Subdomain 4.5: Education, Training, and Feedback 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

Subdomain 4.5: Education, Training, and Feedback Playbook defines the approach for educating and training staff, informing patients, and capturing feedback and incidents related to AI solutions in healthcare delivery organizations (HDOs).

Background Context for Subdomain 4.5: Education, Training, and Feedback

This playbook outlines a flexible framework along with implementation guidance for HDOs to document and implement processes for education, training, and gathering user feedback. Additional context regarding training and feedback is included below.

Training

Several types of training may occur when implementing health AI solutions within HDOs. These include:

1. **General AI Training**
   a. Description: Foundational training provided to all relevant staff that builds baseline AI literacy across the HDO. This includes understanding what AI is and is not, the limitations and potential failure modes of AI systems, organizational policies governing AI use, and how to identify and report concerns. General AI training is typically role-agnostic and is delivered organization-wide on a recurring basis, ensuring that all relevant staff understand the broader context in which these systems operate.
2. **Use Case-Specific Training**
   a. Description: Targeted training designed for staff who directly interact with a particular AI system or workflow. This training covers the specific purpose and intended use of the solution, how to interpret its outputs in clinical or operational context, when and how to override or escalate the AI's recommendation, and the safeguards in place for that particular solution. Use case-specific training is typically delivered at or before system go-live and refreshed whenever the system is updated or performance concerns are identified.
3. **New Hire Training**
   a. Description: A dedicated AI orientation for staff joining the organization who will work with or alongside AI solutions. Rather than waiting for the next scheduled general training cycle, onboarding training ensures new employees develop baseline AI literacy before they encounter these systems in clinical or operational practice.
4. **Leadership and Governance Training**
   a. Description: Targeted training for executives, board members, and/or AI Governance Committee members focused on strategic AI oversight responsibilities rather than day-to-day solution use. This covers risk management, regulatory trends, ethical obligations, and how to evaluate AI governance program performance.
5. **Vendor-Delivered Training**
   a. Description: Training provided directly by the AI vendor on their specific system, typically during pre-deployment, deployment, or following a major update. This is distinct from use case-specific training in that it originates externally and may include access to vendor-supported resources, certification programs, or product roadmap context that the HDO cannot provide independently.
   b. Note: Vendor-provided training support varies significantly in practice. Some vendors offer hands-on, dedicated training assistance, while others provide more limited resources (e.g., documentation-only). HDOs should assess available vendor training support before deployment and plan for internal training capacity accordingly.
6. **Incident Response Training**
   a. Description: Scenario-based training that prepares staff to recognize, escalate, and document AI-related incidents when they occur. This is procedural: teaching staff what to do when an AI system produces unexpected or potentially harmful outputs. Organizations should define minimum training requirements by role and AI risk tier, including required completion, refresh frequency, and competency validation for higher-risk AI solutions.
7. **Patient and Community Training**
   a. Description: Training and educational efforts that build patient, family, caregiver, and broader community capacity to understand and meaningfully engage with AI as it is used in their care and across the HDO. The aim is to develop durable AI literacy so patients can ask informed questions, identify concerns, participate in advisory bodies, and provide substantive input across the AI lifecycle. Examples may include plain-language explainers and FAQs, community town halls or listening sessions, and structured curricula delivered through Patient and Family Advisory Councils (PFACs) or related councils.

Feedback

There are several types of potential feedback outlets when implementing health AI solutions at HDOs. These include:

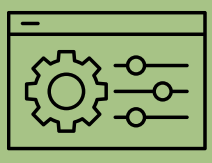
1. **Healthcare Delivery Organization (HDO) to Vendor**
   a. Description: The HDO communicates performance findings, clinical concerns, workflow friction points, and incident data back to the vendor. This feedback loop ensures the vendor understands how the system is performing in real-world clinical and operational contexts and creates shared accountability for ongoing improvement, bug resolution, and model updates.
2. **Vendor to Healthcare Delivery Organization (HDO)**
   a. Description: The vendor communicates proactively to the HDO about known issues, system updates, performance benchmarks, changes to model behavior, and emerging risks identified across their broader customer base. Effective vendor-to-healthcare delivery organization feedback ensures the organization can anticipate changes, update training materials, and adjust governance processes before issues escalate.
   b. Note on Shared Accountability in Contracting: Effective vendor-to-HDO feedback should not be assumed to occur voluntarily. Organizations should ensure vendor contracts include explicit obligations for proactive incident notification, performance reporting, and communication of model changes. Embedding these requirements contractually supports shared accountability and helps manage the costs of ongoing AI oversight.
3. **Staff to Healthcare Delivery Organization (HDO)**
   a. Description: Staff at HDOs may share observations, concerns, and suggestions about AI system performance through internal reporting channels. This includes flagging unexpected outputs, potential bias, workflow disruptions, or patient safety concerns. Creating safe, accessible channels for this type of feedback (e.g., whistleblower protections) is critical for surfacing issues that may not be visible through technical monitoring alone.
4. **Staff to Patient**
   a. Description: In a clinical context, frontline staff may communicate directly with patients about how AI may have been used in their care, answer questions, clarify outputs, and surface non-AI alternatives where appropriate. This bedside-level transparency complements written notices and is often considered the most trusted patient-facing channel for building understanding and trust.
5. **Patient to Healthcare Delivery Organization (HDO)**
   a. Description: Patients and caregivers may provide feedback about their experience when AI informs or influences care. This may include concerns about how AI-generated recommendations were communicated, questions about data use, or requests for human review of AI-assisted decisions. Collecting and acting on patient feedback supports trust-building and informed consent processes.
6. **Healthcare Delivery Organization (HDO) to Healthcare Delivery Organization (HDO)**
   a. Description: HDOs may share AI-related performance data, incident learnings, and governance practices with peer organizations through regional networks, shared governance models, or collaborative learning communities. This peer feedback loop supports sector-wide improvement, enables benchmarking, and helps surface systemic issues not visible within a single organization.
7. **Healthcare Delivery Organization (HDO) to Patient or Community**
   a. Description: HDOs may proactively communicate AI-related information to patients and communities, including how AI was used in their care, updates to privacy practices, changes to AI-enabled services, or opportunities to provide feedback. This direction of communication, which may in some cases originate from vendor-provided information passed along by the organization, supports transparency and patient trust.
   b. Examples of patient or community feedback may include representation on AI advisory committees or governance working groups, structured participation through Patient and Family Advisory Councils (PFACs) or community health advisory boards, inclusion in pilot evaluation panels, and community town halls or listening sessions for higher-risk AI solutions.

