

Copyright in the Era of Generative AI: Mapping Legal Gaps in Indonesian Copyright Protection

Avelyn Pingkan Komuna^{1*}

¹ Faculty of Law, Social and Political Sciences, Universitas Terbuka, Indonesia

*Correspondence avelynkomuna@ecampus.ut.ac.id

ARTICLE HISTORY

Received: 26.10.2025

Accepted: 05.12.2025

Published: 29.12.2025

ARTICLE LICENSE

Copyright © 2025 The Author(s): This is an open-access article distributed under the terms of the Creative Commons Attribution ShareAlike 4.0 International (CC BY-SA 4.0)

ABSTRACT

The rapid advancement of artificial intelligence (AI), particularly generative and data-driven models, has introduced unprecedented challenges to traditional copyright frameworks. This study examines how the technical workflow of AI from data input, text and data mining (TDM), and feature extraction to model building, assessment, and output generation intersects with legally protected creative expressions. By mapping these technical stages against Indonesia's Copyright Law (Law No. 28/2014), the research identifies five critical areas of concern: fair use in TDM, non-literal copying, AI authorship, liability for AI-generated infringements, and dataset transparency. Using a doctrinal legal research approach, the analysis reveals significant normative gaps within the current copyright regime, which remains anthropocentric and ill-equipped to regulate autonomous or algorithmic forms of creativity. The legal gap map developed in this study demonstrates that Indonesian copyright law lacks explicit provisions governing machine-based data processing, algorithmic replication of styles, ownership of AI-generated outputs, allocation of responsibility among AI ecosystem actors, and mandatory disclosure of training datasets. These regulatory deficiencies create substantial uncertainty for creators, AI developers, platforms, and users, ultimately weakening the enforcement of copyright in digital contexts. The study concludes that comprehensive legal reform—potentially through updated exceptions, sui generis protection, and clearer governance of AI datasets—is essential for ensuring that Indonesia's copyright system remains effective, balanced, and future-ready in the age of artificial intelligence.

Keywords: Artificial Intelligence; Copyright Protection; Legal Gap Analysis; Indonesia; Intellectual Property.

1. Introduction

The world has never been the same since artificial intelligence (AI) emerged. In the early 2010s, artificial intelligence started to become a major phenomenon. AlphaGo, a computer developed by DeepMind, won the title of champion in the game of Go by defeating a highly skilled human player. The 2020s saw the rise of OpenAI, a software startup capable of developing generative chatbots that can write, respond to inquiries, and even create poems and articles that resemble human work, Lin, H. Y. (2023). AI systems that could produce paintings for the first time later amazed the world even more; one of these paintings sold for \$432,000, Sun, Y., Yang, C. H., Lyu, Y., & Lin, R. (2022). In 2025, Albania introduced Diella, an AI entity appointed as a cabinet minister responsible for public procurement. The existence of artificial intelligence (AI) is not actually a new concept, as it has been studied and developed for several decades. The term AI refers to the ability of computer systems to interpret the data provided to them, learn from it, and then, based on that data, produce something according to given instructions; in other words, artificial intelligence mimics the capabilities of the human brain.

Its rapid development has transformed AI, once considered merely the stuff of science fiction, into a technology widely regarded as the most influential in shaping human life and even creating new landscapes across various domains. (Pävåloaia & Necula, 2023). In the field of law, AI has exposed the limitations of both international and national legal frameworks worldwide, particularly in the domain of intellectual property. The complexity of AI, which not only functions as a tool for humans but can also imitate human cognitive and intellectual capabilities,

has called into question traditional copyright theories related to creativity, originality, and ownership that have, until now, been considered exclusively human (Hoshier & Kiran, 2024).

The transformation brought by AI technology appears to have outpaced the evolution of legal frameworks. This is evident from the legal gaps concerning intellectual property, particularly AI-generated copyright, worldwide. International agreements that have long served as references for countries in establishing copyright protection standards, such as the Berne Convention and the Trade-Related Aspects of Intellectual Property Rights (TRIPS) Agreement, do not provide specific provisions regarding works created by AI, leaving each country to interpret AI-generated works differently. Continental European countries, for example, regulate copyright attribution for computer-generated works under the Copyright, Designs and Patents Act (CDPA). This is particularly evident in the jurisdiction of English courts, which recognize the work of individuals who develop AI programs (Lee, 2022). In contrast, Anglo-Saxon countries such as the United States maintain their position that copyright protection cannot be granted to works not created by humans as the primary creator. (Rabago, 2024)

The international ambiguity surrounding copyright in the era of AI has also affected Indonesia. As a member of the World Trade Organization, Indonesia has ratified the Berne Convention and the TRIPS Agreement into Law No. 28 of 2014 on Copyright. Similar to the international agreements it has ratified, Indonesian Copyright Law does not yet recognize AI, even though issues surrounding AI copyright protection are highly complex. These issues extend beyond the legal subject as the rights holder and include related matters that currently lack legal regulation in Indonesia. The copyright conflicts arising from the presence of AI are particularly significant, especially regarding the use of human-created works for AI training and determining the extent to which the use of AI as a tool falls under fair use.

