FRAMEWORK FOR DIAGNOSTIC DATA MANAGEMENT IN INDONESIA

2025

Prepared by
PT Castellum Digital Indonesia

https://castellumdigital.id

Table of Content

List of Figures

List of Tables

List of Abbreviation

AI	Articial Intelligence
API	Application Programming Interface
ASIK	Aplikasi Sehat Indonesiaku
CKG	Cek Kesehatan Gratis (Free Health Check)
DICOM	Digital Imaging and Communications in Medicine
DHTS	Digital Health Transformation Strategy
EMR	Electronic Medical Record
FHIR	Fast Healthcare Interoperability Resources
FIND	Foundation for Innovative New Diagnostics
HIS	Health Information System
MoH	Ministry of Health
UU PDP	Undang-Undang Perlindungan Data Pribadi (Personal Data Protection Law)
Pusdatin	Pusat Data dan Teknologi Informasi (Center for Data and Information Technology)
SATUSEHAT	Indonesia's National Health Data Exchange Platform
WHO	World Health Organization

Executive Summary

Indonesia is advancing its digital health transformation through the Digital Health Transformation Strategy (DHTS) (Ministry of Health Indonesia, 2024) 2.0 (draft, unpublished), which envisions a secure and interoperable health data ecosystem. Within this transformation, health data are central to improving clinical accuracy, supporting disease surveillance, and enabling the responsible use of artificial intelligence (AI) in health services.

This Framework for AI-assisted Diagnostic Data Management provides structured guidance to strengthen governance, data quality, and ethical use across the diagnostic data lifecycle. It covers collection, cleaning and validation, annotation and structuring, storage and versioning, sharing and reuse, and continuous monitoring and feedback. The framework aligns with the Personal Data Protection Law (UU PDP 2022) (Republik Indonesia, 2022), the Ministry of Health's Data Governance Policy, and international standards such as the WHO Ethics and Governance of AI for Health (2021) (World Health Organization, 2021).

The Ministry of Health (MoH) leads policy direction, while Pusdatin manages infrastructure, interoperability, and monitoring through SATUSEHAT. Kesprimkom ensures program-level oversight for screening initiatives such as Cek Kesehatan Gratis (CKG). Health facilities and laboratories maintain data accuracy at the source, and system developers ensure compliance and version control.

A key use case demonstrates the integration of Large Language Models (LLMs) into the CKG program, which offers free health checks across CKG Ulang Tahun, CKG Sekolah, and CKG Komunitas. LLMs analyze structured and unstructured data to support doctors and health workers in identifying risks and providing follow-up recommendations.

By strengthening data governance and accountability, this framework supports Indonesia's commitment to ethical, inclusive, and data-driven digital health. It builds the foundation for reliable AI-assisted diagnostics and sustainable health system transformation.

Framework for AI-assisted Diagnostic Data Management in Indonesia

1. Background

Health data management is increasingly recognized as a critical enabler of effective and equitable healthcare, particularly as artificial intelligence (AI) technologies become more integrated into diagnostic processes. The reliability of AI-assisted diagnostics depends heavily on the quality, completeness, and governance of the data that supports them. Poorly curated or fragmented health data can lead to biased algorithms, inconsistent results, and limited scalability of digital health innovations. Beyond technical challenges, institutional and governance factors can further weaken data continuity and quality, limiting the safe and scalable use of AI-assisted diagnostics.

Health and diagnostic data in Indonesia remain fragmented across programs, facilities, and digital platforms, with laboratory, imaging, and screening data stored separately. This fragmentation results in siloed information flows and limits data interoperability, integration, reuse, and quality assurance, constraining the development of reliable AI-assisted diagnostics. It also undermines continuity of care, as clinicians cannot reliably access a complete diagnostic history across facilities and programs, affecting follow-up and clinical decision-making. The Ministry of Health has recognized this gap and placed health data governance as a key component within the Digital Health Transformation Strategy (DHTS) 2.0 Draft, as reflected in the Blueprint of Health Digital Transformation Strategy 2024 and the Health Sector Master Plan[1]. The DHTS 2.0 (DRAFT) outlines Indonesia's national direction for building an integrated, interoperable, and secure digital health ecosystem. It emphasizes data-driven decision-making, interoperability through the SATUSEHAT platform, and the responsible use of advanced technologies such as AI [2].

The development of a Health data Management Framework builds on this national agenda by providing a structured and context-specific approach to ensure that health data collected across Indonesia are reliable, standardized, and ethically managed. The framework aligns with the global principles of trustworthy AI outlined by the **World Health Organization (WHO)** in its Ethics and Governance of Artificial Intelligence for Health (2021), which emphasizes fairness, accountability, transparency, and inclusivity [5].

Indonesia's momentum in adopting AI for diagnostic, such as in radiology, pathology, and digital screening pilots, highlights both opportunities and risks. In the absence of coherent data management and governance standards, AI-assisted diagnostics face heightened risks of bias, poor reproducibility, weakened clinical trust, and constrained sustainability, ultimately limiting the safe and scalable adoption of AI-driven health innovations. Establishing a unified health

[1] The Digital Health Transformation Strategy (DHTS) 2.0 referenced in this section remains a draft document under development and has not been officially released or endorsed by the Ministry of Health.

data framework ensures that innovations are built on sound data governance foundations while maintaining public trust and ethical integrity.

While DHTS 2.0 (Draft) sets a strong national vision for integrated and interoperable digital health systems, further clarification and operational guidance are needed to consistently translate this vision into end-to-end diagnostic data governance practices across the health system. This framework also serves as a bridge between the national and international policy directions. It operationalizes national data regulations and AI ethics guidelines while adapting global best practices, including **FAIR (Findable, Accessible, Interoperable, Reusable)** and **CARE (Collective Benefit, Authority to Control, Responsibility, and Ethics)** principles. In doing so, it ensures that Indonesia's health data ecosystem supports not only technological advancement but also social accountability, inclusivity, and a long-term system resilience.

Objectives

The health data management framework aims to provide a comprehensive foundation for how health data are governed, standardized, and used in Indonesia's health system. It is designed to strengthen the country's digital health ecosystem by ensuring that AI-assisted diagnostic innovations are grounded in sound data management practices and ethical oversight.

The specific objectives of the framework are to:

- Align health data governance with national priorities under the Digital Health Transformation Strategy (DHTS) 2.0 Draft, ensuring coherence across regulatory, technical, and institutional frameworks.
- Establish guiding principles and operational standards for data collection, validation, integration, and sharing across laboratories, imaging systems, and health facilities in alignment with the national health system and the SATUSEHAT platform.
- Support the development and responsible use of AI-assisted diagnostics tools by enabling access to high-quality, representative, and ethically sourced datasets.
- Define roles and responsibilities of key stakeholders involved in health data management, including the Ministry of Health, regional authorities, health facilities, and development partners.
- Ensure compliance with legal and ethical frameworks such as Personal Data Protection Law (UU PDP 2022) and Ministry of Health data governance policies, promoting data privacy, accountability and security
- Enhance data use for decision-making and policy formulation through improved data quality, interoperability, and feedback mechanisms at all levels of the health system.
- Establish monitoring and evaluation mechanisms to assess the effectiveness, equity, and impact of health data use, including periodic reviews and stakeholder feedback loops.
- Strengthen cross-sector collaboration between public health, academia, private sector, and civil society to foster shared ownership and innovation in health data use.

Scope

The framework provides guidance for managing health data throughout its lifecycle, from collection and cleaning to annotation, integration, sharing, continuous monitoring and retention. It applies to both public and private health sector actors who contribute to, manage, or use health data within Indonesia's health system.