AI Incidents

Typically, HDOs expand their existing, non-AI incident category of **Compliance and Regulatory** incidents into three subcategories. These subcategories are People, Data, and AI. Understanding these subcategories helps organizations route incidents appropriately, conduct root cause analyses, and design corrective actions.

- **People**: Incidents driven by human behavior in relation to AI systems. This may include automation bias (over-relying on AI outputs without adequate critical review), failure to escalate concerns, misinterpretation of AI recommendations, or gaps in training.
  - For example, a clinician routinely accepts an AI-generated medication dosing suggestion without cross-referencing patient-specific contraindications, resulting in an adverse drug event.
  - Root cause analysis should examine upstream human-factor contributors (e.g., insufficient onboarding, role-skill mismatch, production pressures that discourage critical review).
- **Data**: Incidents stemming from problems with the data an AI system ingests, was trained on, or produces. This may include population shifts (data drift), biased or non-representative training datasets, data corruption, or adversarial inputs.
  - For example, a risk stratification model trained primarily on one patient population underperforms for a different patient population (e.g., a specific race, ethnicity, age, sex, or primary langiage group), systematically under-triaging patients from that group.
  - Data incidents should trigger escalation through the HDO's data governance channel alongside the AI Governance Committee, with data stewards engaged in root cause analysis. AI incident reports should reference the model version and data vintage active at the time of the incident to enable traceability.
- **AI**: Incidents attributable to the AI model, solution, or system itself. This may include model drift (performance degradation over time), software defects, opaque outputs that cannot be interpreted or verified, or unexpected behavior in edge cases not covered by training data or the AI system.
  - For example, an ambient AI solution begins generating inaccurate transcriptions following a software update, leading to documentation errors that go undetected for several weeks.

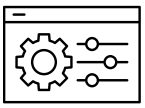
Most HDOs already have cascading event reporting mechanisms in place. The controls in this playbook are generally not intended to require entirely new processes; rather, the controls are often best operationalized by adapting existing patient safety reporting infrastructure to incorporate AI incidents, tags, or escalation pathways. Keeping the above types of training and feedback information in mind, the implementation guidance below describes how HDOs may consider operationalizing each baseline operational control.



Baseline Operational Controls for Subdomain 4.5: Education, Training, and Feedback

Category	Control
Training	Develop a process to ensure all relevant staff receive training on AI solutions, including their purpose, limitations, and related information. Reinforce training via recurring briefings and optional seminars for deeper engagement.
Education	Provide all patients, no later than the date of first service delivery, with a comprehensive Notice of Privacy Practices (NPP), including how protected health information (PHI) may be used by AI solutions, in compliance with HIPAA.
Education	Develop a process for patient transparency and/or consent where appropriate for relevant AI solutions.
Feedback	Develop a process to document, escalate, and notify relevant parties about AI incidents. Ensure it is clearly communicated that all parties involved in an AI incident share responsibility for reporting any voluntary or mandatory incidents.
Feedback	Ensure information included in AI incident communication falls under relevant federal and state laws and protections.
Feedback	Provide whistleblower protections for both internal and external reporting, including anonymity, confidentiality, non-retaliation assurances, and clear disclosure of data retention policies.

Training: Ensure All Relevant Staff Receive Training on AI Solutions



Control 4.5.1: Develop a process to ensure all relevant staff receive training on AI solutions, including their purpose, limitations, and related information. Reinforce training via recurring briefings and optional seminars for deeper engagement.

