Subdomain 4.3: Responsible Data Management and Use Playbook

Guiding responsible AI adoption

Before You Start

For more about the CHAI AI Governance Framework, methods used to develop these playbooks, and information on how to use them, please refer to the CHAI AI Governance Playbooks: Introduction & Background document. Note that **baseline operational controls** were consensus-defined by Healthcare Delivery Organizations (HDOs) as "must-haves" for health AI governance. The **implementation guidance** that accompanies the baseline operational controls are **suggestions/recommendations** based on known practices. It is assumed and appropriate that implementation will differ based on organizational goals, resources, and maturity.

Disclaimer

The implementation guidance, frameworks, and examples contained herein represent generalized practices and considerations drawn from across the field of responsible AI deployment in healthcare. They are offered as one set of perspectives to inform your thinking not as the "right way" or "only way" to build, govern, or operate an AI program.

Every organization operates within its own unique context: differing patient populations, clinical priorities, regulatory environments, resource constraints, technical infrastructure, cultural norms, and stage of AI maturity. While the specific methods, timelines, and depth of implementation will appropriately vary across organizations, the underlying controls, principles, risks, and considerations raised throughout this playbook are intended to be engaged with thoughtfully and in good faith by any organization deploying AI in healthcare.

Organizations should prioritize state and federal regulatory requirements, especially for AI solutions that meet the definition of a Software as a Medical Device (SaMD), as defined by the FDA.

Not a Standard of Care. This playbook does not establish, define, or reflect a legal, regulatory, professional, or industry standard of care, nor is it intended to serve as one. It is a voluntary, aspirational, and educational resource designed to support organizational learning and dialogue. The field of healthcare AI is rapidly evolving, and reasonable organizations and professionals may differ on how to approach the issues discussed herein. The practices described may not be appropriate, feasible, or advisable for every organization, every use case, or every point in time. Adoption of any particular practice described in this playbook is entirely voluntary, and a decision to adopt, adapt, partially implement, or decline to implement any practice should not be construed as evidence of compliance with, or deviation from, any standard of care, duty, or legal obligation. This playbook is not intended to be used, and should not be used, as a benchmark, checklist, or measuring stick against which the conduct of any

organization, clinician, or other party is evaluated in any litigation, regulatory proceeding, accreditation review, or similar context.

General Informational Purpose. The information in this playbook is provided for general informational and internal operational purposes only. It is intended to support organizations in thinking through the design and stewardship of their AI programs and should be used in conjunction with from clinical leaders, technical experts, compliance officers, ethicists, patient and community representatives, and other advisors familiar with the specifics of your organization.

Not Legal Advice. Nothing in this playbook constitutes legal advice, and it should not be relied upon as a substitute for advice from a qualified attorney licensed to practice in the relevant jurisdiction. Nothing in this document creates an attorney-client relationship between the reader and CHAI, its officers, directors, employees, or contributors. The contents reflect general information and observed practices as of the date of publication and may not account for changes in law, regulation, technology, or circumstance that occur thereafter.

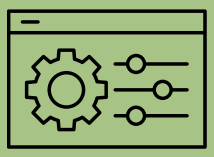
No Warranty. While reasonable efforts have been made to ensure accuracy, no representation or warranty is made as to the completeness, timeliness, accuracy, or suitability of the information for any particular purpose. Laws, regulations, and best practices vary by jurisdiction and evolve over time, and the application of those laws and practices to specific facts can lead to different outcomes.

Illustrative Examples. Any tools, resources, examples, or technology product described or presented are for educational and illustrative purposes only and do not represent endorsement by CHAI or guarantee specific results. Readers should consult qualified legal counsel and relevant experts for specific terms, products, or services.

Independent Judgment Required. Readers should consult with qualified legal counsel, clinical experts, and other appropriate professional advisors before taking, or refraining from taking, any action based on the information in this playbook. The decision to adopt, modify, or disregard any practice described herein rests solely with the reader and their organization. CHAI expressly disclaims any and all liability for any loss, damage, or other consequence, including but not limited to claims of negligence, breach of duty, or failure to meet any standard, arising directly or indirectly from reliance on, use of, citation to, or reference to the contents of this document.

At a Glance

Data is the foundation of every AI system. The quality, governance, and ethical sourcing of data directly determines whether AI systems in healthcare can be trusted to support clinical decisions, improve operations, and serve all patients. Healthcare organizations already manage data under rigorous regulatory frameworks including HIPAA, 21st Century Cures, and state privacy laws. This playbook addresses Responsible Data Management and Use, which builds on these existing obligations while introducing AI-specific recommendations. This playbook outlines the controls and processes required to ensure responsible data management for AI systems in healthcare. HDOs likely already have a data management and use process in place, however, the information below aims to help organizations update and modify processes based on AI specific considerations.



Baseline Operational Controls for Subdomain 4.3: Responsible Data Management and Use

Category	Control
Data Documentation & Stewardship	Establish and maintain a structured data acquisition register for data used in AI systems.
Data Use Agreements & Vendor Compliance	Establish comprehensive agreements defining terms for permissible/non-permissible uses, re-identification and redistribution preferences, requirements for vendor security compliance, audit rights, and data return/destruction terms.
De-identification & Privacy Compliance	Establish written HIPAA-compliant de-identification policies; evaluate PHI datasets to select the appropriate method; validate de-identification before AI training use; maintain oversight of linked datasets.
Data Quality & Bias	Establish a process and owner for recurring AI data quality and bias evaluation to help identify quality gaps and known sources of bias.
Data Preparation Process	Establish a standard operating procedure (SOP) for AI data preparation including rules for data cleaning, encoding practices, normalization/standardization methods, and partitioning procedures

Background: Understanding Health Data Types

Before implementing the controls in this playbook, organizations need to identify what types of data their AI systems use, and what legal agreements apply. The data type determines the governing framework, required contract provisions, and safeguards. The table below summarizes the three primary data classifications relevant to AI in HDOs. Additionally, it is important to have consistent data access governance processes as a key capability, which underpins data access for AI development and use.

Data Type	Description	Legal Agreement Required
Protected Health Information (PHI)	Individually identifiable health information — patient names, dates, contact info, diagnoses, treatment records, lab results, imaging, and any data that could identify an individual.	Business Associate Agreement (BAA) or HIPAA exception
Limited Data Set (LDS)	Partially de-identified data: direct identifiers (names, SSNs, full addresses) are removed, but geographic data (city, state, zip code), dates, and ages may remain.	Data Use Agreement (DUA) or HIPAA exception
De-identified Data (DID)	Data with all 18 HIPAA identifiers removed per Safe Harbor (45 C.F.R. § 164.514(b)) de-identification standard or statistically de-identified via Expert Determination (45 C.F.R. § 164.514(c)).	DUA or other contractual controls (not required by HIPAA, but strongly recommended for AI contexts)

Note: Imaging Data

Imaging files contain extensive embedded metadata beyond the 18 HIPAA Safe Harbor identifiers, including equipment serial numbers, operator IDs, acquisition parameters, and institution names. Standard Safe Harbor de-identification focused on structured data fields misses many of these. Full DICOM tag scrubbing requires specialized tools rather than general-purpose de-identification approaches.

See here for an example tool from Stanford.

Business Associate Agreement (BAA) — Key Provisions for AI using PHI

A BAA is required whenever a vendor creates, receives, maintains, or transmits PHI on behalf of a covered entity. For AI tools, the following provisions are essential:

Permitted Uses and Disclosures: Must be limited to defined functions (e.g., data hosting, analytics). AI-specific uses should be explicitly listed.

Safeguards: Administrative, physical, and technical measures consistent with the HIPAA Security Rule.

Breach Notification: Timely reporting (typically within 10–15 days) of any impermissible use, disclosure, or security incident.

Subcontractors: Flow-down of all BAA obligations to any subcontractors who access PHI.

Return or Destruction of PHI: Required upon contract termination, including AI model weights or embeddings trained on PHI.

Indemnification and Liability Caps: Important for risk allocation; often negotiated heavily with vendors but should not be waived entirely.

Data Use Agreement (DUA) — Key Provisions for Limited Data Sets

A DUA is required under HIPAA before sharing a Limited Data Set. It must specify:

Permitted Uses and Disclosures: Only for research, public health, or health care operations. AI-specific use cases must be explicitly identified.