This study aims to identify copyright issues related to AI-generated works and to examine the gaps in positive law in Indonesia. The findings of the research are expected to result in a legal gap map within Indonesian Copyright Law, which can serve as a foundation for developing comprehensive copyright legal protection in Indonesia.

2. Method Research

This study employs a normative legal approach with a statutory method, examining issues through applicable laws, regulations, and legal theories. Data were collected from primary and secondary sources, including statutes, journals, books, and relevant documents, and analyzed using a qualitative descriptive method to identify and interpret legal gaps.

3. Results and Discussion

3.1 Identify the Issue of Copyright in the AI

To identify and analyze copyright issues arising from AI, it is crucial to first understand the technical workflow through which AI systems operate. In essence, contemporary AI, particularly those employing machine learning and deep learning functions through a structured and sequential pipeline. This pipeline generally consists of four major stages: Data Input, Feature Extraction, Model Building, and Model Assessment, Understanding each of these stages is important for mapping the specific points at which copyright implications may emerge. In simplified form, the workflow of AI can be illustrated in the following figure, inspired by the work of Janiesch, Zschech, and Heinrich (2021).



Picture 1 : The Simple General Workflow of AI

The first stage, Data Input, constitutes the foundation of the entire AI process. AI systems rely heavily on large scale datasets often comprising text, images, audio, video, and other multimodal content sourced from websites, digital platforms, applications, and social media. These data collections frequently involve unstructured numerical representations that must be standardized before further processing.

The second stage, Feature Extraction, refers to the algorithmic process through which raw data are transformed into meaningful patterns and representations. Depending on the technique used, this extraction may occur manually (in traditional machine learning) or automatically through layered neural architectures (in deep learning). At this stage, the machine identifies relationships among data points, builds internal representations, and organizes information into forms suitable for computational learning.

The third stage, Model Building, is the core of AI development. In this phase, the system engages in iterative learning using statistical optimization techniques. The model adjusts its internal parameters through repeated exposure to the extracted features, gradually improving its ability to generate predictions or produce outputs.

The final stage, Model Assessment, evaluates how well the AI performs after the training phase. This assessment typically involves multiple dimensions, including accuracy, generalization, robustness, consistency, safety, and potential algorithmic bias. The purpose is to determine whether the model's behavior aligns with expected outcomes and whether it can reliably operate across diverse contexts.

Based on the analysis of the stages of AI operation outlined above, copyright-related issues can be identified as follows.

