



Health Sector Coordinating Council
Cybersecurity Working Group



Monitor
Threats



Manage
Risks



Measure
Effectiveness



Respond &
Recover



Secure
Medtech

Health Industry Cybersecurity

Third-Party AI Risk and Supply Chain Transparency Guide



APRIL 2026

Table of Contents

About the Health Sector Coordinating Council Cybersecurity Working Group	4
Disclaimer	4
Foreword from the Co-Leads	4
Call to Action	5
Acknowledgments	5
Executive Summary	6
Key Recommendations	8
Implementation Approach	8
Call to Action	8
The Process for AI Third-Party AI Risk and Supply Chain Transparency	9
Terminology and Definitions	9
AI Third-Party Scope	10
Detailed Phase-by-Phase Process	11
Phase 0: AI Use Case Justification & Strategic Assessment	11
Phase 1: Vendor Evaluation and Due Diligence	13
Phase 2: Contract Negotiation & Legal Protections	16
Phase 3: Implementation, Integration & Training	20
Phase 4: Ongoing Monitoring & Performance Management	27
Phase 5: Incident Response & Recovery	30
Phase 6: End-of-Life & Transition Management	35
Conclusion and How to Use This Guide	39

Appendix A. Phase 0 AI Use Case Justification Template and Risk Level Definitions	41
Appendix B: Governance Policy for AI Third-Party Risk	52
Appendix C: Inventory Management	59
Appendix D: RACI Matrix	61
Appendix E: Sample Commercial Contract Language	65
Appendix F: Sample BAA Contract Language	71
Appendix G: AI Vendor Assessment Questions for Procurement and GRC	74
Appendix H: Training Completion Checklist and Curriculum	89
Appendix I: Quality assurance/verification/validation with AI third-party providers	104
Appendix J: References	108

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 more than 480 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.

The CWG's Third-Party AI Risk and Supply Chain Transparency Task Group was convened to address the critical need for enhanced transparency, governance, and strategic risk management of third-party AI solutions and vendors within the healthcare sector. Its objective was to identify critical risk points and challenges in the discovery/disclosure of information to elevate such risks, and develop strategic guidelines and cybersecurity recommendations to ensure robust third-party cybersecurity and risk management practices.

Disclaimer

This document is provided for informational purposes only. Use of this document is neither required nor guarantees compliance with federal, state, or local laws. The information presented may not be applicable to or appropriate for all health sector organizations. This document is not intended to be an exhaustive or definitive source on safeguarding health information from privacy and security risks.

The advice and template materials provided in this guide are neither intended nor offered as legal advice or legal opinions. HSCC-CWG and the authors are not practicing attorneys. This guide and the material herein are intended for educational and information purposes only. The reader should neither act nor fail to act on any legal matter based upon the information or advice provided in this document without first engaging a competent attorney licensed to practice law in their state or territory.

Foreword from the Co-Leads

The healthcare sector's accelerating adoption of artificial intelligence has expanded its dependence on third-party tools and services, introducing complex cybersecurity challenges that traditional risk management tools and models struggle to address. From natural language processing engines embedded in electronic health records (EHRs) to AI-driven remote monitoring devices, many critical functions rely on external vendors whose security postures, data governance practices, and model integrity are difficult to verify. Compounding the risk, healthcare organizations

(HCOs) often lack visibility into the full scope of AI components sourced through layered supply chains, including subcontractors, offshore development, and open-source AI assets. This opacity elevates systemic exposure and risk, further complicating response coordination in the event of a breach or model failure.

Drawing from established frameworks such as the NIST AI Risk Management Framework (NIST AI RMF) and the joint HSCC-HHS Health Industry Cybersecurity Practices (HICP), the Guide adapts best practices to reflect the realities of AI-driven supply chains in healthcare—including data lineage tracking, model auditability, embedded third-party dependencies, and post-deployment monitoring. It outlines critical control areas such as vendor security attestations, model explainability thresholds, and fail-safe requirements for AI-enabled clinical and operational systems. The Guide enables organizations to define accountability expectations and drive performance standards across their extended AI ecosystem.

Crucially, the Guide addresses the growing gaps in discovery and disclosure processes that make AI supply chain risk so difficult to manage. Many HCOs operate with incomplete or outdated vendor inventories, while AI-specific cybersecurity risks—such as synthetic data misuse, training data leakage, and adversarial inference—go unreported by vendors. To counter this, the Guide promotes proactive due diligence, dynamic risk profiling, and contractual transparency. It equips risk managers, compliance teams, and procurement officers with scalable tools to surface hidden dependencies, identify cascading failure points, and align third-party AI vendors and products with mission-critical safety, privacy, and resilience goals.

