

Governing traditional cultural expressions as artificial intelligence training data: Reconstructing communal consent and data transparency in Indonesia

Muhammad Abdul Azis^{1*}, Riski Ananda Kusuma Putri², Alya Synthia Kurniawati³, Muhammad Thufail Ariiq⁴

¹ Universitas Diponegoro, Semarang, Indonesia

^{2,3,4} Universitas Muhammadiyah Cirebon, Cirebon, Indonesia

*Corresponding Author: muhammadabdulazis@students.undip.ac.id

Submission: 28 June 2026

Revision: 10 July 2026

Accepted: 18 July 2026

Abstract

The use of traditional cultural expressions as training data for generative artificial intelligence systems raises legal questions that individual copyright doctrine cannot adequately resolve. This study examines the legal status of traditional cultural expressions, the legitimacy of their use as training data, the position of source communities, and developers' transparency obligations under Indonesian law. It employs doctrinal legal research using statutory, conceptual, and comparative approaches. The findings demonstrate that Indonesian law recognizes state stewardship, inventory obligations, commercial-use authorization, and benefit sharing but does not regulate the classification of model training, procedures for communal consent, source traceability, or legal remedies. This article proposes the Cultural Training Data Legality Test, integrating legal status, community provenance, lawful acquisition, layered consent, purpose limitation, cultural integrity, transparency, benefit sharing, and remedies. For high-risk commercial uses, state authorization and representative community consent should apply cumulatively, with transparency serving as a foundation for accountability.

Keywords: Artificial intelligence; Communal consent; Data transparency; Traditional cultural expressions; Training data.

Abstrak

Pemanfaatan ekspresi budaya tradisional sebagai data pelatihan bagi sistem kecerdasan artifisial generatif menimbulkan persoalan hukum yang belum dapat dijawab secara memadai oleh rezim hak cipta individual. Penelitian ini menganalisis status hukum ekspresi budaya tradisional, legalitas penggunaannya sebagai data pelatihan, kedudukan komunitas asal, serta kewajiban transparansi pengembang dalam hukum Indonesia. Penelitian ini menggunakan metode penelitian hukum doktrinal dengan pendekatan peraturan perundang-undangan, konseptual, dan perbandingan. Hasil penelitian menunjukkan bahwa hukum Indonesia telah mengakui penguasaan negara, kewajiban inventarisasi, izin pemanfaatan komersial, dan pembagian manfaat, tetapi belum mengatur klasifikasi pelatihan model, prosedur persetujuan komunal, keterlacakan sumber, maupun mekanisme pemulihan hukum. Penelitian ini menawarkan Uji Legalitas Data Pelatihan Budaya yang mengintegrasikan status hukum, asal komunitas, legalitas perolehan, persetujuan berlapis, pembatasan tujuan, integritas budaya, transparansi, pembagian manfaat, dan mekanisme pemulihan. Untuk penggunaan komersial yang berisiko budaya tinggi, persetujuan negara dan komunitas perwakilan harus berlaku secara kumulatif, dengan transparansi berfungsi sebagai landasan akuntabilitas.

Kata Kunci: Data pelatihan; Ekspresi budaya tradisional; Kecerdasan artifisial; Persetujuan komunal; Transparansi data.

Introduction

Generative artificial intelligence systems are developed by processing large-scale corpora comprising text, images, music, performance recordings, designs, and audiovisual materials. The training process involves collecting, copying, cleaning, classifying, and transforming source materials into mathematical parameters that enable models to generate new content (Sag, 2023; U.S. Copyright Office, 2025). This reliance on extensive datasets increases the likelihood that digitized cultural materials from museums, libraries, government institutions, researchers, communities, and social media users will be incorporated into training datasets without the knowledge of their source communities (Oruc, 2025). Public accessibility is often treated as implicit permission for unrestricted use, even though online availability does not negate copyright protection, customary restrictions, archival purposes, or the communal interests associated with cultural materials (Buick, 2025; Lewis, 2020).

