Indonesian Health Justice Review
ISSN: XXXX-XXXXX (Online)
Vol. 3, No. 1, 2026, pp. 53-80
Published online since March 25, 2026

Patients or Data? Protecting Patient Rights in the Use of Electronic Medical Records for Artificial Intelligence Training

Ahmad Zulfikar Nuryanto*
Universitas Hasanuddin, Makassar, Indonesia
zulfikarnuryanto@students.unhas.ac.id

DOI: https://doi.org/10.65815/0bxad926

Submitted: 30-08-2025	Revised: 29-11-2025, 27-02-2026	Accepted: 10-03-2026

*Corresponding Author

Abstract

The development of artificial intelligence in healthcare increasingly depends on access to large volumes of clinical data, including electronic medical records. Although such data can support medical research and technological innovation, their secondary use for AI training raises fundamental questions concerning patient autonomy, privacy, consent, and control over health information. This article examines whether and under what circumstances electronic medical records may lawfully be used to train AI systems in Indonesia. Employing a normative juridical methodology, the study analyzes the legal status of health data, patient consent requirements, confidentiality obligations, and the principles governing secondary use of personal data. The study identifies a regulatory tension between the collective interest in advancing healthcare innovation and the individual patient's right to control sensitive health information. De-identification may reduce privacy risks but does not necessarily eliminate all legal and ethical concerns associated with data reuse. The article argues that the legitimacy of AI training using medical records should not depend solely on anonymization but should incorporate purpose limitation, proportionality, transparency, data governance, and appropriate forms of patient participation. A rights-based framework is therefore proposed to reconcile technological innovation with patient autonomy and health justice in Indonesia.

[*Pengembangan kecerdasan buatan (AI) dalam perawatan kesehatan semakin bergantung pada akses ke sejumlah besar data klinis, termasuk rekam medis elektronik. Meskipun data tersebut dapat mendukung penelitian medis dan inovasi teknologi, penggunaan sekundernya untuk pelatihan AI menimbulkan pertanyaan mendasar mengenai otonomi pasien, privasi, persetujuan, dan kendali atas informasi kesehatan. Artikel ini mengkaji apakah dan dalam keadaan apa rekam medis elektronik dapat digunakan secara sah untuk melatih sistem AI di Indonesia. Dengan menggunakan metodologi yuridis normatif, studi ini menganalisis status hukum data kesehatan, persyaratan*

persetujuan pasien, kewajiban kerahasiaan, dan prinsip-prinsip yang mengatur penggunaan sekunder data pribadi. Studi ini mengidentifikasi ketegangan regulasi antara kepentingan kolektif dalam memajukan inovasi perawatan kesehatan dan hak pasien individu untuk mengendalikan informasi kesehatan yang sensitif. De-identifikasi dapat mengurangi risiko privasi tetapi tidak selalu menghilangkan semua kekhawatiran hukum dan etika yang terkait dengan penggunaan kembali data. Artikel ini berpendapat bahwa legitimasi pelatihan AI menggunakan rekam medis tidak boleh hanya bergantung pada anonimisasi tetapi harus mencakup pembatasan tujuan, proporsionalitas, transparansi, tata kelola data, dan bentuk partisipasi pasien yang tepat. Oleh karena itu, kerangka kerja berbasis hak diusulkan untuk mendamaikan inovasi teknologi dengan otonomi pasien dan keadilan kesehatan di Indonesia.]

Keywords: Health data, artificial intelligence, electronic medical records, patient autonomy, data governance

Introduction

Artificial intelligence is increasingly transforming the production of knowledge in healthcare, but the technological capacity of AI is inseparable from the availability of data. Machine-learning systems require extensive datasets to identify statistical patterns, establish predictive relationships, evaluate performance, and improve their capacity to generalize across patient populations. Electronic medical records are particularly valuable because they contain longitudinal and multidimensional information, including diagnoses, laboratory findings, medication histories, clinical observations, treatment outcomes, demographic characteristics, and temporal relationships among health events. These characteristics make EMRs an attractive foundation for predictive models and clinical decision-support systems (Jiang et al. 2017). Yet the same features that make medical records scientifically valuable make them legally and ethically sensitive. Health information is capable of revealing intimate characteristics that individuals may reasonably expect to remain confidential. The use of such information for AI training therefore creates a tension between collective interests in medical innovation and individual interests in privacy and autonomy. The central question is consequently not whether medical data can generate social value, but whether that value justifies using information originally collected for patient care for purposes that patients may not have anticipated.

The concept of secondary use is essential to understanding this tension. Medical information is initially generated for a primary purpose, generally the diagnosis, treatment, monitoring, and administration of healthcare. Secondary use occurs when information is subsequently processed for another purpose, such as

epidemiological research, quality improvement, pharmaceutical development, health-system planning, commercial analytics, or AI model development. The distinction is legally significant because the fact that an individual has provided information for healthcare does not necessarily mean that the individual has agreed to every subsequent use of that information. Contemporary scholarship identifies purpose limitation as a central difficulty in secondary health-data governance because data initially collected in a clinical context may later become useful for purposes that were not fully foreseeable at the time of collection (Becker et al. 2023). AI intensifies this problem because machine-learning systems can generate new purposes through iterative experimentation, model development, and repurposing. Consequently, the legal analysis must address not only whether data are being reused but also whether the new use remains sufficiently connected to the circumstances under which the information was originally obtained.

The value of secondary health data for AI development should nevertheless not be underestimated. AI systems may assist in detecting disease, predicting deterioration, identifying patients at risk, improving resource allocation, supporting clinical decision-making, and discovering previously unrecognized relationships within complex datasets. Research using structured electronic health records has demonstrated the potential of secondary data to support clinical research and health-system improvement, while also identifying concerns regarding data quality, interoperability, and governance (Kushniruk and Borycki 2015; Hripcsak and Albers 2013). AI therefore creates a compelling public-interest argument for responsible access to medical information. Refusing all secondary use would potentially deprive society of significant opportunities to improve healthcare. However, the public benefit argument cannot automatically displace individual rights. Price and Cohen (2019) emphasize that the expansion of medical big data creates a legal and ethical problem precisely because innovation depends upon information that is deeply connected to individual privacy. The appropriate regulatory question is therefore one of proportionality: under what conditions can collective health benefits justify secondary processing while preserving meaningful individual rights?

Patient autonomy provides an important normative starting point. In conventional healthcare, autonomy requires that patients receive sufficient information to make meaningful decisions concerning their bodies and treatment. When medical information is subsequently used for AI development, however, the object of autonomy changes. The patient is no longer deciding whether to undergo a particular treatment but whether information generated through treatment may be repurposed for research, technological development, or commercial innovation. Vayena, Blasimme, and Cohen (2018) argue that machine learning in medicine creates distinctive ethical challenges concerning privacy, transparency, accountability, and control over data. The difficulty is that traditional consent models were primarily developed around identifiable interventions or research projects with relatively defined objectives. AI training may instead involve large

datasets, evolving research questions, multiple downstream users, and unpredictable future applications. This does not necessarily make consent impossible, but it requires reconsideration of what meaningful consent should mean in a data-intensive healthcare environment.

The problem becomes more acute when patients do not know that their medical records are being used for AI training. A patient may reasonably understand that information given to a hospital will be used by doctors and nurses involved in treatment. The patient may also accept certain administrative uses, such as billing or insurance processing, because these purposes are connected to healthcare delivery. Training an AI system, however, can be substantially different. The data may be copied into a research environment, transferred to an external technology company, combined with information from other institutions, or retained after the original clinical relationship ends. In commercial settings, the resulting model may generate significant economic value. Baric-Parker and Anderson (2020) highlight the ethical concerns arising when healthcare organizations share electronic health records with technology companies for AI development, particularly regarding privacy, consent, confidentiality, and corporate transparency. The distinction between clinical care and technological development therefore cannot be ignored simply because both activities may ultimately contribute to healthcare.