Call to Action

HCOs are encouraged to:

- Distribute this document to their senior business and technical leaders and their respective teams to include information technology, cybersecurity, supply chain, third-party risk, vendor management, business continuity, governance, risk and compliance teams, suppliers, and recommend the adoption of these practices, encouraging further understanding and dissemination.
- Evaluate their own third-party and supply chain risk management programs against the best practices outlined in this document.
- Share their experiences and contribute to shaping health-sector AI Governance by engaging the AI Task Group. Please register your interest at <https://healthsectorcouncil.org/contact/>.

Stakeholders consulting this resource are invited to provide any feedback to feedback@healthsectorcouncil.org so that the content can be improved periodically.

Acknowledgments

The HSCC CWG is grateful to its Third-Party AI Risk and Supply Chain Transparency Task Group Co-Leads and by all the authors of this document for their significant investment of personal time in its creation. The authors represent some of the most skilled and experienced experts in their field, and this document would not have been possible without their generosity, leadership, and commitment to a more secure health sector, and the support of their employers.

We are grateful for the leadership and editorial skills of Greg Garcia, Executive Director of the HSCC-CWG and the operational support of Allison Burke.

While many individuals assisted in the development and review of this content, the primary authors across this document and version were:

[Co-Leads](#)

Ed Gaudet, Censinet

Samantha Jacques, McLaren Health

[Contributors](#)

Jonathan Almassi, Columbia Memorial Hospital

Edison Alvarez, BD

Preethi Amurthur, Philips

Gergely Antal, Boston Scientific

Gerry Blass, ComplyAssist

Ryan Brady, Select Health

Dennis Chornenky, Domelabs AI

Brindusa Curcaneanu, NeuroPace

Lacey Harbour, ThermoFisher Scientific

Mike Levin, Solera Health

Robert Maclay, Stanford Children's Health

Kristin Ray, Universal Hospital Services

Mari Savickis, CHIME

Rob Suarez, CareFirst BlueCross BlueShield

Rohit Tandon, Essentia Health

Russell Teague, Fortified Health Security

Executive Summary

The healthcare sector's accelerating adoption of artificial intelligence has dramatically expanded its dependence on third-party tools and services, introducing complex cybersecurity challenges that traditional risk management

models cannot adequately address. From AI-driven clinical decision support systems embedded in EHRs to remote monitoring devices and administrative automation, HCOs face unprecedented risks:

- **Limited visibility** into AI components sourced through layered supply chains, including subcontractors, offshore development, and open-source assets. Opacity in AI supply chains elevates systematic risk exposure and complicates incident response coordination;
- **Difficulty verifying** vendor security postures, data governance practices, and model integrity;
- **Vendors shifting risk to Health Care Organizations (HCOs)** including one-sided contract terms or vendors who are unwilling to sign HCO's HIPAA Business Associate Agreements (BAAs);
- **Incomplete vendor inventories** and unreported AI-specific cybersecurity risks such as synthetic data misuse, training data leakage, and adversarial inference; and
- **Acceleration of change** of AI infrastructure, algorithms, and models at unprecedented rates introduce complexity, steep learning curves, an ever-evolving set of new and updated risks, and an exponentially complex and broad attack surface.

This document provides targeted best practices and implementation guidance that HCOs can adopt regardless of size or sophistication. Drawing from established frameworks including the NIST AI RMF and HICP, it adapts best practices to reflect the realities of AI-driven supply chains in healthcare.

Best Practice Components

1. **Governance Policy Development** – Comprehensive AI governance policies covering accountability, data handling, ethical considerations, security controls, and incident reporting
2. **Procurement Process Enhancement** – Use case justification requirements, enhanced Governance Risk and Compliance (GRC) assessments with AI-specific questions covering data lineage, bias mitigation, security controls, and transparency
3. **Contract and Legal Protections** – Model contract language addressing data ownership, AI training restrictions, product update management, performance standards, and liability provisions; enhanced Business Associate Agreement (BAA) clauses for AI-specific HIPAA compliance
4. **Inventory and Asset Management** – Systematic approaches to discovering existing AI systems and establishing ongoing tracking mechanisms for all AI tools, applications, and embedded capabilities
5. **Quality Assurance and Verification and Validation (QA/VV)** – Structured QA/VV frameworks for third-party AI solutions, including requirements for vendor testing documentation, staging environment validation, and re-validation procedures after updates
6. **Response and Recovery Planning** – Integrated incident response coordination with AI vendors, model rollback procedures, and resilience requirements
7. **End-of-Life Management** – Processes for managing AI system transitions, data migration, and secure decommissioning