- a. Fair Use in Text and Data Mining (TDM). The doctrine of fair use serves as a balancing mechanism between copyright protection and the public's freedom of expression, with the purpose prevent monopolization of creative works (Lin, 2023). At the Data Input and Features Expression stages, AI systems conduct large scale and imperceptible TDM activities without obtaining authorization from creators or copyright holders. The boundaries of fair use become increasingly blurred because copyright-protected content is used by AI to "train" models and produce transformed outputs, yet such use involves acts of reproduction that infringe the creator's exclusive rights. Moreover, AI-generated outputs typically do not provide attribution to original creators, thereby violating moral rights. As a result, the distinction between fair use and infringement becomes uncertain in the context of AI-driven TDM.
- b. Non-Literal Copying Infringement. Non-literal copying refers to the imitation of substantial elements of a work without replicating it in its entirety (Komuna & Alif Arhanda, 2020). During the Model Building and Model Assessment stages, AI algorithms construct models by extracting correlations and aesthetic patterns from training data. Although AI-generated outputs may appear "new," their visual style, composition, or distinctive patterns may closely resemble the works of particular creators, leading to allegations of non-literal copying. This issue is particularly evident in text-to-image models capable of mimicking the "signature style" of illustrators or painters with high fidelity. Such cases complicate copyright enforcement because AI does not reproduce works identically, but nevertheless extracts creative expressions that ought to be protected under copyright law.
- c. AI as a Creator. Another challenge concerns the status of AI as a producer of creative outputs. Under Indonesia's Copyright Law (Law No. 28/2014), a creator is defined as an individual or group of individuals who produce works through intellectual capability, creativity, and human expression. This definition explicitly excludes AI from being recognized as a creator, as AI possesses neither intention, consciousness, nor legal responsibility. However, at the Output Generation stage, AI systems frequently produce works that are aesthetically and substantively similar to human created works (Maguaren, et, all, 2024) raising the question: who owns the copyright over such outputs? Is it the user who provides the prompt, the model developer, the dataset owner, or is no copyright generated at all? This normative gap creates legal uncertainty particularly regarding ownership, protection, and enforcement and opens the potential for conflicts among original creators, AI developers, and users. The debate becomes increasingly relevant as more generative AI platforms commercialize outputs that lack a clearly established legal basis for copyright protection.
- d. Liability for Copyright Infringement by AI. Beyond the issue of authorship, another emerging challenge is the question of liability when copyright infringement involves AI systems Under Indonesia's Copyright Law, liability is generally imposed on natural persons or legal entities, not autonomous systems. In practice, however, infringement may occur at various points within the AI ecosystem: product producers, platform providers, end-users, or third party (Yang, 2023). For example, an AI model trained on copyrighted data without permission may generate outputs that infringe upon the creator's economic rights, yet end-users are often unaware of the training process and thus cannot reasonably bear the responsibility. Meanwhile, AI developers may limit their liability through Terms of Use, reducing opportunities for rights holders to pursue remedies. This ambiguity creates a legal vacuum because Indonesia has no specific rules assigning responsibility for AI-generated infringements, whether arising during training, deployment, or distribution. The absence of a clear liability framework further weakens the effectiveness of copyright enforcement in digital environments.

- e. **Legal Uncertainty Regarding Datasets and Model Transparency.** A critical issue in the copyright debate surrounding AI is the lack of mandatory transparency regarding the sources of data used to train AI models. Most AI developers do not disclose their training datasets either due to trade secret protection or the vast and complex nature of the datasets themselves (Felzmann, Villaronga, Lutz, & Tamó-Larrieux, 2019). This lack of transparency prevents creators from determining whether their works have been used without permission and simultaneously complicates evidentiary processes in copyright disputes. In Indonesia, the Copyright Law does not regulate dataset labeling, dataset registration, or data provenance requirements, allowing AI models to be trained on large volumes of copyrighted content without audit mechanisms. Furthermore, there are no national standards on dataset governance, including how data should be collected, curated, filtered, or validated. This uncertainty is compounded by the absence of rules regarding data contamination situations in which datasets contain a mixture of lawful and unlawful content which may lead to cascading infringements in AI generated outputs. Without regulations mandating dataset transparency, TDM licensing mechanisms, or documentation of training processes, creators are placed at a structural disadvantage, and the enforcement of their copyright becomes increasingly difficult.

3.2 Analysis of Indonesia’s Copyright Law in Addressing AI-Driven Copyright Challenges

Indonesia’s copyright system is anchored in Law No. 28 of 2014 on Copyright, a statute that establishes a comprehensive framework for the protection of works produced through human intellectual effort, creativity, and expression. Drafted in a pre-generative AI context, the law embodies a traditional and anthropocentric view of authorship and originality, offering limited guidance on how its provisions should apply to autonomous or data-driven technological processes. Consequently, while the statute provides robust protection for conventional forms of creative production, it struggles to accommodate the unprecedented modes of creation, reproduction, and dissemination introduced by modern AI systems. Having identified the key copyright issues arising from the technical stages of AI operation from data input, text and data mining, and model training to output generation the next step is to examine the extent to which Indonesia’s current copyright framework can address these challenges. While the previous subsection mapped how each phase of AI workflows interacts with copyright-protected materials and potentially gives rise to infringement risks, this subsection evaluates whether Indonesia’s Copyright Law (Law No. 28/2014) provides adequate legal grounding to regulate such risks. By analyzing the statutory provisions in relation to the five identified problem areas, fair use in TDM, non-literal copying, AI authorship, liability, and dataset transparency, this subchapter highlights the existing normative gaps and structural limitations within the Indonesian copyright regime.

To clearly illustrate the extent of the regulatory shortcomings within Indonesia’s Copyright Law in relation to AI-generated works, the following table provides a structured legal gap map that aligns the identified copyright issues with their corresponding legal provisions, existing gaps, and resulting implications.