Recipient Obligations: No re-identification or contact of individuals; use only for purposes specified in the DUA; implement safeguards to prevent unauthorized use or disclosure.

Reporting Obligations: Recipient must notify the covered entity of any misuse or unauthorized disclosure.

Data Flow Control: Subcontractors or third parties (including AI subprocessors) must be bound by equivalent DUA terms.

De-identified Data

HIPAA does not require a BAA or DUA for de-identified data, but AI contexts introduce risks that make contractual controls essential even when not legally mandated. It is important to limit data shared to reduce the technical possibility of re-identification.

a. Contractual Controls

Include use and re-identification prohibitions even when HIPAA does not require them.

Define ownership and derivative rights: who owns new datasets, trained models, or insights created using the de-identified data?

Require non-re-identification warranties and reserve monitoring and audit rights.

b. Re-identification Risk

Prioritize data minimization to reduce the risk of re-identification through linkage or statistical inference, particularly when de-identified data is combined with other datasets or used to train AI models.

c. State and Other Laws

Several states impose privacy obligations on de-identified or pseudonymized data that go beyond HIPAA. Examples include California (CCPA/CPRA), Washington (My Health Data Act / MHMDA), and Colorado (CPA). Organizations should ensure contracts reflect both federal and applicable state privacy regimes.

Note: AI Specific Risks, Considerations, and Solutions

Non-negligible re-identification risks still exist, even after removal of all 18 identifiers using Safe Harbor methods. This is even more likely with the use of powerful deep learning and language based models. At the same time, statistical de-identification through expert determination is both costly and resource intensive.

When Safe Harbor May not be enough, irrespective of dataset size:

- Rare conditions or procedures: Patients with highly unusual diagnoses combined with age range, state, and year of service can be uniquely identifiable. The rarer the condition, the smaller the dataset needs to be before this becomes a real risk.
- Small geographic areas: For populations fewer than 20,000 people, HHS requires suppressing zip codes. Even state-level geography with a rare condition can be identifying.
- High dimensional data: AI and genomic contexts introduce something called the “mosaic effect”. No single variable is re-identifying, but combinations of variables across a large dataset can be. Much of AI, especially GenAI includes not only high dimensional, but also unique multi-dimensional data.

Resource: Re-identification Risks in HIPAA Safe Harbor Data: A study of data from one environmental health study

These re-identification risks are amplified with GenAI and Agentic solutions. See a non-exhaustive list of risks below.

Generative AI Risks

- Training Data Memorization: LLMs can reproduce verbatim PHI from training data, especially for rare or unusual cases that are inherently identifiable
- Quasi-identifier synthesis: models can generate outputs that combine enough details (age, diagnosis, geography, timeline) to uniquely identify a real patient without reproducing any single record
- Synthetic data risk: AI-generated "synthetic" patient data retains statistical properties of real records, making re-identification feasible particularly for rare conditions or small demographic groups

Agentic AI Risks

- Cross-context aggregation: agents querying multiple systems (EHR, billing, scheduling, labs) can assemble re-identifying profiles even when each individual source is appropriately de-identified
- Persistent memory stores: agentic memory that persists across sessions creates ungoverned data repositories outside normal HIPAA-scoped system
- Tool-call exfiltration: patient context passed to external APIs across multiple tool calls can reconstruct identifiable profiles at third-party services that may not be covered entities

De-identification standards like Safe Harbor and expert determination were originally designed for structured data and were not developed with application to LLMs, RAG retrieval, or agentic reasoning.

While HIPAA does not have specific guidance on this, here are some methods that can help:

k-anonymity (see related article here): each record in a dataset should be indistinguishable from at least k-1 other records across the combination of quasi-identifiers retained. k=5 is a common minimum; k=10 or higher is preferred for sensitive conditions.

Cell suppression — for any subgroup analysis where fewer than 10–15 individuals share a combination of attributes, those cells are typically suppressed or perturbed before release.

Differential privacy is emerging as the appropriate standard for large-scale AI training datasets. It provides a mathematically formal privacy guarantee that limits how much any individual record can influence a model's outputs during fine tuning. Apple, Google, and the U.S. Census Bureau use differential privacy for their large-

Key Tools & Resources

- Health and Human Services (HHS): Guidance Regarding Methods for De-identification of Protected Health Information in Accordance with the Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule

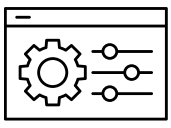
Background: Risk-Based Approach to Responsible Data Governance & Use

As with many of the controls described in these playbooks, it is recommended that organizations take a risk-based approach to responsible data governance and use in the context of AI. This approach may include taking both the risk of the AI solution and the sensitivity of the data sources used in training (if applicable), tuning, testing, or using the model to help prioritize how rigorously to apply a specific control. Additionally, AI solutions come into the HDOs through different pathways, such as through research, internal development/co-development, from a third-party vendor, or embedded into existing platforms (e.g. EHR or care-management platforms) (See Subdomain 4.1 Responsible AI Lifecycle Management Playbook). The implementation of the data management and use controls described in this playbook may differ based on how solutions were brought into use within the HDO. Organizations may want to align responsible lifecycle management standard operating procedures (SOPs) with responsible data management and use SOPs to simplify processes and procedures.

Another big consideration is aligning data management and use procedures with AI Policies. It's one thing to have AI policies in place, but implementing these policies effectively, necessitates methods for monitoring and tracking compliance. This means that both processes and procedures for Responsible Lifecycle Management and Use (see Subdomain 4.1: Responsible Lifecycle Management & Use Playbook), and Responsible Data Management and Use, should enable policy implementation and enforcement.

For example, if a policy prohibits the use of PHI in general purpose models, are organizations able to track metadata associated with prompts used, to identify where and when PHI was inappropriately used?

Data Documentation & Stewardship



Control 4.3.1: Establish and maintain a structured data acquisition register for data used in AI systems.

Before an organization can govern how data is used in AI, it must know what data it has, where it came from, and what it is being used for. Healthcare data is often fragmented across EHRs, claims systems, lab platforms, imaging archives, and third-party vendor environments. Without a data acquisition register, organizations cannot identify privacy risks, check alignment with policies, fulfill audit obligations, respond to patient data requests, or assess whether training data is representative of the populations being served.

A data acquisition register is the foundation of every other data management control in this playbook. It enables de-identification validation (Control 3), DUA compliance tracking (Control 2), and bias assessment (Control 4) by providing a single source of truth about what data exists and how it flows through AI systems. Some organizations choose to prohibit the use of unregistered datasets for training, validation, or inference.

Data used in AI systems refers to data used in developing/co-developing AI solutions (if applicable), data used to tune AI solutions, retrospective data used for testing or validation, and data used for inference (e.g. data inputs into the AI solution that produce relevant outputs).

Implementation Guidance

Note: Why Metadata Matters for AI Governance

Metadata is information about a dataset's origin, scope, structure, and limitations. It makes data interpretable and auditable over time. For AI governance, metadata enables organizations to assess whether training or inference data is representative of the population a model will serve, trace the source of errors or unexpected model behavior, help access appropriate data sources for use as inputs by AI models, and satisfy audit requests from regulators, accreditors, or governance committees. AI also poses distinct challenges for the development of data registries.

Unlike traditional data management where risk is largely contained to the data itself, AI systems can encode patient information into model weights, infer identities from de-identified data through statistical linkage, and generate derivative artifacts that persist long after the original data is returned or destroyed. The controls here are specifically designed to govern risks that don't exist until data is used to train or power an AI.

Data acquisition registers ideally document data across training, fine-tuning, testing, and operational use, but most organizations only reliably capture the first two, since training/validation data tends to be a discrete, bounded extract while inference data (the continuous daily flow through deployed models) is the least documented and carries the biggest governance gaps.

Most feasible now: Inventorying vendor-facing data flows and confirming BAAs/DUAs explicitly cover AI training restrictions.

Most challenging: Documenting operational inference data (requires vendor transparency or technical instrumentation most systems lack), monitoring vendor-side fine-tuning and model updates, and governing unstructured/multimodal data (notes, imaging, genomics) where standard de-identification methods don't hold up.

