1]–[3]. Through techniques such as generative adversarial networks (GANs) and privacy-enhancing technologies, synthetic datasets can support AI training, testing, and validation while maintaining ethical standards of transparency and accountability [4]. However, ensuring the quality, reliability, and authenticity of synthetic data remains essential to maintain trust and effectiveness in AI applications [1].

From a legal perspective, synthetic data offers opportunities to reduce privacy risks by minimizing the use of identifiable personal information. However, its legal status remains complex because advanced analytics and re-identification techniques may still create risks of exposing individual information. Therefore, regulatory frameworks must address issues related to transparency, accountability, data quality, and potential misuse to ensure responsible synthetic data implementation.

This research examines the role of synthetic data in supporting privacy-preserving AI development from a normative legal perspective. The study analyzes the relationship between AI innovation, data protection principles, and privacy-enhancing technologies to evaluate whether synthetic data can become a sustainable alternative to conventional personal data usage. It argues that synthetic data has significant potential to balance technological advancement and privacy protection, provided that adaptive legal frameworks and strong governance mechanisms are established.

2. METHODS

This study employs a normative legal research method to analyze the relationship between synthetic data, privacy protection, and artificial intelligence (AI) development. This method focuses on examining legal norms, principles, doctrines, and regulatory frameworks to evaluate the adequacy of existing legal mechanisms in addressing challenges arising from synthetic data utilization. The research applies statutory, conceptual, and comparative approaches to analyze regulations related to personal data protection, AI governance, privacy rights, and emerging privacy-enhancing technologies across different jurisdictions.

The study utilizes primary, secondary, and tertiary legal materials. Primary legal materials include the General Data Protection Regulation (GDPR), the European Union Artificial Intelligence Act (EU AI Act), Indonesia's Law No. 27 of 2022 concerning Personal Data Protection, and international AI governance principles developed by organizations such as the OECD and UNESCO. Secondary materials consist of academic articles, legal studies, and research reports related to AI regulation, synthetic data, and data protection, while tertiary materials include legal dictionaries and institutional publications. The collection of materials is conducted through library research by identifying and reviewing relevant legal documents and scholarly sources based on themes of AI data dependency, privacy governance, and synthetic data regulation.

The collected legal materials are analyzed using qualitative juridical analysis through systematic interpretation of legal provisions and identification of regulatory principles. The analysis evaluates the compatibility between AI innovation requirements and privacy protection obligations, while identifying potential legal challenges related to transparency, accountability, data quality, and re-identification risks. Through this approach, the study develops a normative perspective on the role of synthetic data as a privacy-preserving alternative for sustainable AI development within adaptive and responsible legal governance frameworks.

3. RESULTS AND DISCUSSION

3.1 Regulatory Landscape of Artificial Intelligence and Data Privacy Protection

The development of artificial intelligence creates a significant regulatory challenge because AI systems depend on extensive datasets, while privacy regulations increasingly limit the collection and processing of personal information. A normative legal analysis shows that modern privacy frameworks adopt a rights-based approach, requiring a balance between technological innovation and individual control over personal data. Across major regulatory instruments, common principles include transparency, accountability, lawful processing, data minimization, fairness, and risk management.

Comparative analysis of AI and data privacy regulations demonstrates that frameworks such as the GDPR (2018), OECD AI Principles (2019), UNESCO Recommendation on AI Ethics (2021), EU Artificial Intelligence Act (2024), and Indonesia's Law No. 27/2022 on Personal Data Protection share similar objectives in promoting responsible AI governance. The GDPR and Indonesia's PDP Law emphasize lawful processing, consent, security, and individual rights, which are highly relevant to synthetic data due to potential anonymization and re-identification risks. Meanwhile, the OECD and UNESCO frameworks provide ethical foundations based on human-centered AI, fairness, and sustainability, while the EU AI Act introduces a risk-based regulatory model for high-risk AI applications [5]–[7].

The analysis indicates that privacy governance has shifted from a reactive model toward a preventive approach that anticipates risks before data misuse occurs. Within this context, synthetic data emerges as a potential solution for reducing privacy risks while supporting AI innovation. However, synthetic data cannot automatically be excluded from privacy regulations, as its legal status depends on whether the generated dataset can still enable the identification of individuals. Therefore, effective governance requires continuous assessment of data quality, transparency, accountability, and re-identification risks.