Public attitudes toward secondary health-data use further demonstrate that legitimacy depends on context. A systematic review of 47 studies involving more than 216,000 respondents found that people are generally supportive of secondary health-data use for research, particularly where data are de-identified, but preferences vary according to purpose, users, sensitivity, and consent mechanisms (Aitken et al. 2021). Another systematic review similarly found that privacy, trust, transparency, and the identity and motivation of data users strongly influence public attitudes toward health-data sharing (Aitken et al. 2020). These findings undermine two extreme assumptions. The first is that patients will always oppose secondary use. The second is that patients will automatically accept secondary use because it serves medical progress. In reality, willingness is conditional. Individuals may support data use for public-interest research while objecting to commercial exploitation, unrestricted sharing, or uses involving particularly sensitive information. Regulation should therefore preserve contextual distinctions rather than treating all secondary uses as legally interchangeable.

The problem is particularly significant for AI because the value of AI depends on scale, diversity, and longitudinal data. A model trained using only a small dataset may perform poorly when deployed across different populations or healthcare settings. Large and heterogeneous datasets can improve statistical robustness and help identify patterns that would not be visible in smaller samples. Yet the same scale creates greater privacy risks. Shilo, Rossman, and Segal (2020) describe healthcare big data as a rapidly expanding resource characterized by diverse sources and large-scale generation, while emphasizing the associated challenges concerning data quality, interpretation, privacy, and governance. The legal

problem is therefore structural: AI innovation rewards aggregation, whereas privacy law often seeks limitation. A regulatory model must reconcile these competing tendencies without reducing patient rights to an obstacle to technological development.

Indonesian law provides an important but incomplete framework for resolving this problem. Law Number 27 of 2022 concerning Personal Data Protection (PDP Law) classifies health and health-information data as specific personal data. This classification recognizes that health information requires stronger protection than ordinary identifying information. The PDP Law also establishes several possible legal bases for processing personal data, including valid explicit consent for specified purposes, contractual necessity, legal obligations, vital interests, public-interest tasks, and certain legitimate interests. (Republic of Indonesia 2022). Article 27 requires processing to be limited, specific, lawful, and transparent, while Article 28 requires processing to conform to the stated purpose. These provisions are highly relevant to AI training because they indicate that legality cannot be assessed merely by asking whether data have been removed of direct identifiers. The purpose and circumstances of processing remain legally significant.

The PDP Law is particularly important because it expressly identifies certain processing activities as potentially high risk. Article 34 requires a Personal Data Protection Impact Assessment where processing may create high risks to data subjects, including processing of specific personal data, large-scale processing, systematic evaluation or monitoring, data matching or combination, use of new technologies, automated decision-making with legal or significant effects, and processing that restricts the exercise of data-subject rights. AI training using large-scale electronic medical records may fall within several of these categories simultaneously. Health data are specific personal data; datasets may be processed on a large scale; AI involves new technologies; and datasets may be combined across sources. The statutory structure therefore supports the argument that AI training should receive enhanced governance rather than being treated as an ordinary data-processing activity. This provides a legal basis for moving beyond a simple consent-versus-no-consent debate toward a broader risk-based framework.

The health-sector framework reinforces these requirements. Minister of Health Regulation Number 24 of 2022 concerning Medical Records establishes electronic medical records as the contemporary model for medical-record administration and expressly emphasizes security and confidentiality. Law Number 17 of 2023 concerning Health similarly recognizes patient rights concerning health information and medical records while imposing obligations relating to confidentiality and security. (Republic of Indonesia 2023). These rules were primarily developed to govern healthcare information rather than AI model training. Nevertheless, their principles remain relevant because AI training represents an additional processing activity involving the same underlying records. The legal question is therefore whether the transition from clinical record to AI training environment respects the original confidentiality relationship and applicable data-

protection principles. This article argues that the answer cannot depend solely on technical anonymization. It requires an assessment of purpose, proportionality, transparency, governance, and patient participation.

The academic literature increasingly supports this broader view. Liaw et al. (2020), in recommendations developed by the Primary Care Informatics Working Group of the International Medical Informatics Association, argue that ethical governance should operate across the entire data lifecycle, including consent, access, sharing, data quality, extraction, transformation, and AI development. Similarly, Murdoch (2021) argues that AI creates new privacy challenges because private companies, healthcare institutions, and public bodies increasingly participate in obtaining, processing, and protecting patient health information. These arguments suggest that the relevant unit of analysis is not merely the dataset but the data lifecycle. Data collected in a hospital can move through extraction, de-identification, storage, annotation, model training, validation, deployment, and further reuse. Each stage may create different risks. A legal framework that approves the initial transfer without governing subsequent stages risks creating an accountability gap.

Against this background, the central research problem of this article is whether electronic medical records may legitimately be used to train AI systems in Indonesia when patients did not specifically anticipate or consent to that use at the time of treatment. The article addresses three related questions. First, what is the legal position of health data contained in electronic medical records under Indonesian law? Second, does de-identification or anonymization sufficiently resolve concerns regarding patient autonomy and secondary use? Third, what governance model can reconcile AI innovation with patient rights without making large-scale health research practically impossible? The article argues that neither unrestricted data use nor universal individual consent provides an adequate solution. Instead, Indonesian law should adopt a layered governance framework in which the required level of patient participation and institutional oversight depends upon the distance between the original clinical purpose and the proposed AI use, the sensitivity and scale of the data, the identity of the recipient, the possibility of re-identification, and the potential consequences for individuals and communities.

The study employs a normative juridical methodology using statutory, conceptual, and comparative approaches. Primary legal materials consist principally of Law Number 27 of 2022 concerning Personal Data Protection, Law Number 17 of 2023 concerning Health, and Minister of Health Regulation Number 24 of 2022 concerning Medical Records. Secondary materials include peer-reviewed journal articles concerning AI ethics, electronic health records, secondary health-data use, consent, privacy, data governance, and privacy-enhancing technologies. Comparative references are used to identify regulatory principles rather than to transplant foreign legal rules directly into Indonesian law. The analysis focuses on the relationship among autonomy, purpose limitation, proportionality, data minimization, transparency, security, and public interest. The

resulting framework treats patient participation as a spectrum rather than a binary requirement. Some uses may legitimately proceed under statutory or public-interest bases without individual consent, whereas high-risk commercial or substantially repurposed uses should require stronger forms of consent or participation.

Electronic Medical Records as Clinical Records and Sources of Secondary AI Data

Electronic medical records occupy a legally distinctive position because they simultaneously function as instruments of healthcare delivery and repositories of personal information. The medical record is generated through interactions between patients and healthcare professionals, but it contains information about the patient rather than merely information about the healthcare institution. The record may include highly intimate information concerning disease, mental health, reproductive health, genetic characteristics, medication use, treatment history, and family circumstances. Digitization does not change the sensitivity of this information; instead, it changes the scale and manner in which the information can be processed. Minister of Health Regulation Number 24 of 2022 recognizes this transformation by requiring electronic medical records and emphasizing security and confidentiality in their administration. The legal significance of the electronic format is therefore not merely technological. It creates an environment in which information can be copied, combined, analyzed, transferred, and reused at a scale that was difficult to achieve with paper records. AI training represents perhaps the clearest example of this transformation because it turns individual clinical information into computational resources for creating generalized models.