Implementation Guidance

HDOs should establish a structured, recurring training program that ensures all relevant staff (across clinical, administrative, technical, and leadership roles) understand the purpose, intended use, and known limitations of each AI system deployed within their environment. Training should be role-differentiated and delivered through multiple modalities, as applicable. These may include general AI literacy training for all staff, use case-specific training for those who interact directly with a particular solution, new hire onboarding, leadership and governance training, vendor-delivered instruction, and incident response preparation. A "just-in-time" approach is

recommended, with training triggered at or before system go-live and refreshed whenever a system is updated or performance concerns arise. Where workflows incorporate multiple AI solutions, organizations should embed AI training within existing role-based education programs and competency frameworks rather than creating parallel tracks for each solution. Supporting resources such as AI fact sheets, model cards (such as the CHAI Applied Model Card), or solution documentation should remain accessible to staff between formal training cycles for self-directed review, if interested. To sustain accountability and consistency over time, training records should be maintained and reviewed on a recurring basis, training responsibilities should be clearly assigned within each department, and organizations should consider designating AI Governance Champions (or equivalent) to provide ongoing peer support and serve as a first point of escalation for staff questions. Completion of required training should be a prerequisite for accessing or using AI systems.

When able, training staff in non-patient environments (e.g., sandbox systems, test EHR instances, or simulation settings) allows organizations to assess system performance, surface known limitations and failure modes, and train end users on how to interact with the solution before patient safety is at stake. Training delivered during this piloting phase most often takes the form of use case-specific training and vendor-delivered training, both of which are effective when conducted in the same environment where testing occurs.

Note

To validate that training is achieving its intended goals, HDOs should implement a tiered competency validation approach calibrated to the risk level and workflow impact of each AI solution. A few examples are included below:

1. For general AI literacy training, a simple attestation of completion or awareness check may suffice.
2. For staff who directly interact with higher-risk or workflow-changing AI solutions, a knowledge assessment or scenario-based check is recommended.
3. Whenever a system undergoes material updates, a re-certification or refresher training should be triggered to ensure continued competency.

Note

Included below are examples of role-specific training focus areas:

- Clinicians: AI system use, updates, override protocols, and known risks.
- AI and IT Staff: maintenance, monitoring, and system performance tracking.
- Compliance and Risk: governance policies, regulatory requirements, and audit processes.
- Operations: workflow integration, escalations, and patient-facing implications.
- Department Lead(s): report system oversight.
- Ethics: understanding the ethical obligations arising from AI use in care delivery, including the organization's principles for ensuring AI use aligns with its mission and values.

Real-World Example: Piloting to Inform Scaling and Training (Lessons from KPNC's Advance Alert Monitor Program)

This example is drawn from a published peer-reviewed source and is therefore an attributable Real-World Example.

Kaiser Permanente Northern California (KPNC) piloted an AI-based early warning system at two hospitals before expanding to 19 more. The pilot was essential not just as a proof of concept, but as a diagnostic phase that revealed what scaling would actually require.

Key pilot findings that shaped scaling and training:

The pilot exposed significant variation in rapid response team (RRT) staffing, training standards, and documentation practices across hospitals, making clear that a standardized operational infrastructure had to be built before expansion. It also surfaced alert fatigue as a major adoption barrier, leading to the architectural decision to interpose a remote Virtual Quality Team (VQT) of nurses between the algorithm and frontline clinicians; a solution that emerged from observing clinician behavior during the pilot.

How training was designed from pilot learnings:

Training was role-specific (VQT nurses, RRT nurses, hospitalists, social workers) and grounded in the workflow realities the pilot made visible. Important to note, KPNC instituted a mandatory two-week shadowing period before each hospital's go-live, during which alerts were shown to selected clinicians for workflow practice runs (functioning as a simulation environment where staff could rehearse the system before patient safety was at stake). Daily debriefs continued for at least two weeks post-launch to capture feedback and refine workflows.

No hospital was authorized to go live without meeting formal readiness criteria spanning IT, leadership, administrative, and training domains which ensured deployment authorization was a deliberate gate.

Source: Paulson SS, Dummett BA, Green J, et al. What do we do after the pilot is done? Implementation of a hospital early warning system at scale. *The Joint Commission Journal on Quality and Patient Safety.* 2020;46(4):207-216.

Key Tools & Resources

International Association of Privacy Professionals (IAPP) Artificial Intelligence Governance Professional (AIGP)

- *The AIGP credential demonstrates competency in AI development, the ability to conduct ethical AI deployment, and the ability to utilize best practices in AI management to ensure safety and trust. AIGP is for any professional tasked with developing AI governance and risk management in their operations.*

Association of American Medical Colleges (AAMC) AI Competencies Across the Learning Continuum

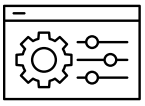
- *In-development, targeted publication in Fall 2026: AAMC is developing a national set of competencies in AI designed to be relevant for undergraduate, graduate, and continuing medical education.*

Implementation Guiding Questions

Below are standard questions that HDOs may ask to guide discussion for **control 4.5.1**:

- Have you inventoried which staff roles interact with or are affected by each deployed AI system?
- Does your training program cover the purpose, intended use cases, and known limitations of each AI solution (e.g., via an AI fact sheet, model or solution card, or similar documentation)?
- Is training differentiated by role? Do clinicians, administrative staff, IT staff, and leadership each receive content appropriate to their level of exposure and accountability?
- Does training include both knowledge-based content and scenario-based exercises with testing or knowledge checks?
- Is training delivered using a "just-in-time" approach (e.g., conducted when the AI system is about to be used and refreshed when the system has major updates)?
- Are training records maintained and reviewed on a recurring basis (e.g., annually)?
- Are training audits maintained and reviewed on a recurring basis to ensure users are adherent and the training itself is effective?
- Is a process in place to trigger supplemental training sessions before or at go-live for new AI deployments?
- Are AI fact sheets, model or solution cards, or similar documentation accessible to staff for self-directed refreshers between formal training cycles?
- Have you designated AI Governance Champions (or equivalent) within departments to provide ongoing peer support and serve as first-point-of-escalation for questions?
- Is training responsibility clearly assigned (e.g., led by a senior member of each department) to reinforce accountability?
- Are end users involved in piloting before full rollout? Is their feedback formally collected and acted on?
- Are known limitations and failure modes documented and communicated to staff prior to go-live?