Traditional cultural expressions differ from individually authored works because they are typically created collectively rather than by identifiable individuals. These expressions are preserved, adapted, transmitted, and inherited across generations by communities (Aroian, 2025; Setiyono et al., 2024). Their significance extends beyond economic value. Traditional motifs, music, dances, rituals, oral traditions, architecture, and cultural symbols may embody historical memory, spiritual significance, social functions, and culturally prescribed restrictions governing access and use (United Nations, 2007; WIPO, 2024a). Their incorporation into generative artificial intelligence systems may result in cultural harm through decontextualization, the conflation of distinct cultural identities, misleading representations, the reproduction of sacred elements, and commercial exploitation without equitable benefit sharing for source communities (Dharmawan et al., 2024; Oruc, 2025).

Indonesian law provides multiple layers of legal protection for traditional cultural expressions. Article 38 of Law Number 28 of 2014 on Copyright stipulates that copyright over traditional cultural expressions is vested in the state and requires the state to inventory, preserve, and safeguard them (Republic of Indonesia, 2014). Law Number 5 of 2017 on the Advancement of Culture recognizes oral traditions, customs, rites, traditional knowledge, traditional technology, arts, languages, folk games, and traditional sports as objects of cultural advancement (Republic of Indonesia, 2017). Government Regulation Number 87 of 2021 further strengthens the integrated cultural data system and regulates the protection and utilization of cultural objects (Republic of Indonesia, 2021).

Government Regulation Number 56 of 2022 on Communal Intellectual Property establishes a more comprehensive legal framework by regulating the classification, inventory, protection, preservation, information systems, and utilization of communal intellectual property (Republic of Indonesia, 2022). The regulation permits utilization

while requiring respect for the values upheld by source communities. Commercial use is also linked to licensing and benefit-sharing arrangements, although the procedures governing community participation, consent, cultural impact assessment, and benefit distribution remain insufficiently developed (Dharmawan et al., 2024). These provisions demonstrate that traditional cultural expressions do not constitute an unrestricted public domain. The unresolved legal issue is whether artificial intelligence model training constitutes a form of utilization and, if so, what legal requirements should govern that activity.

Indonesia's artificial intelligence governance remains primarily based on the Circular Letter of the Minister of Communication and Informatics Number 9 of 2023 on Artificial Intelligence Ethics. The Circular establishes principles of inclusiveness, humanity, security, accessibility, transparency, accountability, personal data protection, sustainable development, and intellectual property protection (Ministry of Communication and Informatics of the Republic of Indonesia, 2023). However, its legal status does not establish enforceable communal rights, mandatory training-data disclosure, specific sanctions, or effective legal remedies. Interministerial discussions concerning a draft presidential regulation on artificial intelligence ethics and a national artificial intelligence roadmap were completed in May 2026, but the proposed instruments had not entered into force when this article was prepared (Ministry of Communication and Digital Affairs of the Republic of Indonesia, 2026).

Previous studies have examined various dimensions of this issue. Dharmawan et al. (2024) discussed the risks arising from the transformation of traditional cultural expressions through generative artificial intelligence systems. Aroian (2025) identified the structural tension between conventional copyright doctrine and the communal, intergenerational, and dynamic nature of traditional culture. Buick (2025) argued that training-data transparency facilitates oversight; however, disclosure alone remains ineffective in the absence of enforceable substantive rights. Oruc (2025) further contended that the use of cultural heritage as artificial intelligence training data requires careful consideration of provenance, accessibility, community interests, and institutional responsibility. Nevertheless, these studies have not developed a comprehensive doctrinal framework integrating state stewardship, communal consent, data provenance, purpose limitation, cultural integrity, transparency, benefit sharing, and legal remedies.