Table 1 Framework scope

Institutional Scope	**Data Scope**
• National and subnational health authorities, including the Ministry of Health, provincial, and district health offices • Public and private health facilities such as hospitals, laboratories, and primary health centers (Puskesmas) • Technology partners, academic institutions, and development organizations supporting health data systems and AI-assisted innovations	• Health datasets including the free health check related variables • Data collected through the national platform (SATUSEHAT ASIK CKG) • Metadata associated with health data, such as collection method, timestamps
Regulatory Scope	**Technical Scope**
• Aligned with the Personal Data Protection Law (UU PDP 2022), Minister of Health Regulation No. 24/2022 on Medical Records, and DHTS 2.0 (DRAFT) principles for interoperability and digital governance. • Guided by the Kominfo AI Ethics Guidelines (2025) (Kementerian Komunikasi dan Informatika, 2023)and international frameworks such as WHO AI Ethics Guidance (2021) and FAIR/CARE data principles.	• Covers system integration with national data exchanges, use of standardized terminologies (ICD-10/11, SNOMED CT, LOINC, DICOM), and compliance with data interoperability standards (HL7 FHIR). • Addresses both structured and unstructured health data used in AI-assisted analysis, decision support, and public health surveillance.
User Scope	
• **Clinicians (Doctors, Nurses, Lab Staff)**: Clinicians accurately record patient data, use their expertise to help label it for AI training (like marking X-rays), and then use the finished AI tools to help make faster, better diagnoses. • **Data Analysts & AI Developers**: The technical person cleans, secures, and anonymizes all the health data, then uses it to build, train, and test the AI models to ensure they are accurate and safe. • **Policy Makers & Health Administrators**: Sets the rules for data privacy and ethics, governs who can access the data, and uses the insights from the AI to plan national health strategies and improve public health.	

- **Patients & Caregivers**: Individuals receiving care and their authorized caregivers. They provide accurate information and consent preferences, can review/correct their records, and may contribute patient-reported outcomes for AI evaluation. They retain rights to privacy, informed consent, and transparency about AI use, with accessible channels to control data sharing.

The framework focuses on operationalizing existing national policies rather than introducing new systems. It serves as a practical guide for implementing responsible health data practices across Indonesia's evolving digital health landscape.

2. Local Context: Health Data and Digital Transformation in Indonesia

2.1 National Digital Health Transformation

Indonesia's health sector is undergoing a comprehensive transformation driven by the Digital Health Transformation Strategy (DHTS) 2.0 Draft *(draft, unpublished),* which serves as the national roadmap for achieving a connected, secure, and data-driven health ecosystem. The strategy outlines a phased approach to strengthening governance, infrastructure, and human resources to ensure that digital systems can deliver better health outcomes across all levels of care.

According to the DHTS 2.0 draft, the strategy focuses on five strategic pillars:

1. Governance and leadership, to ensure coordination, accountability, and alignment between national and subnational actors
2. Integrated and interoperable data systems, to enable secure and efficient data exchange across programs and facilities
3. Digital health infrastructure and standards, to establish foundational platforms and technical guidelines that support interoperability and data integrity
4. Human resources and institutional capacity, to build digital literacy and strengthen the skills of health workers, data managers, and policy makers.
5. Ethical and sustainable innovation, to promote responsible use of emerging technologies such as artificial intelligence (AI), big data analytics, and the Internet of Things (IoT) solutions

At the core of this transformation is SATUSEHAT, the national health data exchange platform managed by the Center for Data and Information Technology (Pusdatin) under the Ministry of Health (MoH). SATUSEHAT provides the interoperability backbone that connects electronic medical records, laboratory systems, and programmatic databases, enabling the secure and standardized flow of health data. It also supports patient-centric care by linking individual health records across facilities and services.

Complimentary digital tools like ASIK reinforce data capture and reporting at the community level for maternal and child health, tuberculosis, and other priority programs. Together, SATUSEHAT and ASIK operationalize the DHTS 2.0 draft vision of a unified health data ecosystem.

The draft DHTS 2.0 also emphasizes governance alignment with other national initiatives driven by the Ministry of Communication and Digital, National Cyber and Crypto Agency, and cross-sectoral partners. These alignments ensure that health data initiatives are in accordance with the national cybersecurity, privacy, and AI ethics regulations.

Importantly, the DHTS 2.0 draft positions diagnostic and clinical data as a strategic asset necessary to improve health outcomes and enable innovation. Further, it recognizes that there is a need for standardized data structure, legal charity around ownership and use of data, and

robust ethics for AI-assisted analytics. In this way, the strategy acknowledges that only using AI on top of solid, high-quality data will improve clinical decision-making, enable early disease detection, and improve service efficiency.

2.2 Current Challenges in Health Data Management

Though Indonesia has achieved significant success in digital transformation, the process of the implementation of the DHTS 2.0 (draft, unpublished) has identified a set of challenges related to the effective management of health and diagnostic information. This set of challenges can reportedly be classified as multidimensional in Figure 1.

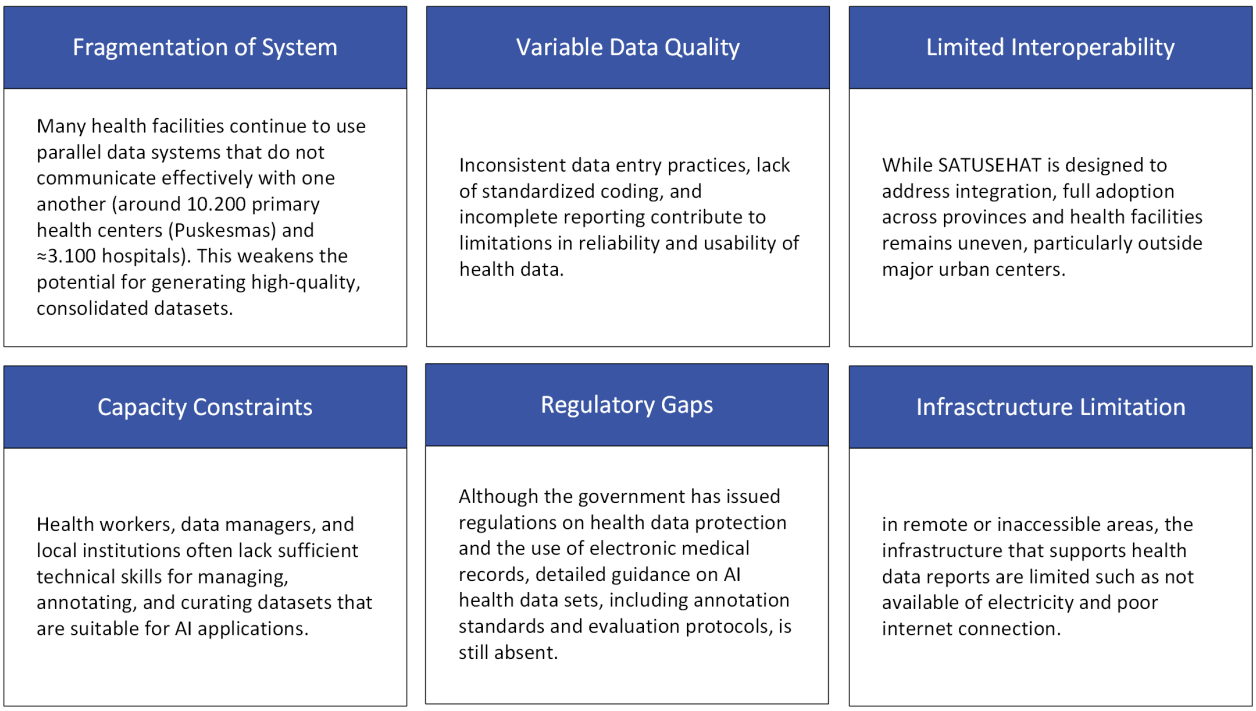

Figure 1 Current Challenges of Health Data Management in Indonesia

2.3 Opportunities for AI-assisted Diagnostics

The Indonesian health sector has shown increasing interest in AI applications, including symptom checkers, imaging analysis tools, and digital assistants for health workers. AI has the potential to support rapid diagnosis, triaging, and disease surveillance, particularly in resource-constrained areas. However, the success of AI tools depends on the availability of high-quality, contextually relevant data. Without standardized practices for collection, management, and governance of health data, the risks of algorithmic bias, inequitable access, and ineffective outcomes increase.

2.4 Institutional Stakeholders and Policy Landscape

Key actors in shaping health data governance shows in Table 1.