The distinction between primary and secondary use is therefore foundational. Primary use concerns purposes closely connected to the circumstances for which the information was initially collected, such as diagnosis, treatment, referral, medication management, and clinical documentation. Secondary use occurs when the same information is processed for a different purpose. The distinction is not necessarily absolute because healthcare institutions routinely use records for quality improvement, billing, public-health reporting, education, and research. Nevertheless, AI training can represent a substantial departure from ordinary clinical use, particularly when data are transferred to an external company or combined with datasets from unrelated institutions. Becker et al. (2023) demonstrate that the legal classification of secondary health-data use raises difficult questions concerning purpose limitation and further processing because the relationship between original and subsequent purposes is not always straightforward. The challenge is greater for AI because the ultimate application of a model may not be fully determined when training begins. The legal system must consequently examine not only the current purpose but also the foreseeable trajectory of data use.

AI training also differs from ordinary statistical analysis because the resulting model may itself become a new technological asset. A dataset can be used temporarily to answer a specific research question, while a trained AI model may be retained, commercialized, licensed, integrated into products, or further trained on additional datasets. The transformation of patient data into a model therefore creates questions concerning whether the patient's privacy interest disappears once raw data are no longer directly accessible. The answer should be negative. Privacy interests can continue because models may encode information derived from training data and may potentially reveal or reproduce characteristics of individual records under certain conditions. The legal issue is therefore not limited to whether a researcher can directly identify a patient from a database. It also concerns whether the broader processing ecosystem creates unreasonable risks to the patient. Mittelstadt et al. (2016) emphasize that algorithmic systems can create ethical problems concerning privacy, opacity, traceability, and autonomy even when individual decisions are not directly made by humans.

The classification of health information as specific personal data under the PDP Law reinforces the need for heightened protection. Article 4 explicitly includes health and health-information data within specific personal data. This means that hospitals and researchers cannot treat medical records in the same way as ordinary demographic datasets. The sensitivity of the data affects the proportionality of security, governance, and processing arrangements. A database containing thousands of medical records used to train an AI model may create risks not merely because it contains names or identity numbers but because the underlying information itself can be sensitive. Even if identifiers are removed, combinations of clinical characteristics can sometimes permit inference about individuals or small communities. Price and Cohen (2019) argue that medical big data challenges conventional assumptions about privacy because the increasing ability to combine datasets makes information more difficult to regard as permanently anonymous.

The legal consequence is that de-identification should be treated as a risk-reduction mechanism rather than an automatic termination of legal and ethical obligations. The nature of AI also creates a data-quality dimension that is directly relevant to patient rights. Medical records are not neutral representations of reality. They are produced within institutional practices and may contain missing values, inconsistent terminology, documentation errors, socioeconomic disparities in access to care, and differences in clinical practice. When such information is used to train AI, these characteristics can become embedded in algorithmic systems. Liaw et al. (2020) emphasize that data curation and quality assessment should be integrated into ethical governance because poor data management can generate downstream risks in AI systems.

This is significant for Indonesia because hospitals and healthcare facilities may have substantial differences in infrastructure, coding standards, documentation practices, and patient populations. If data from technologically advanced urban hospitals dominate training datasets, an AI system may perform differently for

populations whose health information is less represented. Thus, patient rights in AI training should include not only privacy but also protection against discriminatory consequences arising from unrepresentative data.

Consent, Autonomy, and the Limits of Anonymization

Consent is frequently presented as the most direct mechanism for protecting patient autonomy, but its application to AI training is more complicated than simply asking patients to sign a form. The PDP Law establishes several legal bases for processing personal data, and valid explicit consent is only one of them. Article 20 allows processing based on consent for one or several specific purposes communicated to the data subject, while also recognizing other legal bases such as legal obligations, vital interests, public-interest functions, and certain legitimate interests. (Republic of Indonesia 2022). This statutory structure means that Indonesian data protection law does not adopt an absolute consent model. Nevertheless, when consent is used as the legal basis, Article 22 requires that consent be given in accordance with statutory requirements, and Article 24 requires controllers to be able to demonstrate the consent obtained. (Republic of Indonesia 2022). For AI training, the crucial issue is therefore whether a patient's consent to healthcare or research can reasonably encompass subsequent AI development. The answer should depend on specificity, transparency, foreseeability, and the nature of the subsequent use.

A conventional consent form can be problematic because AI research often involves uncertain future purposes. Researchers may know that they intend to develop predictive models but may not know which variables will prove useful, which algorithms will perform best, or which future applications may emerge. This uncertainty has encouraged debates concerning broad consent, dynamic consent, and alternative governance mechanisms. Zenker et al. (2022) describe a national German approach to broad consent for secondary use of health data and biosamples, emphasizing that broad consent can be ethically and legally acceptable when supported by safeguards, independent review, transparency, and appropriate governance. Broad consent therefore demonstrates that consent need not always identify one narrow research question. However, broad consent is not equivalent to unlimited permission. Its legitimacy depends on institutional safeguards and meaningful information concerning the general categories of future use. The Indonesian context could therefore recognize carefully designed broad consent for defined categories of AI research while rejecting vague clauses that effectively permit unrestricted commercialization or transfer.

Dynamic consent offers a different model by allowing patients to exercise continuing control over data use. The approach can provide information about who is using data, for what purpose, and under what conditions, while permitting participants to modify their preferences. Kaye et al. (2015) proposed dynamic consent as a means of improving trust and communication in large-scale electronic health-record research. Later empirical research indicates that individuals

appreciate the possibility of managing changing preferences and receiving information about data use and research outcomes (Muller et al. 2023). Dynamic consent may be particularly useful for AI because data-processing purposes can evolve. However, it also has limitations. Requiring patients to respond to every new project may create participation fatigue and could disproportionately burden individuals with limited digital literacy or access. A legal framework should therefore treat dynamic consent as one possible mechanism rather than a universal requirement. Its appropriateness should depend on the risk and complexity of the proposed AI use.

The argument for consent becomes stronger as the secondary use becomes more distant from the original healthcare purpose. A hospital using de-identified records to evaluate whether a clinical pathway improves patient outcomes is conceptually closer to healthcare quality improvement than a technology company using identifiable or linkable records to develop a commercial AI product. Public-attitude research supports this contextual approach. Grande et al. (2013) found that individuals' preferences concerning secondary health-information use varied according to the purpose, the organization using the information, and the sensitivity of the data. Nakada et al. (2019) similarly found that Japanese respondents evaluated secondary uses differently depending on the perceived public benefit and the identity of the user, with greater sensitivity toward commercial applications.

These findings indicate that legal legitimacy should not depend merely on whether a use is labeled “research.” Commercial AI development may involve genuine medical benefits, but the commercial character of the activity can affect reasonable patient expectations and therefore should be transparent. The role of anonymization requires particular attention. Anonymization can substantially reduce the risk that a dataset will directly identify individuals, and it is therefore an important privacy-preserving technique. Yet anonymization should not be treated as a legal magic word. Modern data environments permit datasets to be combined with other information, potentially increasing the possibility of re-identification. Furthermore, even where re-identification is technically difficult, patients may object to the principle of unrestricted secondary use. The European legal literature distinguishes between data that are genuinely anonymous and data that are merely pseudonymized or de-identified but remain potentially linkable. Zenker et al. (2022) explicitly describe secondary use of de-identified but not necessarily anonymized health data as raising continuing legal and ethical questions.

This distinction is highly relevant to Indonesia. If identifiers are removed but the hospital retains a key capable of reconnecting records to individuals, the underlying information should continue to be treated as personal data. The obligations of the controller therefore cannot simply disappear. Anonymization also fails to resolve questions of collective privacy. Health datasets can reveal information about identifiable groups even when no particular individual can be identified. A dataset showing unusually high prevalence of a stigmatizing disease within a small community may affect the community's reputation or social standing.

