Health Sector Coordinating Council
Cybersecurity Working Group

Measure Effectiveness | Manage Risks | Respond & Recover | Secure Medtech | Monitor Threats

Health Industry Cybersecurity

Playbook for A.I. Clinical Resilience (HIC-PAICR)

OCTOBER 2026

Table of Contents

About the Health Sector Coordinating Council Cybersecurity Working Group 3

Acknowledgements 3

Executive Summary 3

- Purpose 3
- Audience 4
- Scope 6
- Determining the Number and Types of Playbooks Required 7

Documenting AI Use in Clinical Workflows 10

- Workflow Inventory 10

Methods to Determine the Number and Types of Playbooks 12

Playbook Design 14

- Foundational Design Principles 15
- Core Components of the Playbook 15
- AI Clinical Workflow Risk Playbook Matrix 17
- Risk Classification 18

Appendix A 21

Appendix B AI Threat Table 22

Appendix C: Security Framework Crosswalk (OWASP Top 10 for LLM Applications × MITRE ATLAS) 23

About the Health Sector Coordinating Council Cybersecurity Working Group

The Healthcare and Public Health Sector Coordinating Council (HSCC) is a coalition of private-sector critical healthcare infrastructure entities organized under a national public-private partnership framework to advise the government in the identification and mitigation of strategic threats and vulnerabilities facing the sector's ability to deliver services and assets to the public. The HSCC Cybersecurity Working Group (CWG) is the largest HSCC working group of almost 500 healthcare providers, pharmaceutical and medtech companies, payers and health IT entities partnering with government to identify and mitigate cyber threats to health data and research, systems, manufacturing and patient care. The CWG membership collaboratively develops and publishes freely- available healthcare cybersecurity best practices, policy/procedure recommendations, and outreach/ communications programs emphasizing the imperative that cyber safety is patient safety.

Acknowledgements

The HSCC Cybersecurity Working Group established the Artificial Intelligence Cybersecurity Task Group in March 2025 and split into 5 subgroups to address the multi-layered complexities of the AI cybersecurity issue: 1) Education and Nomenclature; 2) Governance; 3) Third Party Risk; 4) Secure by Design; and 5) – the topic of this guide, Threat Operations.

This guide could not be possible without the dedicated leadership and hands-on work of co-leaders of the Threat Operations Subgroup:

Janette Arencibia
CDR, MSC, USN
Chief, Combat Support Agency Integration
Chair, NATO Communication & Medical Information Management
Lead Planner Mission Relevant Terrain - Cyber
PEO Medical Systems/CIO (J6)

David Bailey
Vice President, Consulting Solutions and Strategy
Clearwater

Executive Summary

Purpose

As artificial intelligence becomes increasingly embedded in clinical operations, healthcare organizations (HDOs) must proactively prepare for the safety, security, and operational risks that accompany its use. Inpatient workflows involve complex, multi-day coordination across multidisciplinary teams from admission through discharge; acute care workflows demand rapid, high-stakes decision-making in time-sensitive environments such as emergency departments and intensive care units.

Outpatient workflows prioritize efficiency, continuity, and high patient throughput in ambulatory settings. Because AI systems may influence triage decisions, diagnostic assessments, monitoring, documentation, scheduling, and care coordination across these environments, organizations must understand how workflow context, decision criticality,

and level of automation shape risk exposure. This document provides a framework for identifying AI touchpoints, assessing threat vectors, and designing practical, context-specific playbooks that safeguard patient safety, data integrity, regulatory compliance, and operational resilience across the full spectrum of clinical care delivery. It is designed to help healthcare leaders, clinical teams, IT professionals, and risk managers develop structured AI playbooks tailored to the unique threat profiles of inpatient, acute care, and outpatient clinical workflows.

The *Playbook for AI Clinical Resilience* is intended to serve as a practical guide for healthcare organizations to build playbooks to effectively detect, analyze, and respond to threats that impact AI-integrated clinical operations. The playbook should guide healthcare organizations to proactively manage cyber and operational threats targeting AI-enabled clinical workflows.

Audience

The intended audience for this guide includes a cross-functional team that spans clinicians, governance leaders, technical experts, and compliance officers. This ensures AI threats are addressed holistically — from the moment an issue is detected, through patient safety review, to regulatory response and system remediation.

Audience profiles include:

Clinical Leadership and Care Teams

- Chief Medical Officer (CMO), Chief Nursing Officer (CNO), clinical directors
- Physicians, nurses, and allied health professionals involved in AI-enabled workflows

These stakeholders need to understand how AI failures may impact patient care and what actions to take during clinical incidents.

Health IT and Informatics Leaders

- Chief Information Officer (CIO), Chief Medical Information Officer (CMIO), Chief Nursing Informatics Officer (CNIO)
- Clinical informatics and Electronic Health Record (EHR) integration teams: Within HDO-controlled workflows and interfaces, these teams are responsible for implementation, system integration, and safe local operation. Note: An HDO integration team can validate the local workflow and interfaces, but it cannot assure the intrinsic safety of a vendor-controlled or FDA-regulated product. Product safety, regulated device functionality, and manufacturer-controlled configurations remain the responsibility of the manufacturer of record, with shared validation of end-to-end integrations.

Cybersecurity and Risk Management Teams

- Chief Information Security Officer (CISO), security operations, enterprise risk management
- Third-party/vendor risk management teams: These groups focus on threat detection, incident response, and managing risks introduced by AI systems and external vendors.

Data Science, AI, and Analytics Teams

- AI/Machine Learning (ML) engineers, data scientists, model governance teams. For HDO-developed AI, internal data science teams may own model development, validation, monitoring, and remediation. For manufacturer- or vendor-controlled AI, those activities must be performed or authorized by the responsible developer/manufacturer under its own change-control and quality processes. The HDO's role is to monitor local performance and escalate suspected issues, not to retrain or recalibrate the model itself.