Table 1 Stakeholders and Policy Landscape

Stakeholder	Roles and Responsibilities
Ministry of Health (MoH)	Acts as the primary authority for national policy direction, coordination, and oversight of digital health implementation. Through its directorates working groups, MoH ensures policy coherence across DHTS 2.0 (DRAFT), SATUSEHAT, and AI-assisted diagnostic initiatives
Center for Data and Information Technology (Pusdatin)	Leads on the development and management of SATUSEHAT, the national health data exchange platform. Responsible for ensuring interoperability, version control, metadata management, and system integrity. Pusdatin provides national standards for data formats, coding, and exchange protocols, as well as technical oversight for secuirty and quality assurance.
Directorate of Primary Health and Community Communication (Kesprimkom)	Plays a key role in strengthening community-level health data systems and digital engagement. Oversees implementation of ASIK (Aplikasi Sehat IndonesiaKu) and other digital tools used for promotive and preventive health services, including Posyandu digitization and routine screening programs. Ensures that community-based diagnostic and screening data are captured, validated, and linked to SATUSEHAT for integration into national datasets.
Ministry of Communication and Digital Affairs (KOMDIGI)	Provides national direction and regulatory coordination for digital governance that underpins health data exchange. KOMDIGI supports cross-sector interoperability (including alignment with national digital identity and trust services), promotes a secure and responsible digital ecosystem, and helps ensure health data governance is consistent with broader national digital policies and standards.
Satu Data Indonesia (SDI) Ministries/Agencies (cross-sector stakeholders)	Serve as cross-government enablers for integrated data governance and interoperability. They provide foundational services and oversight such as population identity and civil registration, national coordination, cybersecurity and information assurance, health financing and claims data, budgeting and fiscal governance, and subnational administration so health datasets can be linked, protected, and responsibly shared in line with Satu Data Indonesia principles and mechanisms.
Provincial and district health offices (Dinas Kesehatan)	Translate national digital health and data governance policies into operational activities at the regional level. Responsible for data completeness, validation, and local system interoperability, as well as capacity building for health workers and data officers.
Health facilities	Serves as the main producers and custodians of health data. Responsible for accurate data collection, validation, and secure storage. Maintain version-controlled health datasets and ensure operational accountability for data integrity and privacy.
Private sector and technology providers	Support the design and integration of digital health solutions, including laboratory and AI-assisted diagnostic systems. Required

Stakeholder	Roles and Responsibilities
	to comply with national interoperability and data protection standards such as HL7 FHIR, DICOM, and UU PDP 2022.
Academic and research institution	Contribute to research, model development, and evidence-based evaluation of diagnostic and AI tools. Work under approved data-sharing agreements and ethical clearance to ensure responsible use of de-identified data.
Development partners	Provide technical assistance, capacity building, and funding to strengthen digital health infrastructure, interoperability, and health data management. Support implementation of DHTS 2.0 (DRAFT) at national and subnational levels.
Konsil Kesehatan Indonesia (KKI)	KKI provides the main regulatory guardrails for an AI-assisted health data framework. It ensures that the use of AI respects national standards for professional practice, patient safety, and accountability, and that digital tools never replace but support professional judgment. KKI's role is to make sure AI systems, documentation, and data use remain legally and ethically compliant.
Kolegium (Medicine, Nursing, etc.)	Kolegium sets the scientific and competency standards that shape how AI can be safely used in each profession. They define what data are clinically relevant, what level of accuracy and explainability AI tools need, and what skills health workers must have to interpret AI outputs. Their involvement ensures that the framework is clinically sound and aligned with each profession's scope of practice.
Professional Stakeholders (IDI, PPNI, IBI, etc.)	Professional organizations represent frontline practitioners who will use AI in daily work. They provide feedback on usability and workflow, advocate for fair responsibilities and safeguards, and help develop ethical guidance for AI use. They also support training and socialization, making the AI-assisted health data framework more acceptable, trusted, and workable in real clinical settings.
Ethical oversight and legal bodies	Oversee compliance with ethical, legal, and data protection frameworks. This includes institutional review boards, data protection authorities, and national ethics committees that ensure fairness, transparency, and accountability in AI and data use.
Patient/ Public	Act as data subjects and rights-holders whose trust and experiences shape legitimate health data and AI use. They provide consent, preferences, and feedback on how their data is collected, shared, and used, including in AI-assisted services. Through complaints about mechanisms, community fora, and patient organizations, they highlight risks, inequities, and unacceptable practices. Their engagement helps ensure that data governance and AI frameworks remain people-centered, transparent, and socially accountable.

These institutions play an important role in ensuring that any health data management framework aligns with both national health priorities and international standards.

3. Governing Principles

The governing principles of this data framework for AI-assisted diagnostics are aligned with Indonesia's DHTS 2.0 (DRAFT) draft and are designed to ensure data quality, security, ethical use, and long-term sustainability. These include the following principles:

- **Data Integrity & Storage:** Hybrid storage (local + SATUSEHAT cloud) for medical records, lab data, and screening results. Dataset versioning to support reproducibility. Pusdatin oversees metadata and version control, while Dinas Kesehatan conducts quality audits to ensure accountability and traceability.
- **Data Privacy & Governance:** Aligned with **UU PDP (2022)**, SATUSEHAT regulations, and Technical Guidelines for the Free Health Check Program (Juknis CKG). Consent mechanisms are integrated in SATUSEHAT Mobile and ASIK. De-identification applied for AI model training. Pusdatin and Komdigi jointly ensure compliance with data protection and secure sharing standards across systems.
- **Infrastructure Readiness:** Standardized equipment across facilities (newborn oximeter, Hb-meters, spirometry, USG, etc. per Juknis). Reliable connectivity for real-time reporting into SATUSEHAT.
- **Ethical Oversight:** Oversight by MoH, Dinas Kesehatan, and facility review boards to ensure inclusivity, especially for vulnerable groups (children, elderly, disabled).

Together, these principles establish a secure, compliant, and equitable environment for developing and deploying AI in diagnostics, especially in resource-limited settings as shown in Figure 2 as pillars of governing principles of AI-assisted diagnostic. In Indonesia context the current challenges and suggested actions about governing principle explain on Table 2.

Figure 2 Governing Principle of Data Framework for AI-assisted Diagnostic

Table 2 Pillar of Governing Principles

1. Data storage and versioning	**Current challenges:** 1. Fragmented and unconnected health data systems 2. No version control or data provenance tracking 3. Inconsistent use of data standards (DICOM, HL7, FHIR) 4. Limited infrastructure for storing large-scale imaging data 5. Lack of integrated and coordinated data backup between local systems and national servers.
	Suggested Actions: 1. Establish a national health data repository 2. Apply version-controlled and auditable data storage systems 3. Define and enforce national standards for health data formats and metadata 4. Adopt hybrid cloud architecture combining local and centralized storage 5. Establish a monitoring and alert system to track backup performance, version control, and synchronization status, coordinated by Pusdatin and verified by Kesprimkom MoH in collaboration with Dinas Kesehatan.
2. Data governance and privacy	**Current challenges:** 1. Uneven implementation of health data regulations 2. Lack of clear guidance for health data use in AI applications 3. Manual and fragmented consent and anonymization processes 4. Potential bias in datasets due to uneven data representation 5. Absence of a clear ethical framework tailored to the Indonesian digital health ecosystem

	Suggested Actions: 1. Develop a unified national framework for health data governance 2. Strengthen data privacy mechanisms by enforcing UU PDP (2022) and secure data sharing across national platforms such as SATUSEHAT and ASIK 3. Adapt and adopt ethical reference frameworks such as the WHO Guidance on Ethics and Governance of AI for Health (2021), contextualized to Indonesia's health data governance system. 4. Establish ethical review and quality assurance protocols for AI datasets 5. Implement digital consent management integrated with national platforms 6. Assign oversight responsibilities to Pusdatin for technical compliance, Kesprimkom for operational verification, and Komdigi for cross-sectoral coordination on privacy and AI ethics.
3. Infrastructure	**Current challenges:** 1. Unequal digital readiness across regions and facilities 2. Limited computing resources for AI model training and validation 3. Disconnected diagnostic systems without interoperability layers 4. High cybersecurity risks and insufficient protection mechanisms
	Suggested Actions: 1. Build a federated infrastructure model with regional data centers 2. Invest in AI-ready computing infrastructure at strategic hospitals or centers 3. Promote API-based architecture for system integration and scalability 4. Strengthen cybersecurity standards, audits, and response systems

4. Data Framework

A data framework refers to the structured set of standards, practices, and tools that govern how data is collected, stored, processed, and shared across its lifecycle. In the context of AI-assisted diagnostics, a robust data framework ensures that data used to train, validate, and deploy AI models is of high quality, ethically sourced, secure, and interoperable shows on Figure 3. This facilitates the development of reliable diagnostic systems that can be integrated into diverse healthcare settings, particularly in low-and middle-income countries.

The core objective of a data framework for AI diagnostics is to promote the trustworthy and efficient use of data to evaluate and monitor equitable and high-performing AI solutions in health. It acts as a guide to streamline data management across multiple stages including data acquisition, cleaning, annotation, integration, sharing, and ongoing monitoring. Importantly, it incorporates considerations of privacy, security, infrastructure, and ethical governance, which are essential in healthcare data ecosystems.