Key Recommendations

HCO's should:

- **Establish AI governance bodies** appropriate to organizational size and complexity, with clear accountability for oversight, security attestations, risk categorization, approval processes, and training requirements;
- **Implement shared responsibility models** with vendors through contractual transparency requirements, advance notification of changes, and joint validation activities;
- **Enhance procurement workflows** to identify AI early in the acquisition process and require comprehensive vetting before deployment;
- **Proactively manage the AI lifecycle** from initial assessment through end-of-life, with particular attention to update management and configuration validation;
- **Strive for vendor transparency** for model training data, potential biases, and dependencies given the appropriate use case, risk level and impact to the business; and
- **Surface hidden dependencies** through creating and managing an active inventory and utilizing dynamic risk profiling and scalable due diligence tools.

Implementation Approach

The Guide enables scalable implementation:

- **Small/rural organizations:** Baseline requirements and minimum question-sets to establish foundational protections
- **Mid-sized organizations:** Enhanced controls and more comprehensive vendor assessments
- **Large/complex organizations:** Advanced risk stratification, extensive validation protocols, and sophisticated governance structures
- **All organizations:** Share strategies amongst HCOs and with the HSCC AI Task Group.

Call to Action

HCO's are encouraged to:

1. Distribute this document to relevant teams and leadership
2. Evaluate existing third-party and supply chain risk management programs against these best practices
3. Engage with the HSCC AI Task Group to share experiences and contribute to evolving health-sector AI governance
4. Provide feedback to improve this living document as AI technologies and regulations continue to evolve

The Process for AI Third-Party AI Risk and Supply Chain Transparency

Terminology and Definitions

Consistent use of AI terminology is essential for effective governance. This guide uses terms as defined in the HSCC AI Cyber Glossary, the authoritative reference for AI-related terminology across all HSCC AI Task Group publications. Readers encountering unfamiliar terms or seeking precise definitions for governance, procurement, or policy purposes should consult the glossary directly. The AI Cyber Glossary is updated as new Task Group publications are released and as AI technologies and regulatory requirements evolve. It is available at <https://healthsectorcouncil.org/ai-cyber-glossary/>.

Managing third-party AI risk requires a structured, lifecycle-based approach that recognizes the unique characteristics of artificial intelligence systems. Unlike traditional software, AI systems introduce dynamic risks through model drift, training data dependencies, algorithmic bias, and complex supply chain relationships that may span multiple vendors, open-source components, and cloud service providers.

This process establishes a shared responsibility model between HCOs and third-party AI vendors, ensuring transparent management of AI-specific risks throughout the entire technology lifecycle. The guide aligns with the [Health Industry Cybersecurity Practices for Supply Chain Risk Management](#) (HIC-SCRiM) as a baseline and incorporates AI-specific controls. This process should be considered as an enhancement to, not a replacement of, a vendor risk process.



Figure 1: The AI Third-Party Supplier Risk Management Lifecycle

Each phase includes both standard third-party risk management activities (per HIC-SCRiM) and AI-specific enhancements that address unique machine learning risks. See Appendix C for discussion of creating an inventory of AI systems and examples of AI-enabled solution categories.

AI Third-Party Scope

This framework applies to all third-party vendors and their AI-powered solutions, medical devices, and/or services in procurement or usage in HCOs including but not limited to:

Clinical Systems

- Clinical decision support systems (CDSS)
- Diagnostic and treatment tools (radiology, pathology, etc.)
- Predictive analytics for patient outcomes
- AI-enabled medical devices (regulated under FDA)
- Home use AI-Enabled wearable monitors and apps (non-FDA regulated)
- Ambient listening solutions (non-FDA regulated)

Operational & Administrative

- Revenue cycle management
- Patient scheduling and flow optimization
- Supply chain forecasting, fraud detection and compliance monitoring
- Customer relationship management (CRM)

Data & Analytics

- Population health analytics platforms
- Research and development models
- Natural language processing (NLP) engines for documentation
- Data mining and pattern recognition tools

Infrastructure & Embedded

- AI capabilities embedded within EHR systems
- Chatbots and virtual assistants
- Meeting transcription and documentation tools
- Email filtering and security AI
- Network monitoring and threat detection AI

Exclusions (Separate Consideration)

While important, the following are addressed separately and not within the primary scope of this third-party vendor guidance:

- Internally developed AI applications (separate governance framework required)
- AI used solely for personal productivity by individual staff (subject to acceptable use policies)
- Research-only AI not touching production systems or patient data

Detailed Phase-by-Phase Process

Phase 0: AI Use Case Justification & Strategic Assessment

Before committing resources to AI vendor evaluations, HCOs must determine whether AI is truly the appropriate solution. “Phase 0” serves as a critical gate-keeping function that prevents unnecessary AI adoption, ensuring organizations only pursue AI when there is clear strategic alignment, demonstrable value, and acceptable risk.