Compliance, Legal, and Privacy Officers

- Chief Compliance Officer, General Counsel, Privacy Officer
- Regulatory and audit teams: Legal, privacy, quality, and regulatory teams determine which binding requirements apply to the organization and affected function — including Health Insurance Portability and Accountability Act (HIPAA) Privacy Rule and Security Rule obligations, device user-facility obligations, and manufacturer obligations for regulated medical devices. Draft regulations, final guidance, accreditation expectations, and voluntary AI frameworks should be identified and tracked separately from binding legal requirements, since "emerging AI regulations" may be proposed, voluntary, jurisdiction-specific, or not yet applicable.

Quality, Patient Safety, and Clinical Risk Leaders

- Patient safety officers, quality improvement teams, clinical risk managers: They evaluate the impact of AI on patient outcomes and integrate AI risks into broader safety and quality frameworks.

Operations and Administrative Leadership

- Chief Operating Officer (COO), service line leaders, hospital administrators: They are responsible for maintaining workflow continuity, especially during AI downtime or disruptions.

Vendor Management and Procurement

- Contracting, sourcing, and vendor oversight teams: They manage third-party AI relationships, Service Level Agreements (SLAs), and escalation pathways during vendor-related incidents.

Communications

- Internal and external communications leads: Needed for media, patient/family, and regulator-facing outreach during an AI-related incident (see the Communication Plan, below).

Human Resources

- HR/workforce relations leads: Needed for workforce training, accountability, and labor-relations issues connected to an AI incident (for example, if the root cause involves policy violations, phishing, or changes to clinical staff workflows that intersect with union agreements or working conditions).

For AI-Enabled Medical Devices or Vendor-Controlled Clinical AI

Where AI is embedded in or integrated with a regulated medical device, or is otherwise vendor-controlled, include:

- The manufacturer of record
- Product security / PSIRT (Product Security Incident Response Team)

- Clinical engineering / biomedical engineering (biomed)
- Manufacturer product quality and regulatory affairs
- Medical safety
- The hosting provider and relevant system integrators

Each party is responsible only for the systems and functions within its control, and for the obligations established by applicable law and agreement — participation and decision rights follow control of the affected component, not organizational hierarchy alone.

Scope

To effectively respond to AI-related risks in the clinical workflow, a healthcare organization needs incident response and governance playbooks that target AI-specific failures, ethical risks, security threats, and regulatory concerns within the following types of clinical workflows:

In-patient clinical workflows describe the coordinated processes and handoffs that occur from the time a patient is admitted to a healthcare facility until discharge, ensuring safe, efficient, and high-quality care.

Acute care clinical workflows describe the coordinated, time-sensitive processes used to assess, stabilize, treat, and discharge patients who present with urgent or emergent medical needs, often in hospital-based settings such as emergency departments (ED), urgent care (UC), intensive care units (ICU), or specialized acute care units.

Outpatient clinical workflows describe the processes that guide patient care in settings where patients receive diagnosis, treatment, and follow-up without being admitted to a hospital, or post-discharge. These workflows emphasize efficiency, continuity, and patient-centered care while accommodating high patient volume.

Across care settings, AI introduces distinct risk considerations aligned to the clinical tempo and decision intensity of the workflow (see table 1.2 below). In inpatient environments, where care unfolds over multi-day stays with continuous monitoring and multidisciplinary coordination, AI tools support risk prediction and care planning but introduce risks such as model drift in predictive scores (e.g., sepsis alerts), alert fatigue, bias in care recommendations, and data integration failures between clinical systems and AI platforms. In acute care, where rapid triage, stabilization, and protocol-driven interventions are critical to preventing deterioration, AI-related threats center on triage misclassification, false negatives or false positives in time-sensitive imaging (e.g., stroke), over-reliance on automated recommendations under pressure, and system downtime during emergencies. In outpatient settings, which emphasize efficiency, preventive care, and chronic disease management, AI is often embedded in scheduling, patient engagement, and result management workflows; here, risks include missed abnormal findings due to filtering errors, misinformation from chatbots or virtual assistants, predictive analytics bias affecting care management, and privacy vulnerabilities in AI-enabled communication tools. Together, these variations highlight the need to tailor AI threat mitigation strategies to the clinical context, operational pace, and patient safety impact of each workflow type.

Table 1.2 Risk Considerations for Clinical Workflow

Workflow Type	Unique Steps	Desired Outcomes	Potential AI RIsks
Inpatient	Admission & intake from referral, or direct admit; daily multidisciplinary rounds; continuous monitoring & nursing care; discharge planning with post-acute coordination	Maintain patient stability; deliver comprehensive care across multi-day stay	Model drift affecting predictive risk scores (e.g., sepsis); alert fatigue from AI-generated monitoring alerts; bias in care recommendations for certain patient groups; data integration errors between Clinical Systems and AI tools
Acute Care	Rapid triage & prioritization; immediate stabilization; protocol-driven emergency interventions; rapid disposition to ICU, surgery, or discharge	Save life / prevent rapid deterioration; minimize time-to-treatment for critical conditions; ensure safe transitions	AI triage misclassification leading to delayed treatment; over-reliance on AI during high-pressure decision-making; imaging AI false negatives or false positives in time-sensitive conditions (e.g., stroke); system downtime during emergencies
Outpatient	Appointment scheduling & pre-visit prep; intake & vitals by medical assistant/nurse; short, focused provider evaluation; referral to medical specialist; post-visit follow-up for results & chronic care	Efficient, high-quality encounters; preventive care & chronic disease management; enhance patient engagement	Missed abnormal results due to AI filtering errors; chatbot or virtual assistant misinformation; predictive analytics bias in chronic disease management; privacy breaches from AI-driven patient engagement tools

Determining the Number and Types of Playbooks Required

Determining the number and types of playbooks needed to address AI threats to clinical workflow depends on several structural, clinical, and technical factors:

Clinical Criticality and Patient Safety Impact: Workflows with high acuity and low tolerance for error (e.g., emergency triage, ICU monitoring, surgical decision support) warrant distinct, highly detailed playbooks because AI failures can result in immediate patient harm. Lower-acuity or administrative use cases (e.g., scheduling optimization) may be grouped under broader operational AI playbooks.