Figure 3 Data Framework for AI based Diagnostics

This framework is built around six interconnected pillars, each representing a critical stage in the AI data lifecycle. Each component is aligned with a comprehensive set of guiding principles and technical standards drawn from globally recognized data governance models, including the FAIR Principles, CARE Principles, WHO AI Ethics Guidelines, and national health data standards. The framework spans critical areas such as infrastructure, storage and versioning, data governance, and privacy, and integrates key technical processes like data curation, validation, with current best practices but also empowers practitioners and users to implement data equity seamlessly in their operations.

4.1. Data Collection

Data collection refers to the systematic process of gathering, demographic, diagnostic, and health-related information from individuals, laboratories, and health facilities for the purposes of screening, diagnosis, monitoring, and decision-making. In Indonesia, data collection forms the foundation of digital health transformation efforts led by the Ministry of Health through SATUSEHAT, which aims to unify data from multiple health service points, primary health centers (Puskesmas), hospitals, laboratories, and community-based programs.

For AI-assisted diagnostics, structured and high-quality data collection is crucial not only for clinical and public health use but also for developing and validating machine learning models. Data must be representative of Indonesia's diverse population, including various age groups, regions, and disease burdens, ensuring that AI tools are safe, equitable, and locally relevant.

In Indonesia, health data collection supports multiple life-stage screening packages as described in Ministry of Health technical guidelines (Juknis), including:

- newborn screening (thyroid, G6PD, adrenal, congenital heart disease, growth),
- child screening (tuberculosis, thalassemia, developmental milestones, dental and vision),
- adult screening (non-communicable disease risks, cancer, infectious diseases, and mental health), and
- elderly care (geriatric screening).

One of the flagship initiatives within these screening programs is the Cek Kesehatan Gratis (CKG), which promotes early detection and preventive health across various population groups. According to official Ministry of Health communications, the program is divided into three categories:

- CKG Ulang Tahun, which provides annual health checks aligned with an individual's birth month. This category focuses on the early detection of non-communicable diseases and key health risk factors.
- CKG Sekolah, targeting students aged 7 to 17 years across more than 282,000 educational units, including primary, junior, and senior secondary schools, madrasahs, and pesantren. The program aims to reach approximately 53 million students nationwide by 2025, providing physical examinations, nutritional assessments, vision and dental checks, and mental health screening.
- CKG Komunitas, which serves community groups in workplaces, public events, and health posts (Posyandu), provides free screenings for adults and the general population to strengthen early detection and continuous health monitoring.

All categories follow the Juknis CKG 2024 and relevant technical guidelines. Data are gathered following *Juknis* (technical guidelines) through a combination of anthropometric measurements, laboratory tests, and structured questionnaires collected via *SATUSEHAT Mobile (*for individual and facility-based health data entry), *ASIK (*for community-based and Puskesmas reporting), and digital chatbots such as WhatsApp-based reporting tools. These tools ensure consistency and traceability, allowing data to be captured at the point of service and automatically synchronized with SATUSEHAT for analysis and reporting.

Key Issues	Suggested Actions
- Fragmented data sources across programs and facilities that limit interoperability and reuse. - Incomplete and inconsistent reporting is caused by variation in data entry practices and non-standardized forms. - Unequal digital readiness across regions affects timely and reliable digital reporting. - Limited representation of certain populations, particularly rural, elderly, or marginalized groups, increases potential bias in AI applications. - Weak metadata and traceability because metadata such as time, source, location, or device ID are not consistently captured, reducing auditability and reliability for AI model development.	- Standardize collection tools and formats by applying nationally harmonized digital forms and coding systems such as ICD-10/11, SNOMED CT, and LOINC across SATUSEHAT and related systems. - Ensure inclusivity and representativeness by planning data collection to cover diverse demographic and geographic groups in all provinces. - Enhance digital entry mechanisms through wider use of SATUSEHAT Mobile, ASIK, and chatbots to capture data directly at the point of service while reducing manual input. - Strengthen metadata capture by recording key information for each diagnostic record including source, timestamp, and collection method to support audit trails and transparency. - Integrate data equity monitoring within Pusdatin dashboards to track coverage and completeness by region and population group. - Build capacity among health workers and data officers to use digital tools correctly, follow Juknis guidance, and maintain data quality assurance standards.

Figure 4 Key issues and suggestions for data collection

4.2. Data Cleaning & Validation

Data cleaning and validation refer to the systematic processes used to identify and correct inaccuracies, inconsistencies, and errors in health data collected from various sources. These processes ensure that the data used in AI systems, analytics, and policy decisions are accurate, complete, and standardized according to national and international reference standards.

In Indonesia, data cleaning and validation are critical for maintaining the integrity of health data collected through SATUSEHAT, ASIK, and related digital health applications, particularly within large-scale screening programs such as the Cek Kesehatan Gratis (CKG) initiative. The CKG program collects data from millions of individuals across schools, communities, and annual birthday-based screenings, which

are then used by doctors and health workers to inform diagnosis, follow-up actions, and public health planning. For AI-assisted diagnostics, these data also serve as training and validation inputs for machine learning models that support early disease detection and risk assessment.

Therefore, ensuring that the data is clean and validated is essential. If the data are incomplete, duplicated, or incorrectly coded, clinical decisions may be compromised, and AI model performance could become biased or unreliable. Clean and validated data enhance trust, accountability, and evidence-based decision-making, forming the foundation for the responsible use of AI in Indonesia's digital health ecosystem.

Ensuring data integrity requires coordinated roles across all levels of the health system.

Table 3 Stakeholders for data cleaning and their roles

Stakeholder	Role
Pusdatin	Leads the establishment of national validation standards, manages automated validation pipelines within SATUSEHAT, and conducts data quality audits.
Kesprimkom	Oversees program-level validation for community/screening initiatives (e.g., CKG); ensures data collection aligns with technical guidelines (Juknis).
Health Facilities & Puskesmas	Perform first-level verification at point of data entry.
District & Provincial Health Offices	Monitor data quality across facilities; provide feedback; coordinate corrective actions.
System Developers & Digital Platform Teams	Embed automated validation rules; maintain version control; ensure interoperability with the national data pipeline.
All Actors (Together)	Create a continuous quality assurance loop that sustains accuracy, completeness, and reliability of health data.

Key Issues	Suggested Actions
- Inconsistent validation mechanisms across facilities, with some relying on manual checks rather than automated systems. - Data entry errors, including incomplete records, duplicate submissions, and incorrect codes, which reduce reliability for analysis and AI use. - Limited use of automated data quality pipelines in regional systems, resulting in delays in error detection and correction. - Lack of feedback loops between data collectors and program managers to inform continuous improvement of data accuracy. - Uneven capacity among health workers to perform validation and quality assurance activities.	- Implement automated validation processes within SATUSEHAT and related systems to detect errors such as duplicates, outliers, and missing data in real time. - Integrate validation rules based on Juknis and WHO reference standards to ensure consistency in diagnostic coding and measurement ranges. - Provide visual dashboards and data quality reports to facilities to highlight completeness, timeliness, and accuracy indicators. - Establish feedback mechanisms so that health workers receive error summaries and correction requests directly from central or district data teams. - Conduct periodic data quality audits to identify systemic issues and track improvements over time. - Build capacity at the facility and district levels on data validation techniques, including the use of digital tools for automated checks and corrections. - Promote accountability by ensuring that each data collection point has designated personnel responsible for data quality assurance.

Figure 5 Key issues and suggestions for data cleaning and validation

4.3. Data Annotation & Structuring

Data annotation and structuring involve labeling, organizing, and formatting health data so that they are consistent, interoperable, and usable for analytical or AI purposes. This process ensures that the datasets are standardized across facilities and systems, allowing them to be integrated into national and international health data ecosystems.

In Indonesia, data annotation and structuring are central to enabling AI-assisted diagnostics that rely on well-labeled and coded data, such as medical images, laboratory test results, and questionnaire-based assessments. Proper annotation ensures that datasets are of high quality, consistent, and suitable for supervised machine learning, while also maintaining compatibility with the SATUSEHAT platform and other digital health systems. The use of standardized data models and terminologies allows Indonesia's health data to be interoperable across health programs and regions, improving traceability and enabling data reuse for research, clinical decision-making, and policy development.

Ensuring consistent annotation and structuring requires coordinated roles and clear accountability across multiple institutions.