Similarly, AI systems trained on datasets representing particular ethnic, socioeconomic, geographic, or disease populations may create inferences about those populations. Floridi et al. (2018) emphasize that AI governance should incorporate principles of beneficence, non-maleficence, autonomy, justice, and explicability rather than focusing exclusively on individual data protection.

This broader perspective is important for Indonesia because health-data governance can have distributive consequences. Data derived from vulnerable communities should not become a resource for technological development without considering who benefits and who bears the risks. The argument that patients should have no interest once data are anonymized is therefore normatively inadequate. Patient autonomy is not reducible to the ability to prevent one's name from appearing in a dataset. Autonomy also concerns the ability to understand and meaningfully influence how information generated through one's interaction with healthcare is used. Tosoni et al. (2021) found that patients' preferences concerning use of personal health information outside the circle of care were influenced by the nature of the proposed use, the identity of the recipient, and the safeguards applied.

The implication is that governance should provide patients with meaningful information even when individual consent is not legally required. Transparency can therefore serve as an intermediate mechanism between unrestricted institutional control and individualized authorization. This supports a model in which some low-risk secondary uses may proceed without individual opt-in while still requiring notice, public transparency, independent oversight, and meaningful objection mechanisms.

The concept of social license further strengthens this argument. A healthcare system may possess a formal legal basis to use data but still lose public trust if patients believe that information is being used secretly or primarily for private gain. Patient trust is particularly important because healthcare depends on individuals being willing to disclose sensitive information. If patients become concerned that information given to physicians may later be transferred to technology companies without meaningful safeguards, they may withhold information or avoid healthcare services. Such behavior could undermine both individual treatment and public-health objectives. A 2020 Nature Medicine editorial emphasized that secondary use of patient data can generate major benefits but depends upon maintaining patient trust and treating patients as partners in the research enterprise.

The legal significance of trust is therefore practical as well as ethical. A data-governance model that formally permits extensive secondary use but destroys public confidence may ultimately reduce the quality and availability of the data required for AI innovation. The following table summarizes the proposed differentiation between forms of AI-related secondary use. The table is not intended to create a new statutory classification but to provide an analytical framework for interpreting proportionality under existing Indonesian law. The underlying principle is that the greater the distance between the original clinical purpose and the proposed AI activity, the greater the justification for enhanced governance. The

framework considers purpose, identifiability, recipient, commercial orientation, risk, and appropriate patient participation. It therefore moves beyond the simplistic distinction between "consented" and "non-consented" data use.

TABLE 1. Proposed Risk-Based Classification of AI Training Using Electronic Medical Records in Indonesia

Category	Typical Purpose	Data Condition	Main Beneficiary	Proposed Governance	Patient Participation
Level I: Clinical-support use	Clinical decision support, internal quality improvement	Identifiable or coded	Patient and healthcare institution	Institutional authorization, security, access control	Notice generally sufficient where legally authorized
Level II: Public-interest research	Epidemiology, disease surveillance, health-system research	Preferably de-identified	Public/health system	Ethics review, data minimization, security, transparency	Opt-out or broad consent where feasible
Level III: Academic AI research	Development and validation of AI models for research	De-identified/ pseudonymized	Academic/research community	Ethics review, data-access committee, DPIA, controlled access	Broad or tiered consent preferred; alternatives require justification
Level IV: Commercial AI development	Development of proprietary AI products/services	Preferably de-identified; strict restrictions on linkage	Private/commercial actor	Enhanced contractual, technical, ethical, and regulatory safeguards	Explicit or appropriately granular consent should generally be preferred
Level V: High-risk AI development	AI involving sensitive inference, automated decisions, or highly consequential applications	Highly restricted	Mixed/private/public	DPIA, independent oversight, strict purpose limitation, auditability	Specific and enhanced participation/consent

Source: Various sources, analyzed by author

The Table 1 demonstrates that consent should not be treated as a binary legal switch. Some processing may legitimately rely on other legal bases recognized by the PDP Law, particularly where processing is necessary for public-interest functions or legal obligations. However, the absence of an individual-consent requirement does not imply the absence of governance. Article 34's requirement for impact assessment in high-risk processing is particularly significant because large-scale AI training involving health data may simultaneously involve specific data, new technologies, data combination, and large-scale processing.

The Table 1 therefore proposes graduated safeguards rather than universal opt-in consent. This approach attempts to preserve the benefits of health-data research while recognizing that commercial and high-risk uses warrant stronger patient participation. It also reflects empirical evidence showing that people generally support beneficial secondary research but place greater emphasis on transparency, control, and purpose when data move outside ordinary healthcare contexts (Aitken et al. 2021; Muller et al. 2023).

The framework further implies that consent should be designed according to the information necessary for meaningful choice. Patients should understand that their records may be used for AI training, the general category of AI application, whether the data will be transferred outside the healthcare institution, whether commercial entities will participate, whether data will be retained, and what safeguards will apply. Patients do not necessarily need to understand technical details such as neural-network architectures or optimization functions. However, they should understand the social and legal consequences of the proposed use. This distinction is important because overly technical consent forms can create the appearance of transparency without producing genuine understanding. Gerke, Minssen, and Cohen (2020) identify informed consent, transparency, data privacy, safety, and accountability as interconnected challenges in AI-driven healthcare. The appropriate consent model should therefore communicate material consequences rather than merely disclose technical terminology.

Purpose Limitation, Proportionality, and the Problem of AI's Unpredictable Future Uses

Purpose limitation presents one of the most difficult problems for AI training because machine learning thrives on discovering relationships that researchers may not have predicted in advance. Article 28 of the Indonesian PDP Law requires controllers to process personal data in accordance with the purpose of processing. At first glance, this requirement appears difficult to reconcile with exploratory machine learning, where researchers may begin with a broad objective such as improving disease prediction and subsequently discover unexpected correlations. Yet purpose limitation does not necessarily require researchers to know every future finding. Rather, it requires them to establish a sufficiently defined legitimate purpose and prevent uncontrolled repurposing beyond that purpose. Becker et al. (2023) demonstrate that similar questions arise under the GDPR because secondary use and further processing create difficult boundaries between legitimate scientific reuse and incompatible new purposes. Indonesia should therefore interpret purpose limitation flexibly enough to accommodate legitimate scientific discovery but strictly enough to prevent unrestricted exploitation.

A useful distinction can be made between purpose uncertainty and purpose expansion. Purpose uncertainty exists when researchers do not know what specific findings an AI model will generate but remain within a defined research domain.

Purpose expansion occurs when data originally collected for one domain are subsequently used for materially different activities. For example, using hospital records to develop an AI model for predicting hospital readmission may remain closely connected to healthcare research. Using the same records to develop an employee-surveillance system, targeted advertising platform, or unrelated consumer-profiling system would represent a much greater departure from the original context. The legal analysis should therefore ask whether the new use falls within a reasonably foreseeable and legitimate category rather than requiring precise prediction of every algorithmic output. This approach would preserve scientific flexibility while maintaining the normative boundary imposed by purpose limitation.

The proportionality principle provides a second mechanism for reconciling innovation and rights. Proportionality asks whether the proposed processing is suitable for achieving a legitimate purpose, necessary in light of available alternatives, and appropriately balanced against the rights and interests affected. This approach is particularly appropriate for AI because the same scientific objective may sometimes be achieved using less intrusive datasets. For instance, researchers may not require full longitudinal records if a limited set of variables is sufficient for a particular model. They may not require identifiable records if de-identified information is adequate. They may not require centralization if federated learning can produce equivalent results without transferring raw records. Proportionality therefore shifts the question from "Can the hospital technically provide the data?" to "How much data are actually necessary for the legitimate AI objective?"