For any AI system where the full scope of data it acts on shifts over time or at runtime, comprehensive logging of data access (see Control 5, Section 8) is an important complement to the register. The register documents what the system is authorized to access and under what conditions; logs record what it actually accessed.

Types of Metadata

There are different categories of metadata that may be important to capture in a data register. The maturity of the organization and data documentation methods, the risks of the AI system, and sensitivity of the data used to train, tune, test, or use the AI solutions, should be considered when determining the breadth and depth of information is captured in a data register.

Descriptive metadata: Type of data and where it comes from.

- Dataset name, description, and purpose
- Data source (e.g., EHR system, claims database, wearable device)
- Date of creation or acquisition and collection method
- Geographic or institutional scope (e.g., which facilities, which patient populations)
- Data sensitivity classification: Note that data sensitivity may refer to differentiating between whether data is PHI, de-identified, synthetic, publicly available, etc. It may other include sensitive data categories such as payment or financial data, data subject to heightened protections under state or federal law (e.g., behavioral health records, substance use disorder treatment records, reproductive health information,

HIV/AIDS-related data, or data about minors), or other confidential categories defined in the organization's data classification policy.
- Linked AI systems or models using this dataset

Structural metadata: How data is organized

- File format, schema, and field definitions
- Variable types, units of measurement, and coding standards used (e.g., ICD-10, SNOMED, LOINC)
- Number of records and time period covered

Administrative metadata: Ownership, access, and lifecycle

- Data owner and steward
- Access controls and authorized users
- Legal basis for use (BAA, DUA, IRB approval, patient consent)
- Retention period and disposition plan
- Version history and date of last update

Provenance and lineage metadata: The history of transformations

- Where the data originated and how it moved through systems
- What transformations, cleaning steps, or de-identification processes were applied
- Which models or AI systems have used the dataset
- Links to related datasets or derived outputs

Quality metadata: Known characteristics and limitations

- Known gaps, missingness rates, or data quality flags
- Population represented and any known representativeness limitations
- Validation status
- Cautioned uses

Minimum Recommended Fields to Include in a Data Acquisition Register

In practice, documenting and monitoring all of these metadata fields can be challenging. Less mature organizations may just need a place to start, while more mature organizations may face challenges developing processes for metadata capture for cutting edge technologies, where methods are not yet well defined. For AI governance purposes, the most critical starting point is usually elements of provenance (where data came from, how it was collected), structural basics (what it contains, time period, population), and administrative fields (legal basis, who owns it, what AI systems use it). The rest can be built out as organizational maturity grows. See the following examples of what a basic data acquisition register might look like, noting that organizations should adapt and update existing processes, and that more mature organizations may want to build on this structure.

The following are important to document :

- Data source name and type (e.g., EHR, claims, lab, imaging, external registry, wearable device)
- Date of acquisition and method (e.g., EHR extract, vendor API, manual upload)
- Data owner (the organizational unit or individual responsible for the source system)
- Data steward (the individual(s) accountable for governance of this dataset)
- Legal basis for use (e.g., Business Associate Agreement, Data Use Agreement, de-identification, patient consent)
- Sensitivity classification (PHI, de-identified, synthetic, publicly available)

    - Other types of sensitive data categories may include: payment or financial data; data subject to heightened protections under state or federal law (e.g., behavioral health records, substance use disorder treatment records, reproductive health information, HIV/AIDS-related data, or data about minors); or other confidential categories defined in the organization's data classification policy.
- Linked AI systems or models using this dataset
- Retention period and disposition plan (when will data be destroyed or archived?)

In some cases, such as for an AI system that continuously learns, outputs from the AI system (e.g. predictions, coded claims, clinical notes, etc.) may be used as inputs for model retraining, calibration, or performance scoring. If feedback loops exist, organizations may want to additionally include information on the secondary data flow, document its legal basis, and note whether the feedback loop generates new types of data.

Note: Data Registers for AI Systems with Ongoing or Real-Time Data Flows

Traditional AI data acquisition registers are designed around discrete, development-time datasets, data assembled for training, validation, or testing before a model is deployed. Many AI systems, however, acquire and process data continuously after deployment. This includes predictive ML models that ingest real-time EHR feeds, models that retrain or update on accumulating data, and generative or agentic systems that retrieve and process data dynamically at runtime. Organizations deploying any of these system types should consider expanding their register practice to capture the following:

- **Real-time inference inputs:** For AI systems that ingest live data at runtime, such as a predictive model drawing continuously from EHR fields, lab values, or device streams, the register should document which data sources and fields are accessed, the legal basis for that access (noting that operational BAAs may not automatically cover all AI-specific uses), and what happens when expected data fields are missing or anomalous. The live data distribution feeding the model at runtime may also inform ongoing data quality and bias monitoring under Control 4.
- **Continuously learning or periodically retrained models:** For models that update their parameters on new data, whether through scheduled retraining or ongoing learning, the register should capture not only the original training dataset but also the mechanism by which new data enters the model, the cadence and criteria for updates, and who is responsible for approving those updates. This is a distinct governance obligation from the original development-time data documentation.
- **Retrieval corpora (for RAG-based systems):** The document or record corpus used to ground generative AI outputs is functionally equivalent to training data in terms of its influence on model behavior. Register entries for RAG systems should capture what sources are included, the criteria for inclusion, and the update cadence — since a stale or poorly governed retrieval corpus introduces the same quality and bias risks as a poorly prepared training dataset.
- **Data categories permitted in prompts or context windows:** For generative AI systems, what patient data may be included in prompts or passed to the model as context is itself a data governance decision. This should be documented, including the data categories permitted, any categories that are prohibited, and the legal basis for including PHI where applicable.
- **Persistent memory stores:** If an agentic system maintains memory across sessions (e.g. retaining clinical context, user preferences, or prior interactions) that memory store should have its own register entry, with access controls, a defined retention period, and BAA coverage confirmed.
- **Authorized runtime data sources for agentic systems:** For agentic AI that queries multiple systems or APIs to complete tasks, the register should document which systems, databases, and external APIs the agent is authorized to access — not only the datasets used to develop it. This authorization scope should be reviewed whenever the system's capabilities or integrations change.

Steps for Implementation

1. Inventory existing AI-related data. Begin by documenting data currently used in higher-risk AI tools or tools that touch sensitive data sources. This should include vendor-deployed tools where data is transmitted to the vendor. Many organizations are surprised to find how many data flows exist without documentation. A risk-based approach to defining data registers and necessary fields may help reduce burden and allow for the development of scaleable processes over time.
2. Select a register format appropriate to your organization's size, capacity, and industry best practices. A structured spreadsheet is sufficient for smaller or less mature organizations; dedicated data catalog tools offer automation and integration for larger or more mature organizations.
3. Assign data stewards. Each dataset entry should have a named individual responsible for keeping the record accurate. This role may be held by an existing IT lead, data analyst, or compliance officer and does not need to be a dedicated position.
4. Integrate with the AI Inventory Registry. The data acquisition register should link to the AI Inventory Registry maintained under Domain 3: Organizational Resources Playbook. Each AI solution in the inventory, or at minimum, each higher risk tool, should reference the datasets it touches.
5. Establish a review cadence and specify triggers for data register updates. Review and update the register when new AI tools are procured, when vendors change their data practices, or when new data sources are accessed. Consider aligning review frequency to the risk level of the AI tool and data involved: higher-risk tools (e.g., tools supporting clinical decision-making, tools with access to large volumes of PHI, etc.) may warrant more frequent data governance review, while lower-risk administrative tools may be reviewed annually or upon significant operational change.
6. Establish a retention period for the data register after the AI solution itself is decommissioned.

Illustrative Examples by Organization Size

The following scenarios are illustrative only. They represent plausible approaches based on resource level and are not drawn from specific documented organizations or published case studies.