This study adopts a conceptual framework integrating state stewardship, communal consent, cultural integrity, and data transparency. State stewardship is understood as a public responsibility to safeguard communal interests rather than as an expression of unrestricted state ownership. Communal consent refers to authorization obtained through representative and culturally legitimate decision-making mechanisms. Cultural integrity concerns the preservation of historical meaning, spiritual

significance, social function, and community-based restrictions. Data transparency refers to the traceability of the provenance, legal status, purpose, and processing of cultural materials. This study examines how Indonesian law protects traditional cultural expressions used as artificial intelligence training data and explores how communal consent and data transparency should be reconstructed. It proposes a Cultural Training Data Legality Test encompassing lawful acquisition, community provenance, layered consent, purpose limitation, cultural integrity, transparency, benefit sharing, and legal remedies. This study contributes to legal scholarship by developing a doctrinal standard for cultural data governance while also contributing to the humanities and civilization studies by protecting collective memory, cultural representation, historical meaning, and community authority over technological reproduction.

Method

This study employed doctrinal legal research using statutory, conceptual, and comparative approaches. The statutory approach examined the legal framework governing traditional cultural expressions, communal intellectual property, cultural advancement, and artificial intelligence governance in Indonesia. The conceptual approach was applied to analyze state stewardship, communal consent, cultural integrity, data transparency, benefit sharing, and legal remedies. The comparative approach examined selected international legal developments concerning copyright compliance, training-data disclosure, cultural heritage protection, and the governance of generative artificial intelligence.

The research relied on primary and secondary legal materials. Primary legal materials included Law Number 28 of 2014 on Copyright, Law Number 5 of 2017 on the Advancement of Culture, Government Regulation Number 87 of 2021, Government Regulation Number 56 of 2022 on Communal Intellectual Property, and the Circular Letter of the Minister of Communication and Informatics Number 9 of 2023 on Artificial Intelligence Ethics. Comparative legal materials comprised relevant international legal instruments and foreign regulatory frameworks concerning traditional cultural expressions and artificial intelligence. Secondary legal materials consisted of academic journal articles, books, institutional reports, and legal commentaries addressing copyright, communal cultural rights, data governance, and generative artificial intelligence.

Legal materials were collected through structured documentary research using official legal databases, academic databases, institutional repositories, and publications issued by relevant international organizations. The research instruments consisted of a legal-material classification matrix and an analytical framework designed to identify the legal status of traditional cultural expressions, the respective authority of the state and source communities, licensing requirements, transparency obligations, benefit-sharing

mechanisms, and available legal remedies. The materials were selected based on their legal authority, relevance to the research questions, scholarly credibility, and contribution to contemporary legal scholarship.

The analysis employed grammatical, systematic, teleological, and comparative methods of legal interpretation. Grammatical interpretation examined the meaning of legal terms such as state-held copyright, utilization, authorization, protection, and communal intellectual property. Systematic interpretation analyzed the interrelationship between copyright law, cultural law, communal intellectual property regulation, and artificial intelligence governance. Teleological interpretation assessed whether the existing legal framework fulfilled the objectives of cultural preservation, community participation, legal accountability, and equitable benefit sharing. Comparative interpretation was used to identify regulatory models that could inform the reconstruction of Indonesian law.

The credibility of the analysis was ensured through source triangulation, hierarchical verification of legal authorities, and consistency checks between statutory provisions, legal principles, and scholarly interpretations. Primary legal materials were prioritized over secondary sources, while recent academic literature was used to confirm the current state of scholarship. Conflicting interpretations were evaluated by comparing their normative foundations and their compatibility with the objectives of cultural protection. The research process culminated in the formulation of the Cultural Training Data Legality Test as a prescriptive doctrinal framework for evaluating the lawful use of traditional cultural expressions in artificial intelligence training.

Results and Discussion

Legal construction and normative gaps in the protection of traditional cultural expressions as training data

Article 38 of Law Number 28 of 2014 assigns copyright over traditional cultural expressions to the state and mandates their inventory, protection, and preservation. When read together with Government Regulation Number 56 of 2022, this authority is more appropriately understood as state stewardship exercised in the interests of source communities rather than as an expression of unrestricted state ownership. The regulation recognizes communities of origin and links commercial utilization to authorization and benefit sharing (Republic of Indonesia, 2014, 2022). Accordingly, state authorization should protect and administer communal rights but cannot automatically supersede the cultural authority of communities that have preserved and transmitted these expressions across generations (Setiyono et al., 2024).