Type of AI Function (Decision Support vs. Automation vs. Engagement): Different AI roles create different risk profiles:

- Diagnostic/predictive decision support: risks of misclassification, bias, model drift
- Generative AI (documentation, chatbots): misinformation, hallucinations, PHI (Protected Health Information) leakage

- Operational automation (scheduling, throughput optimization): data integrity and workflow disruption risks
- Monitoring/alerting systems: alert fatigue, missed alerts, false reassurance
- Agentic and tool-using AI: systems that can place orders, route cases, send messages, update records, or call other tools introduce excessive agency risk — the AI takes action, not just a recommendation. Where this applies, playbooks should call for least-privilege tool permissions, allow-listed actions, step-up (human) approval before any patient-impacting write-back to clinical systems, execution logs, rate limits, and an emergency stop.

Degree of Clinical Integration: The tighter the integration with EHRs, order entry, imaging systems, or device feeds, the more complex the risk landscape. Deeply embedded tools may require dedicated incident response and downtime playbooks, while loosely coupled tools may be addressed in broader AI governance guidance.

Regulatory Status and Manufacturer Control: Classification by clinical criticality, function, integration, and autonomy alone does not distinguish a regulated medical-device function from non-device AI. The same containment or recovery action may be a permissible local adjustment for an HDO-developed administrative model but an unauthorized modification for a regulated device. For each AI function, also identify whether it is:

- an FDA-regulated device function
- part of a regulated medical device
- non-device clinical software
- administrative AI
- the manufacturer of record
- intended use
- authorized configuration
- software/model version; and
- indicating which party can lawfully and technically modify the function.

These distinctions affect every downstream playbook decision, from who leads the response to who authorizes return to service.

Automation Level and Human Oversight Model: Fully automated or semi-autonomous systems (e.g., auto-triage, auto-order generation) require different response protocols than tools that provide advisory output only. Playbooks should reflect where human override exist and where it does not. The level of autonomy embedded in an AI system fundamentally changes both the risk profile and the required response structure within a clinical AI playbook. Systems that generate advisory outputs only (e.g., risk scores, documentation suggestions, imaging annotations) may operate within a human-in-the-loop (HITL) model, where a clinician reviews, interprets, and decides whether to act. In contrast, fully automated or semi-autonomous systems (e.g., auto-triage routing, automated order generation, protocol-triggered interventions, autonomous monitoring escalations) can initiate clinical or operational actions without real-time human validation. These are better described as human-on-the-loop (HOTL), where a human monitors and retains override authority but does not approve each action before it happens — a distinction that matters because humans generally cannot review machine-speed decisions one at a time. This distinction has major implications for governance, monitoring, and incident response.

Data Sensitivity and Regulatory Exposure: AI tools processing PHI, behavioral health data, genomic data, or pediatric records may necessitate distinct privacy and compliance playbooks. Regulatory expectations (HIPAA, FDA oversight, state AI laws) may also influence playbook segmentation. Data sensitivity increases the severity of harm, while regulatory exposure increases the complexity of response. When both are high, e.g., AI used in pediatric behavioral health for clinical decision support, the justification for a distinct, specialized playbook becomes strong.

The number and type of AI threat playbooks should therefore scale with:

- Sensitivity of the data processed
- Number of overlapping regulatory regimes
- Reporting and remediation divergence
- Long-term impact of compromise
- Civil rights and discrimination risk

As AI becomes more embedded in care delivery, segmentation of playbooks along data sensitivity and regulatory lines is a recommended governance practice where it improves operational clarity. It is not itself a legal requirement unless incorporated into an applicable law, accreditation standard, policy, or agreement.

Threat Source and Failure Mode Categories: Playbooks can be structured around:

- Technical failures (model drift, data corruption, downtime)
- Security threats (data poisoning, prompt injection, adversarial inputs)
- Operational misuse (automation bias, over-reliance, insider misuse)
- Third-party/vendor risk

Organizational Scale and AI Maturity: A large health system with dozens of AI tools may need:

- Enterprise AI incident response playbook
- High-risk clinical AI playbooks (acute, inpatient)
- Generative AI governance and misuse playbook
- Vendor AI risk management playbook

Smaller organizations may consolidate these into fewer, modular playbooks.

Dependency on Third Parties and Cloud Services: Heavy reliance on external AI vendors or cloud-hosted models warrants the development of dedicated supply chain and vendor outage playbooks. When clinical AI capabilities, such as predictive analytics, imaging interpretation, automated documentation, patient engagement chatbots, or triage engines are delivered through third-party platforms, the organization's risk exposure extends beyond its internal controls. The attack surface expands to include vendor infrastructure, model development pipelines, sub-processors, Application Programming Interface (API) integrations, and cloud service dependencies. In these environments, clinical performance and patient safety become directly intertwined with vendor security posture, uptime commitments, change management practices, and incident response timelines.

Dedicated playbooks help organizations anticipate and manage failures — including but not limited to vendor-related failures, such as HDO-side implementation errors and data corruption — by clearly defining escalation pathways, contractual notification triggers, clinical impact assessment procedures, and workflow reversion strategies. Vendor and manufacturer duties in these playbooks should be defined in, and limited to, applicable

agreements and the components each party actually controls; a manufacturer may not control the cloud, EHR, network, or a customer's local configuration, and playbooks should not assume support, access, or notification speed the contract does not provide. They establish roles and responsibilities across IT, clinical leadership, compliance, and vendor management teams, ensuring coordinated response during outages, security incidents, or unexpected model changes. Playbooks should identify contracted support hours, escalation contacts, evidence-access limitations, notification triggers, and supported fallback modes. Most importantly, these playbooks preserve patient safety and operational continuity when direct control over the underlying AI system is limited, providing structured mechanisms for rapid containment, communication, and safe transition to manual or alternative processes.