Small Clinic, Critical Access Hospital, or Community Health Center	Mid-Size or Regional System	Large Health System or Academic Medical Center
Maintains a shared spreadsheet register with 10–15 fields per dataset. Data steward role is held by the EHR coordinator or IT lead. Register is reviewed quarterly and updated when new AI pilots are initiated independently or with HCCN and/or state PCA partnerships. Focus is on documenting data flows to external AI vendors, particularly ensuring BAAs are in place for any data leaving the organization.	Uses a data catalog integrated with the EHR platform. Department-level data stewards assigned per clinical or operational domain. Automated metadata capture supplements manual entries for structured data. Unstructured data (clinical notes, imaging) requires additional documentation steps. Register reviewed with IT governance team quarterly; major updates triggered by new AI procurement.	Enterprise data catalog fully integrated with the MLOps platform and AI inventory. AI Domain Stewards maintain records per AI model, with lineage tracked from source system to model training and deployment. Register linked to IRB protocols, model cards, and data use agreement status. Reviewed in real-time via automated dashboards.

Real-World Example

A large multi-state health system has implemented a tiered data governance structure with designated Domain Trustees, Domain Stewards, Technical Domain Stewards, and a newly created AI Domain Steward role. The AI Domain Steward is responsible for assessing data requirements for classification and quality, facilitating model adoption by partnering with use case owners, ensuring proper documentation of use cases and implementation procedures, and coordinating compliance with responsible AI principles.

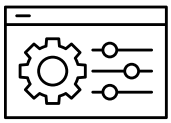
Their governance domains span Consumer, Clinical, Finance, Payer, Provider, Workforce, Location, and Technical Artifact categories, which are coordinated through a central Data Governance Council. Inflow includes Trustees, Stewards, and SMEs; outflow includes policies, standards, training, and reporting to the organization.

Source: Health System's Data Management Strategy

Key Tools & Resources

- TEFCA
- Example Data Cataloging Tools (not endorsement):
  - Microsoft Purview
  - Collibra Data Catalog
  - Alation Data Catalog
  - Databricks Unity Catalog
  - Acryl DataHub
- FAIR Data Principles
- Illustrative example of Data Acquisition Registers for a sample of AI solutions

Data Use Agreements & Vendor Compliance



Control 4.3.2: Establish comprehensive Data Use Agreements and/or Business Associate Agreements that includes information on permitted and prohibited uses, compliance with organization's security and privacy requirements, rights to audit, consequences of non-compliance, and data return/destruction terms at end of use.

In healthcare AI, virtually every tool deployment involves some form of data sharing. Even solutions that appear to operate only within the organization's environment may transmit data to vendor cloud platforms, model APIs, or third-party analytics services. Data Use Agreements are the legal mechanism for governing these flows, defining what vendors can and cannot do with healthcare data.

Note: Contracting Challenges

Vendor contract enforcement is one of the most significant challenges in healthcare AI data governance. Vendors often resist strict data use terms, including audit clauses and penalties for non-compliance, making negotiations challenging. This becomes even more challenging for smaller health centers or for less resourced HDOs. It is rare for HDOs to successfully negotiate shared liability for data misuse. This makes getting the DUA right at the outset critical.

Implementation Guidance

Note: Data Sharing is a Decision, Not a Default

Before negotiating a Data Use Agreement, organizations benefit from first confirming that sharing data with a vendor is the right approach for the specific use case. Some AI tools can operate effectively with de-identified or synthetic data, or without requiring access to patient-level data at all. Others may require data sharing but only for a subset of the data elements initially requested.

Asking 'what is the minimum data this vendor actually needs for the contracted service?' before engaging legal review often results in a narrower data scope, lower compliance risk, and stronger negotiating leverage. The DUA should codify the outcome of this decision — it should not be the first place the decision gets made.

One practical starting step is to ask vendors requesting patient data to document, in writing, the specific data elements needed and the reason each element is required for the contracted service. This creates a useful baseline for legal review and sets appropriate expectations before negotiations begin.

Key DUA Provisions for AI

Data management between HDOs and third party vendors is a critical component of the formal relationship. While the Subdomain 4.4: Third Party Management Playbook covers the full set of roles, responsibilities, and accountability structures that need to be defined and documented between the organization and third party, the below information includes the specific data components to keep in mind. HDO's should consider a risk-based approach to the implementation of this control. AI-specific DUAs and/or BAAs addenda should address:

- Permissible use definition: Explicitly list what the vendor is permitted to do with the data. General language such as "for services purposes" is insufficient and may permit unauthorized model training.
- AI training restriction: If the vendor is not permitted to train or fine-tune models on your data, this must be stated explicitly. Conversely, if training is permitted, define its scope and any resulting IP rights.
- Re-identification prohibition: Contractually forbid the vendor from attempting to re-identify de-identified data or linking datasets in ways that could enable re-identification.
- Data Minimization: Require vendors to use only the minimum data necessary for the specified purpose. Prohibit the vendor from retaining data beyond what is required.
- Security requirements: Specify minimum standards including encryption at rest and in transit, role-based access controls, and regular independent security assessments (e.g., SOC 2 Type II, HITRUST).
- Audit rights: Reserve the right to audit vendor compliance with the DUA, including access logs, security practices, and data handling procedures.
- Penalties for non-compliance: Define remedies for breach, including the right to terminate the agreement and seek financial damages.
- Data return and destruction: Require vendors to return or certifiably destroy all raw datasets, stop further training, and delete any models fine-tuned on your organizations data (if feasible) — upon contract termination.
- Subcontractor disclosure: Require vendors to disclose any subcontractors or downstream processors who will access data, and extend DUA obligations to them.
- Incident Response Plan: Develop and maintain an incident response plan to address data breaches and other security incidents promptly.

Note: Considerations on Derivative Rights

AI training restrictions govern whether a vendor may use your data to build or improve AI models. A related and distinct question is what rights, if any, your organization retains in models, features, or analytical outputs that are derived from your data, sometimes called derivative rights or data-derived IP. This is worth understanding even where IP ownership is not a primary goal.

There is no standard approach. Common arrangements range from organizations having no rights in vendor-developed models, to retaining the right to audit model behavior, to negotiating licensing terms or revenue-sharing when the vendor's product is materially improved using your data. The enforceability and practical value of these terms varies significantly with organizational size, legal capacity, and the uniqueness of the data asset.

For most healthcare organizations, the practical priority is ensuring the vendor cannot use your data in ways that could harm patients or your organization (e.g. by training models used to disadvantage your patient population, or by selling insights derived from your data to market competitors). IP co-ownership or revenue-sharing terms **may** be worth pursuing for organizations with unique data assets and the legal capacity to enforce complex provisions, but should be approached with realistic expectations about negotiating leverage, particularly for smaller HDOs or community health organizations. More work is needed to understand how smaller HDOs or CHCs, who often serve unique populations in unique regions of the country, might approach this.

When reviewing AI training and derivative rights terms, useful questions to consider include:

- How do HDOs and vendors define what constitutes a derivative? Are models themselves considered derivatives? Can vendors can retain embeddings, feature vectors, or synthetic data generated from organizational data?
- What specific models or features is the vendor developing using my data?
- Would the vendor's product improvement rely materially on my organization's data?
- What happens to derivatives if the contract is terminated?

Steps for Implementation

1. Audit existing vendor contracts. Review all current AI tool agreements for the provisions above. Identify gaps and prioritize remediation based on risk level of the tool and volume of data involved.
2. Develop a standard AI DUA and/or BAA addendum. Rather than renegotiating all contracts, create a standard addendum that can be appended to existing DUAs/BAAs for vendors handling data used in AI systems.
3. Establish/modify a vendor security assessment process. Create a standardized questionnaire that all AI vendors must complete before contracting. Review responses with IT security and legal.
4. Create a contract review checklist. Give procurement and legal teams a checklist of AI-specific provisions to verify before signing any new AI vendor agreement.
5. Track DUA/BAA compliance in the data acquisition register. Add DUA/BAA status, expiration date, and last audit date as fields in the data register so compliance is monitored on an ongoing basis.

Illustrative Examples by Organization Size

The following scenarios are illustrative only. They represent plausible approaches based on resource level and are not drawn from specific documented organizations or published case studies.