Table 1. Legal Gap Map of Indonesia’s Copyright Law in the AI Era

Issue Area	Relevant AI Technical Stage	Current Legal Provision (Law No. 28/2014)	Regulatory Gap	Implications
1. Fair Use in Text and Data Mining (TDM)	Data Input; Feature Extraction	Copyright exceptions (Arts. 43–51) do not include TDM; reproduction requires authorization.	No specific TDM exception; no distinction between human and machine copying; no licensing mechanism for TDM.	AI training activities may be unlawful by default; legal uncertainty for developers; creators lack mechanism to protect works used in TDM.
2. Non-Literal Copying	Model Building; Model Assessment	Protection against reproduction of “substantial parts,” but no standard for substantial similarity tests.	No criteria for assessing non-literal or algorithmic similarity; no rules for style imitation.	Difficult to prosecute AI-mediated copying; creators' distinctive styles unprotected; text-to-image models can mimic artists without clear infringement boundaries.

3. AI as a Creator (Authorship & Ownership)	Output	“Creator” must be a natural person (Arts. 1(2) & 8).	No category for AI-generated works; no rules for determining ownership when human input is minimal; no sui generis protection.	Legal uncertainty in commercial exploitation; disputes among users, developers, and dataset owners; AI outputs risk falling into legal limbo.
4. Copyright Liability in AI-Generated Infringement	Input, feature extraction, model building, output	Liability attaches only to natural persons or legal entities; no rules for autonomous systems.	No allocation of liability between developers, dataset curators, platforms, and end users; no safe harbor provisions; limited enforcement pathways.	Rights holders lack effective remedy; platforms may disclaim liability; end-users exposed to unpredictable legal risk.
5. Dataset Governance & Model Transparency	Input, Feature Extraction	No provisions on dataset provenance, labeling, registration, or transparency.	No requirements for dataset disclosure; no standards for data governance; no mechanism to detect contaminated datasets.	Creators cannot verify unauthorized use; evidentiary challenges in disputes; AI systems may perpetuate infringement at scale.

The legal gap map above illustrates that, although Indonesia’s Copyright Law (Law No. 28/2014) provides a comprehensive framework governing human-centered creative processes, it remains structurally limited in addressing copyright challenges arising from AI-driven production. In the context of text and data mining (TDM), the statute does not expressly recognize TDM as a distinct lawful use, nor does it provide limitations or exceptions comparable to those found in jurisdictions that have modernized their copyright regimes to accommodate large-scale computational analysis. As a result, the legality of TDM activities performed during AI model development remains ambiguous and overly dependent on traditional doctrines of permission and infringement, which were not designed for machine-based processing of massive datasets.

Similarly, the current statute is insufficient to address issues of non-literal copying that arise when AI outputs reproduce structural, stylistic, or conceptual elements from training data without directly copying expression. The law’s focus on substantial or verbatim reproduction does not account for algorithmic extraction of patterns or latent features, leaving a normative vacuum in determining whether such outputs constitute derivative works or lawful transformations.

The most significant doctrinal gap emerges in relation to AI authorship. Indonesian copyright law adheres strictly to a human-centered notion of authorship, requiring human intellectual effort as the basis of protection. Consequently, AI-generated content falls outside the scope of protectable works, yet the statute also provides no legal classification or regulatory mechanism for such outputs. This absence creates legal uncertainty for creators, users, and industries relying on AI systems.

Further, the law does not explicitly identify or allocate liability for AI-generated infringements. The Copyright Law’s infringement provisions presuppose human actors capable of intent, negligence, or direct participation. AI systems that autonomously produce infringing outputs challenge the applicability of existing liability doctrines, creating uncertainty regarding the responsibility of developers, deployers, users, or intermediaries.

Finally, the framework lacks requirements concerning dataset transparency, an increasingly important dimension of AI governance. Without statutory obligations for dataset disclosure, provenance documentation, or rights verification, copyright holders lack mechanisms to ascertain whether their works have been included in AI training. This gap weakens enforcement, undermines accountability, and limits the ability of the legal system to safeguard authorial rights in the context of opaque, large-scale data processing.

4. Conclusion

The examination of AI’s operational workflow from data input and feature extraction to model building, assessment, and output generation reveals that copyright issues are embedded at nearly every stage of the AI development cycle. Modern AI systems rely on vast datasets and computationally intensive text and data mining processes that directly intersect with the exclusive rights granted to creators under copyright law. The identification of key challenges fair use in TDM, non-literal copying, AI authorship, liability attribution, and dataset

transparency demonstrates that the technical nature of AI amplifies longstanding doctrinal tensions while simultaneously exposing new regulatory blind spots.