This interpretation highlights the limitations of conventional copyright doctrine. Copyright law generally presumes an identifiable author, a fixed work, clearly defined exclusive rights, and a limited term of protection, whereas traditional cultural

expressions are collective, dynamic, and embedded within customary practices (Aroian, 2025). Consequently, the central issue extends beyond questions of ownership. It also concerns the allocation of authority among the state, institutions that hold digital copies, and communities possessing cultural legitimacy. A museum or archive may control a photograph or recording without having the authority to license the underlying ritual, motif, oral tradition, or performance for commercial artificial intelligence model training (Dharmawan et al., 2024; Oruc, 2025).

Public accessibility must likewise be distinguished from lawful authorization. Cultural materials may be digitized for preservation, education, tourism, or research while remaining subject to customary restrictions, attribution requirements, or limitations on the use of sacred materials. WIPO's Draft Articles recognize the continuing interests of source communities in attribution, integrity, control, and equitable benefit sharing, even when traditional cultural expressions are publicly accessible (WIPO, 2024). Digitization therefore alters the mode of access but does not sever the relationship between a traditional cultural expression and its source community. Indonesian law should accordingly distinguish access to a digital copy from legal authorization to process the underlying cultural expression for a new technological or commercial purpose.

Artificial intelligence model training should be classified as a form of utilization when cultural materials are intentionally processed to develop or enhance a model's capabilities. Training may involve the collection, copying, cleaning, classification, conversion into machine-readable representations, parameter optimization, and continuous model updating. These activities have both economic and legal significance, even when the resulting model does not retain each source in the form of an individually retrievable file (Sag, 2023; U.S. Copyright Office, 2025). Restricting the concept of utilization solely to direct display or reproduction would overlook the economic function of training data in generating commercial value and would undermine the protective objectives of Government Regulation Number 56 of 2022.

Classifying model training as utilization does not require identical legal treatment for every activity. Public preservation initiatives, limited academic research, community-led language revitalization, and the commercial development of foundation models present substantially different levels of risk. A proportionate assessment should therefore consider the purpose of processing, cultural sensitivity, processing scale, anticipated commercial gain, model openness, and the likelihood of reproducing recognizable cultural expressions. The CARE Principles support this differentiated approach by linking data use to collective benefit, authority to control, responsibility, and ethics (Carroll et al., 2020). Such proportionality prevents unrestricted extraction while preserving legitimate cultural, educational, and research activities.

The existing licensing framework does not yet establish a comprehensive procedure for obtaining communal consent. Government Regulation Number 56 of 2022 does not clearly specify how representative bodies should be identified, how internal community deliberations should be conducted, what information developers must disclose, or whether authorization extends to subsequent model updates and secondary uses (Dharmawan et al., 2024). Consequently, permission granted by an uploader, researcher, cultural institution, or local government official may be formally documented yet remain invalid from the perspective of the source community. Effective communal consent requires verifiable representation, adequate disclosure of relevant information, authorization that is specific to the intended purpose, and opportunities for communities to impose cultural restrictions or renegotiate authorization when proposed uses change materially.

Transparency constitutes the principal evidentiary bridge between substantive rights and effective enforcement. Source communities cannot challenge unauthorized processing or claim benefit sharing if they are unable to determine whether their traditional cultural expressions have been incorporated into training datasets. The European Union Artificial Intelligence Act requires providers of general-purpose artificial intelligence models to maintain copyright compliance policies and publish summaries of training data, although disclosure alone does not constitute legal authorization (European Parliament & Council of the European Union, 2024). Indonesian ethical guidance recognizes the principles of transparency and intellectual property protection, but it does not establish binding disclosure obligations, audit mechanisms, or community rights of access (Ministry of Communication and Informatics of the Republic of Indonesia, 2023). Existing legal remedies likewise remain focused on conventional copyright infringement rather than addressing culturally appropriate corrective measures, model retraining, or restrictions on generated outputs.