Education: Provide a Notice of Privacy Practices for AI Use



Control 4.5.2: Provide all patients, no later than the date of first service delivery, with a comprehensive Notice of Privacy Practices (NPP), including how protected health information (PHI) may be used by AI solutions, in compliance with HIPAA.

Implementation Guidance

HDOs have a legal obligation under HIPAA to inform patients how their protected health information may be used or disclosed through the Notice of Privacy Practices (NPP), which must be provided no later than the first service encounter. As AI systems become more embedded in clinical and operational workflows, organizations should consider updating their NPP and related patient communications to clearly explain how patient data may be used by advanced analytics or AI-enabled systems, what safeguards are in place, and what rights patients retain under HIPAA.

The HIPAA requirement to provide an NPP no later than the date of first service delivery may not apply in emergency treatment situations, where the NPP should generally be provided as soon as reasonably practicable after the emergency. Organizations should consult with legal counsel regarding applicable exceptions.

Note

HIPAA Compliance and AI Disclosure: HIPAA requires covered entities to provide patients with a Notice of Privacy Practices (NPP) no later than the date of first service delivery, and within 30 days of a patient's request. As AI systems increasingly use PHI to generate recommendations, predict risk, or support clinical decisions, organizations should review and update their NPP to explicitly describe AI-related data use. Failing to disclose AI use in the NPP may constitute a HIPAA violation if AI systems process, analyze, or generate outputs using patient health information without appropriate disclosure.

Source: Notice of Privacy Practices for Protected Health Information

Key Tools & Resources

HIPAA Notice of Privacy Practices (NPP) template

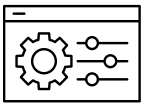
- *The HIPAA Privacy Rule requires health plans and covered health care providers to develop and distribute a notice that provides a clear, user friendly explanation of individuals' rights with respect to their personal health information and the privacy practices of health plans and health care providers. HHS developed these model notices to help improve patient experience and understanding and to assist health care entities to meet new requirements.*

Implementation Guiding Questions

Below are standard questions that HDOs may ask to guide discussion for **control 4.5.2**:

- Has your organization updated its Notice of Privacy Practices (NPP) to describe how patient PHI may be accessed, processed, or used by AI-enabled systems?
- Is a process in place to review and update the NPP when privacy practices materially change, including when new technologies are deployed?
- Are staff trained, or is there standard documentation in place, to understand and explain organizational privacy practices related to AI-enabled systems to patients who have questions about the written notice?

Education: Patient Transparency and/or Consent for AI Solutions



Control 4.5.3: Develop a process for patient transparency and/or consent where appropriate for relevant AI solutions.

Implementation Guidance

Transparency and consent are distinct. Most AI-enabled workflows require transparency (i.e., clear, accessible disclosure of how AI may inform care) and do not require formal patient consent. Formal consent should be reserved for AI uses that materially change the standard of care, introduce discrete clinical risk, or require

collection of a new category of patient data. Consistent with longstanding industry best practice for other clinical technologies, organizations are not expected to enumerate every AI component embedded in care delivery within a consent form. Where an AI solution materially influences a clinical recommendation, patients should also be informed of any reasonable non-AI alternatives available to them.

Transparency with patients about AI use in their care works well when it is woven naturally into existing touch points of the patient journey. Organizations can start by identifying where patients are most likely to encounter AI-informed decisions or AI-generated outputs, and ensuring that materials already in use, such as printed handouts, consent forms, and patient portal content, accurately reflect how AI is being used in those contexts. Where possible, existing workflows and documents can be updated incrementally to include plain-language AI descriptions, rather than building new materials from scratch.

In settings where patients may interact directly with AI solutions, such as scheduling assistants or triage support, a simple "Request Human Assistance" option can go a long way toward building trust and accommodating patients who prefer a different experience. Keeping materials current may not require a heavy lift if a lightweight review process is built into existing deployment and update workflows. Staff play a natural role here as well: use case-specific training helps clinical and front-line staff answer patient questions about AI in their care setting, drawing on the same knowledge they already use day-to-day. Where capacity allows, consulting Patient and Family Advisory Councils or similar groups can help ensure that transparency materials reflect what patients actually want to know, which could lead to simpler, more effective communication than organizations might develop on their own.

Note: When to Escalate to Formal Consent

Formal patient consent should generally be reserved for AI uses that meet one or more of the following thresholds:

- The AI materially changes the standard of care or substitutes for clinician judgment in a discrete decision.
- The AI introduces a new clinical risk not previously contemplated by general consent to treatment.
- The AI requires collection of a new category of patient data (e.g., voice or biometric capture).
- The AI use is investigational or part of a research protocol.
- Applicable state or specialty regulation requires explicit consent for the use case (e.g., certain genomic or behavioral health applications).