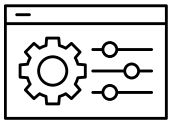
Small Clinic, Critical Access Hospital, or Community Health Center	Mid-Size or Regional System	Large Health System or Academic Medical Center
Uses a template DUA with an AI-specific addendum developed with support from the state Primary Care Association or HCCN. Reviewed by part-time or shared legal counsel. Negotiation leverage is limited, but may be strengthened with group purchasing (e.g. network-level, with hospital system affiliations, or with Clinically Integrated Networks when related to clinical integration or value-based performance) ; focus is on ensuring vendors cannot use data for model training beyond the specific contracted purpose, and that BAAs are in place for all data transmissions.	Dedicated legal and compliance review process for all AI vendor contracts. A master vendor registry tracks DUA status and renewal dates. Standard AI addendum developed internally, updated annually to reflect evolving regulatory guidance. Security questionnaire completed by all new AI vendors prior to contracting. Annual compliance reviews for high-risk or high-data-volume vendors.	Comprehensive AI procurement governance program with dedicated AI legal counsel and privacy engineering team. DUAs include model training restrictions, model weight controls (preventing vendors from retaining AI models trained on institutional data), and right to audit at any time. Vendors may be required to demonstrate HITRUST certification or SOC 2 Type II compliance before contracting. Joint security reviews for highest-risk tools.

Key Resources & Tools

Note: The following does not constitute legal advice.

- Data Use Agreement Discussion Worksheet
- HHS Model Business Associates Agreement

De-identification & Privacy Compliance



Control 4.3.3: Establish written HIPAA-compliant de-identification policies (Safe Harbor or Expert Determination) considering applicable state privacy laws; where relevant, evaluate each PHI-containing dataset to select the method that minimizes re-identification risk; validate de-identification before any dataset is used in AI training or model development; and maintain oversight of de-identified datasets, reassessing re-identification risk whenever datasets are linked or new identifiers emerge.

De-identification is one of the most technically demanding requirements in healthcare AI. HIPAA establishes two methods for de-identification, but AI-specific challenges go well beyond standard research de-identification. AI models trained on patient data can learn to re-identify individuals from seemingly non-identifiable patterns. Language models and embedding models are particularly sensitive. Even text stripped of explicit identifiers can encode patient-identifiable combinations of clinical history, demographics, and treatment patterns. Furthermore, linking of multiple sources of de-identified data can elevate the risks of re-identification. Note that some AI solutions may not require de-identification for use, but it is important to have a policy around de-identification requirements, contexts, and procedures.

For HDOs and vendors, the stakes are high: a de-identification failure in an AI context could mean patient data is exposed through model outputs, incorporated into vendor training pipelines without consent, or made accessible through model inversion attacks. Strong de-identification controls, combined with Data Use Agreements (Control 2) are crucial for maintaining the privacy of health data.

When is data de-identification necessary?

The short answer is that HIPAA's de-identification requirements are triggered by **use and disclosure**, not by the sensitivity of the data itself. Whether de-identification is necessary depends on what you're doing with the data and who receives it.

Does the recipient of this data (or the use of this data) fall within the covered entity's existing authorization framework, treatment, payment, operations, or a BAA?

If yes, de-identification is generally not required but may still be advisable for risk management.

If no, de-identification or explicit patient authorization is the necessary condition for the activity to be permissible.

Implementation Guidance

HIPAA De-identification Methods

HIPAA provides two accepted pathways for de-identification:

Safe Harbor Method

Remove all 18 HIPAA-specified identifiers, including: names, geographic subdivisions smaller than state, dates more specific than year for individuals over 89, telephone and fax numbers, email addresses, Social Security numbers, medical record and health plan beneficiary numbers, account and certificate/license numbers, vehicle and device identifiers, URLs, IP addresses, biometric identifiers (fingerprints, voice prints), full-face photographs, and any other unique identifying numbers, codes, or characteristics.

Safe Harbor is operationally simpler but may be insufficient for AI training data, particularly for unstructured clinical notes where combinations of clinical details can re-identify individuals even after explicit identifiers are removed.

Expert Determination Method

A qualified statistician or scientific expert analyzes the dataset and applies statistical or scientific principles to determine that the risk of identifying an individual is very small. The expert documents the methods and results that justify this determination.

Expert Determination is more rigorous and better suited to AI training data, but is more costly and requires access to qualified expertise. For smaller organizations, this may be available through regional health information exchanges, state HIEs, or contracted privacy consultants.

AI-Specific De-identification Considerations

- Text embeddings and language models: Text embeddings produced from clinical notes can encode patient-identifiable information in ways that are not detectable through standard de-identification review. Any text embeddings or LLM fine-tuning inputs derived from patient data should be treated as patient-level data and governed accordingly.
- Synthetic data as an alternative: Synthetic data generation — creating statistically representative but entirely artificial patient records — is an increasingly viable alternative to de-identification for AI training. Synthetic data can reduce re-identification risk while preserving statistical properties of the data.
- Linkage risk: Re-identification risk increases substantially when de-identified datasets are linked. Two datasets that individually meet de-identification standards may become re-identifiable in combination. Each linkage event should trigger a fresh risk assessment.
- State privacy laws: Several states have enacted privacy laws that go beyond HIPAA (e.g., California CMIA, Washington My Health MY Data Act). De-identification policies should account for the privacy laws of all states in which the organization operates or in which patients reside.
- Unstructured data and medical imaging: Clinical notes, radiology reports, and other free-text data often need specialized NLP-based de-identification tools; standard rule-based removal of explicit identifiers is insufficient for unstructured text. Medical images (e.g., DICOM files) need particular care: metadata embedded in image headers often contains patient identifiers beyond the 18 HIPAA Safe Harbor fields (such as institution name, acquisition date, equipment identifiers, and operator names) and must be reviewed and stripped before images are shared or used for AI training. Pixel-level data in certain image types, including full-face photographs and some dermatology, ophthalmology, or plastic surgery images, may directly re-identify patients and should be treated accordingly.
- Vulnerable and Sensitive Populations: Data involving individuals with heightened privacy protections (e.g. behavioral health, substance use disorder, HIV/AIDS, reproductive health, minors, or small geographic or demographic groups) may carry elevated re-identification risk even after standard de-identification.

Organizations using such data for AI training or sharing it with vendors should consider more conservative de-identification approaches, stricter access controls, and/or explicit contractual restrictions on downstream use. This is an evolving area; consult applicable federal, state, and local laws and legal counsel.

Steps for Implementation

1. Establish a written de-identification policy. Document which method (Safe Harbor, Expert Determination, or synthetic data) will be used for which scenarios, and define who has authority to approve de-identification for AI training use.
2. Create a dataset assessment process. For each new AI use case involving patient data, evaluate the dataset and intended use to select the appropriate de-identification method. Document this assessment.
3. Validate de-identification before use. Use automated tools for initial validation, supplemented by privacy officer or qualified expert review. Document the validation results and link them to the data acquisition register (Control 1).
4. Track and manage linkage risk. When de-identified datasets are combined or linked, conduct a formal re-assessment of re-identification risk and update HIPAA risk documentation accordingly.
5. Address unstructured data separately. Develop specific procedures for clinical notes, imaging reports, and other free-text data, using NLP-based de-identification tools and additional validation steps.

Illustrative Examples by Organization Size

The following scenarios are illustrative only. They represent plausible approaches based on resource level and are not drawn from specific documented organizations or published case studies.

Small Clinic, Critical Access Hospital, or Community Health Center	Mid-Size or Regional System	Large Health System or Academic Medical Center
Relies primarily on Safe Harbor de-identification for any AI training data. Uses the EHR vendor's built-in de-identification tools where available. For complex use cases requiring Expert Determination, leverages de-identification services available through the state HIE, regional HCCN, or contracted privacy consultant rather than building internal capability. Privacy officer reviews all AI data use cases before any data is shared with a vendor or used for training.	Written de-identification policy covers both Safe Harbor and Expert Determination. Uses tools for unstructured text de-identification. Validation is performed by the IT data governance team in consultation with the Privacy Officer. All validated datasets are logged in the data acquisition register. Linkage risk assessments documented for any datasets that are combined or joined.	Dedicated privacy engineering team manages de-identification for AI use cases. Expert Determination used for all AI training data, with independent statistical validation. Explores synthetic data generation (e.g., Syntegra, Synthea) for highest-risk use cases involving sensitive populations. Text embedding and LLM inputs treated as patient-level data and prohibited from leaving the secure data environment, consistent with Mayo Clinic Platform's approach.