This phase establishes the foundation for subsequent evaluation by documenting the specific problem, assessing whether AI is required versus non-AI approaches, and analyzing benefits, risks, costs, and organizational readiness. The outcome is well-documented justification that defines intended use cases, classifies the AI system by safety impact level (Low, Medium, High, or Critical), identifies key stakeholders and governance requirements, and establishes clear success criteria.

By carefully assessing whether AI delivers real value given its complexity and risk, organizations can avoid costly mistakes, streamline vendor evaluations, and build consensus on goals, risk tolerance, and accountability. This disciplined approach protects organizations from adopting AI for innovation's sake rather than solving clearly defined problems with measurable outcomes.

For a proposed template to review Use Cases see [Appendix A](#). For a proposed policy for AI Governance See [Appendix B](#). For proposed roles and responsibilities in a RACI diagram see [Appendix D](#).

Key Activities:

1) Problem Definition & Solution Assessment

- a) Document the specific problem or opportunity the AI solution will address
- b) Evaluate whether AI is required or if non-AI approaches will suffice
- c) Consider the added complexity and risk AI introduces versus potential benefits
- d) Assess return on investment (ROI) and total cost of ownership (TCO)

2) Use Case Documentation

- a) Define intended use cases with clinical or operational precision
- b) Identify user populations and frequency of use
- c) Document expected workflows and integration points
- d) Determine criticality to patient care and business operations

3) Risk Classification & Categorization

- a) Classify AI solution by patient and operational safety impact: Low, Medium, High, Critical
 - i) **Low Impact:** Minimal or no safety or financial impact
 - (1) AI failure would not affect patient care, safety, or significant financial outcomes
 - (2) Errors are easily detected and corrected by users
 - (3) Examples: Word prediction tools, email autocomplete, meeting scheduling assistants
 - ii) **Medium Impact:** Moderate safety or financial impact
 - (1) AI supports decision-making but humans retain full control and review all outputs
 - (2) Errors could affect individual patients or transactions but are caught through validation processes; clear mechanisms for human override and correction required
 - (3) Examples: Clinical decision support where clinicians make final decisions, coding assistance tools, supply chain forecasting
 - iii) **High Impact:** Significant safety or financial impact
 - (1) AI substantially influences important clinical or financial decisions
 - (2) Errors could affect multiple patients, department-level or service-line operations, and/or significant financial outcomes
 - (3) Examples: Automated medication dosing recommendations, AI-driven triage systems, fraud detection with automatic claim denials, predictive models influencing treatment planning
 - iv) **Critical Impact:** Life-threatening safety risk or enterprise-critical financial/operational impact
 - (1) AI may result in autonomous or near-autonomous decisions affecting patient life/safety, multiple services, departments, or organization-wide operations
 - (2) Failure could result in serious harm, death, or catastrophic financial/operational consequences
 - (3) Examples: Autonomous diagnostic AI without physician review, AI-controlled medical devices (ventilators, insulin pumps), fully automated clinical pathways, enterprise-wide operational AI (staffing, capacity management), AI controlling critical infrastructure
- b) Categorize by data sensitivity (Protected Health information (PHI)/ Personally Identifiable Information (PII) access requirements)
 - i) Determine regulatory classification (FDA medical device, Software as a Medical Device (SaMD), Clinical Decision Support Software, non-device health IT, etc.)
- c) Determine regulatory classification (FDA medical device, non-device software, Payment Card Industry(PCI) implications, etc.)