Documenting AI Use in Clinical Workflows

Workflow Inventory

An essential step to help an organization understand what playbooks to develop is to systematically "map" how AI is used across the care delivery lifecycle. This involves identifying each AI-enabled system, its function (e.g., triage support, diagnostic augmentation, documentation generation, scheduling optimization), the data it ingests and produces, its integration points with EHRs and other clinical systems, and the human decision-makers who rely on its outputs. The mapping should trace data flows, model dependencies, automation levels, and override mechanisms to clarify where AI meaningfully influences clinical, operational, or administrative decisions.

For AI-enabled medical devices, or any AI managed across a diverse installed base, the inventory should also capture: the manufacturer of record; product and model; device identifier (where applicable); software, firmware, and model version; supported configuration; relevant local customization; hosting and service dependencies; the remote-service pathway; end-of-support status; and the manufacturer's support or escalation contact.

Table 2.1 provides an example of a clinical workflow inventory and how AI is woven throughout the care continuum, influencing decision-making, documentation, monitoring, and patient engagement across multiple workflow touchpoints. The completion of workflow and AI integration inventory enables the insight to develop the number and types of playbooks to develop.

Table 2.1 Sample Workflow and AI Integration Inventory

Workflow	Activities	Key Outcomes	AI Integrations
Admission & Intake	Patient registration and demographic data capture; verification of insurance and consent forms; initial nursing assessment (vitals, allergies, medical	Admission orders, initial care plan, electronic health record (EHR) initiation	Automated patient triage in the ER or intake (AI-driven symptom checkers prioritizing urgency); Optical Character Recognition (OCR) & Natural Language

	history); assignment of patient to a bed and care team		Processing (NLP) for registration to auto-fill forms from prior records or scanned documents
Initial Assessment & Care Planning	Comprehensive history and physical exam; diagnostic orders (labs, imaging); initial medication reconciliation; creation of individualized care plan (clinical goals, interventions)	Problem list, treatment orders, baseline diagnostics	Clinical decision support (CDS): AI suggests differential diagnoses and diagnostic orders based on patient history and presenting symptoms; risk stratification models predict likelihood of complications (e.g., sepsis risk scoring); speech-to-text documentation: ambient AI scribe records and structures physician-patient interactions into the EHR
Daily Care & Monitoring	Medication administration; vital signs and symptom monitoring; patient hygiene, nutrition, mobility assistance; documentation of care in the EHR; physician workflow: daily rounds, progress notes, order adjustments; coordination with specialists and multidisciplinary team	Updated care plan, daily progress reports, order changes	Predictive deterioration alerts: continuous analysis of vitals and lab trends to alert nurses before patient status declines; smart infusion pump monitoring: AI detects anomalies in medication delivery; fall risk prediction: real-time AI monitoring from in-room cameras and movement sensors
Diagnostics & Interventions	Lab and imaging studies; procedures (e.g., central line placement, wound care); specialty consults	Results in EHR, modified treatment plans	Radiology image analysis: AI assists in detecting fractures, pneumonia, stroke, tumors faster; lab result interpretation: AI flags abnormal values in context of patient's condition; procedure guidance: AI-enhanced imaging overlays during surgery or interventions
Care Coordination & Communication	Interdisciplinary rounds (IDR) to align on care goals; handoff communication during shift changes;	Consistent care plan, reduced errors, aligned expectations	Automated shift handoff summaries based on EHR notes, ensuring critical info isn't missed; natural language processing to

	family/patient education and updates		summarize multidisciplinary team notes into patient-friendly language; scheduling optimization for therapy sessions and consults
Discharge Planning	Review of discharge criteria; prescriptions and medication education; patient/family teaching on home care, wound care, diet, follow-up; coordination with post-acute care providers (home health, rehab, skilled nursing)	Discharge summary, follow-up appointments, patient instructions	Readmission risk prediction to guide additional interventions before discharge; automated discharge summaries generated from hospitalization record; tailored patient education - AI customizes instructions based on literacy level and language preference
Post-Discharge Follow-Up	Transitional care phone call within 24-72 hours; monitoring for readmission risk; coordination of outpatient services	Continuity of care, reduced readmissions	Chatbots or virtual nurses to check on symptoms, medication adherence, and recovery progress; predictive analytics to flag patients at high risk for ER return; natural language analytics on follow-up calls to detect patient distress or non-compliance cues

One method for applying an AI filter through a broader range of clinical workflows – such as home health, claims and payments, retail pharmacy and 14 others - is through use of the HSCC CWG Health Industry Cybersecurity Sector Mapping and Risk Toolkit (SMART). SMART provides healthcare organizations with templates and a methodology to visualize, identify and measure systemic risk posed by third party technology, software and communications services essential to clinical, administrative and manufacturing workflows.

Methods to Determine the Number and Types of Playbooks

The number and type of AI playbooks should reflect:

- Clinical acuity differences
- Autonomy level differences
- Regulatory divergence
- Vendor dependency
- Failure mode similarity

Playbook segmentation should improve clarity during crisis, not add complexity.

The goal is operational precision: ensuring that when an AI-driven risk emerges in admission triage, inpatient monitoring, imaging interpretation, or post-discharge chatbots, the organization knows exactly which playbook applies, who leads, how to contain risk, and how to protect patient safety without hesitation. Too many playbooks can create confusion during incidents. Use the smallest number of playbooks and modular annexes that enables responders to identify the correct actions and authorities during an incident. The appropriate number depends on organizational scale, clinical risk, architecture, and existing incident-response structures; the Table 3.1 range referenced below is illustrative only, not a hard rule.

The following table outlines a phased approach to determine the number and type of playbooks you need for your organization.