HDOs should also address how patient opt-out requests are handled. While opt-out may not be possible for all AI applications, HDOs should:

1. **Identify which AI uses allow opt-out.** The HDO maintains an AI inventory of its AI solutions with a clear designation for each: opt-out eligible or not.
2. **Communicate proactively to patients.** Admission paperwork and the patient rights notice include plain-language statements informing patients that AI solutions may be used in their care and that they may request to opt out of certain uses. Staff are trained to briefly explain this during the admissions process.
3. **Establish a consistent honoring process.** When a patient opts out, the request is documented in the internal system(s) and flagged for the care team.
4. **Acknowledge what cannot be opted out.** The HDO is transparent that some AI uses are not eligible for opt-out, and explains why in accessible terms.

Illustrative Example

A regional HDO updated its patient intake packet to include a one-page plain-language insert titled "How We Use Artificial Intelligence in Your Care." The insert explained three AI solutions in use at the organization and how patients could ask questions or opt out where applicable. The same content was added to the patient portal as a static resource page, with a link from the appointment confirmation emails.

Key Tools & Resources

American Hospital Association (AHA) Patient and Family Advisory Councils (PFAC) Blueprint

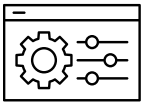
- *To capture learnings and help the field build and sustain high performing PFACs, the AHA gathered a group of patient and family engagement leaders to share their insights. Use the Patient and Family Advisory Councils Blueprint to help guide your organization through the process of forming and maintaining a successful PFAC.*

Implementation Guiding Questions

Below are standard questions that HDOs may ask to guide discussion for **control 4.5.3**:

- Have you identified touchpoints in the patient journey where AI-informed care or AI-generated outputs may be encountered?
- Do your printed materials, patient portals, and consent forms accurately and plainly describe AI use in the relevant care context?
- Is there a consistent, accessible mechanism for patients to ask questions or raise concerns about AI in their care at any point during their visit or episode?
- Are visual cues (e.g., signage, posters in waiting areas or clinical settings) present where AI is actively used in patient-facing interactions?
- Are patient transparency materials reviewed and updated whenever new AI systems are deployed or when existing systems are materially updated?
- Are Patient and Family Advisory Councils (PFACs) or equivalent bodies consulted in the design or review of patient-facing AI transparency materials?
- Are transparency materials available in accessible formats for patients with disabilities and for patients with limited AI literacy?

Feedback: AI Incident Reporting and Notification



Control 4.5.4: Develop a process to document, escalate, and notify relevant parties about AI incidents. Ensure it is clearly communicated that all parties involved in an AI incident share responsibility for reporting any voluntary or mandatory incidents.

Implementation Guidance

When an AI incident occurs in an HDO, timely, structured notification is essential to contain risk and preserve accountability. HDOs should establish that incident reporting is a shared responsibility: all parties involved, not just the most senior staff member present, are expected to report, and department head(s) should be formally notified following any mandatory reporting event. In practice, HDOs can build this into existing incident management workflows by adding AI-specific incident categories to current reporting systems, developing standard communication templates for common AI incident types to ensure speed and legal consistency, and defining clear internal service level agreements; for example, internal notification within 24 hours and regulatory notification within 72 hours, where applicable. Equally important is closing the loop: staff who report incidents should be informed of outcomes and corrective actions taken, which reinforces a proactive reporting culture rather than a reactive one. By anchoring AI incident notification within existing patient safety and quality improvement infrastructure, HDOs can operationalize this control without building parallel processes from scratch.

Included below is an *example* of how notification may be differentiated by incident type.

Notification Audience	Typical Trigger	Responsible Party
Department Head	Any AI anomaly or unexpected output affecting patient care.	Staff member who identifies the issue.
AI Governance Committee (or similar governance body)	All formally classified AI incidents.	Department Head, Safety Officer, or similar role.
Patient, Family, or Caregiver	AI-related adverse event affecting a specific patient, and where the patient was aware of the AI's involvement; material near-misses or workflow changes that affect care. Notification may be paired with plain-language context that reinforces the patient's role as a partner in detection and reporting.	Treating clinician.
Legal or Compliance	Any incident with potential risk or regulatory implications.	AI Governance Committee, or similar role.
Vendor	Suspected software defect or performance issue.	IT, Technical team, or similar role.
Patient & Family Advisory Council (PFAC) or Community Advisory Body	AI incidents or material changes affecting the broader patient population, access, or community trust; periodic AI governance updates regardless of incident status; pre-deployment, pilot, and sunset decisions for higher risk AI solutions.	AI Governance Committee, with support from a designated patient-engagement or community-engagement lead.

Note

HDOs may consider whether existing incident management infrastructure (e.g., an EHR-integrated safety reporting module, a risk management platform) can be extended to capture AI-specific incident fields rather than building a parallel system from scratch. Adding an “AI involvement” tag or category to existing workflows can significantly reduce implementation burden.