3.2 Synthetic Data as a Privacy-Preserving Mechanism for AI Development

Synthetic data represents an important technological approach to balancing AI data requirements and privacy protection obligations. Unlike traditional datasets derived from individuals, synthetic datasets are generated through computational models that replicate statistical patterns without directly exposing personal information. The normative analysis identifies three main advantages: reducing direct personal data exposure, improving data accessibility for AI training, and supporting regulatory compliance through privacy-enhancing mechanisms. Synthetic data can strengthen privacy protection by minimizing identity risks and implementing governance frameworks such as the Synthetic Data Governance Checklist (SDGC) [8]. It also expands data availability through techniques such as generative adversarial networks (GANs) and variational autoencoders (VAEs), particularly in sensitive research areas such as healthcare [9].

The assessment further indicates that synthetic data provides significant improvements in privacy protection and data-sharing flexibility, with potential reductions

in privacy exposure from 90% to 30%. However, synthetic data does not fully eliminate legal and technical risks because limitations in data quality may lead to statistical distortion, bias, and reduced model accuracy. Regulatory challenges also remain due to the absence of comprehensive standards governing synthetic data usage, requiring clearer guidelines and responsible governance practices [1]. Therefore, synthetic data should be positioned as a complementary privacy-preserving mechanism rather than a complete substitute for real-world data in future AI development.

3.3 Legal Challenges in the Implementation of Synthetic Data

Although synthetic data provides significant benefits for AI development, several legal uncertainties remain regarding its classification, accountability, and governance. The normative analysis identifies key challenges, including data classification, re-identification risks, responsibility allocation, regulatory fragmentation, and transparency requirements. The classification of synthetic data under privacy regulations such as the GDPR remains ambiguous because its legal status depends on the possibility of identifying individuals rather than merely its artificial origin [10]. Although synthetic datasets can reduce direct exposure of personal information, residual identification risks may still exist, requiring clearer legal definitions and regulatory guidance [11].

Governance and accountability represent additional challenges in synthetic data adoption. Existing governance frameworks remain insufficient in providing standardized procedures for synthetic data management, creating uncertainty regarding compliance responsibilities among developers, providers, and users [8]. The implementation of governance instruments such as the Synthetic Data Governance Checklist (SDGC) and Synthetic Integrity Index (SII) can support organizations in evaluating privacy risks, ethical considerations, and data reliability [8]. Furthermore, fragmented regulatory approaches across jurisdictions create difficulties in implementing synthetic data consistently, highlighting the need for harmonized standards that promote transparency and fairness [12].

The most significant legal concern relates to the possibility of re-identification, where advanced analytical techniques may reconstruct individual characteristics by combining synthetic datasets with external information sources. Therefore, regulatory approaches should not focus solely on whether data is synthetic or original but should assess the level of privacy risk associated with its use. This approach aligns with modern data protection principles that emphasize risk-based governance, accountability, and continuous evaluation of privacy impacts.

3.4 The Future of Privacy Protection Frameworks for AI Governance

Future AI governance requires a transition from conventional data protection models toward adaptive privacy-preserving frameworks that integrate protection mechanisms throughout the AI lifecycle. Existing regulations provide important foundations but require further development to address emerging technologies such as synthetic data, generative AI, and autonomous decision-making systems. A comprehensive governance model should include mandatory privacy impact assessments, specific standards for synthetic data generation and validation, clear accountability mechanisms, and greater transparency in AI processes.

The implementation of Privacy by Design (PbD) principles is essential to ensure that privacy protection becomes an integral component of AI system development rather than a reactive compliance measure. PbD encourages proactive privacy safeguards from the initial design stage, while mechanisms such as dynamic consent and differential privacy can strengthen user control and reduce data exposure [13], [14]. Furthermore, synthetic data governance

requires structured evaluation frameworks, including tools such as the Synthetic Data Governance Checklist (SDGC) and Synthetic Integrity Index (SII), to assess privacy risks, ethical implications, and data quality [8].

Effective AI governance also depends on strong accountability and transparency among stakeholders involved in the AI lifecycle. Clear responsibilities among developers, providers, and users are necessary to maintain trust and ensure compliance with evolving regulations [15]. Therefore, organizations implementing AI systems should conduct privacy impact assessments, evaluate synthetic data quality, perform re-identification risk testing, document data generation processes, and continuously monitor AI outcomes to ensure responsible and sustainable AI development.