Real-World Example

A large academic medical center affiliated organization maintains strict patient-level data restriction policies across four categories governing what can leave their secure data environment:

1. Patient-level or row-level data: No patient-level data of any kind may be exported, regardless of perceived re-identification risk.

2. PII data elements: All PII elements — including those previously obfuscated or pseudonymized — are prohibited from export (names, IDs, dates, locations, etc.).

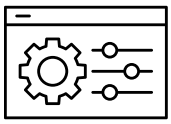
3. Aggregated metrics: Only metrics representing 10 or more patients may be exported; no proportions that could derive a value less than 10.

4. Text embeddings and LLM ingredients: Text embeddings or any ingredients used to train or tune language models are prohibited from export, as they can encode sensitive patient information.

The organizations export review process requires manual review of every data export request by their data team. A consistent export denial rate above 25% results in a temporary block from data access and export submission. All exported files must include a READ ME file documenting column names, source, description, and patient count per row.

Source: Organization's Technical Offboarding — Artifact Export Process, 2026.

Data Quality & Bias



Control 4.3.4: Establish a process and owner for recurring AI data quality and bias evaluation to help identify quality gaps and known sources of bias.

Data quality and bias are two sides of the same patient safety concern. Poor data quality produces unreliable AI models; biased training data produces models that perform poorly for some, and better for others. When thinking about subgroups, note the in addition to subgroups that include demographic breakdowns, disability status, insurance status, and preferred language, there may also be other relevant subgroups where performance differences may matter. Some of this may be use-case dependent, but may include features such as different hospital locations, different device companies, outpatient vs. inpatient, or pediatric vs. adult care departments. In healthcare, a model trained predominantly on data that is not representative of the population you serve may mispredict risk, misclassify conditions, or document differently than expected.

It is rare for healthcare datasets to be truly unbiased, and not all bias leads to harmful outcomes. Every dataset reflects choices about what to measure, who to include, when to collect, and how to label. EHR data, for example, reflects who sought care, who had access to care, what clinicians chose to document, and what billing codes are incentivized. The goal of this control is therefore not to produce bias-free data, but to ensure that organizations know what data they have and for whom. This way, the biases present in AI data are identified, disclosed, and

managed so their effects on model outputs are understood and minimized over time. It also allows organizations to compare the representativeness of their data to what vendors report about their training data.

For safety-net organizations, this control carries particular weight. These organizations disproportionately serve populations that are systematically underrepresented in AI training datasets developed at larger institutions. Deploying AI tools built on non-representative data without quality and bias evaluation creates real risk of worsening existing health disparities.

Implementation Guidance

Sources of Bias in Healthcare AI Data

Understanding where bias originates is the first step to managing it. The following types are the most commonly encountered in healthcare AI:

Bias	Description
Representation Bias	Training data that over-represents certain patient populations (e.g., insured, English-speaking, urban, or those who regularly engage with the health system) and underrepresents others. Models trained on such data may perform well on average but poorly for marginalized groups.
Label Bias	In healthcare we often use proxies when ground truth data is either costly to acquire, or subjective (like risk). For example, a model trained to predict "health need" using health care cost as a proxy will inherit the historical inequities in care access that drove those costs (see Obermeyer et al., 2019).
Temporal Bias	Health data reflects past clinical practice. Models trained on historical data may reinforce outdated care patterns, failing to account for changes in treatment guidelines, disease prevalence, or population demographics.
Survivorship & Selection Bias	Patients with complete, longitudinal records are not representative of all patients. The sickest patients, those who disengage from care, or those who experience gaps in coverage are frequently underrepresented, which can cause models to underestimate risk for high-need populations.
Measurement Bias	Variables used as proxies for non-medical drivers of health such as, zip code for socioeconomic status, insurance type for income, area deprivation index for individual circumstance, can introduce systematic error that compounds across model iterations.

Dimensions of Data Quality for AI

- **Completeness:** Are all required data fields populated or missing data dealt with appropriately? Note that some missing data can be systematic (predictable), and this can introduce bias and reduce reliability for models.
- **Accuracy:** Is the data correctly recorded? Errors in diagnoses, medication orders, lab values, or demographic data directly affect model performance.
- **Consistency:** Is data recorded consistently across facilities, providers, time periods, and EHR systems? Inconsistency is a common challenge in multi-site healthcare organizations.
- **Timeliness:** Is the data current enough for the intended use case? Models trained on pre-pandemic data may not generalize to post-pandemic clinical patterns.
- **Representativeness:** Does the training data reflect the actual patient population the model will be applied to?
    - Example: Datasets drawn predominantly from commercially insured populations may not generalize to Medicaid, dual-eligible, or uninsured patients, and vice versa. Payer type correlates with care access patterns, clinical documentation practices, and social risk factor capture, contributing to some of the bias types noted above, and meaning that models trained on a skewed payer distribution may systematically under- or over-estimate risk for patient groups with a different payer mix. This is particularly relevant for safety-net organizations and FQHCs that deploy AI tools trained primarily on commercially insured populations.

Steps for Implementation

1. Define data quality standards before model development or procurement. Establish acceptable thresholds for completeness, accuracy, and consistency for each dataset used in AI. Document these standards and include them in procurement requirements for vendor tools.
2. Conduct a representativeness assessment. Before using any dataset for AI training or validating/piloting any vendor tool, compare the demographic distribution of the dataset to the population the model will serve. Document any significant gaps--do not assume a mismatch disqualifies a tool, but ensure it is understood and disclosed.
3. Identify and document sources of bias. For each dataset or AI tool, explicitly map which types of bias (representation, label, temporal, selection, measurement) are likely present and how they may affect model outputs.
4. Having data biases does not mean there will be differences in model outputs or performance. Evaluate and monitor AI tool performance across relevant subgroups. Do not rely solely on aggregate performance metrics.
5. Establish a recurring bias monitoring schedule. Bias assessment is not a one-time gate. Assessments should occur at data ingestion, during model training or validation (if applicable), at deployment, and at pre-defined post-deployment intervals. This is because bias can emerge or worsen as patient populations change or data drift occurs. The frequency and rigor of post-deployment bias monitoring should be proportionate to the risk level of the AI tool or the sensitivity of the data.

Real-World Example

A large multi-state health system's Data Governance principles include "Quality" as a core pillar — stating that "all data must be of the appropriate quality for its use within the organization and the quality should be measured and monitored." Their AI Guiding Principles further include "Human-Centric & Societal Beneficial": AI use will serve the goals of healthcare services consumers and be designed with considerations of systematic bias, ethical concerns, and societal well-being at the forefront, with a commitment to continuously mitigating bias and ethical concerns. Their "Actively Monitored" principle ensures all AI solutions are actively monitored for performance and intended outcomes.

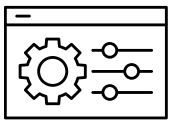
Source: Organization's Strategy to Privacy Analytics Board on AI Governance — Data Management, 2024.

Note: Start Somewhere

If you are overwhelmed about where to start, the answer is somewhere--there is a lot we don't know about how to evaluate data and models for bias. However, understanding your data helps make more informed choices about which solutions are right for your organization and what steps need to be taken to make them right. At the end of the day this should not be a barrier to adoption, but a facilitator.

Consider which subgroups are particularly important for your population and use-case, and start there. For example, a CHC that serves primarily non-English speakers, may want to look at how their Ambient AI solution performs by language subgroups, and may want to capture metadata associated with data language.

Data Preparation Process



Control 4.3.5: Establish a standard operating procedure (SOP) for AI data preparation.

How data is prepared for AI training has a profound effect on model reliability, reproducibility, and governance. In many healthcare organizations, data preparation is ad hoc: different data scientists make different choices about handling missing values, encoding categorical variables, or splitting training and test data. These choices are rarely documented, making it impossible to reproduce model results, audit development decisions, or hand models off between teams.

For traditional ML models, SOPs for AI data preparation may covers data cleaning rules, encoding practices, normalization/standarization methods, and train/validate/test partitioning with appropriate patient-level and temporal separation. For generative and agentic AI, this covers prompt and context management, retrieval corpus

preparation, fine-tuning data curation where applicable, and governance of data sources accessible to agentic systems.