4) Stakeholder Identification

- a) Identify business owners and clinical champions
- b) Engage privacy, security, legal, and compliance teams early
- c) Establish governance committee review requirements based on risk classification

5) Alternatives Analysis

- a) Document alternative approaches considered (non-AI solutions, internal development, different vendors)
- b) Justify why AI approach is optimal for the specific use case versus the alternative

AI-Specific Considerations:

- Does the AI solution process PHI/PII, and if so, what type of AI model (classical ML, generative AI, LLM)?
- What is the level of automation and human oversight in the AI workflow?
- Are there regulatory requirements specific to this AI application (FDA clearance, etc.)?
- What level of model transparency or explainability is required for this use case based on clinical impact, regulatory requirements, and/or organizational governance?
- Where are AI models trained? (Publicly (e.g. OpenAI, Anthropic) or Privately (e.g., internal high compute platforms/foundation models))
- Where is the data hosted? (public cloud, vendor infrastructure, internal system/on premises, etc.).
- What controls exist to ensure data protection, model integrity, and regulatory compliance?

Deliverables:

- Use Case Justification Document
- Initial Risk Classification
- Stakeholder Identification Matrix
- Business Case with ROI/TCO Analysis

Key Takeaway: Not every problem requires an AI solution. This phase prevents unnecessary AI adoption and ensures that when AI is pursued, there is clear strategic alignment and understanding of the risk profile before resources are committed to vendor evaluation.

Phase 1: Vendor Evaluation and Due Diligence

Phase 1 represents the critical decision point where HCOs determine which AI vendors merit trust with patient data, clinical workflows, and operations. Unlike traditional software evaluation focused on functionality and basic security, AI vendor assessment demands deeper scrutiny into training data provenance and quality, algorithmic bias mitigation, model transparency and explainability, external AI service dependencies, and responsible AI governance structures. Organizations relying solely on standard vendor questionnaires and security certifications without probing AI-specific risks may inevitably onboard vendors whose systems later exhibit performance degradation, bias, security vulnerabilities, or opacity undermining clinical safety and regulatory compliance.

A fundamental Phase 1 challenge is assessing not only new AI vendors during active procurement, but also conducting retroactive evaluation of existing vendors and AI-enabled products already deployed. Many organizations discover through asset inventory initiatives that AI capabilities have proliferated without formal oversight—medical devices received AI-enabled firmware updates, EHR systems activated AI clinical decision support features, and administrative tools incorporated AI—all without triggering AI-specific risk assessments. Organizations must implement tiered assessment frameworks: essential baseline questions for all AI vendors, enhanced assessments for Medium/High-impact systems, and comprehensive evaluation for critical-impact AI affecting patient life and safety. Effective execution requires cross-functional collaboration among procurement, security, privacy, compliance, legal, clinical leadership, and business owners.

For example vendor assessment questions, see [Appendix G](#).

Key Activities:

1) Procurement and Asset Inventory Discovery and Intake

- a) Identify an inventory of AI-enabled third-party vendors, products, solutions, devices, and services
 - i) Those that are new and in the procurement process
 - ii) Those that are existing through an asset inventory and discovery process
- b) Conduct intake of relevant data
 - i) Needed for the due diligence and vendor evaluation process
 - ii) In support of the business impact analysis

2) Standard Third-Party Risk Assessment (per [HIC-SCRiM](#))

- a) Vendor financial stability and business continuity
- b) General cybersecurity practices and certifications (e.g. NIST, Soc 2 Type II, etc)
- c) Data residency and privacy compliance
- d) Standard contractual terms and insurance coverage
- e) Reference checks and market reputation
- f) Willingness to sign HCOs BAA

3) AI-Specific Governance, Risk, and Compliance (GRC) Assessment

a) Data Lineage, Quality & Bias

- i) Model training methodology and data sources
- ii) Training data quality assurance processes
- iii) Bias assessment and mitigation strategies (race, gender, age, socioeconomic factors)
- iv) Data traceability from source to model output
- v) Use of synthetic data and validation processes
- vi) PHI/PII redaction and de-identification methods

b) Model Transparency & Explainability

- i) Model architecture documentation
- ii) Software Bill of Materials (SBOM) and AI Bill of Materials (AI BOM) listing which feature/functions are AI capable and enabled, model dependencies and third-party services or equivalent documentation describing model architecture, training data, dependencies, external services, and third-party components
- iii) Explainability capabilities appropriate to use case criticality
- iv) Decision boundaries and confidence thresholds
- v) Human oversight mechanisms and override capabilities
- vi) Version control and model lineage tracking

c) Security Controls & AI-Specific Risks

- i) Review of data required for model (PHI/PII, sensitive, redacted, etc) and locations of processing of that data (on premises, in cloud, etc)
- ii) Infrastructure hardening against AI-specific attacks including membership inference
- iii) Protections against prompt injection and adversarial input attacks
- iv) Model extraction/model theft, model inversion/training data leakage and reverse engineering protections
- v) Anomalous AI behavior monitoring capabilities