Indonesia's Copyright Law (Law No. 28/2014), despite providing a comprehensive and well-established framework for human centered creative production, is not yet equipped to accommodate the complexities introduced by AI-driven content generation. The legal gap map illustrates that the statute lacks explicit provisions addressing machine-based data processing, algorithmic imitation of stylistic elements, non-human authorship, allocation of responsibility for autonomous infringements, and transparency in dataset governance. These normative deficiencies hinder the law's ability to protect creators, regulate emerging AI industries, and provide legal certainty for stakeholders engaged in AI development and use.

Taken together, the findings underscore an urgent need for regulatory adaptation. Without updating the copyright framework to address the unique characteristics of AI, Indonesia risks falling behind jurisdictions that have already modernized their copyright regimes to balance innovation with the protection of intellectual property. The unresolved gaps not only weaken enforcement mechanisms but also constrain the ethical and sustainable growth of the national AI ecosystem. As AI technologies continue to evolve rapidly, a forward-looking legal reform whether through targeted amendments, sui generis protection, or integrated AI governance mechanisms will be essential to ensure that Indonesia's copyright system remains robust, relevant, and resilient in the digital era.

5. Speech Thank You

The author also acknowledges the valuable assistance of digital research tools, including artificial intelligence-based language systems, which were utilized to support the translation, linguistic polishing, and technical editing processes of this manuscript. All analytical interpretations, legal assessments, and substantive conclusions remain solely the responsibility of the author.

6. Bibliography

- Eviani, N. Y., Maskun, M., & Ahmad Fachri Faqi. (2024). Legal Challenges of AI-Induced Copyright Infringement: Evaluating Liability and Dispute Resolution Mechanisms in Digital Era. *Jambura Law Review*.
- Felzmann, H., Villaronga, E., Lutz, C., & Tamó-Larrieux, A. (2019). Transparency you can trust: Transparency requirements for artificial intelligence between legal norms and contextual concerns. *Big Data & Societ*.
- Hoshiar, S., & Kiran, S. (2024, October). Copyright in the Age of Artificial Intelligence: Unravelling the Complexities For the Protection of AI-Generated Work. In *2024 ITU Kaleidoscope: Innovation and Digital Transformation for a Sustainable World (ITU K)* (pp. 1-7). IEEE.
- Janiesch, C., Zschech, P., & Heinrich, K. (2021). Machine learning and deep learning. *Business & Information Systems Engineering*, 63(6), 685–695. <https://doi.org/10.1007/s12599-021-00711-1>
- Jordan, M. I. (2019). Artificial intelligence—the revolution hasn't happened yet. *Harvard Data Science Review*, 1-9.
- Kear, A., & Folkes, S. L. (2020). The Hierarchy of AI. *International Journal of Advanced Computer Science and Applications*.
- Komuna, A. P., & Putra, A. A. (2020). Pelanggaran Hak Cipta Nonliteral Terhadap Karya Sinematografi Di Indonesia. *Alauddin Law Development Journal*, 2(3), 465-472.
- Komuna, A. P., Salam, S., Wirawan, A. R., Slamet, A., & Sanjaya, I. M. G. (2024). Legal Protection of Regional Local Potential Through Registration of Geographical Indications of Ponda Mats to Strengthen the Community Economy In Buton Regency. *JHR (Jurnal Hukum Replik)*, 12(2), 508-524 <http://dx.doi.org/10.31000/jhr.v12i2.11938>.
- Latif, E., Gengchen, M., Nyaaba, M., Wu, X., Liu, N., Lu, G., . . . Zhai, X. (2023). Artificial General Intelligence (AGI) for Education. *arXiv preprint arXiv:2304.12479*.
- Lee, J. (2022). *Artificial intelligence and international law*. Dordrecht: Springer.
- Lin, H. Y. (2023). Standing on the Shoulders of AI Giants. *Computer*, 56(01), 97-101.
- Păvăloaia, V. D., & Necula, S. C. (2023). Artificial intelligence as a disruptive technology—a systematic literature review. *Electronics*, 12(5), 1102.

- Rabago, G. (2024). Can AI Have a Signature: Legal Ownership and Authorship of Creative Materials Involving Artificial Intelligence. *UC Merced Undergraduate Research Journal*, 16(2).
- Sun, Y., Yang, C. H., Lyu, Y., & Lin, R. (2022). From pigments to pixels: a comparison of human and AI painting. *Applied Sciences*, 12(8), 3724.
- Yang, Y. (2023). Attribution of Liability for Copyright Infringement by Artificial Intelligence Generated Content. *Lecture Notes in Education Psychology and Public Media*.