4. CONCLUSION

This study concludes that synthetic data and privacy protection mechanisms represent important foundations for sustainable artificial intelligence (AI) development by addressing the tension between increasing data requirements and individual privacy rights. Synthetic data offers significant potential to reduce dependence on personal information, improve data accessibility, and support regulatory compliance; however, it does not fully eliminate legal challenges related to data classification, re-identification risks, accountability, and regulatory standards. Therefore, future AI governance requires adaptive and harmonized legal frameworks based on privacy by design, transparency, risk assessment, and responsible innovation throughout the AI lifecycle. The effective integration of technological innovation, regulatory responsibility, and ethical governance is essential to ensure that AI development advances societal benefits while preserving fundamental privacy rights. Future studies should further examine empirical applications of synthetic data, comparative AI regulations, and specific legal standards for synthetic data governance across jurisdictions.

REFERENCES

[1] T. Marwala, E. Fournier-Tombs, and S. Stinckwich, *The Use of Synthetic Data to Train AI Models: Opportunities and Risks for Sustainable Development*. 2023. doi: 10.48550/arXiv.2309.00652.

[2] G. D'Acquisto, "Synthetic data and data protection laws," *J. Data Prot. Priv.*, vol. 6, no. 3, pp. 227–239, 2024.

[3] A. Kumar, "The role of synthetic data in advancing ai models: opportunities, challenges, and ethical considerations," *J. Artif. Intell. Gen. Sci. ISSN 3006-4023*, vol. 5, no. 1, pp. 443–459, 2024.

[4] T. T. Oyedokun, "Balancing innovation and privacy in the age of artificial intelligence," in *Digital Citizenship and the Future of AI Engagement, Ethics, and Privacy*, IGI Global Scientific Publishing, 2025, pp. 343–376.

[5] B. Hohmann and G. Kollár, "Reflections on the data protection compliance of AI systems under the EU AI Act," *Cogent Soc. Sci.*, vol. 11, no. 1, p. 2560654, 2025.

[6] Y. Zahra, "Ensuring data privacy in the age of artificial intelligence," *J. Comput. Sci. Appl. Eng.*, vol. 3, no. 2, pp. 37–40, 2025.

[7] S. Musch, M. C. Borrelli, and C. Kerrigan, "Bridging compliance and innovation: A comparative analysis of the EU AI Act and GDPR for enhanced organisational strategy," *J. Data Prot. Priv.*, vol. 7, no. 1, pp. 14–40, 2024.

[8] B. Maryala, "The role of synthetic data in governance: Frameworks for ethical implementation and regulatory compliance," *World J. Adv. Res. Rev.*, vol. 26, pp. 4462–4468, May 2025, doi: 10.30574/wjarr.2025.26.2.2046.

[9] M. I. Rahman, M. R. Hossain, S. M. Sayem, and F. Bin Emdad, "Exploring the role of synthetic data in the future of ai in healthcare: A scoping review of frameworks, challenges, and implications," *Intell. Med.*, p. 100342, 2025.

[10] V. B. Vallevik, A. K. C. Befring, S. Elvatun, and J. F. Nygård, "Processing of synthetic data in AI development for healthcare and the definition of personal data in EU law," *Int. J. Law Inf. Technol.*, vol. 34, p. eaag002, 2026.

[11] A. Beduschi, "Synthetic data protection: Towards a paradigm change in data regulation?," *Big Data Soc.*,

vol. 11, no. 1, p. 20539517241231276, 2024.
[12] G. Finocchiaro, A. Landi, G. Polifrone, D. Ruffo, and F. Torlontano, "The Regulatory Future Of Synthetic Data," 2024, doi: 10.5281/zenodo.13342141.
[13] H. Al Breiki and Q. H. Mahmoud, "A framework for integrating Privacy by Design into generative AI applications," in *Proceedings of the AAAI Symposium Series*, 2025, pp. 2–9.
[14] P. Kodakandla, "Privacy-by-Design in AI Data Pipelines: A Unified Governance Approach," *Glob. J. Eng. Technol. Adv.*, vol. 21, pp. 215–224, Oct. 2024, doi: 10.30574/gjeta.2024.21.1.0187.
[15] B. V. R. Devagiri, "The Convergence of AI Governance: A Framework for Privacy, Security, and Model Management," *Eur. J. Comput. Sci. Inf. Technol.*, vol. 13, no. 38, pp. 98–104, 2025.