Table 2 Stakeholders and responsibilities for data annotation and structuring

Party	Roles / Responsibilities
Ministry of Health	Sets policy direction and technical guidance via national standards and Juknis for data formats, coding, and reference values.
Pusdatin	Establishes/maintains interoperability standards (e.g., HL7 FHIR for diagnostics, DICOM for imaging); manages metadata registries and validation pipelines within SATUSEHAT.
Kesprimkom	Ensures annotation/structuring align with operational workflows in screening & diagnostic programs (e.g., CKG); oversees data consistency and program-specific reference ranges.
Health Facilities, Laboratories & Implementing Partners	Apply standards at data collection and annotation; ensure diagnostic results, imaging files, and questionnaires are accurately coded before submission.
System Developers & Digital Health Application Teams	Embed annotation standards; ensure interoperability with national APIs; support automated data structuring within their platforms.

Key Issues	Suggested Actions
- Limited adoption of standard data formats and controlled vocabularies across different diagnostic systems. Inconsistent labeling of laboratory results, imaging data, and screening questionnaires that reduces interoperability. Lack of national guidelines for annotation standards and reference thresholds used in AI model development. Variations in software tools and annotation quality among facilities and implementing partners. Insufficient capacity among data managers and health workers to apply standardized coding and annotation practices.	Use standardized data formats such as HL7 FHIR for health data exchange and DICOM for medical imaging. Apply globally recognized coding systems including ICD-10/11, SNOMED CT, and LOINC to ensure semantic consistency. Define national reference standards for diagnostic thresholds, such as Hb < 11 g/dL for anemia, and ensure their consistent application across all facilities. Provide annotation tools and training modules adapted to Indonesia's diagnostic workflows, including laboratory and imaging data annotation. Develop technical guidance under the Ministry of Health for consistent labeling and metadata recording to support AI model training. Encourage cross-program coordination between SATUSEHAT, ASIK, and other digital systems to align annotation and data structuring protocols. Establish quality control processes to review and verify annotated datasets before use in AI development or national reporting.

Figure 6 Key issues and suggestions for data annotation and structuring

4.4. Data Integration & Harmonization

Data integration and harmonization refer to the process of combining health data from multiple systems and sources into a unified and consistent structure. The objective is to ensure that information from laboratories, imaging systems, electronic medical records, and community programs can be accessed, compared, and analyzed together.

In Indonesia, integration and harmonization are critical steps to achieve the vision of SATUSEHAT as a national health data exchange. Health data, including laboratory results, imaging studies, and screening information, are often collected in different formats and stored in separate systems. Harmonizing these data ensures that each

patient's information is connected across services, and that datasets can support both clinical and AI-assisted applications.

A harmonized system allows health data to flow securely across facilities and administrative levels, improves interoperability between public and private providers, and enables the development of comprehensive datasets for analysis, prediction, and health planning.

Key Issues	Suggested Actions
- Fragmented diagnostic systems that store data separately and lack integration pathways. - Limited interoperability between laboratory information systems, radiology systems, and electronic medical records. - Absence of common identifiers or patient indexing mechanisms that prevent accurate record linkage. - Technical challenges in aligning different data formats, coding standards, and metadata structures. - Insufficient coordination between national and local systems, which leads to data duplication or data loss.	- Promote the use of open standards such as HL7 FHIR, DICOM, and API-based communication for all health data systems to ensure interoperability. - Strengthen the role of SATUSEHAT as the main integration layer by ensuring that laboratory and imaging data are automatically linked with electronic medical records. -Develop and maintain a national patient identifier system to support longitudinal data tracking and prevent duplication. -Create a unified data mapping and harmonization protocol to align diagnostic codes, variable names, and metadata structures across platforms. -Encourage regional data centers or federated data nodes to facilitate integration while maintaining data ownership at the local level. - Provide technical guidance and training for IT staff and data managers to implement data harmonization tools and methods. - Establish monitoring mechanisms under Pusdatin to regularly assess interoperability performance, data completeness, and accuracy across connected systems.

Figure 7 Key issues and suggestions for data integration and harmonization

4.5. Data Sharing & Reuse

Data sharing and reuse refer to the responsible and secure exchange of health datasets for secondary purposes such as research, model development, public health monitoring, and policy formulation. Effective data sharing ensures that collected information continues to provide value beyond its original use while maintaining privacy, consent, and data protection standards.

In Indonesia, data sharing and reuse are essential to accelerate innovation, promote collaboration, and strengthen the use of AI-assisted diagnostic solutions. By enabling controlled access to de-identified datasets through platforms such as SATUSEHAT, the Ministry of Health and its partners can facilitate evidence-based policymaking, clinical improvements, and AI model development. Responsible reuse also supports transparency and reproducibility in health research, ensuring that health data contribute to long-term national health goals.

The implementation of these processes involves shared institutional responsibilities. The Ministry of Health (MoH) provides national policy direction and issues of regulations that define ethical, legal, and procedural requirements for secondary data use. To address patient consent for data reuse, healthcare facilities and other data controllers should explicitly include clear clauses in consent forms explaining that de-identified patient data may be used for secondary purposes such as machine learning, AI development, and health research. Patients should be informed—in simple, accessible language—about what types of secondary use are possible, how their privacy will be protected, and who may access the data. They must have the right to decline or limit secondary use without any impact on the quality of care they receive, as well as a practical way to opt out later if they change their mind. Embedding these options into standard consent workflows and digital registration processes ensures that data reuse for AI remains transparent, voluntary, and aligned with patients' rights.

Table 3 Stakeholders and responsibilities for data sharing and use

Party	Roles / Responsibilities
Pusdatin	Manages technical infrastructure and secure data-sharing environments within SATUSEHAT; enforces version control, de-identification, and verified credentialed access.
Kesprimkom	Ensures data from community/screening programs (e.g., CKG) meet programmatic and ethical standards before reuse.
Health Facilities, Laboratories & Implementing Partners	Ensure accuracy and de-identification at the point of submission.
Research Institutions, Universities & Private Developers	Comply with data use agreements; obtain ethical clearance prior to accessing datasets.

A well-governed sharing mechanism ensures that sensitive health information remains protected, access is role-based, and data use complies with national regulations such as the Personal Data Protection Law (UU PDP 2022) and Ministry of Health data governance policies.

Key Issues	Suggested Actions
- Limited availability of clear national policies and licensing frameworks for the secondary use of health data. - Inconsistent application of de-identification and anonymization processes before data are shared. - Lack of centralized repositories or version-controlled systems for dataset access and tracking. - Unclear procedures for researchers and developers to request and obtain access to de-identified data. - Concerns over data misuse, ownership disputes, and lack of trust between data producers and users.	- Develop a national data sharing and reuse policy under the Ministry of Health that defines governance structures, licensing models, and ethical oversight. - Ensure that all shared datasets are de-identified or anonymized according to national and WHO-recommended standards. - Create centralized or federated repositories within SATUSEHAT to manage dataset versions and enable secure access through role-based permissions. - Establish clear procedures for data access requests, including ethics review and data use agreements for researchers and developers. - Apply the FAIR (Findable, Accessible, Interoperable, and Reusable) principles to guide dataset documentation, discoverability, and transparency. - Introduce open or restricted-use licensing models that specify conditions of use, attribution requirements, and data ownership responsibilities. - Promote collaboration between government, academia, and industry to encourage responsible data reuse for AI development and public health innovation. - Build public awareness and institutional capacity on data protection, responsible reuse, and open science practices.

Figure 8 Key issues and suggestions for data sharing and use

4.6. Continuous Monitoring & Feedback

Continuous monitoring and feedback ensure that health data and AI tools remain accurate, safe, and relevant over time. In Indonesia, these processes help maintain data integrity, improve diagnostic accuracy, and enable responsible AI use within the national digital health ecosystem.

In the Indonesian context, responsibility for monitoring AI solutions is shared across multiple actors: the Ministry of Health (especially units overseeing SATUSEHAT and medical devices/Digital Health), health facility management and clinical leaders who

oversee how tools are used in practice, and AI solution providers who must track model performance, safety, and compliance over time. Together, they are expected to implement routine audits, performance reviews, and incident reporting, with support from data managers, professional associations, and academic or independent evaluators to provide external validation.

Monitoring focuses on tracking data completeness, timeliness, accuracy, and interoperability across facilities and programs. For AI-assisted diagnostics, it also detects data drift, bias, and performance degradation to maintain model reliability. Feedback loops connect health workers, data managers, and administrators, supporting corrective actions, system updates, and improved decision-making.