Data minimization follows directly from this analysis. Although AI researchers often seek the largest possible dataset because more observations can improve model performance, collecting everything simply because it might become useful conflicts with a rights-based approach. The legal and ethical objective should be to identify the minimum information reasonably necessary for the defined model-development purpose. This does not mean that every variable must be justified individually, because some exploratory research legitimately requires broader datasets. It does mean that controllers and researchers should document why particular categories of information are included and whether less sensitive alternatives could achieve substantially similar objectives. Price and Cohen (2019) emphasize the importance of balancing innovation against privacy in medical big-data environments. Data minimization should therefore be understood as a governance principle rather than a rigid mathematical formula.

The PDP Law's impact-assessment requirement provides an especially important mechanism for operationalizing proportionality. Article 34 requires an assessment where processing creates potentially high risks, including processing specific data, large-scale processing, new technologies, data combination, and automated decision-making with significant consequences. AI training based on electronic medical records can trigger several of these factors simultaneously. A

Data Protection Impact Assessment should therefore identify what information will be processed, why it is needed, who will receive it, what risks of re-identification exist, whether the model can generate sensitive inferences, how long information will be retained, and what safeguards will be implemented. Such assessment should occur before data are transferred to the AI-development environment rather than after the model has already been trained. The preventive character of impact assessment makes it particularly valuable in protecting patient rights.

AI development also requires attention to downstream model use. A dataset may initially be processed for a low-risk research purpose but later become the foundation of a clinical decision-support system. The risk profile changes when an experimental model becomes operational technology. Patients may therefore be affected not only by the original data-processing activity but by the consequences of the model developed from their information. WHO's guidance emphasizes that AI for health should protect autonomy, promote safety, ensure transparency and explainability, establish responsibility, and promote equity. This lifecycle perspective suggests that governance should continue after model training. Researchers and developers should monitor whether models are used consistently with the original purpose, whether additional datasets are incorporated, whether commercial applications emerge, and whether significant risks arise during deployment. Consent and governance should therefore be understood as processes rather than single events.

The issue of secondary use becomes particularly complex where commercial entities participate. Commercial AI development can create substantial public benefits, including improved diagnostic tools, drug discovery, and personalized medicine. It can also generate private economic value from information originally provided by patients within a healthcare relationship. Baric-Parker and Anderson (2020) identify corporate transparency as one of the central concerns in patient-data sharing for AI development. The legal system should not assume that commercial involvement makes research illegitimate. Rather, commercial involvement should trigger greater transparency concerning the identity of the recipient, the intended use, intellectual-property arrangements, data retention, and potential commercialization. Patients should be able to understand whether their data are being used solely for academic research or are contributing to the development of proprietary products. This information can materially affect reasonable expectations and therefore should form part of meaningful participation.

Commercial benefit also raises the question of whether patients have a right to share in the economic value generated from their data. Indonesian law does not establish a general property right in personal data equivalent to ownership of physical property. Treating patients as conventional “owners” of every derivative model may therefore be conceptually problematic. Nevertheless, rejecting property language does not mean patients have no control interests. Data protection law protects rights, interests, and processing conditions rather than simply assigning ownership. The more appropriate framework is therefore one of data rights and

governance rather than absolute data ownership. A patient may not own an AI model merely because the model was trained using records concerning that patient, but the patient may possess legally protected interests concerning whether the underlying information could be processed, transferred, and repurposed. This distinction prevents both extremes: unrestricted institutional ownership and unrealistic individual property claims over complex derivative technologies.

Purpose limitation should also be interpreted in light of contextual integrity. Information disclosed within a clinical relationship carries contextual expectations. A patient may disclose information because it is necessary for diagnosis and treatment, not because the patient wishes to participate in commercial technology development. Helen Nissenbaum's theory of contextual integrity provides a useful conceptual lens for understanding why privacy can be violated even when information is not publicly disclosed: privacy depends on whether information flows appropriately within the norms of a particular context (Nissenbaum 2004). Although not an Indonesian legal doctrine, contextual integrity helps explain why secondary AI use can raise legitimate concerns even when direct identifiers are removed. The relevant issue is not simply secrecy but whether the movement of information from patient care to algorithm development is consistent with the expectations and norms governing the healthcare context.

Data Governance, Patient Participation, and a Rights-Based Indonesian Framework

A rights-based framework should begin by recognizing that patient participation can take multiple forms. Individual consent is one mechanism, but it is not the only mechanism through which patient interests can be protected. Transparency, independent ethics review, data-access committees, public consultation, opt-out mechanisms, auditability, security, contractual restrictions, and post-use accountability can also contribute to legitimate governance. Cumyn et al. (2021) found that citizens, researchers, and ethics-committee members hold diverse views concerning consent for secondary health-data use, demonstrating that no single consent model universally satisfies all stakeholders.

A sophisticated regulatory framework should therefore combine individual and institutional mechanisms. The objective is not to maximize the number of signatures obtained from patients but to create a governance system capable of ensuring that data use remains legitimate throughout its lifecycle. Consent should be treated as one component of accountability rather than a substitute for institutional governance.

The first institutional requirement should be data governance by design. Before an EMR dataset is transferred for AI training, the healthcare institution should establish clear rules concerning permissible purposes, access, retention, data quality, security, sharing, and deletion. Liaw et al. (2020) emphasize that ethical data management must extend throughout the data lifecycle, including extraction

and transformation processes that may otherwise be overlooked. This is especially relevant because the process of preparing medical records for AI often requires extraction, cleaning, labeling, normalization, and integration. Each stage can create opportunities for unauthorized disclosure or alteration. Data governance should therefore begin before the dataset leaves the clinical system. The hospital should maintain documentation showing how the dataset was created, what legal basis supports its use, who approved access, what transformations occurred, and what restrictions apply to downstream processing.

The second requirement should be independent review for high-risk AI data use. Research ethics committees have traditionally focused on research involving human subjects, but large-scale secondary use of existing medical records can create risks even when researchers never interact directly with patients. The risk arises from the information itself and the potential consequences of its processing. Zenker et al. (2022) identify independent review as an important safeguard accompanying broad consent for secondary health-data research. Indonesia could similarly require independent review for high-risk AI training involving large-scale health data. Such review could examine whether the proposed purpose is legitimate, whether the dataset is proportionate, whether de-identification is adequate, whether commercial interests are transparent, and whether patient participation is appropriate. The review should not function as a bureaucratic obstacle but as a mechanism for assessing risks that individual patients cannot reasonably evaluate themselves.

The third requirement should be transparency to patients and the public. Transparency is particularly important where individual consent is not obtained. Patients should have accessible information concerning whether their healthcare institution participates in AI research, what broad categories of data may be used, what categories of organizations may receive data, what safeguards exist, and how complaints can be submitted. Public transparency can also take the form of registries describing approved AI-data projects. Such registries would allow society to see how health data are being used without disclosing individual patient information. Public transparency is consistent with findings that trust depends strongly on understanding who uses health information and for what purposes (Aitken et al. 2020). Transparency is therefore not merely a communication strategy. It is an accountability mechanism that makes invisible data flows more visible and allows patients to evaluate whether institutional practices align with social expectations.

The fourth requirement should be controlled access rather than unrestricted data transfer. Where AI development requires external researchers or companies, institutions should provide only the access necessary for the approved purpose. Data-access committees can determine whether a requester meets defined requirements and whether the requested variables are proportionate. Technical measures should include role-based access, authentication, encryption, logging, secure research environments, and restrictions on downloading raw data. The

objective is to create a controlled research environment in which data can be analyzed without unnecessarily distributing copies. Privacy-preserving technologies can further reduce exposure. Salim and Park (2023), for example, describe a federated-learning approach that permits AI models to be trained across electronic health records while keeping raw data within participating institutions. Such technologies cannot replace legal governance, but they demonstrate that technological architecture can support the legal objective of minimizing unnecessary data movement.