- vi) Data poisoning and adversarial input controls
- vii) Multi-factor authentication and access controls for AI systems

d) AI Governance & Risk Management Framework

- i) Vendor's internal AI governance structure
- ii) Model development lifecycle and approval processes
- iii) Risk assessments for AI system failure modes
- iv) Model drift and degradation monitoring approaches
- v) Quality management system for AI in clinical settings
- vi) Alignment with NIST AI RMF, FDA guidance, or other frameworks

e) Third-Party & Supply Chain Dependencies

- i) Use of third-party models, application programming interfaces (APIs), or data services
- ii) Open-source component inventory and management
- iii) Cloud service provider relationships (especially for LLMs like OpenAI, Anthropic, etc.)
- iv) Existence of BAAs with all AI subcontractors (if required)
- v) [HIC-SCRiM](#) or equivalent supply chain risk assessments
- vi) Vendor's process for evaluating and monitoring their own suppliers (fourth party suppliers)

f) Regulatory Compliance & Validation

- i) FDA clearance or approval status (if regulated as medical device, SaMD, etc.)
- ii) Clinical validation studies and evidence
- iii) Compliance with applicable regulations including but not limited to: Health Insurance Portability and Accountability Act (HIPAA); Health Information Technology for Economic and Clinical Health Act (HITECH); Protecting and Transforming Cyber Health Care (PATCH) Act of 2022; Federal Information Security Modernization Act (FISMA)' Section 5 of the FTC Act; and state privacy and data laws
- iv) Compliance with applicable international regulations (EU AI Act, General Data Protection Regulation (GDPR), etc.)
- v) Quality management system documentation
- vi) Adverse event reporting procedures

g) Operational Readiness

- i) Integration capabilities (API, HL7, FHIR compatibility)
- ii) System maintenance procedures (patches, upgrades, updates)
- iii) Service Level Agreements (SLAs) for uptime and performance
- iv) Training and support services provided
- v) User permission and role-based access configuration
- vi) Sandbox or test environment availability

h) Ethical & Responsible AI Practices

- i) AI ethics principles and implementation
- ii) Intended and prohibited uses in healthcare settings
- iii) Human oversight requirements in AI decision-making
- iv) Internal AI ethics board or review processes
- v) Disclosure practices for generative AI and LLM use

i) Security Risk Assessment (SRA)

- i) Comprehensive security assessment aligned with organizational requirements
- ii) Penetration testing and vulnerability assessment results
- iii) Incident response capabilities and history (including prior model failures, bias incidents and model recalls)
- iv) Audit logging and system activity tracking
- v) Encryption standards for data at rest and in transit

j) Reference Checks & Validation

- i) Contact existing healthcare customers
- ii) Verify vendor claims about capabilities and performance
- iii) Assess vendor responsiveness and support quality
- iv) Review any publicly available incident or breach history

AI-Specific Considerations:

- Prioritize questions based on organizational size and AI maturity (implement tiered question sets: essential/enhanced/comprehensive)
- Rural and smaller organizations should focus on essential baseline questions
- Larger organizations should implement comprehensive assessments for high-risk AI systems as defined the governance of the organization

Deliverables:

- Completed GRC Assessment with AI-specific questions
- Security Risk Assessment Report
- Vendor Scorecard with risk rating
- Gap analysis identifying unmet requirements
- Recommendation for approval, conditional approval, or rejection

Key Takeaway: AI vendor evaluation requires both traditional IT security assessment and specialized AI risk evaluation. Organizations should develop tiered assessment approaches scaled to their size, resources, and the risk profile of specific AI solutions.

Phase 2: Contract Negotiation & Legal Protections

Standard software licensing agreements and BAAs are insufficient for AI systems in healthcare. Unlike conventional software, AI systems evolve through model updates, drift as data distributions change, and exhibit unpredictable behaviors. Without appropriate contractual protections, organizations face liability for AI failures, inability to access critical model information, vendor lock-in, and inadequate security recourse.

Effective AI contracting establishes a shared responsibility framework delineating accountability for governance, risk management, security, and compliance. It creates enforceable transparency obligations—requiring disclosure of model architectures/design, training data sources, limitations, and dependencies—while providing audit rights, update approval processes, and termination protections. These terms enable organizations to enforce governance

policies through vendor participation in quality assurance, incident response, and performance validation essential for patient safety and compliance.