Table 4 Stakeholders and responsibilities for continuous monitoring and feedback

Party	Roles / Responsibilities
Ministry of Health (MoH)	Defines policies and indicators for national data quality and digital performance.
Pusdatin	Technical lead managing SATUSEHAT monitoring infrastructure; ensures real-time tracking, interoperability, and operational continuity; maintains national dashboards and validation pipelines for central/regional decisions.
Kesprimkom	Oversees program-specific monitoring (e.g., CKG); ensures screening data completeness and verification per technical guidelines.
Health Facilities & Puskesmas	Perform first-line monitoring; review local reports; provide feedback to district/provincial offices.
Developers & AI Solution Providers	Track model performance internally; report results to Pusdatin for validation and calibration.

Through SATUSEHAT, continuous monitoring and feedback enhance accountability, transparency, and trust. These mechanisms ensure that health data and AI models remain reliable and aligned with evolving health priorities.

Key Issues	Suggested Actions
- Limited capacity for real-time monitoring of health data completeness, timeliness, and accuracy at facility and district levels. - Lack of automated systems to detect data drift and AI model performance changes after deployment. - Weak feedback channels between central authorities and frontline health facilities for continuous improvement. - Absence of standardized indicators for evaluating data quality and AI system reliability. - Limited integration of monitoring results into national dashboards and policy reviews.	- Establish real-time monitoring dashboards within SATUSEHAT and Pusdatin to track data quality indicators such as completeness, consistency, and timeliness. - Implement automated alerts to detect irregularities or performance drift in AI diagnostic systems. - Develop standardized metrics and reporting templates for evaluating the accuracy, fairness, and reliability of AI models in clinical use. - Create structured feedback loops between health facilities, district offices, and the Ministry of Health to review data quality and system performance on a regular basis. - Provide periodic feedback reports to data collectors and managers to inform training needs and recognize quality improvements. - Integrate monitoring results into national health reviews and strategic planning processes to inform policy adjustments and digital health investments. - Conduct regular evaluations of AI models using updated datasets to ensure continued relevance and ethical performance. - Build institutional capacity for monitoring and evaluation, including training on the use of digital dashboards, analytics tools, and performance reporting.

Figure 9 Key issues and suggestions for continuous monitoring and feedback

5. Use Case: LLM-based Health Check for CKG

The use case of the LLM-based Health Check (CKG) demonstrates how large language model (LLM) technology can support promotive and preventive health services in Indonesia by analyzing self-reported data and screening results. The objective is to enable early detection of potential health risks, provide personalized feedback, and assist frontline health workers in follow-up actions. This use case aligns with the Ministry of Health's agenda to strengthen primary healthcare and digital transformation under the SATUSEHAT initiative.

The CKG use case focuses on creating an accessible digital service that helps individuals and families conduct initial health assessments using structured inputs such as questionnaire responses, anthropometric data, and basic laboratory or screening results. The integration of LLM technology allows the system to interpret both structured and unstructured information, categorize risk levels, and generate follow-up recommendations that are aligned with clinical protocols and Juknis guidance.

Objective
Support promotive and preventive services by analyzing patient inputs (symptoms, questionnaire responses, screening results) through LLM models.

Workflow

1. Patient registers via ASIK CKG (SatuSehat)
2. Inputs self-reported data (questionnaires, symptoms).
3. LLM processes structured and unstructured inputs, providing risk categorization and follow-up recommendations.
4. Data stored in SATUSEHAT

Data Standards
Structured in FHIR; annotated with ICD/SNOMED codes.

Benefits
Early detection, equitable access, reduced burden on Primary Health Care Facility (FKTP), patient empowerment.

Risks & Mitigations
Bias (ensure diverse datasets), privacy (strict de-identification), reliability (human oversight for final decisions).