Table 3.1 Phased Approach to Playbook Determination

Method	Determination
Phase 1 Create an AI Workflow Risk Map	For each of the AI integrations in the AI clinical workflow inventory document the following: AI System Name; Vendor; Clinical Function; Data inputs and outputs; Level of autonomy (advisory vs semi-autonomous vs autonomous); Regulatory classification, including whether it is an FDA-regulated device function, and the manufacturer of record where applicable.
Phase 2 Classify AI systems by risk	For each AI integration, assess across five core risk dimensions: 1. Clinical Criticality — High acuity (ED triage, deterioration alerts, imaging interpretation); moderate acuity (CDS suggestions, lab interpretation); operational/administrative (OCR registration, scheduling optimization). 2. Level of Autonomy — Advisory only; semi-autonomous (auto-populated orders, auto-prioritization); fully autonomous (closed loop or automated execution). 3. Data Sensitivity & Regulatory Exposure — Data Classification (PHI, behavioral health, genomic, etc.). 4. Technical Architecture — Embedded in EMR; device-integrated; cloud-hosted SaaS; external API dependent; vendor managed model updates. 5. Failure Mode Pattern (group tools how they fail) — Misclassification / false negatives or positives; alert fatigue / threshold miscalculation; hallucination / misinformation; model drift; data integration corruption; downtime / vendor outage; bias amplification.
Phase 3 Determine Playbook Segmentation Model	Using the risk mapping from Phase 2, choose one of three structured approaches: OPTION 1: Workflow-Based Playbooks — segment by clinical environment: Inpatient AI Response Playbook; Acute Care AI Response Playbook; Outpatient & Post-Discharge AI Response Playbook. Best when organizational structure is workflow-centric, clinical leaders need environment-specific guidance, and AI risk differs significantly by acuity. OPTION 2: Threat-Type Playbooks — segment by failure mode: AI Misclassification & Clinical Error Playbook; AI Downtime & Workflow Reversion Playbook; Model Drift & Performance Degradation Playbook; Bias & Algorithmic Fairness Playbook; AI Data Breach & Privacy Incident

	Playbook; Vendor/Supply Chain Incident Playbook. Best when many AI tools share similar technical architectures, security and IT teams drive response, and the organization prefers cross-cutting controls. OPTION 3: Hybrid Approach — most mature healthcare systems adopt a layered structure: Tier 1 – Enterprise AI Governance Playbook (applies to all AI: incident classification criteria, escalation matrix, regulatory reporting framework, communication protocols); Tier 2 – High-Risk Clinical AI Playbooks (acute triage AI, predictive deterioration monitoring, imaging AI, procedure guidance); Tier 3 – Functional Risk Playbooks (vendor & cloud outage, model drift & revalidation, bias & discrimination response, generative AI misinformation, data privacy & PHI incident). This structure avoids duplication while maintaining clarity.
Phase 4 Quantitative Playbook Threshold Assessment	Use a structured scoring model to determine how many playbooks are justified. For each AI system, score 1–5 across: Patient harm potential; Autonomy level; Regulatory exposure; Vendor dependency; Scale of impact (number of patients affected if failure occurs). If multiple AI systems cluster at high scores in the same category, create a dedicated playbook. In addition to inherent AI risk, also score response-feasibility and control-boundary dimensions: the risk introduced by suspension or isolation itself; availability of a validated manual/downtime workflow; time to clinical degradation; detectability and confidence of the signal; installed-base/version variability; and which party has the authority and technical control to implement each response. A high score alone must not be read as authorization to modify or shut down a device — it flags where dedicated planning is warranted. Any scoring model, weighting, and threshold used as an example in this guidance is a starting example only; organizations should define and validate their own criteria.

Playbook Design

AI risk response playbooks are predefined, role-based action guides that enable rapid, consistent, and compliant handling of incidents where AI systems in clinical workflows present risks to patient safety, care quality, privacy, or compliance. They ensure:

- Timely detection of AI-related failures or threats.
- Clear accountability for immediate actions.
- Minimized patient harm and operational disruption.
- Regulatory alignment with HIPAA, FDA, and Joint Commission requirements. These are legally and operationally distinct: FDA regulation applies only to regulated device functions and regulated entities; HIPAA governs privacy, security, and breach obligations; and Joint Commission requirements are accreditation and patient-safety expectations. Playbooks should use separate pathways for each, plus contractual commitments and voluntary frameworks, rather than treating them as one undifferentiated "compliance" category.

Foundational Design Principles

When structuring AI risk playbooks, organizations should:

- Integrate with existing incident response frameworks (clinical safety events, IT security, quality management).
- Use a trigger-based activation model so playbooks launch only when predefined, observable and instrumentable performance, safety, or compliance thresholds are breached, not solely on events that can only be discovered through human review or after-the-fact reporting. Triggers such as "policy breach" or "conflicts with clinical guidelines" are difficult behavioral classifications even for human reviewers, and nearly impossible to detect at machine speed without dedicated instrumentation; wherever a trigger in this document depends on human reporting, treat it as a signal that specialized monitoring is needed before the playbook can run reliably.
- Assign multi-role responsibilities (Data Science, Clinical Governance, IT Security, Compliance, Leadership, Communications, and — for regulated or safety-critical products — Clinical Engineering/Biomed and the manufacturer of record, including product security, quality, regulatory, and medical-safety representatives as applicable. HDO leaders retain authority over local clinical operations; the manufacturer retains authority over manufacturer-controlled product changes and supported configurations).
- Build for explainability & transparency so actions are defensible to clinicians, patients, and regulators.
- Ensure version control & ongoing testing through tabletop exercises and functional exercises. Tabletop exercises help leadership, cybersecurity, privacy, legal, communications, and clinical/operations teams walk through decision-making and coordination; functional exercises test whether response procedures actually work in practice (detection, escalation, account controls, vendor notification, communications, documentation, continuity of operations). For exercises involving regulated or safety-critical products, include clinical engineering and the manufacturer or service provider, and use tabletop or controlled test environments unless the manufacturer and HDO have jointly approved live testing — do not alter production devices or validated configurations solely to run an exercise.
- Measure response times, recurrence, and corrective actions with metrics and post-incident review.