Table 1. Mapping of existing legal protection and normative gaps

Aspect	Existing legal protection	Gap in relation to model training
Status and authority	State stewardship, inventories, and recognition of communities of origin	No clear relationship between state authorization and communal decision-making
Utilization	Commercial-use and benefit-sharing framework under Government Regulation Number 56 of 2022	Model training, parameter optimization, and model updating are not expressly classified as forms of utilization
Data systems	Integrated cultural and communal intellectual property databases	No rights metadata, sensitivity labels, permitted-purpose fields, or links to artificial intelligence models

Technology governance	Ethical principles of transparency, accountability, and intellectual property protection	No binding disclosure obligations, audit authority, sanctions, or community access procedures
Remedies	General injunctions and compensation for established infringement	No explicit provisions for correction, model retraining, output restrictions, or culturally specific remedies

Source: Author's analysis based on Republic of Indonesia (2014, 2017, 2021, 2022) and the Ministry of Communication and Informatics of the Republic of Indonesia (2023).

Table 1 demonstrates that the principal challenge lies in normative fragmentation rather than the complete absence of legal protection. Previous studies have identified tensions between communal culture and copyright doctrine, the risks associated with generative artificial intelligence, and the importance of training-data transparency (Aroian, 2025; Buick, 2025; Dharmawan et al., 2024). This study advances the existing scholarship by identifying the model-training stage as the decisive legal event. State stewardship, communal consent, transparency, and legal remedies should operate as an integrated sequence because authorization cannot be properly evaluated without reliable provenance, while transparency has limited legal value in the absence of enforceable communal rights. This integrated interpretation provides the doctrinal foundation for the legal reconstruction proposed in the following section.

Reconstruction of communal consent and training-data transparency

The proposed reconstruction begins by conceptualizing digitized traditional cultural expressions as relational cultural data. Their conversion into digital files or datasets does not sever their connection to collective memory, social functions, spiritual significance, or customary authority. Accordingly, state stewardship and communal consent should operate cumulatively in cases involving commercial exploitation or culturally high-risk uses. The state is responsible for verifying legal status, safeguarding the public interest, and ensuring regulatory compliance, whereas representative communities determine whether the proposed use respects cultural meaning and established customary restrictions. This layered governance model avoids exclusive state control while preventing individual licensors or data holders from authorizing uses that exceed their legitimate cultural mandate.

Communal consent should satisfy four minimum requirements. First, the representative mandate must be verifiable and inclusive of all affected groups. Second, the information provided should clearly identify the developer, source materials, model, intended purpose, duration of use, commercial nature, potential output risks, and benefit-sharing arrangements. Third, authorization must remain limited to the specific processing activities approved and should not automatically extend to unrelated models, data transfers, or subsequent model updates. Fourth, communities should have access to effective mechanisms for review, complaint, and renegotiation. These requirements translate the CARE Principles and collective cultural rights into a

procedural framework that can be implemented within Indonesia's communal intellectual property regime (Carroll et al., 2020; United Nations, 2007).

A risk-based framework is necessary to preserve proportionality. Open and non-sensitive cultural materials used for limited public-interest research may warrant simplified notification and compliance requirements. In contrast, commercial applications, sacred materials, restricted traditional knowledge, identity-bearing cultural symbols, and models capable of reproducing recognizable cultural expressions should trigger enhanced consent requirements, cultural impact assessments, attribution safeguards, and benefit-sharing obligations. Prohibiting certain forms of processing may also be justified where disclosure or reproduction would violate fundamental customary restrictions. A risk-based classification therefore avoids imposing a blanket prohibition while recognizing that the cultural consequences of data processing cannot be assessed solely in terms of economic harm.