Contract management extends beyond execution. Organizations must monitor vendor compliance to the contract, track renewals, document performance issues, and update terms as technology and regulations evolve.

For BAAs, AI processing of Protected Health Information (PHI) requires AI-specific amendments addressing model training restrictions, data minimization, and security safeguards aligned with HIPAA and organizational governance policies.

For examples of contract language, see [Appendix E](#). For examples of additional clauses for BAAs, see [Appendix F](#).

Key Activities:

1) Standard Commercial Contract Terms

- a) Scope of license and permitted uses
- b) Pricing, payment terms, and renewal conditions
- c) Service Level Agreements (SLAs) for system availability
- d) Support and maintenance provisions
- e) Warranties and representations
- f) Limitation of liability (ensure mutual protections)
- g) Indemnification (ensure mutual provisions)
- h) Termination rights and notice periods
- i) Dispute resolution mechanisms

2) AI-Specific Contract Clauses

a) Data Ownership & Control

- i) Organization retains ownership of all input data, output data, and derived insights
- ii) Rights/ownership/IP of model improvements must be explicitly defined in the contract
- iii) Vendor cannot use organizational data to train or improve models without explicit written consent
- iv) Clear delineation of what constitutes "training data" versus "operational data"
- v) Restrictions on model reuse across other clients including restrictions on using the organization's proprietary data or PHI to train shared models without explicit authorization

b) Confidentiality & Intellectual Property

- i) Protection of trade secrets, algorithms, and training methodologies (mutual)
- ii) Confidentiality of test results and performance metrics
- iii) Clear ownership of any custom-developed AI models
- iv) Protection of prompt data and user interactions

c) Security & Compliance Requirements

- i) Specific security controls based on GRC assessment findings
- ii) Adherence to applicable healthcare cybersecurity regulations including HIPAA, HITECH, PATCH Act, and applicable state privacy laws
- iii) Explicit prohibition on using PHI or PII to train public or third-party AI models without prior written consent
- iv) Encryption standards for data in transit and at rest

- v) Access control and authentication requirements
- vi) Compliance with organizational security policies
- vii) Regular security assessment and penetration testing requirements
- viii) Right to audit security controls

d) Updates, Patches & Change Management

- i) Defined cadence and approval process for updates, patches, and feature changes
- ii) Advance notification requirements (e.g., 30 days for non-critical, 72 hours for security patches)
- iii) Provision of detailed release notes and change documentation
- iv) Compliance with approved Predetermined Change Control Plans (PCCP) for AI-Enabled medical devices, where applicable.
- v) Requirement for sandbox/test environment deployment before production
- vi) Testing and verification/validation support from vendor
- vii) Organization's right to delay or reject updates
- viii) Rollback procedures and vendor support for regression issues
- ix) Documentation of validation testing completed by vendor prior to release

e) Model Performance & Quality Assurance

- i) Documented Quality Assurance and Verification/Validation (QA/VV) plan
- ii) Performance metrics and accuracy thresholds
- iii) Model drift monitoring and revalidation requirements
- iv) Notification procedures for performance degradation
- v) Bias monitoring and mitigation commitments
- vi) Clinical validation documentation (if applicable)

f) Transparency & Explainability

- i) Requirements for model documentation appropriate to risk level
- ii) Explainability features aligned with intended use
- iii) Access to model decision rationale for high-risk applications
- iv) Disclosure of third-party models, APIs, or data sources used
- v) Notification of changes to model architecture or training data

g) Incident Response & Breach Notification

- i) Mandatory communication plan and defined timelines (e.g., 24-48 hours for security incidents) including obligations under the FTC Health Breach Notification Rule (if any)
- ii) Vendor obligations to assist in incident investigation and recovery
- iii) Requirements for coordinated response to AI-related incidents (model failures, bias events, data breaches)
- iv) Post-incident reporting and root cause analysis
- v) Notification of near-harm or harm events to appropriate authorities (FDA reporting under Safe Medical Devices Act where applicable)

h) Data Return, Destruction & Portability

- i) Requirements for data return or secure destruction at contract end
- ii) Timeline for data return (e.g., 30 days post-termination)
- iii) Format requirements for returned data (interoperable, non-proprietary)