Core Components of the Playbook

Title & Scope – Defines the AI risk scenario (e.g., AI Model Drift, Bias Detection, Data Breach in AI Training Data).

Purpose Statement – Why this playbook exists and the specific risk it addresses.

Trigger Conditions – Measurable events or thresholds that initiate the playbook (e.g., drop in model accuracy below 85%, patient safety event linked to AI). Triggers must use model- and use-case-specific metrics established before deployment and approved by the appropriate clinical and product owner — for example, sensitivity, specificity, calibration, false-negative rate, subgroup performance, override rate, or a clinically significant deviation from the validated baseline. No universal threshold (such as a flat 85% accuracy figure) is implied by this guidance.

- Document critical clinical workflows and AI Integration Points:

Immediate Actions – Containment steps to protect patients and maintain care continuity, split into what the HDO can decide locally and what requires the manufacturer or system owner.

- **Detection & Verification**
  - Automated monitoring alerts or clinician reports.
  - Validate performance changes using predefined metrics.
  - Where feasible, add runtime telemetry (model/version, input source, retrieved documents, tool calls, user/session, data classification, override rate, error/escalation rate, API volume, latency, and safety-filter events) so issues surface before clinical impact rather than only after.
- **Immediate Containment** (HDO-controlled)
  - Before suspending, isolating, or reverting a safety-critical or regulated AI function, perform and document a clinical risk assessment confirming the fallback workflow can be operated safely — except where immediate action is necessary to prevent imminent harm.
  - Suspend or limit AI usage if safety risk is present.
  - Switch to manual or alternative workflows.
  - HDO personnel may implement only actions within their authority and supported operating procedures.
- **Investigation**
  - Root cause analysis (data quality, integration issues, model changes).
  - Engage Data Science, Clinical Governance, IT, and Compliance.
  - Notify the manufacturer through the contracted escalation path.
- **Remediation & Recovery**
  - Changes to manufacturer-controlled or regulated functionality (retraining, recalibrating, changing thresholds or data sources) must be performed or expressly authorized by the manufacturer of record under its own change-control and quality processes — an HDO does not retrain or recalibrate a vendor-controlled model unilaterally.
  - Return to service requires two distinct gates: (1) confirmation by the manufacturer or system owner that the affected product/model version and configuration are supported and technically released, and (2) HDO approval that the local clinical workflow, interfaces, downtime recovery, and patient-safety controls have been validated. Clinical engineering, quality/regulatory, product security, and clinical leadership participate as applicable.
- **Communication & Documentation**
  - Notify clinicians and stakeholders of status and interim procedures.
  - Maintain incident log for compliance and quality review, including an evidence and audit trail sufficient to reconstruct the event — model version, input data snapshot, prompt and response logs where applicable, clinician override history, EHR transaction records, vendor tickets, API responses, alert history, affected patient cohort, investigation timeline, decisions made, and approval records.
- **Risk Acceptance & Compensating Controls**
  - If full remediation cannot be completed immediately, document a temporary risk-acceptance decision with a clinical safety rationale, compensating controls, accountable executive approval, an expiration date, enhanced monitoring, and clear criteria for returning to normal operation.

Roles & Responsibilities – Named positions or teams for each action step. Accountability assignments may vary by playbook type — a patient-harm event is typically anchored to clinical leadership, while a data breach is typically tied to security/privacy officers.

Escalation Path – Who to notify and in what sequence, including governance and executive escalation.

Investigation & Remediation Steps – Root cause analysis, retraining, recalibration, or vendor engagement (see gating requirements above for manufacturer- or vendor-controlled AI).

Communication Plan – Internal (clinician/staff updates) and external (patients, regulators) communication. Use a notification decision matrix tied to event classification and each party's own legal obligations, since a suspected anomaly, a confirmed safety event, a privacy breach, an MDR (Medical Device Reporting)-reportable event, a recall, and a public disclosure have different standards, owners, and clocks:

- internal clinical safety communication (clinician/staff updates);
- HIPAA breach notification; device user-facility reporting obligations;
- manufacturer MDR/field-action obligations;
- contractual notice to vendors/partners;
- law-enforcement coordination, where relevant; and
- public communications.

Each party remains responsible for its own legal obligations and should coordinate on factual statements without delaying urgent patient-safety actions. Patient notification specifically should be a multidisciplinary decision — clinical leadership/patient safety, legal, privacy, risk management, and communications, based on verified facts and applicable law — not a privacy-only determination.

Post-Incident Review – Lessons learned, process updates, retraining needs.

AI Clinical Workflow Risk Playbook Matrix

Map data flows, APIs, third-party integrations and risk classification. Because sensitive information can be disclosed through prompts, retrieved context, logs, embeddings, API calls, and outputs — not just through a database breach — this mapping should include a data-flow view for AI-touched workflows showing PHI/PII entry points, storage locations, retrieval sources, model/provider calls, logs, outputs, and downstream recipients.

(Sample Template Below)

Workflow	Data Flow	API	Third-Party Integrations	Risk Description	Risk Classification
Admission & Intake					
Initial Assessment & Care Planning					
Daily Care & Monitoring					
Diagnostics & Interventions					
Care Coordination & Communication					
Discharge Planning					
Post-Discharge Follow-Up					

Risk Classification

Prioritize workflows based on patient impact and system criticality. Treat any numeric threshold used below as an illustrative starting point for your organization to validate, not a fixed rule.