Transparency should operate at both the public and regulatory levels throughout the entire artificial intelligence model life cycle. Public disclosures should identify the principal categories of source materials, relevant cultural collections, acquisition methods, and compliance measures. More detailed records should remain accessible to regulators and affected communities, including information on provenance, licences, dataset modifications, safeguards, and the models in which the data are used. Before model training, developers should conduct provenance verification and legal rights assessments. During development, they should maintain comprehensive records and evaluate the risk of reproducing protected cultural expressions. Following deployment, developers should maintain accessible complaint mechanisms, investigate substantiated claims, and document any corrective measures undertaken. This framework establishes transparency as a continuous obligation of accountability rather than as a one-time disclosure requirement (Buick, 2025; UNESCO, 2021).

These requirements are consolidated in the Cultural Training Data Legality Test. The framework evaluates legality through six interrelated elements: legal status and community provenance, lawful acquisition and traceability, layered consent, purpose limitation and cultural integrity, transparency and auditability, and benefit sharing accompanied by effective remedies. Non-compliance at an earlier stage affects the validity of all subsequent processing. A developer cannot rely on benefit sharing to legitimize unlawfully acquired data, nor can transparency substitute for valid communal consent. Accordingly, the framework differs from conventional ethics checklists because it establishes legal consequences while integrating evidentiary requirements, authorization, substantive protection, and corrective measures within a single doctrinal framework.

Table 2. Cultural training data legality test

Element		Core legal requirement	Consequence of non-compliance
Status and provenance		Identify the traditional cultural expression, source community, competing claims, and level of cultural sensitivity	Suspend processing until legal authority is established
Lawful acquisition		Maintain records of data sources, licences, legal authority, copies, dataset modifications, and provenance	Data deletion, audit, and adverse evidentiary inference
Layered consent		Obtain approval from both the state and representative community for commercial or high-risk uses	Invalid authorization, suspension, compensation, or renegotiation
Purpose and integrity		Apply purpose limitations, cultural safeguards, attribution requirements, and restrictions on sacred materials	Correction, output restriction, or prohibition of processing
Transparency and audit		Publish training-data summaries, retain records, and provide targeted access for audit purposes	Disclosure orders, administrative penalties, and burden shifting
Benefit sharing and remedies		Establish benefit-sharing arrangements, reporting obligations, correction procedures, deletion, retraining, and compensation	Damages, profit sharing, model retraining, or restrictions on model use

Source: Author's construction based on Carroll et al. (2020), Buick (2025), WIPO (2024), and Government Regulation Number 56 of 2022.

Benefit sharing under this framework should reflect both the scale and the nature of the intended use. Direct licensing arrangements may be appropriate for identifiable datasets and representative communities, whereas collective funding mechanisms may be more suitable for datasets derived from multiple communities. Non-financial benefits may include digital infrastructure, support for digitization initiatives, community-controlled archives, capacity-building programs, language technologies, and access to technological tools developed from the datasets. Such measures should complement, rather than replace, monetary compensation where commercial value is substantial. Legal remedies should likewise address cultural harm in addition to economic loss. Depending on the severity of the infringement, available remedies may include correction, removal of data from active datasets, restrictions on generated outputs, model retraining, attribution, profit sharing, financial compensation, or independent monitoring. Technical infeasibility should not constitute a general defence where inadequate record-keeping is the source of the problem.

The proposed framework makes a theoretical contribution by integrating state stewardship, communal consent, cultural integrity, and data transparency into a unified doctrinal model. It extends the CARE Principles within the Indonesian legal

context without assuming that all cultural communities should be treated as legally equivalent to Indigenous peoples. The essential insight is relational: legal authority derives from the continuing social and cultural relationship between a traditional cultural expression and its source community (Carroll et al., 2020; Lewis, 2020). The framework also advances previous scholarship on transparency by conceptualizing disclosure not as an independent solution but as the evidentiary foundation that enables consent, benefit sharing, accountability, and legal remedies to operate effectively.