Key Tools & Resources

Example Template: Closing the Feedback Loop on AI Incident Reporting

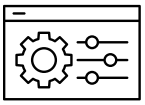
- HDOs should implement and maintain a documented process for closing the feedback loop on AI incident reporting, including communicating outcomes and visible follow-up actions to staff. This will help to sustain trust and preserve early identification of AI-related risks. Examples included are illustrative, derived from industry best practice, and not prescriptive.

Implementation Guiding Questions

Below are standard questions that HDOs may ask to guide discussion for **control 4.5.4**:

- Has your organization defined what constitutes a "mandatory" vs. "voluntary" reportable AI incident, and are those definitions clearly communicated to relevant staff?
- Is it explicitly communicated to all parties involved in an AI incident that they share individual responsibility for reporting (not only the most senior person in the chain)?
- Is the head of the relevant department(s) notified after every mandatory reporting incident, and is this notification documented?
- Is there a documented internal notification chain, including escalation pathways if the department head is unavailable? This may vary depending on centralized, decentralized, or distributed organizational and governance structure.
- Does your incident reporting system capture information, such as: (a) severity of the incident; (b) frequency; (c) type of incident; (d) how AI was involved; and (e) contact information for follow-up by a supervisor?
- Are time-bound notification requirements defined and enforced (e.g., internal notification within 24 hours, regulatory notification within 72 hours for applicable incidents)?
- Are standard communication templates available for common incident types to ensure consistency, appropriate legal framing, and speed of response?
- Is incident reporting designed to be proactive rather than reactive (e.g., are staff empowered to report early concerns before they escalate to formal incidents)?
- Does your incident management process include a feedback loop so that the staff who reported an incident are informed of the outcome and any corrective actions taken?

Feedback: Compliance in AI Incident Communication



Control 4.5.5: Ensure information included in AI incident communication falls under relevant federal and state laws and protections.

Implementation Guidance

When AI incidents occur in HDOs, the data captured in those reports, such as clinical outcomes, patient identifiers, model outputs, and workflow logs, is subject to the same federal and state privacy frameworks that govern all healthcare information, plus an expanding layer of AI-specific regulation. HDOs should ensure their legal and compliance teams have reviewed what gets documented in AI incident reports and confirmed it aligns with HIPAA

and applicable state law. For AI systems classified as Software as a Medical Device (SaMD), U.S. Food and Drug Administration (FDA) reporting obligations add another layer. Much of this work can be embedded into existing compliance review cycles: legal teams can add AI incident data categories to their standard annual HIPAA risk assessments, and staff training on incident communication can be updated to include a brief module on what can, and cannot, be shared externally.

Note

Mandatory incident reporting timelines and data handling requirements vary by state. This variability may persist across AI-enabled systems. Two illustrative examples:

- **Pennsylvania**: The Medical Care Availability and Reduction of Error (MCARE) Act of 2002 requires hospitals to report serious events to the Pennsylvania Department of Health within 24 hours of confirmation, while incidents are reported through the Pennsylvania Patient Safety Reporting System administered by the Pennsylvania Patient Safety Authority.
- **New York**: The New York Patient Occurrence Reporting and Tracking System (NYPORTS) requires reporting of certain adverse events to the New York State Department of Health.

Note: this content is informational only and does not constitute legal advice.

Source: Medical Care Availability and Reduction of Error Fund

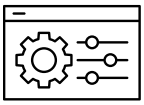
Source: The New York Patient Occurrence Reporting and Tracking System (NYPORTS)

Implementation Guiding Questions

Below are standard questions that HDOs may ask to guide discussion for **control 4.5.5**:

- Has your legal and compliance team reviewed what categories of data are captured in AI incident reports, and confirmed they are handled appropriately under HIPAA and applicable state law?
- Have you identified your state-specific mandatory AI or patient safety reporting obligations (e.g., MCARE Act in Pennsylvania, NYPORTS in New York, and similar state regimes)?
- Are staff trained on their legal obligations, and limitations, when communicating about AI-related incidents internally and externally?
- Are external regulatory reporting requirements identified and documented (e.g., FDA for AI systems classified as Software as a Medical Device)?
- Is there a process to review and update incident communication protocols as new AI-specific regulations emerge at the state or federal level?

Feedback: Whistleblower Protections



Control 4.5.6: Provide whistleblower protections for both internal and external reporting, including anonymity, confidentiality, non-retaliation assurances, and clear disclosure of data retention policies.

Implementation Guidance

HDOs should ensure a robust, accessible incident reporting system is in place. This system should be one that staff, patients, and other relevant parties can use to flag AI-related incidents and concerns. Most organizations can build on existing event reporting infrastructure (e.g., patient safety reporting platforms (PSOs), compliance hotlines) by adding AI-specific categories or tags, rather than creating an entirely separate system.

Note: How to Build AI Incident Reporting into Existing Infrastructure

Audit existing reporting channels. Review current patient safety platforms, compliance hotlines, and peer review processes to identify which systems staff, patients, and other relevant parties already use and trust.

Add AI-specific categories rather than building a new system. Incorporate an AI-related concern tag or category into existing reporting tools, with a few structured prompts such as which AI solution was involved, what the concern was, and any surrounding context.

Ensure accessibility across reporter types. Confirm that reporting pathways are available to all relevant HDO staff and patients.

Communicate the pathway broadly. Inform staff of how and where to report AI-related concerns during training, onboarding, and through supporting documentation such as AI fact sheets or model cards.