Federated learning is particularly relevant to the Indonesian context because healthcare institutions may wish to collaborate without centralizing large volumes of sensitive medical information. Instead of transferring raw records to a central AI developer, hospitals can retain data locally and share model parameters or controlled computational outputs. This can reduce certain privacy risks, although it does not eliminate them. Federated systems can still face model-inversion attacks, membership-inference risks, malicious participants, or leakage through shared parameters. Consequently, federated learning should be treated as a privacy-enhancing technology rather than an automatic guarantee of compliance. Its deployment should be accompanied by security assessments, access controls, appropriate aggregation methods, and legal agreements. The broader principle is that data governance should encourage technological designs that reduce unnecessary exposure. Privacy law should therefore not merely regulate data after collection but should influence the architecture through which AI systems are developed.

Another relevant technology is differential privacy, which seeks to limit the ability to infer whether a particular individual contributed to a dataset or model. Differential privacy has been increasingly examined as a means of allowing useful statistical analysis while reducing individual disclosure risks. However, privacy-enhancing technologies involve trade-offs. Stronger privacy protection can reduce data utility, while excessive optimization for accuracy can increase disclosure risks. AI governance should therefore require institutions to document the balance between utility and privacy rather than assuming that one can be maximized without affecting the other. The principle of proportionality provides an appropriate legal framework for evaluating this trade-off. If a small reduction in model accuracy produces a substantial reduction in privacy risk, the privacy-preserving alternative may be preferable. The question should be whether the chosen technical architecture is reasonably necessary for the legitimate purpose, not whether it achieves the absolute highest predictive performance.

The sixth requirement should be data-quality and representativeness governance. Patient rights are threatened not only when information is disclosed but also when inaccurate or biased information is used to build systems that later affect patients. AI trained on incomplete or systematically unrepresentative medical records may generate discriminatory predictions. Obermeyer et al. (2019) demonstrated that an algorithm used in healthcare could reproduce racial

disparities because it predicted healthcare costs rather than underlying health needs. Although the case arose in the United States, it demonstrates a general principle: the choice of target variable and training data can reproduce existing structural inequalities. Indonesia should therefore require AI developers and healthcare institutions to evaluate whether training data adequately represent relevant populations. The rights-based approach should include a commitment to fairness because privacy protection is incomplete if data are securely used to produce discriminatory healthcare technologies.

Patient participation should also extend beyond consent to include benefit sharing and feedback. Patients may reasonably ask what happens after their information contributes to AI research. Do they receive information about research outcomes? Does the hospital benefit? Does the community benefit? Are successful AI products reinvested into healthcare? These questions need not imply a direct financial entitlement for every patient. Rather, they concern whether data governance produces reciprocal social benefits. Muller et al. (2023) found that participants valued communication and the return of meaningful research results in large-scale health-data reuse. This supports the development of a governance model in which patients are treated as partners rather than passive sources of data. Public reporting of research results, explanations of societal benefits, and institutional commitments to reinvestment may strengthen the legitimacy of secondary AI use.

The seventh requirement should be strict control over secondary commercialization. An institution that initially receives consent for academic research should not automatically assume that the resulting dataset may later be transferred to a private company for commercial AI development. Commercialization may represent a new purpose requiring reassessment. A broad consent model can potentially cover categories of commercial research if this is transparently explained and appropriately governed, but a vague authorization clause should not be sufficient. The distinction is important because patients may accept data use for public-interest medical research while objecting to unrestricted commercial exploitation. Nakada et al. (2019) found that public attitudes toward pharmaceutical use of patient records depend partly on perceptions of public benefit and the particular purposes involved. Indonesian regulation should similarly require disclosure of commercial involvement and prohibit institutions from using generic healthcare consent as a substitute for meaningful authorization of materially different commercial purposes.

The eighth requirement should be clear allocation of responsibility among institutions. The hospital, AI developer, academic researcher, cloud provider, and data processor may each participate in the data lifecycle. Contracts should specify which party determines the purpose, which party processes the data, which party has access, what security measures apply, how incidents are reported, and how data are deleted. However, contractual allocation should not deprive patients of their rights. The hospital should remain accountable for the legality of transferring

patient information into an external AI environment where it determines the purpose of the processing. Technology companies should be responsible for security and processing activities within their control. This reflects the accountability model of the PDP Law, which distinguishes controllers and processors while requiring appropriate governance. The institutional relationship should therefore be transparent enough that patients and regulators can identify where responsibility lies.

The ninth requirement should be retention and model governance. Data used to train AI systems should not necessarily be retained indefinitely merely because future research may become useful. Retention should correspond to legitimate research and legal requirements. At the same time, deleting raw data after training may not fully eliminate privacy risks because models can continue to embody information derived from training datasets. AI governance should therefore consider whether trained models can be audited, whether they may be further trained, whether their weights are transferred to third parties, and whether model outputs could reveal sensitive information. This is another reason why data protection should continue throughout the AI lifecycle. Gerke, Minssen, and Cohen (2020) argue that AI governance must address data protection alongside safety, liability, cybersecurity, and transparency. A model should therefore not be treated as legally irrelevant merely because the original medical records are no longer directly accessible.

The tenth requirement should be effective rights and remedies. Patients should have accessible mechanisms to ask whether their data are being used for secondary AI purposes, obtain relevant information, challenge unlawful processing, and complain about violations. The PDP Law provides a framework for data-subject rights, administrative sanctions, dispute resolution, and liability. (Republic of Indonesia 2022). Yet practical effectiveness depends on institutional procedures. Hospitals should designate responsible personnel, maintain data-processing records, provide accessible privacy notices, and establish complaint channels. Where patients are unable to understand how their information was used, formal legal rights may be ineffective in practice. Health institutions should therefore treat patient participation as part of service governance rather than merely as a legal document completed at admission.

The resulting Indonesian model can be described as Rights-Based Layered AI Data Governance. At its foundation are legality and purpose limitation. The second layer consists of proportionality, data minimization, and security. The third layer consists of institutional oversight, ethics review, impact assessment, and controlled access. The fourth layer consists of transparency and patient participation. The fifth layer consists of accountability, auditability, and remedies. The model is deliberately pluralistic because no single mechanism can adequately address the complexity of AI training. Consent protects autonomy but cannot guarantee security. Anonymization reduces identification risk but cannot resolve questions of purpose or commercial exploitation. Ethics review provides institutional scrutiny

but cannot replace patient rights. Technical safeguards reduce security risks but cannot determine whether a purpose is legitimate. Only an integrated governance model can address the full lifecycle.

TABLE 2. Rights-Based Layered Governance Model for AI Training Using Electronic Medical Records

Governance Layer	Core Principle	Main Legal/Ethical Function	Institutional Mechanism
Layer 1	Legality and purpose limitation	Establish lawful basis and legitimate purpose	Legal assessment and documented purpose
Layer 2	Necessity and proportionality	Prevent excessive data collection	Data minimization and variable justification
Layer 3	Privacy and security	Reduce unauthorized disclosure	De-identification, encryption, access control
Layer 4	Independent oversight	Evaluate high-risk processing	Ethics committee and DPIA
Layer 5	Transparency and participation	Protect autonomy and trust	Notice, broad/dynamic consent, opt-out where appropriate
Layer 6	Accountability	Ensure traceability	Audit logs, governance records, processor contracts
Layer 7	Remedy and benefit	Protect affected patients and society	Complaints, compensation, corrective action, public reporting

Source: Various sources, analyzed by author

The significance of this table is that patient rights are distributed across the entire data lifecycle rather than concentrated in the moment of consent. This is important because AI training creates risks that may emerge after data are initially collected. A patient may provide valid consent, but a later security failure may expose the data. Conversely, a legally permissible public-interest research project may not require individual consent, but the institution may still violate patient rights if it uses excessive data or transfers them to unauthorized commercial actors. The layered framework therefore avoids equating consent with legality. It treats consent as one mechanism within a broader system of accountability. This approach is consistent with WHO's position that AI governance should protect autonomy, safety, transparency, accountability, inclusiveness, and sustainability throughout the design and deployment process.