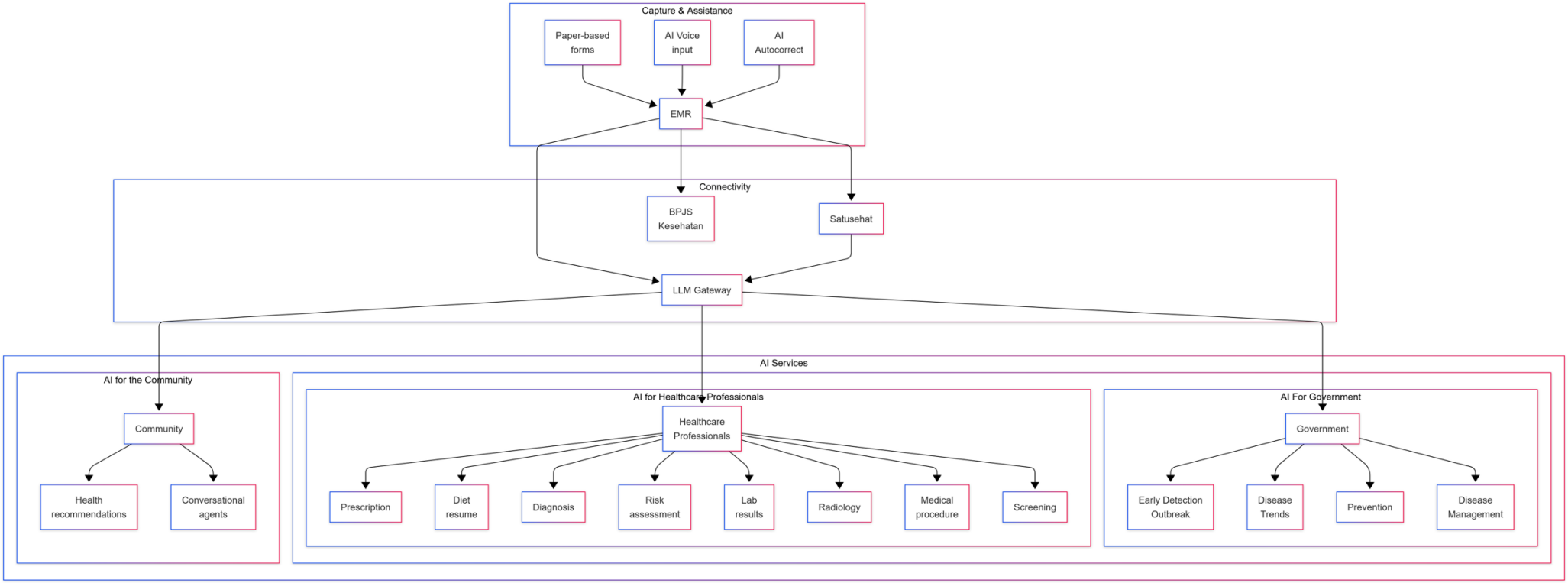

Figure 10 LLM workflows for free health check

5.1. Workflow Overview for Free Health Check

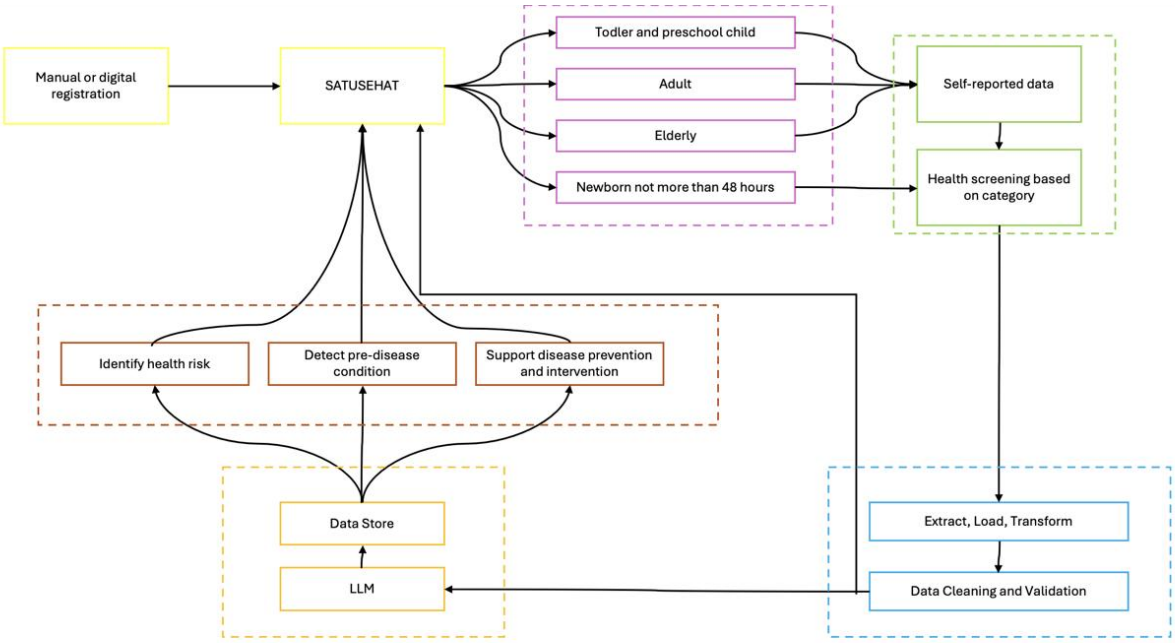

Figure 11 Overview for free health check patient

Figures 11 show that the workflow of LLM can support promotive and preventive health services through free health checks. The process begins on the SATUSEHAT platform, where users are grouped by age (like Newborn or Adult). Most users (except newborns) provide Self-reported data, which triggers specific Health Screening. This screening data is then processed, first using ETL (Extract, Load, Transform) to format it, followed by Data Cleaning and Validation to ensure quality. The clean data is then fed into the system's core, an LLM, which uses a Data store to conduct its analysis. Finally, the LLM's analysis generates the key outcomes—to Identify health risk, detect pre-disease conditions, and Support disease prevention—with these results being sent back to the user on the SATUSEHAT platform.

5.2. Workflow Overview: Calibration Process for Deployment of LLM for Free Health Check

The Calibration workflow explains how the AI-Assisted Diagnostic Data Framework converts raw model outputs into clinically interpretable and trustworthy risk categories. It describes how thresholds are selected and adjusted for different conditions and population segments, how performance is evaluated using calibration metrics, and how feedback from real-world use and clinicians is fed back into the system. The following section describes the calibration process in detail, including each activity and decision point involved.

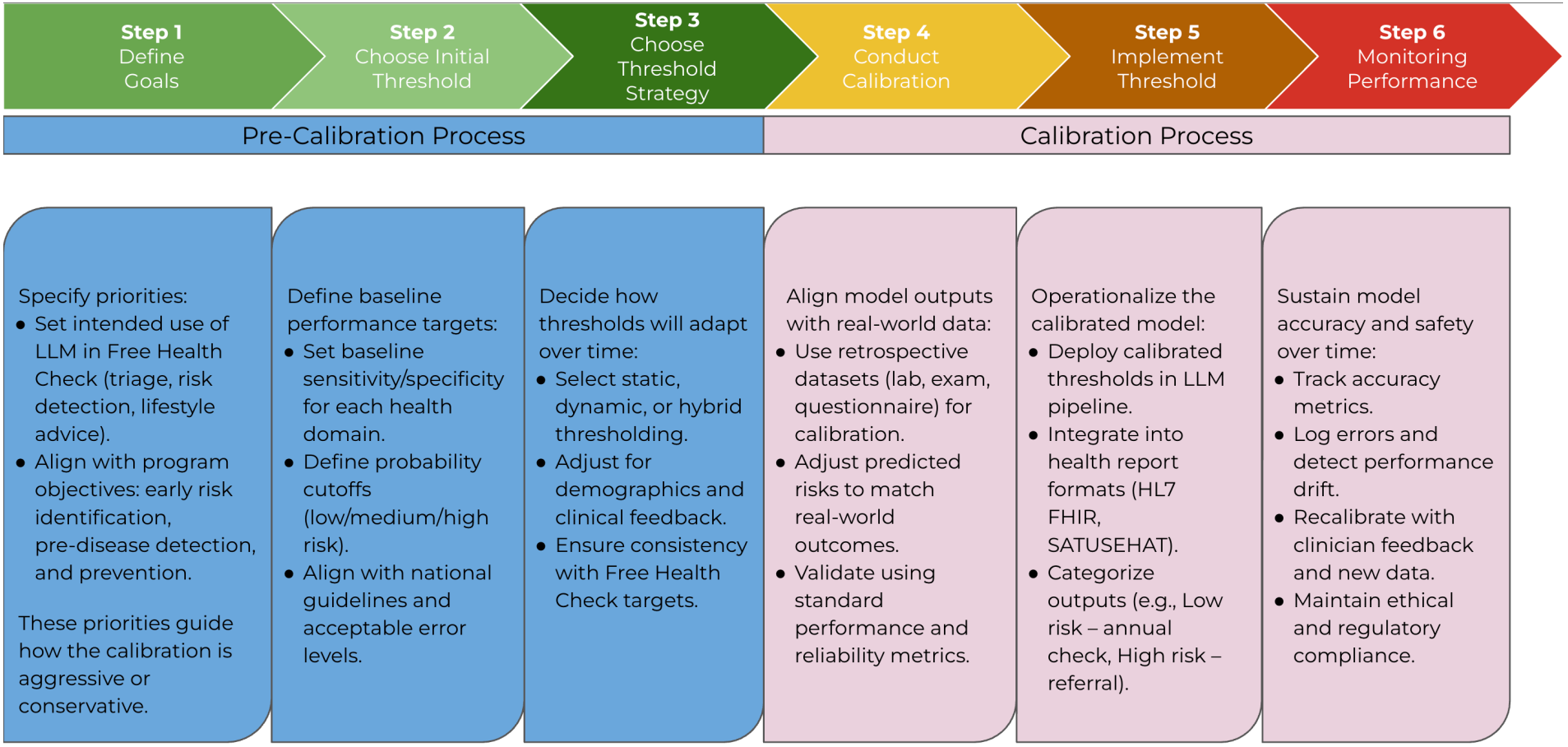

Figure 12 Pre-calibration and calibration process

Step 1: Define Goals

- Establish intended use of the LLM in Free Health Check workflows (e.g., triage, risk factor identification, lifestyle recommendation, referral flagging).
- Align with clinical objectives of the Free Health Check program:
    - Identify health risks early
    - Detect pre-disease conditions
    - Support disease prevention and timely intervention

Step 2: Choose Initial Threshold

- Select baseline sensitivity/specificity trade-offs for different health domains (cardiovascular, metabolic, respiratory, mental health).
- Define decision boundaries (e.g., probability scores for risk classification: low, medium, high).
- Consider national guidelines, epidemiological data, and acceptable false positive/negative levels.

Step 3: Choose Threshold Strategy

- Strategy options:
    - **Static thresholds** (fixed probability cutoffs for disease flags).
    - **Dynamic thresholds** (adaptive to population characteristics, age group, comorbidities).
    - **Hybrid thresholds** (risk scores adjusted by clinician feedback or historical data).
- Incorporate stratification by demographic group (e.g., infants, adults, elderly) consistent with Free Health Check targets

Step 4: Conduct Calibration

- Use representative datasets from Free Health Check programs, including laboratory results, anthropometric data, and questionnaire records, to calibrate the model's risk estimates.
- Adjust probability outputs so that predicted risks correspond accurately to observed outcomes.
- Validate calibration quality using established performance and reliability metrics.

Step 5: Implement Threshold

- Deploy calibrated thresholds into LLM pipeline for Free Health Check:
    - Risk classification embedded in structured reports.
    - Outputs mapped to standard health record formats (HL7 FHIR, SATUSEHAT integration).
    - Actionable categories (e.g., “Low risk – annual checkup”, “High risk – immediate referral”).

Step 6: Monitoring Performance

- **Accuracy Tracking**: monitor sensitivity, specificity, and AUROC for each health domain over time
- **Error Logging**: flag false positives/negatives, track by demographic subgroup.
- **Performance Drift Detection**: identify shifts in population health data or LLM output distributions.
- **Continuous Learning**: incorporate clinician feedback, new guidelines, and updated datasets into periodic recalibration.
- **Ethical/Compliance Monitoring**: ensure adherence to health data privacy (SIKN, SATUSEHAT standards) and WHO/FDA AI ethics frameworks.

5.3. Workflow Overview for Training AI-Assisted Diagnostic Tools

The Training workflow outlines how data flows through the AI-Assisted Diagnostic Data Framework to produce robust and reliable AI models for diagnostic support. It describes how data is collected, governed, cleaned, standardized, labeled, and then used to train and evaluate models in a way that respects clinical guidelines, interoperability standards, and fairness constraints. The following section provides a detailed breakdown of the training pipeline, from data ingestion through model development to evaluation and deployment.