The following summary describes the types of AI risk response playbooks a healthcare provider would need for inpatient, acute care, and outpatient clinical workflows, reflecting the different environments, care priorities, and operational challenges in each setting:

AI Risk Type	Recommended Playbook	Trigger	AI Stakeholders
Model Drift	Human in the Loop (HITL)	Predefined, clinically validated performance or safety threshold exceeded (e.g., accuracy, sensitivity, calibration, false-negative rate) vs. baseline	Data Science, IT, Clinical Governance; manufacturer-specified criteria for manufacturer-controlled systems
Clinical Decision Error	Clinical Decision Support (CDS) Error Response Playbook	AI recommendation conflicts with guidelines or causes harm/near miss (requires instrumented detection or a clinician reporting pathway)	Physician Lead, Quality & Safety, IT
Data Breach	AI Data Breach Response Playbook	Unauthorized access to AI training or inference data	IT Security, Compliance, Privacy Officer
Data Poisoning	Prompt Injection / Data Poisoning Playbook	Malicious manipulation of AI inputs or datasets	Data Science, IT Security, Vendor Support
Bias in AI Outputs	Bias Detection (HITL) & Mitigation Playbook	Higher error rates for specific demographics	Data Science, DEI Committee, Compliance
Inappropriate AI Use	Inappropriate Use and/or Correlation of AI Output & Application Playbook	AI used outside intended scope (e.g., denial of care)	Compliance, Legal, Clinical Governance, HR
System Downtime	AI System Downtime & Contingency Playbook	AI service unavailable or degraded	IT, Clinical Operations
Integration Failure	Integration Failure with EHR/Devices Playbook	AI output not integrating with clinical systems	IT, Vendor Support, Clinical Operations
FDA Regulatory Non-Compliance	Regulatory Non-Compliance Playbook (device pathway)	Failure to meet FDA device requirements	Compliance, Legal, Quality & Safety, Manufacturer
HIPAA Non-Compliance	Regulatory Non-Compliance Playbook (privacy/security pathway)	Failure to meet HIPAA Privacy/Security Rule requirements	Compliance, Legal, Privacy Officer, IT Security

Accreditation Non-Compliance	Regulatory Non-Compliance Playbook (accreditation pathway)	Failure to meet Joint Commission or similar accreditation expectations	Compliance, Quality & Safety, Clinical Leadership
Transparency Request	Transparency & Explainability Request Playbook	Clinician or patient requests rationale for AI decision or diagnosis	Data Science, Clinical Lead, Patient Advocacy
Patient Harm	Patient Harm or Near-Miss Due to AI Playbook (HITL)	Adverse patient outcome linked to AI	Physician Lead, Quality & Safety, Compliance
Hallucination/Misinformation	Misinformation / Hallucination Containment Playbook (HITL)	AI generates clinically false or misleading content	Clinical Governance, Data Science, IT, IT Security
Excessive Agency (Agentic AI)	Agentic AI Containment Playbook	AI-initiated action (order, message, routing, tool call) occurs outside allow-listed scope or without required approval	Data Science, Clinical Governance, IT Security

Appendix A

AI can be integrated into every stage of the in-patient clinical workflow to improve speed, accuracy, coordination, and decision-making as outlined in the examples in the table below:

Workflow	Activities	Key Outcomes
Admission & Intake	Patient registration and demographic data capture; verification of insurance and consent forms; initial nursing assessment (vitals, allergies, medical history); assignment of patient to a bed and care team	Admission orders, initial care plan, electronic health record (EHR) initiation
Initial Assessment & Care Planning	Comprehensive history and physical exam; diagnostic orders (labs, imaging); initial medication reconciliation; creation of individualized care plan (clinical goals, interventions)	Problem list, treatment orders, baseline diagnostics
Daily Care & Monitoring	Core Nursing Workflow: medication administration; vital signs and symptom monitoring; patient hygiene, nutrition, mobility assistance; documentation of care in the EHR; Physician Workflow: daily rounds, progress notes, order adjustments; coordination with specialists and multidisciplinary team	Updated care plan, daily progress reports, order changes
Diagnostics & Interventions	Lab and imaging studies; procedures (e.g., central line placement, wound care); specialty consults	Results in EHR, modified treatment plans
Care Coordination & Communication	Interdisciplinary rounds (IDR) to align on care goals; handoff communication during shift changes; family/patient education and updates	Consistent care plan, reduced errors, aligned expectations
Discharge Planning	Review of discharge criteria; prescriptions and medication education; patient/family teaching on home care, wound care, diet, follow-up; coordination with post-acute care providers (home health, rehab, skilled nursing)	Discharge summary, follow-up appointments, patient instructions
Post-Discharge Follow-Up	Transitional care phone call within 24-72 hours; monitoring for readmission risk; coordination of outpatient services	Continuity of care, reduced readmissions

Appendix B AI Threat Table

Threat	Description	Impact
Data Poisoning	Malicious manipulation of training data corrupts the model's behavior.	Causes the model to learn incorrect patterns, leading to biased or harmful outputs.
Prompt Injection	Maliciously crafted inputs to generative AI systems (ambient scribes, chatbots, summarization tools) that cause the model to produce clinically dangerous, privacy-violating, or misleading outputs.	Fabricated clinical documentation, inappropriate medication orders, PHI disclosure, or bypassing of safety guardrails.
Adversarial Attacks	Inputs crafted to deceive AI models into making incorrect predictions. Healthcare-specific examples include adversarial perturbations to radiology images causing misclassification, ECG waveform manipulation causing arrhythmia-detection failures, or manipulated patient-identification images.	Can bypass security systems (e.g., facial recognition, malware detection). In a clinical setting, this can bypass imaging AI or patient-identification safeguards.
Model Inversion	Reconstructing sensitive training data by querying the model.	Breaches/data privacy, including the risk of reconstructing protected health information (PHI) from a clinical AI model's outputs.
AI Supply Chain Attacks	Compromising third-party components (e.g., pre-trained models, libraries).	Introduces backdoors or vulnerabilities into AI systems.
Model Drift	Degradation in model performance due to changes in data over time.	This leads to inaccurate predictions and poor decision-making.
Unauthorized Model Access or Modification	Gaining access to modify or steal AI models.	Tampering with model logic or stealing proprietary algorithms.
Denial of Service (DoS) on AI Pipelines	Overloading AI systems degrades performance or availability.	Disrupt real-time AI services like autonomous driving or fraud detection.
Shadow AI	Unauthorized or unmanaged AI systems deployed within an organization.	Increases attack surface and compliance risks. In healthcare specifically: PHI transmission to unvetted third-party AI providers without authorization; lack of audit trail; potential HIPAA violation; and AI outputs used in clinical decisions without validation.