The framework also provides a basis for determining when individual consent should be strengthened. Explicit or granular consent should be preferred where AI training involves identifiable or readily linkable health information, substantial commercial interests, highly sensitive categories of data, significant downstream consequences, or purposes substantially removed from the original clinical context.

Conversely, where data are effectively de-identified, processing serves a legitimate public-interest research purpose, risks are low, independent review exists, and patients receive meaningful transparency and objection mechanisms, the law may reasonably permit processing on another legal basis where the statutory requirements are satisfied. This differentiated approach is more consistent with the PDP Law than a universal consent rule because the statute itself recognizes multiple legal bases. It also avoids making socially valuable health research impossible simply because obtaining individual consent from every historical patient would be impracticable.

A particularly important question is whether patients should possess an absolute right to withdraw their data after AI training has occurred. Dynamic consent models can provide mechanisms for withdrawal, but withdrawal becomes technically difficult once information has been integrated into a trained model or aggregated with other datasets. This difficulty should be disclosed in advance. A patient should not be promised a form of control that the technical system cannot realistically provide. At the same time, the inability to reverse every downstream processing step should not become an excuse for eliminating patient choice altogether. Governance should distinguish between withdrawal from future processing, deletion of identifiable records where legally and technically feasible, and withdrawal from already completed research. Clear communication about these distinctions is essential for meaningful consent. Otherwise, patients may believe that withdrawal guarantees complete erasure when the actual technical and scientific process cannot provide that result.

Finally, Indonesian AI governance should recognize that health data are a collective social resource without treating patients as resources to be exploited. Health systems depend on information generated by millions of individuals, and AI research may generate benefits that cannot be attributed to individual data contributions. At the same time, the social value of data does not eliminate individual rights. Мьller (2022) examines whether there may be a civic duty to support medical AI development by sharing electronic health records but recognizes the continuing importance of informed consent and competing values.

The more defensible position is therefore neither individual absolutism nor collective data collectivism. Patients should participate in a governance system that recognizes legitimate public interests in health innovation while imposing meaningful constraints on how their information is collected, combined, transferred, commercialized, and transformed into AI systems. This balance is central to health justice because the benefits of AI should be broadly distributed while the privacy risks should not disproportionately fall upon vulnerable populations.

Conclusion

The use of electronic medical records to train artificial intelligence systems presents one of the most significant governance challenges arising from digital healthcare. Medical records possess enormous scientific value because they contain longitudinal, multidimensional, and clinically meaningful information that can support predictive models, disease research, clinical decision support, and health-system innovation. Yet their value cannot be separated from the rights and interests of the individuals whose experiences generated the data. Indonesian law already establishes important protections. The PDP Law classifies health information as specific personal data, establishes lawful bases for processing, requires purpose limitation, recognizes accountability, and requires impact assessments for high-risk processing. The Health Law and Minister of Health Regulation Number 24 of 2022 further establish obligations concerning medical records, security, and confidentiality. The central legal challenge is therefore not whether Indonesian law protects health data, but whether existing principles are interpreted adequately when clinical information is repurposed for large-scale AI development.

This article has argued that neither universal individual consent nor anonymization alone provides an adequate solution. Consent remains an important expression of patient autonomy, particularly when AI training involves identifiable information, commercial actors, substantial repurposing, or high-risk applications. However, the PDP Law itself recognizes multiple legal bases for processing, meaning that some public-interest and legally authorized research may proceed without individual consent when appropriate safeguards exist. Similarly, de-identification can substantially reduce privacy risks but does not automatically eliminate legal and ethical concerns concerning purpose, governance, commercial use, collective privacy, or future processing. A rights-based framework should therefore combine legality, purpose limitation, proportionality, data minimization, privacy-enhancing technologies, independent oversight, transparency, patient participation, accountability, and effective remedies. This approach recognizes that privacy is not merely the absence of a name from a database; it is a continuing interest in how information flows through social, institutional, and technological contexts.

For Indonesia, the appropriate policy direction is a Rights-Based Layered AI Data Governance model. Under this framework, the degree of governance increases according to the sensitivity, scale, identifiability, commercial orientation, and potential consequences of the proposed AI use. Lower-risk public-interest research may rely on de-identified data, transparency, independent review, and proportionate participation mechanisms. Academic AI research should generally involve stronger ethics review, data-access controls, impact assessment, and broad or tiered consent where feasible. Commercial and high-risk AI development should receive enhanced scrutiny and, where appropriate, more specific patient authorization. Hospitals should remain accountable for establishing lawful and

secure data-governance systems, while AI developers and processors should bear responsibility for processing activities within their control. The ultimate objective should not be to choose between patients and data. Rather, Indonesian healthcare governance should recognize that data have value precisely because they originate from patients, and technological innovation is legitimate only when that value is pursued consistently with the dignity, autonomy, privacy, and interests of those patients.

References

Adamu, Jibril, Raseeda Hamzah, and Marshima Mohd Rosli. 2020. "Security Issues and Framework of Electronic Medical Record: A Review." *Bulletin of Electrical Engineering and Informatics* 9 (2): 565–572.

Aitken, Mark, et al. 2021. "A Systematic Literature Review of Attitudes towards Secondary Use and Sharing of Health Administrative and Clinical Trial Data: A Focus on Consent." *Systematic Reviews* 10: 132.

Aitken, Mark, Judith Cunningham-Burley, Sarah Pagliari, and Barbara L. M. M. "A Systematic Literature Review of Health Consumer Attitudes towards Secondary Use and Sharing of Health Administrative and Clinical Trial Data: A Focus on Privacy, Trust, and Transparency." *Systematic Reviews* 9 (2020).

Baric-Parker, Jean, and Emily E. Anderson. 2020. "Patient Data-Sharing for AI: Ethical Challenges, Catholic Solutions." *Linacre Quarterly* 87 (4). https://doi.org/10.1177/0024363920922690.

Becker, Regina, Davit Chokoshvili, Giovanni Comandй, Edward S. Dove, Alison Hall, Colin Mitchell, Fruzsyna Molnбr-Gбbor, Pilar Nicolбs, Sini Tervo, and Adrian Thorogood. 2023. "Secondary Use of Personal Health Data: When Is It 'Further Processing' under the GDPR, and What Are the Implications for Data Controllers?" *European Journal of Health Law* 30 (2): 129–157.

Cumyn, Annabelle, Roxanne Dault, Adrien Barton, Anne-Marie Cloutier, and Jean-Franзois Ethier. 2021. "Citizens, Research Ethics Committee Members and Researchers' Attitude Toward Information and Consent for the Secondary Use of Health Data: Implications for Research Within Learning Health Systems." *Journal of Empirical Research on Human Research Ethics* 16 (3): 165–178.

Floridi, Luciano, Josh Cowls, Monica Beltrametti, Raja Chatila, Patrice Chazerand, Virginia Dignum, Christoph Luetge, Robert Madelin, Ugo Pagallo, Francesca Rossi, Burkhard Schafer, and Effy Vayena. 2018. "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations." *Minds and Machines* 28: 689–707. https://doi.org/10.1007/s11023-018-9482-5.