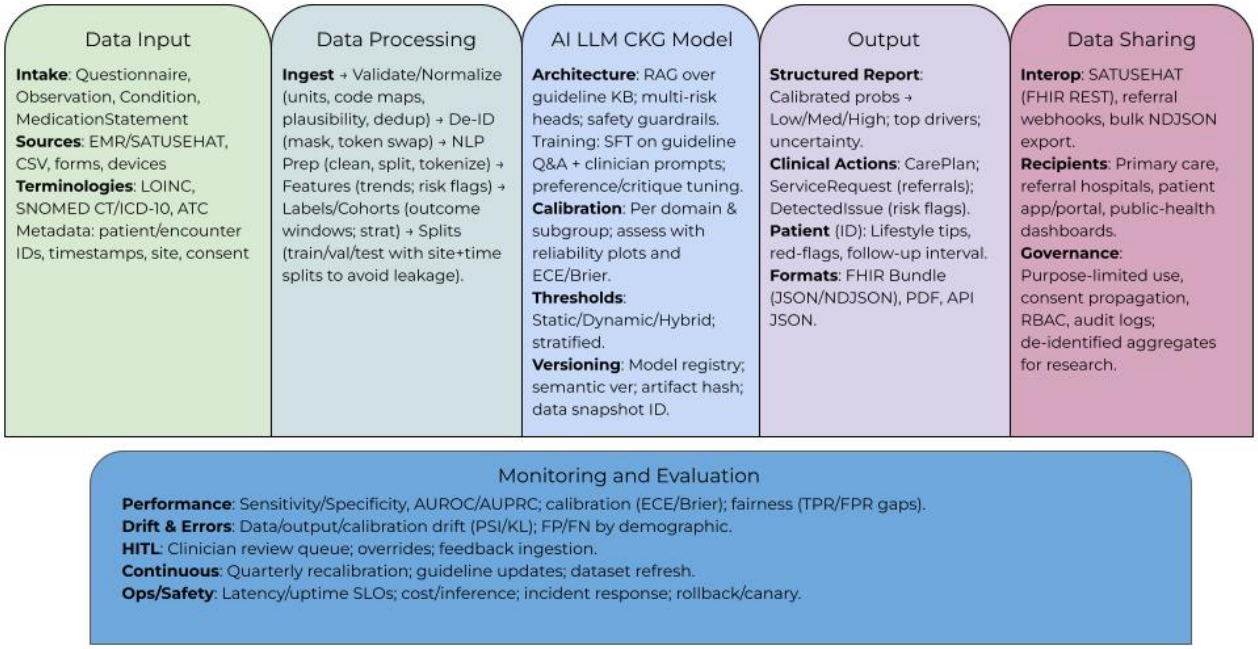

Figure 13 Workflow overview for training AI-assisted diagnostic tool

Step 1 – Data Input

- Collect multi-modal data from routine care: demographics, vitals, anthropometrics, labs, questionnaires, medical history, medications, and device streams.
- Represent all inputs in interoperable formats (e.g. FHIR resources, standard code systems) to support integration with national health data infrastructures.
- Apply governance from the outset: consent, de-identification for model development, access control, and basic data quality checks (completeness, plausibility, duplicates).

Step 2 – Data Processing

- Clean and harmonise data: unit normalisation, coding alignment, removal of duplicates, and handling of implausible values.
- Construct clinically meaningful features (e.g. trends over time, risk-factor aggregates) and prepare text fields through domain-aware NLP preprocessing.
- Define labels and cohorts with explicit time windows; split into training/validation/test sets using rules that reduce leakage and preserve real-world conditions.

Step 3 – AI LLM CKG Model

- Specify the diagnostic-support tasks: triage, risk stratification, extraction of risk factors, guideline-based advice, and referral suggestion.
- Train and fine-tune the model on curated datasets and guideline-aligned instructions, integrating structured predictors and clinical knowledge sources.
- Evaluate along multiple dimensions: discrimination, calibration, clinical utility, hallucination rates, fairness across subgroups, and guideline adherence.

Step 4 – Output

- Convert continuous risk scores into discrete categories (e.g. Low/Medium/High) using static, adaptive, or hybrid thresholds appropriate to each condition.
- Align thresholds with clinical and public health trade-offs (e.g. acceptable false-positive/false-negative rates in screening contexts).
- Generate structured outputs (risk scores, categories, key contributing factors, uncertainty statements) that can be consumed by clinical systems and user interfaces.

Step 5 – Data Sharing

- Package outputs and relevant input summaries in standard exchange formats (e.g. FHIR Bundles) for clinical systems, dashboards, and patient applications.
- Preserve terminology mappings end-to-end to ensure consistent interpretation across institutions and software vendors.
- Enforce policies for purpose-limited use, role-based access, audit logging, and de-identified aggregate sharing for research and quality improvement.

Step 6 – Monitoring and Evaluation

- Monitor performance over time with domain-specific metrics (AUROC, AUPRC, calibration indices) broken down by key demographic and clinical subgroups.
- Track data, output, and calibration drift; log false positives/negatives and safety incidents for systematic review.
- Incorporate clinician and user feedback into periodic data refresh, retraining, recalibration, and controlled model rollout using formal model registry and versioning.

6. Conclusion

As artificial intelligence (AI) becomes more embedded in diagnostic processes, a comprehensive and equitable data framework is no longer optional, but a critical enabler of effective healthcare. The reliability of AI-assisted diagnostics depends heavily on the quality, completeness, and governance of the data that supports them. This "Framework for Health data Management in Indonesia" recognizes that poorly curated or fragmented data can lead to biased algorithms, inconsistent results, and limited scalability. It provides a structured, context-specific approach to ensure health data is reliable, standardized, and ethically managed, aligning directly with the national Digital Health Transformation Strategy (DHTS) 2.0 Draft.

Furthermore, this framework operationalizes these principles throughout the entire data lifecycle. It aligns with global frameworks such as FAIR (Findable, Accessible, Interoperable, Reusable) and CARE while leveraging trusted technical standards like SNOMED CT, HL7 FHIR, DICOM, and WHO ICD-11 to reinforce interoperability. The success of AI tools depends not just on algorithmic precision, but on the ethical handling and diversity of the data that fuels them. Therefore, ensuring inclusivity in data collection to avoid bias , standardizing annotation for supervised learning , and harmonizing data across systems via the SATUSEHAT platform are all critical.

This framework is not only a technical guide, but also a strategic tool for stakeholders, including the Ministry of Health, health facilities, and technology providers. It is demonstrated through a specific use case for an LLM-based Free Health Check (CKG), which details the data flow from collection via ASIK CKG to the LLM's analysis. Critically, it specifies a rigorous 6-step calibration and monitoring process to sustain model accuracy, track for "performance drift," and ensure safety over time.

By embedding these practices, from robust data governance and privacy controls aligned with UU PDP 2022 to continuous AI model monitoring, this framework serves to build a health system that is digitally empowered, inclusive, and resilient. It provides the necessary foundation to build public trust, foster innovation, and realize the full potential of AI to transform Indonesia's health landscape.

References

1. Kementerian Komunikasi dan Informatika Republik Indonesia. (2023). *Surat Edaran Menteri Komunikasi dan Informatika Nomor 9 Tahun 2023 tentang Etika Kecerdasan Artifisial*. https://jdih.komdigi.go.id/produk_hukum/view/id/883/t/surat%2Bedaran%2Bmenteri%2Bkomunikasi%2Bdan%2Binformatika%2Bnomor%2B9%2Btahun%2B2023
2. Ministry of Health of the Republic of Indonesia. (2024). Indonesia digital health transformation strategy 2024. World Health Organization, Country Planning Cycle Database. https://extranet.who.int/countryplanningcycles/planning-cycle-files/digital-health-transformation-strategy-2024
3. Ministry of Health. (2025). Technical Guidelines for Free Health Check. Ministry of Health, Republic of Indonesia.
4. Republik Indonesia. (2022). *Undang-Undang Republik Indonesia Nomor 27 Tahun 2022 tentang Pelindungan Data Pribadi* (Lembaran Negara Republik Indonesia Tahun 2022 Nomor 196, Tambahan Lembaran Negara Republik Indonesia Nomor 6820). https://peraturan.bpk.go.id/Details/229798/uu-no-27-tahun-2022
5. World Health Organization. (2021, June 28). Ethics and governance of artificial intelligence for health (ISBN 9789240029200). https://www.who.int/publications/i/item/9789240029200