The practical implication of this study is the need for a two-tier regulatory reform. Government Regulation Number 56 of 2022 should explicitly classify artificial intelligence model training as a form of utilization, establish risk categories, formalize layered consent procedures, and incorporate rights metadata and cultural sensitivity indicators into communal intellectual property databases. In addition, binding artificial intelligence legislation should impose provider-level obligations concerning provenance records, training-data summaries, auditing procedures, complaint mechanisms, sanctions, and model-specific remedies. Effective implementation requires coordination among intellectual property authorities, cultural institutions, digital regulators, regional governments, and representative communities. Each institution should perform a clearly defined role rather than relying solely on state authorization to resolve complex cultural and technological issues.

The novelty of this study rests on four interrelated propositions. First, artificial intelligence model training is conceptualized as a distinct legal event. Second, public accessibility is distinguished from lawful authorization. Third, state approval and representative community consent operate cumulatively in cases involving commercial or culturally sensitive processing. Fourth, transparency functions as a mechanism of evidence and accountability rather than as a substitute for legal authorization. The Cultural Training Data Legality Test translates these propositions into a prescriptive doctrinal framework applicable to legislation, licensing, cultural inventories, auditing, and judicial review. Its principal contribution to legal scholarship and the humanities lies in protecting collective memory, cultural meaning, and community authority while enabling proportionate research, cultural preservation, and responsible technological innovation.

Conclusion

Indonesian law recognizes traditional cultural expressions through the interconnected regimes of copyright, cultural advancement, and communal intellectual property. Together, these legal frameworks establish state stewardship, inventory obligations, protective measures, licensing mechanisms, and benefit-sharing arrangements. However, their application to artificial intelligence training remains incomplete because the existing legal framework does not explicitly classify model training as a

form of utilization, distinguish public accessibility from lawful authorization, regulate communal consent procedures, require source traceability, or provide effective legal remedies. Accordingly, state stewardship should be understood as a public responsibility to safeguard communal interests, while source communities retain authority over culturally legitimate consent, restrictions on use, and the preservation of cultural integrity.

This study proposes the Cultural Training Data Legality Test as a doctrinal framework encompassing legal status, community provenance, lawful acquisition, layered consent, purpose limitation, cultural integrity, transparency, benefit sharing, and legal remedies. For commercial uses involving a high level of cultural risk, state authorization and representative community consent should operate cumulatively. Transparency should ensure the disclosure of data provenance, legal status, processing purposes, and the scope of data use, but it cannot substitute for valid authorization. By conceptualizing artificial intelligence model training as an independent legal event, the proposed framework provides a normative foundation for the future development of Indonesian regulations governing artificial intelligence and cultural data governance.

References

- Aroian, M. (2025). Traditional cultural expressions and copyright law: Denoting the long-lasting confusion. *The Journal of World Intellectual Property*, 28(2), 568–588. <https://doi.org/10.1111/jwip.12345>
- Buick, A. (2025). Copyright and AI training data: Transparency to the rescue? *Journal of Intellectual Property Law & Practice*, 20(3), 182–192. <https://doi.org/10.1093/jiplp/jpae102>
- Carroll, S. R., Garba, I., Figueroa-Rodríguez, O. L., Holbrook, J., Lovett, R., Materechera, S., Parsons, M., Raseroka, K., Rodriguez-Lonebear, D., Rowe, R., Sara, R., Walker, J. D., Anderson, J., & Hudson, M. (2020). The CARE principles for Indigenous data governance. *Data Science Journal*, 19, Article 43. <https://doi.org/10.5334/dsj-2020-043>
- Dharmawan, N. K. S., Kasih, D. P. D., Samsithawrati, P. A., Dwijayanthi, P. T., Salain, M. S. P. D., Mahaswari, M., Ustriyana, M. G., & Moisa, R. V. (2024). Quo vadis traditional cultural expressions protection: Threats from personal intellectual property and artificial intelligence. *Law Reform*, 19(2), 321–343. <https://doi.org/10.14710/lr.v19i2.58639>
- European Parliament & Council of the European Union. (2024). 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 certain Union