Gariépy-Saper, Katherine, and Nicholas Decarie. 2021. "Privacy of Electronic Health Records: A Review of the Literature." *Journal of the Canadian Health Libraries Association* 42.

Gerke, Sara, Timo Minssen, and I. Glenn Cohen. 2020. "Ethical and Legal Challenges of Artificial Intelligence-Driven Healthcare." In *Artificial Intelligence in Healthcare*, 295–336. Academic Press.

Grande, David, Nandita Mitra, Anand Shah, Fei Wan, and David A. Asch. 2013. "Public Preferences About Secondary Uses of Electronic Health Information." *JAMA Internal Medicine* 173 (19): 1798–1806.

Hripcsak, George, and David J. Albers. 2013. "Next-Generation Phenotyping of Electronic Health Records." *Journal of the American Medical Informatics Association* 20 (1): 117–121.

Jiang, Fei, Yong Jiang, Hui Zhi, Yi Dong, Hao Li, Sufeng Ma, Yilong Wang, Qiang Dong, Haipeng Shen, and Yongjun Wang. 2017. "Artificial Intelligence in Healthcare: Past, Present and Future." *Stroke and Vascular Neurology* 2 (4): 230–243. https://doi.org/10.1136/svn-2017-000101.

Kaye, Jane, et al. 2015. "Dynamic Consent: A Patient Interface for Twenty-First Century Research Networks." *European Journal of Human Genetics* 23.

Keshta, Ismail, and Alaa Odeh. 2021. "Security and Privacy of Electronic Health Records: Concerns and Challenges." *Egyptian Informatics Journal* 22 (2): 177–183.

Kruse, Clemens Scott, R. Frederick, T. Jacobson, and D. Monticone. 2017. "Cybersecurity in Healthcare: A Systematic Review of Modern Threats and Trends." *Technology and Health Care* 25 (1): 1–10.

Kushniruk, Andre W., and Elizabeth M. Borycki. 2015. "Health Information Systems Research and the Future of Healthcare." *Healthcare Informatics Research*.

Liaw, Siaw-Teng, Harshana Liyanage, Craig Kuziemsky, Amanda L. Terry, Richard Schreiber, Jitendra Jonnagaddala, and Simon de Lusignan. 2020. "Ethical Use of Electronic Health Record Data and Artificial Intelligence: Recommendations of the Primary Care Informatics Working Group of the International Medical Informatics Association." *Yearbook of Medical Informatics* 29 (1): 51–57. https://doi.org/10.1055/s-0040-1701980.

Mittelstadt, Brent Daniel, Patrick Allo, Mariarosaria Taddeo, Sandra Wachter, and Luciano Floridi. 2016. "The Ethics of Algorithms: Mapping the Debate." *Big Data & Society* 3 (2). https://doi.org/10.1177/2053951716679679.

Moulaei, Khadijeh, Saeed Akhlaghpour, and Farhad Fatehi. 2025. "Patient Consent for the Secondary Use of Health Data in Artificial Intelligence (AI) Models: A Scoping Review." *International Journal of Medical Informatics* 198: 105872. https://doi.org/10.1016/j.ijmedinf.2025.105872.

Muller, Sam H. A., Ghislaine J. M. W. van Thiel, Menno Mostert, and Johannes J. M. van Delden. 2023. "Dynamic Consent, Communication and Return of Results in Large-Scale Health Data Reuse: Survey of Public Preferences." *Digital Health*. https://doi.org/10.1177/20552076231190997.

Müller, Sebastian. 2022. "Is There a Civic Duty to Support Medical AI Development by Sharing Electronic Health Records?" *BMC Medical Ethics* 23: 134. https://doi.org/10.1186/s12910-022-00871-z.

Murdoch, Blake. 2021. "Privacy and Artificial Intelligence: Challenges for Protecting Health Information in a New Era." *BMC Medical Ethics* 22.

Nakada, Haruka, Yusuke Inoue, Keiichiro Yamamoto, Kenji Matsui, Tsunakuni Ikka, and Shimon Tashiro. 2019. "Public Attitudes Toward the Secondary Uses of Patient Records for Pharmaceutical Companies' Activities in Japan." *Therapeutic Innovation & Regulatory Science* 54. https://doi.org/10.1177/2168479019872143.

Nissenbaum, Helen. 2004. "Privacy as Contextual Integrity." *Washington Law Review* 79 (1): 119–158.

Obermeyer, Ziad, Brian Powers, Christine Vogeli, and Sendhil Mullainathan. 2019. "Dissecting Racial Bias in an Algorithm Used to Manage the Health of Populations." *Science* 366 (6464): 447–453.

Price, W. Nicholson II, and I. Glenn Cohen. 2019. "Privacy in the Age of Medical Big Data." *Nature Medicine* 25: 37–43. https://doi.org/10.1038/s41591-018-0272-7.

Republic of Indonesia. 2022. Law Number 27 of 2022 concerning Personal Data Protection. State Gazette of the Republic of Indonesia 2022, No. 196.

Republic of Indonesia. 2022. Minister of Health Regulation Number 24 of 2022 concerning Medical Records.

Republic of Indonesia. 2023. Law Number 17 of 2023 concerning Health.

Salim, Mikail Mohammed, and Jong Hyuk Park. 2023. "Federated Learning-Based Secure Electronic Health Record Sharing Scheme in Medical Informatics." *IEEE Journal of Biomedical and Health Informatics* 27 (2): 617–624. https://doi.org/10.1109/JBHI.2022.3174823.

Shilo, Smadar, Hagai Rossman, and Eran Segal. 2020. "Axes of a Revolution: Challenges and Promises of Big Data in Healthcare." *Nature Medicine* 26: 29–38. https://doi.org/10.1038/s41591-019-0727-5.

Tosoni, Sarah, Indu Voruganti, Katherine Lajkosz, Flavio Habal, Patricia Murphy, Rebecca K. S. Wong, Donald Willison, Carl Virtanen, Ann Heesters, and Fei-Fei Liu. 2021. "The Use of Personal Health Information Outside the Circle of Care: Consent Preferences of Patients from an Academic Health Care Institution." *BMC Medical Ethics* 22: 29.

Vayena, Effy, Alessandro Blasimme, and I. Glenn Cohen. 2018. "Machine Learning in Medicine: Addressing Ethical Challenges." *PLoS Medicine* 15 (11): e1002689. https://doi.org/10.1371/journal.pmed.1002689.

Viberg Johansson, Jennifer, Heidi Beate Bentzen, and Deborah Mascalzoni. 2022. "What Ethical Approaches Are Used by Scientists When Sharing Health Data? An Interview Study." *BMC Medical Ethics* 23: 41.

World Health Organization. 2021. *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance.* Geneva: World Health Organization.

Zenker, Sven, Daniel Strech, Kristina Ihrig, Roland Jahns, Gabriele Мьller, Christoph Schickhardt, Georg Schmidt, Ronald Speer, Eva Winkler, Sebastian Graf von Kielmansegg, and Johannes Drepper. 2022. "Data Protection-Compliant Broad Consent for Secondary Use of Health Care Data and Human Biosamples for (Bio)Medical Research: Towards a New German National Standard." *Journal of Biomedical Informatics* 131: 104096. https://doi.org/10.1016/j.jbi.2022.104096.

***

Acknowledgment
None

Funding Information
None

Conflicting Interest Statement
The authors state that there is no conflict of interest in the publication of this article.

Publishing Ethical and Originality Statement
All authors declared that this work is original and has never been published in any form and in any media, nor is it under consideration for publication in any journal, and all sources cited in this work refer to the basic standards of scientific citation.

Generative AI Statement
N/A