Assign clear ownership for review. Route AI-related reports to a designated individual or governance body responsible for reviewing concerns, identifying patterns, and ensuring follow-up action.

Health AI systems introduce a new category of reportable concern (such as model bias, unexplained model outputs, and patient safety risks tied to AI-driven recommendations) that may not be fully captured under an HDO's existing incident reporting infrastructure. This control calls on HDOs to extend their whistleblower protection frameworks to explicitly cover AI-related disclosures, ensuring that both HDO staff and external parties (including patients and vendors) can report concerns safely, anonymously, and without fear of retaliation. Rather than building entirely new infrastructure, much of this can be embedded into what HDOs already have: non-retaliation language can be updated to reference AI specifically within existing HR policies, AI-related reporting categories can be added to current compliance hotline intake forms, and whistleblower protections can be incorporated into the AI modules of onboarding and annual training programs. Practical additions (e.g., assigning reporters a case code for anonymity, establishing clear service level agreements by urgency level, and defining escalation pathways to the Chief AI Officer or similar staff for unresolved reports) strengthen accountability without requiring a standalone system.

Illustrative Example: Urgency Levels

HDOs may develop semi-automated priority systems for routing and responding to incident reports. One example:

- Urgency 1 (Critical): Respond within 24 hours.
- Urgency 2-3 (Moderate): Respond within 3 business days.
- Urgency 4-5 (Low): Respond within 5 business days.

Note: Defining Whistleblower Protections

The four below terms have specific operational meanings within a whistleblower protection program.

- **Anonymity** means the reporter's identity is never collected, recorded, or knowable by anyone. A truly anonymous channel (for example, a third party hotline or a no-login web form) accepts a report about an AI system without capturing names, employee IDs, IP addresses, badge data, or any metadata that could re-identify the source. In a healthcare setting this matters because staff may fear professional or contractual consequences for flagging an AI safety issue, and anonymity removes the need to trust the organization's discretion.
- **Confidentiality** is weaker than anonymity but operationally distinct: the reporter's identity *is* known to a narrowly scoped group but shielded from disclosure to certain parties (e.g., supervisors). Confidentiality is what allows investigators to follow up with the reporter for clarification. Confidentiality protections typically need to specify that identifying information will not be shared with the AI vendor during root-cause analysis, since vendor contracts often create indirect channels back to a reporter's employer.
- **Non-retaliation assurances** are formal, written commitments by the HDO that no adverse action will be taken against anyone who in good faith reports a concern about a health AI system, attempted to report one, participates in an investigation, or refuses to use an AI solution they reasonably believe is unsafe. "Adverse action" is defined broadly and typically covers termination, demotion, reassignment, or related actions. Strong programs tie these assurances to existing legal frameworks (e.g., False Claims Act, HIPAA's anti-retaliation provision at 45 CFR 164.530(g), state whistleblower statutes).
- **Clear disclosure of data retention policies** means the HDO publishes, in plain language and before a report is made, exactly what happens to the data generated by a whistleblower submission. This data or information includes: what is collected (e.g., the report text, attachments, investigation notes, audit logs), where it is stored, who can access it under what conditions, how long each submission is retained, and the method of destruction.

Together, these four protections function as a layered system: anonymity and confidentiality control who knows, non-retaliation assurances control what can be done if identity becomes known, and retention disclosures control how long the risk persists.

Implementation Guiding Questions

Below are standard questions that HDOs may ask to guide discussion for **control 4.5.6**:

- Does your organization have a formal, written whistleblower protection policy that explicitly covers AI-related reporting (not just general healthcare incident reporting)?
- Are multiple channels available for submitting reports (including online forms, telephone, and in-person options) to accommodate reporters with different preferences or accessibility needs?

- Is anonymity protected throughout the reporting and investigation process (e.g., reporters assigned a case code rather than identified by name to staff handling the report)?
- Is non-retaliation explicitly stated in the policy and in the reporting interface itself, covering demotion, dismissal, harassment, and other adverse actions?
- Are reporters provided with a tracking portal or mechanism to follow the status of their report, including estimated timelines and escalation paths?
- Is there a defined escalation pathway if a report is not addressed within a specified timeframe (e.g., automatic escalation to the Chief AI Officer or similar staff if not responded to within 14 days)?
- Are data retention timelines clearly disclosed to reporters (e.g., records archived for up to 7 years for legal compliance and audit purposes)?
- Does the reporting form include an explicit acknowledgment of protections and reporter obligations (e.g., "I accept" checkbox)?
- Are whistleblowing protocols and legal protections covered in staff AI training programs, including at onboarding?
- Are both internal parties (employees, contractors) and external parties (patients, vendors) covered under the whistleblower protections?
- Is there a defined response service level agreement (SLA) by urgency level (e.g., critical reports responded to within 24 hours)?
- Has your legal team confirmed that your whistleblower policy addresses applicable federal anti-retaliation statutes (including the False Claims Act) particularly where AI tools are used in billing, coding, clinical documentation, or prior authorization workflows?

Real-World Examples

Explore these real-world examples of how HDOs are successfully approaching education, training, and feedback. These examples offer valuable insights and practical strategies that you can adapt to your own organization.