Note that not all organizations will be using data to develop or train models, but must still have data preparation SOPs for data that will be input into the model for use, used to fine-tune the model, or shared with third-party vendors.

Implementation Guidance

Key Elements of an AI Data Preparation SOP

Note that the information presented below is just a snapshot and some may not be applicable given the specific use-case.

1. Data Cleaning Rules (Traditional ML)

Data cleaning should be done to ensure data integrity while maintaining alignment with the intended model deployment data environment.

Traditional methods of dealing with missing data or outliers don't always work well against live, production data, and can impact performance.

How to resolve inconsistencies across data sources: document rules for reconciling conflicting values from different EHR modules, facilities, or encounter types.

2. Encoding Practices (Traditional ML)

Categorical variables: define when to use one-hot encoding vs. label encoding vs. learned embeddings.

Clinical codes (ICD, CPT, SNOMED, RxNorm): document code mapping, aggregation, and normalization rules.

Temporal features: define how dates, times, and time intervals are encoded, including handling of dates for patients over age 89 (HIPAA Safe Harbor requirement).

3. Normalization & Standardization (Traditional ML)

Document which continuous variables are normalized or standardized and by what method (min-max scaling, z-score standardization, log transformation).

Critical rule: normalization parameters (min, max, mean, standard deviation) must be calculated only on training data and then applied to validation and test data. Calculating on the full dataset before splitting constitutes data leakage.

4. Partitioning Rules (Train / Validate / Test Splits) (Traditional ML)

Random splits are generally insufficient for healthcare data.

SOP Considerations:

- Temporal separation: The test set should use data from a later time period than training, simulating the real-world deployment condition where the model will be applied to future patients.
- Patient-level separation: The same patient's data must not appear in both training and test sets. This is a common error that inflates performance metrics because the model has effectively "seen" patients in the test set during training.
- Site-level separation: For multi-site organizations, consider holding out data from one or more sites for external validation.

Document split ratios (e.g., 70/15/15 train/validate/test) and the rationale for the chosen approach.

5. Prompt and Context Window Management (GenAI/Agentic AI)

What data is included in prompts or context windows should be treated as a data preparation decision. For many LLM deployments, the prompt is the only "data" the model ever receives — making its construction the functional equivalent of a training data pipeline in traditional ML.

SOP Considerations:

- Define what categories of clinical data may be included in prompts or context — including whether PHI is permitted and under what conditions.
- Establish prompt templates for common clinical use cases; treat templates as versioned, governed artifacts rather than ad hoc text.
- Document rules for truncation and summarization when clinical records exceed context window limits — truncation decisions affect what information the model acts on.
- Specify logging requirements for prompt content, particularly for any prompts that include PHI, to support audit and incident investigation.

6. Retrieval Corpus Preparation (RAG Systems) (GenAI/Agentic)

For Retrieval-Augmented Generation (RAG) systems, where a model retrieves documents or records to ground its outputs, the quality and governance of the retrieval corpus is the functional equivalent of training data quality. Poorly curated retrieval corpora introduce the same reliability and bias risks as poorly prepared training data, but are often less visible because they sit outside the model itself. These SOP considerations should apply to solutions developed in house or co-developed with vendors, but may also inform vendor discussions. For example, for a vendor providing a RAG-based clinical decision support tool, it may be important to know the metadata of the data sources it relies on and pulls from, and the minimum necessary data it needs to access from within your organizations, to successfully run.

SOP Considerations:

- Define criteria for which documents or records are eligible/ineligible for inclusion in the retrieval corpus (e.g., only verified clinical guidelines, only records from defined date ranges).
- Document chunking and embedding strategies — how documents are divided and indexed affects what the model retrieves and ultimately generates.
- Establish a freshness and update cadence — clinical guidelines, formularies, and protocols change, and stale retrieval corpora produce outdated outputs.
- Define access controls for the retrieval corpus — what patient data, if any, is indexed, and who can query it through the AI system.
- Document metadata tagging standards to support filtering, auditing, and corpus management over time.

7. Fine-Tuning Data Curation (GenAI/Agentic)

When organizations fine-tune models on local data, traditional ML data preparation requirements apply with several important additions specific to the fine-tuning context.

Additional SOP Considerations:

- Instruction-response pair quality: For instruction-tuning, document the process for creating, reviewing, and validating instruction-response pairs. Clinical pairs should be reviewed by domain experts before inclusion.
- Data mix ratios: Document the ratio of domain-specific to general data in the fine-tuning set. Over-reliance on domain-specific data can cause the model to lose general reasoning capabilities.
- Memorization risk assessment: Apply heightened scrutiny to rare-condition cases in fine-tuning data. Models fine-tuned on small clinical datasets are more susceptible to memorization and re-identification risk (see Re-identification Risks, Control 3).
- Consent and use authorization: Document the legal basis for using patient data in fine-tuning. Fine-tuning constitutes a new use of data that should be explicitly covered by existing consents, BAAs, or DUAs.

8. Agentic Tool and Data Access Governance (GenAI/Agentic)

For agentic AI systems — which can query multiple data sources, call external APIs, and take actions autonomously, data preparation governance extends beyond inputs to the model itself. The data sources an agent can access, and how it processes their outputs, constitute a form of runtime data environment that requires the same intentional governance as a training dataset.

SOP Considerations

- Define and document which systems, APIs, and data sources an agentic AI system is authorized to query — this is the agentic equivalent of a training data inventory.
- Establish least-privilege data access: agents should only be able to query systems relevant to the specific task they are authorized to perform. Cross-system data aggregation should require explicit justification.
- Document how tool-call outputs are validated before being used in further reasoning — unvalidated external outputs can introduce errors or adversarial content into the agent's reasoning chain.
- Specify logging requirements for all tool calls and data queries made by agentic systems — not just final outputs. This is essential for audit, incident investigation, and governance review. Access to different data sources, human approvals/override (where relevant), should be logged for audit purposes.
- Govern agentic memory: if the agent maintains persistent memory across sessions, that memory store must be brought under formal data governance (access controls, retention policies, BAA coverage).
- Establish type of human oversight requirements and access controls: Human oversight may need to take risk of the system and sensitivity of the data in mind. For agentic AI systems operating on or with access sensitive data or that fall under higher-risk (e.g. access to patient data/PHI, payment data, or integrated into clinical workflows) define which types of actions/access require human review or confirmation before execution. Even where full automation is the operational goal, initial deployment phases should include meaningful review to validate that the system is behaving as expected before reducing oversight.
- Establish escalation and halt procedures: Define conditions under which an agentic system should pause, escalate to a human reviewer, or decline to take action. (e.g. data operations or access that could lead to security/privacy vulnerables or alter source information, taking actions with potentially irreversible clinical or operational consequences; encountering data quality flags, unexpected inputs, or out-of-distribution conditions; etc.) Document the escalation path clearly, and ensure that frontline users and operators know how to initiate escalation or halt the system.

Implementation Guidance

1. Document current practices. Interview data scientists and analysts to understand how data preparation is currently done. Identify where practices differ across teams, tools, or use cases.
2. Draft an SOP template. Start with the four elements above and adapt to your organization's most common use cases. A one-page checklist is sufficient for smaller organizations; larger organizations may need more detailed playbooks by use case type (e.g., EHR-based prediction models vs. imaging models).
3. Review with clinical or other end-user experts. For example, rules for handling missing lab values, encoding diagnosis codes, or selecting date windows should be reviewed with clinical staff to ensure they reflect clinical reality. This is also true of data that is clinically meaningful as a time-series or trend versus a single time-point.
4. Version-control all data preparation code. Each model version submitted for review or deployment should reference a specific, documented version of the preparation code.
5. Link preparation documentation to model cards. The data preparation SOP version and any deviations should be recorded in the model card (Domain 3: Organizational Resources Playbook, Control 3), enabling future

audits and model updates.

Related Challenges

Below are challenges that organizations identified in by HDOs when working to implement responsible data management controls.