Insider Threat	Authorized users exploiting AI system access for unauthorized purposes, including curiosity-driven access to celebrity records, data exfiltration, or manipulation of AI outputs for personal or financial gain.	HIPAA violations, patient harm, liability.

Appendix C: Security Framework Crosswalk (OWASP Top 10 for LLM Applications × MITRE ATLAS)

To help security teams align detection, testing, and response language to industry standards, the table below maps each threat from Appendix B to the OWASP Top 10 for LLM Applications (2025, v2.0) and to MITRE ATLAS (Adversarial Threat Landscape for Artificial-Intelligence Systems). Where a threat is not primarily an adversarial/security risk, or has no clean one-to-one match in either framework, that is noted rather than forced into a category.

Table C.1 Appendix B Threats Mapped to OWASP LLM Top 10 and MITRE ATLAS

Threat	OWASP Top 10 for LLM Applications (2025)	MITRE ATLAS (Tactic/Technique)	Notes
Data Poisoning	LLM04:2025 – Data and Model Poisoning	ML/AI Attack Staging tactic; AML.T0020 – Poison Training Data	Direct match in both frameworks.
Prompt Injection	LLM01:2025 – Prompt Injection	Initial Access tactic; AML.T0051 – LLM Prompt Injection	Direct match; ATLAS distinguishes direct vs. indirect (RAG/tool-content) injection.
Adversarial Attacks	No direct 2025 category; closest analogs are LLM04 (Data and Model Poisoning) and LLM08 (Vector and Embedding Weaknesses), depending on where the perturbation is introduced	ML/AI Attack Staging tactic; AML.T0043 – Craft Adversarial Data	Classic evasion/perturbation attacks predate the LLM-specific list and sit more naturally in ATLAS.
Model Inversion	LLM02:2025 – Sensitive Information Disclosure	Exfiltration tactic; AML.T0024 – Exfiltration via AI Inference API	Direct match; risk is PHI/training-data reconstruction from model outputs.
AI Supply Chain Attacks	LLM03:2025 – Supply Chain	Initial Access / Resource Development tactics (ML supply chain compromise — validate current	Direct match at the category level.

		technique ID against the live ATLAS release)	
Model Drift	Not a security category in the 2025 list — an operational/performance risk, not an adversarial one. Indirectly touches LLM04 if the drift is adversarially induced, or LLM09 (Misinformation) if degraded outputs read as confident but wrong	Out of scope unless the drift is adversarially induced (see Data Poisoning, above)	Keep this on your quality/model-governance track, not your security-incident track, unless poisoning is suspected.
Unauthorized Model Access or Modification	LLM03:2025 – Supply Chain (if via a compromised pipeline) and/or LLM02 (if access enables data extraction)	Persistence / Exfiltration tactics (model theft and unauthorized modification — validate current technique IDs against the live ATLAS release)	Partial match; exact ATLAS technique numbering has changed across recent releases.
Denial of Service (DoS) on AI Pipelines	LLM10:2025 – Unbounded Consumption	Impact tactic (resource-exhaustion/denial techniques — validate current technique IDs against the live ATLAS release)	Direct match at the category level.
Shadow AI	No dedicated 2025 category — a governance/asset-management gap rather than a model-layer vulnerability; increases exposure to LLM02 (disclosure) and LLM03 (unmanaged supply chain)	Reconnaissance / Discovery tactics (unmanaged AI as an attacker entry point)	Primarily an inventory and governance control, not a single technical mitigation.
Insider Threat	No dedicated 2025 category — an access-control/governance issue; overlaps LLM02 (disclosure) and LLM06 (Excessive Agency, if the insider abuses agentic permissions)	Valid Accounts-style access misuse under Initial Access / Collection tactics	Mitigate primarily through access control, logging, and least-privilege — not a model-layer fix.

Note: OWASP Top 10 for LLM Applications category numbers (LLM01–LLM10) reflect the 2025 (v2.0) edition published by the OWASP GenAI Security Project. MITRE ATLAS is actively maintained and technique IDs (AML.Toxxx) have been renumbered and expanded across recent releases; validate current technique IDs against genai.owasp.org and atlas.mitre.org before using this table in a client-facing assessment.

Table 1.1 Accountability Matrix

Role	Responsible (R)	Accountable (A)	Consulted (C)	Informed (I)
Chief Medical Officer (CMO)	A	A	C	I
Clinical Governance Committee	R	A	C	I
Department Chiefs/Service Line Leaders	R	A	C	I
Physicians	R		C	I
Nurses & Nurse Practitioners	R		C	I
Allied Health Professionals	R		C	I
Chief Information Officer (CIO)		A	C	I
Chief Data/AI Officer	R	A	C	I
Data Science & Analytics Teams	R		C	I
IT Security & Privacy Officers	R	A	C	I
Patient Safety Officers	R	A	C	I

Quality Improvement Teams	R		C	I
Compliance & Legal Counsel	R	A	C	I
Clinical Informatics Leaders	R		C	I
Risk Management	R		C	I
Training & Education Teams	R		C	I
Medical Supplies Logistics	R		C	I
Communications	R		C	I
Clinical Engineering/Biomed (regulated devices)	R	A	C	I
Manufacturer of Record (regulated/vendor-controlled AI)	R	A	C	I

####