Real-World Example #1: Parkland Health (Dallas, TX)

Parkland was among the first U.S. safety-net system to embed a predictive analytics platform PIECES (Parkland Intelligent e-Coordination and Evaluation System) directly into routine clinical workflow, beginning with an EHR-driven 30-day heart-failure readmission and mortality model published by Amarasingham et al. in *Medical Care* (2010) and later expanded to other risk domains, including a rapid-response COVID-19 severity tool in 2020.

Parkland's approach to education, training, and feedback reflects what a smaller safety-net system with constrained resources can realistically operationalize. Rather than rolling out broad "AI literacy" curricula, training was scoped tightly to the case managers and hospitalists who actually acted on the risk score, and was reinforced in existing multidisciplinary rounds. Models were piloted in silent mode against live EHR data before alerts were surfaced to clinicians, satisfying the playbook's pre-deployment testing expectation without a dedicated sandbox. Staff feedback was captured in the same rounding and case-management workflow where the score appeared, and concerns were triaged by Parkland Center for Clinical Innovation (PCCI) engineers back into retraining cycles. AI-related safety concerns were routed through Parkland's existing patient-safety event-reporting system, which is consistent with the playbook's explicit guidance that most organizations should add an AI tag to existing infrastructure rather than build a parallel system.

The broader lesson is that smaller HDOs can meet Subdomain 4.5 recommendations and guidance by scoping training narrowly to the people who touch the model, aligning to existing patient-safety reporting, and using a small dedicated research partner (academic affiliate or regional collaborative) as the peer-to-peer feedback channel.

Source: Home | PCCI Innovation

Real-World Example #2: UC San Diego Health (San Diego, CA)

UC San Diego Health prospectively deployed COMPOSER, an in-house-developed deep-learning sepsis early-warning model, across its emergency departments. The deployment and its outcomes were published by Boussina et al. in *npj Digital Medicine* (2024), which reported associations with reduced in-hospital sepsis mortality following rollout, one of the most rigorously evaluated prospective deployments of a homegrown clinical ML model at an HDO.

The implementation illustrates several of the playbook's education, training, and feedback expectations. Training was role-differentiated: ED nurses were trained on the specific alert triage pathway, physicians received use-case-specific guidance on how COMPOSER differed from the prior rules-based screen, and the model was accompanied by plain-language documentation describing its inputs and known limitations. The research and operational teams ran COMPOSER in silent mode before activation, used a phased go-live, and established a structured clinician feedback loop in which alert disposition (acted on, deferred, or overridden) was logged and reviewed by the development team to recalibrate thresholds and retrain staff where needed. AI-related concerns were surfaced through the existing patient-safety reporting channel, with the clinical informatics group designated as the review body. This aligns with Subdomain 4.5 and the guidance that AI-specific categories be added to existing reporting infrastructure rather than built separately.

Source: Boussina A, Shashikumar SP, Malhotra A, et al. Impact of a deep learning sepsis prediction model on quality of care and survival. *npj Digital Medicine*. 2024;7:14. doi:10.1038/s41746-023-00986-6

Tools & Resources

Included below are tools and resources that may be helpful for developing and operationalizing your own education, training, and feedback processes.

Professional Certificate in Nursing Essentials of Responsible AI

- *The Nursing Essentials of Responsible AI professional certificate, developed by the Florida State University College of Nursing in partnership with the Coalition for Health AI (CHAI), prepares nurses at all levels to lead the ethical integration of artificial intelligence in healthcare.*

CHAI Partner Ecosystem

- *Collaborate with trusted partners delivering excellence in AI governance, testing, data management, and advisory services – all aligned with the CHAI AI Governance Framework.*

International Association of Privacy Professionals (IAPP) Artificial Intelligence Governance Professional (AIGP)

- *The AIGP credential demonstrates competency in AI development, the ability to conduct ethical AI deployment, and the ability to utilize best practices in AI management to ensure safety and trust. AIGP is for any professional tasked with developing AI governance and risk management in their operations.*

Association of American Medical Colleges (AAMC) AI Competencies Across the Learning Continuum

- *In-development, targeted publication in Fall 2026: AAMC is developing a national set of competencies in AI designed to be relevant for undergraduate, graduate, and continuing medical education.*

HIPAA Notice of Privacy Practices (NPP) template

- *The HIPAA Privacy Rule requires health plans and covered health care providers to develop and distribute a notice that provides a clear, user friendly explanation of individuals' rights with respect to their personal health information and the privacy practices of health plans and health care providers. HHS developed these model notices to help improve patient experience and understanding and to assist health care entities to meet new requirements.*

American Hospital Association (AHA) Patient and Family Advisory Councils (PFAC) Blueprint

- *To capture learnings and help the field build and sustain high performing PFACs, the AHA gathered a group of patient and family engagement leaders to share their insights. Use the Patient and Family Advisory Councils Blueprint to help guide your organization through the process of forming and maintaining a successful PFAC.*

Example Template: Closing the Feedback Loop on AI Incident Reporting

- HDOs should implement and maintain a documented process for closing the feedback loop on AI incident reporting, including communicating outcomes and visible follow-up actions to staff. This will help to sustain trust and preserve early identification of AI-related risks. Examples included are illustrative, derived from industry best practice, and not prescriptive.

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