Challenge / Gap	Details	Possible Mitigation Strategies
Vendor Contract Enforcement	Vendors often resist strict data use terms, including audit clauses and penalties for non-compliance. It is rare for HDOs to get vendors to share liability risk. Contracts are often ambiguous about when modified or derived data counts as redistribution.	Use a standard AI DUA addendum as a non-negotiable baseline rather than negotiating from a blank page. Pursue coalition-level negotiations through state hospital associations, HCCNs, or GPOs to increase leverage. Scope audit rights narrowly (e.g., access logs and data handling records) to reduce vendor resistance. Define "modification" and "derivative data" explicitly in contracts to eliminate ambiguity. Document all vendor deviations from standard terms in the AI Inventory Registry.
Provenance and Lineage Tracking	Creating and maintaining data provenance logs and implementing integrated lineage tracking software is technically difficult and resource-intensive. Existing data lineage tools often lag in capturing AI model lineage effectively.	Start with a manual data acquisition register (Control 1) before investing in automated lineage tools. Adopt platforms with native lineage support (e.g., Databricks Unity Catalog, Microsoft Purview) when upgrading data infrastructure. Require vendors to provide data lineage documentation as a procurement condition. Use model cards (Domain 3 Organizational Resources Playbook, Control 3) to capture model-level lineage even when infrastructure-level automation is not yet in place.
Data Quality and Bias Validation	Obtaining representative datasets and conducting recurring validation and testing pipelines — especially alongside third-party recipients — is burdensome. Incomplete data is a common challenge in healthcare, particularly for race, ethnicity, language, and Non-medical drivers of health variables.	Require vendors to provide subgroup performance data before contract execution rather than post-deployment. Use open-source tools (IBM AIF360, Fairlearn, Aequitas) to reduce assessment costs. Partner with regional data networks or state HIEs for access to more representative datasets. Accept and document known limitations explicitly rather than skipping assessment — transparency about bias is more defensible than an unsubstantiated claim of unbiased data. Participate in UDS+ or equivalent data quality initiatives to address upstream gaps.

De-identification Expertise & Cost	Engaging qualified, independent experts for validation of HIPAA Expert Determinations is costly and these experts are scarce, which may slow projects and innovation. Smaller organizations often lack internal expertise for Safe Harbor de-identification of unstructured text.	Default to Safe Harbor de-identification for most AI use cases — it is more accessible and does not require expert validation. Reserve Expert Determination for cases where Safe Harbor is too restrictive and the use case clearly justifies the cost. Explore shared service models through HCCNs, regional HIEs, or state hospital associations for pooled de-identification expertise. Work with vendors to accept de-identified data rather than PHI wherever technically feasible, reducing the overall burden.
Minimum Necessary Data	Organizations often export or share more data than is essential for AI development or use, creating compliance, privacy, and governance risks. Both internal teams and vendors frequently default to requesting broader datasets, requiring strong discipline to limit data to what is strictly needed.	Implement a formal data request review process requiring documented justification for every data element requested. Embed minimum necessary assessments into the data acquisition register — each entry should document why each field is needed for the intended purpose. Use the AI DUA addendum to contractually require vendors to document and justify the data elements they request. Establish a privacy officer or data steward review gate for any AI data export exceeding a defined volume or sensitivity threshold.
Keeping Pace with State Privacy Law Changes	The state health privacy law landscape is rapidly evolving. New laws in Washington (MHMDA), California (CPRA), Colorado (CPA), and other states impose obligations on de-identified or pseudonymized data that go beyond HIPAA, requiring ongoing contract and policy updates.	Designate a privacy officer or legal counsel owner responsible for monitoring state law developments. Use a modular DUA addendum structure so that state-specific provisions can be updated without renegotiating the full agreement. Subscribe to resources such as the IAPP State Privacy Law Tracker or HHS OCR guidance updates. Build state law review into the annual contract monitoring cycle described in the AI DUA addendum.

Tools & Resources

Included below are national and international standards (sector-agnostic) related to data management. These resources are shared for reference and do not signify promotion or endorsement.

Core Data Governance and Maturity Frameworks

- EDM Council – DCAM (Data Management Capability Assessment Model)
  - 🌐 Data Management – DCAM – EDM Council
  - Industry-standard capability model for data governance, quality, metadata, and lineage.
- CMMI Institute – Data Management Maturity (DMM / CMMI Data)
  - https://dev12.cmmiinstitute.com/products/dmm/
  - Formal staged maturity model for enterprise data management process
- DAMA International – DAMA-DMBOK (Data Management Body of Knowledge)
  - 🌐 DAMA® Data Management Body of Knowledge (DAMA-DMBOK®)
  - Canonical definitions and functional framework for data governance, stewardship, and quality.

Healthcare-specific Governance and Maturity Context

- HIMSS – Analytics Maturity Assessment Model (AMAM)
  - 🌐 Analytics Maturity Assessment Model (AMAM)
  - Healthcare-focused maturity model linking data governance, analytics, and outcomes.

Research Data Governance

- GO FAIR Initiative – FAIR Principles
  - 🌐 FAIR Principles – GO FAIR
  - Standards for making data findable, accessible, interoperable, and reusable.

Responsible Use of Health Data Certification

- Joint Commission – Voluntary RUHD Certification
  - 🌐 Responsible Use of Health Data
  - Certification aims to objectively evaluate whether an organization is following best practices in their secondary use of **de-identified data**. It promotes the responsible use of healthcare data by ensuring that protocols are in place for transparency, limitations of use, and patient engagement.

Next Steps

Links below will connect you to other AI governance playbooks in this series.

Domain 1: AI Policy

Domain 2: Organizational Structures

Domain 3: Organizational Resources

Domain 4: Organizational Processes

- Subdomain 4.1: Responsible AI Lifecycle Management and Use
- Subdomain 4.2: Risk & Impact Assessments
- Subdomain 4.3: Responsible Data Management and Use
- Subdomain 4.4: Third Party Management
- Subdomain 4.5: Education, Training, and Feedback

Attribution

CHAI AI Governance Playbooks were developed through the contributions and collaboration of more than 150 organizations across the healthcare ecosystem. CHAI extends its sincere appreciation to all participating organizations and contributors for their time, expertise, and thoughtful feedback throughout the development process. We would also like to provide special recognition to the individuals below who requested attribution by name for their contributions to these efforts:

- Josh Hjelmstad, MHA, AVIA Health
- Sandra Gregg, DNP, MHA, RN, PMP, LSSBB, CCMP, HCD, Booz Allen Hamilton, Inc.
- Vidhi Vinay Kothari, Carnegie Mellon University
- Stephen McLaughlin, PhD, Children's National Hospital
- Ryan Marshall Felder, PhD HEC-C, Cleveland Clinic
- Po-Hao Chen, MD, MBA, Cleveland Clinic Foundation
- Katherine Mulready, JD MPS, cliexa
- Holden Groves, MD, Columbia University/NewYork-Presbyterian
- Shakira J. Grant, MBBS, MSCR, CROSS Global Research & Strategy
- Namrata Rastogi, MBBS MRCGP MS, Enigma Ventures
- Aleocidio Sette Balzanelo, MD, MBA, FOLKS
- Rocco Orlando, MD, Hartford HealthCare
- Michael Blumental, HealthLeap
- Romanus M. Joseph, Innovative Health Accountable Care Organization (ACO)
- Rebecca G. Mishuris, MD, MS, MPH, Mass General Brigham
- Nasibeh Zanjirani Farahani, PhD, CSM, Mayo Clinic
- Jason Gillman, MD, FIDSA, MCG Health
- Douglas Graham, Individual Contributor
- Nicole Petroski, Mercy Health Services
- Haris Shuaib, Newton's Tree
- Rémy Ibarcq, Individual Contributor
- Srikar Nallan, MS, Stanford Healthcare
- Jaimie Weber, MD, Tampa General Hospital
- R. Brandon Hunter, MD, FAAP, Texas Children's Hospital
- Angela L. Lipscomb, JD, LL.M, The University of Texas MD Anderson Cancer Center
- Christopher H. Lee, MBA, BSN, RN-BC, UCLA Health
- Abby Steele, Individual Contributor
- Shiba Kuanar, University of Minnesota
- Joshua Miller, JD, CHC, University of Rochester Medicine
- Anne Travis, MD, MSc, Wolters Kluwer
- Michael L. Shaw, JD, ZS

License