- legislative acts. Official Journal of the European Union, L 2024/1689. <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- Lewis, J. E. (Ed.). (2020). Indigenous protocol and artificial intelligence position paper. Indigenous Protocol and Artificial Intelligence Working Group & Canadian Institute for Advanced Research. <https://doi.org/10.11573/spectrum.library.concordia.ca.00986506>
- Ministry of Communication and Digital Affairs of the Republic of Indonesia. (2026, May 6). Bangun tata kelola AI yang etis dan bertanggung jawab, pemerintah rampungkan pembahasan RPerpres. JDIH Kementerian Komunikasi dan Digital. <https://jdih.komdigi.go.id/berita/view/116>
- Ministry of Communication and Informatics of the Republic of Indonesia. (2023). Surat Edaran Menteri Komunikasi dan Informatika Republik Indonesia Nomor 9 Tahun 2023 tentang Etika Kecerdasan Artifisial [Circular Letter Number 9 of 2023 on Artificial Intelligence Ethics]. https://jdih.komdigi.go.id/produk_hukum/view/id/883
- Oruc, P. (2025). Cultural heritage as AI training data. IIC: International Review of Intellectual Property and Competition Law, 56, 1455–1459. <https://doi.org/10.1007/s40319-025-01659-5>
- Republic of Indonesia. (2014). Undang-Undang Republik Indonesia Nomor 28 Tahun 2014 tentang Hak Cipta [Law Number 28 of 2014 on Copyright]. <https://peraturan.bpk.go.id/Details/38690>
- Republic of Indonesia. (2017). Undang-Undang Republik Indonesia Nomor 5 Tahun 2017 tentang Pemajuan Kebudayaan [Law Number 5 of 2017 on the Advancement of Culture]. <https://peraturan.bpk.go.id/Details/37642/uu-no-5-tahun-2017>
- Republic of Indonesia. (2021). Peraturan Pemerintah Republik Indonesia Nomor 87 Tahun 2021 tentang Peraturan Pelaksanaan Undang-Undang Nomor 5 Tahun 2017 tentang Pemajuan Kebudayaan [Government Regulation Number 87 of 2021 on the Implementation of the Advancement of Culture Law]. <https://peraturan.bpk.go.id/Details/177044/pp-no-87-tahun-2021>
- Republic of Indonesia. (2022). Peraturan Pemerintah Republik Indonesia Nomor 56 Tahun 2022 tentang Kekayaan Intelektual Komunal [Government Regulation Number 56 of 2022 on Communal Intellectual Property]. <https://peraturan.bpk.go.id/Details/233504/pp-no-56-tahun-2022>
- Sag, M. (2023). Copyright safety for generative AI. Houston Law Review, 61(2), 295–347. <https://houstonlawreview.org/article/92126-copyright-safety-for-generative-ai>

- Setiyono, S., Keumala, D., Sabirin, A., & Nursantih, N. (2024). Has Indonesia safeguarded traditional cultural expressions? *Jambura Law Review*, 6(2), 206–239. <https://doi.org/10.33756/jlr.v6i2.24106>
- UNESCO. (2021). Recommendation on the ethics of artificial intelligence. <https://unesdoc.unesco.org/ark:/48223/pf0000381137>
- United Nations. (2007). United Nations Declaration on the Rights of Indigenous Peoples (A/RES/61/295). <https://docs.un.org/en/A/RES/61/295>
- U.S. Copyright Office. (2025). Copyright and artificial intelligence, Part 3: Generative AI training. <https://www.copyright.gov/ai/Copyright-and-Artificial-Intelligence-Part-3-Generative-AI-Training-Report-Pre-Publication-Version.pdf>
- World Intellectual Property Organization. (2024). The protection of traditional cultural expressions: Draft articles (WIPO/GRTKF/IC/49/5). https://www.wipo.int/edocs/mdocs/tk/en/wipo_grtkf_ic_49/wipo_grtkf_ic_49_5.pdf