- iv) Certification of data destruction
- v) Transition assistance to replacement vendor
- vi) No data retention by vendor after termination without explicit agreement
- i) Third-Party Dependencies & Supply Chain**
 - i) Inventory of all third-party AI services, models, and data providers
 - ii) Flow down security and privacy requirements to subcontractors
 - iii) BAAs with all entities handling PHI
 - iv) Notification of changes to third-party relationships
 - v) Organization's right to approve or reject specific subcontractors
- j) Regulatory Compliance & Liability**
 - i) Vendor representations regarding FDA clearance/approval (if applicable)
 - ii) Compliance with evolving AI regulations (EU AI Act, state AI laws, etc.)
 - iii) Liability allocation for AI-generated errors or harm
 - iv) Vendor's obligation to maintain regulatory compliance throughout contract term
 - v) Notification requirements for regulatory inquiries or enforcement actions
 - vi) No limitation on vendor's obligation to report safety events to regulators
- k) End-of-Life & Transition Support**
 - i) Advance notification requirements for end-of-life (12-18 months minimum)
 - ii) Vendor's obligations to support migration to replacement systems
 - iii) Continued security support during transition period
 - iv) Data migration assistance and validation
 - v) Documentation and knowledge transfer
- 3) Business Associate Agreement (BAA) Enhancements for AI**
 - a) Explicit prohibition on using PHI for model training, testing, or improvement outside the contracted services without explicit written consent
 - b) Clear definition of permitted uses and disclosures specific to AI processing
 - c) Minimum necessary standard enforcement for AI data access
 - d) AI-specific safeguards (technical and administrative)
 - e) Breach notification timelines specific to AI systems (e.g., 2 days for AI-related incidents)
 - f) Flow-down requirements to all AI subcontractors and service providers
 - g) Patient access and amendment rights for AI-processed PHI
 - h) Post-termination data handling for AI models containing embedded PHI
 - i) Audit rights specific to AI Governance, AI data flows and processing, and other validation documentation where appropriate.
- 4) Service Level Agreements (SLAs)**
 - a) System uptime and availability targets
 - b) Response time for support requests
 - c) Resolution timeframes for critical issues
 - d) Model performance baselines (accuracy, precision, recall as applicable)
 - e) Penalties for SLA breaches
 - f) Escalation procedures

AI-Specific Considerations:

- Consider whether a separate AI-specific BAA or BAA addendum is appropriate for high-risk AI systems
- Ensure no limitations on reporting safety events to FDA or other regulatory bodies
- Balance vendor's need for model improvement with organization's data protection requirements
- Address the unique challenge of AI models that may "learn" or adapt, requiring periodic reassessment

Deliverables:

- Master Services Agreement with AI-specific clauses
- Business Associate Agreement or BAA Addendum for AI
- Service Level Agreement
- Data Processing Agreement (if applicable)
- Statement of Work defining specific deliverables and milestones

Key Takeaway: Standard vendor contracts are insufficient for AI systems. Organizations must negotiate AI-specific protections addressing data use for training, model updates, explainability, bias management, and shared responsibility for incident response and quality assurance.

Phase 3: Implementation, Integration & Training

Following vendor selection and contract negotiation, Phase 3 focuses on the critical transition to production deployment, one of the highest-risk periods in the AI lifecycle. Organizations must safely integrate AI systems into existing workflows while ensuring security, compliance, and user readiness. Rushed implementation can result in system failures, security vulnerabilities, and patient safety incidents.

Unlike traditional software, AI implementations require additional validation of model performance, bias mitigation, clinical accuracy, and fail-safe mechanisms unique to machine learning. Organizations must verify that systems perform as expected in their specific environment, which may differ from vendor demonstrations due to variations in data quality and workflow patterns. AI systems can behave unpredictably with real-world data and create unintended consequences when integrated with existing systems. For details on quality assurance, verification and validation see [Appendix I](#).

Traditional application security approaches are insufficient for AI systems. Traditional static code analysis techniques are insufficient for detecting vulnerabilities in machine learning models — AI vulnerabilities are often behavioral, not syntactic. This means organizations must go beyond conventional vulnerability management and conduct AI-specific threat modeling that accounts for prompt injection, data poisoning, model manipulation, and the unique attack surfaces created when AI systems interact with clinical environments and patient data.

Equally important is ensuring end users understand AI capabilities and limitations, recognize when outputs require scrutiny, know how to override recommendations, and continuously monitor errors. Inadequate training leads to workarounds bypassing safety controls, over-reliance without oversight, or user rejection. By investing in role-specific training, hands-on practice, and competency assessment, organizations build the foundation for safe and effective AI operations that deliver value while managing risks. For a full training checklist and curriculum, see [Appendix H](#).

Key Activities

1) Pre-Implementation Planning

- a) Develop detailed implementation project plan with milestones
- b) Assign project roles and responsibilities (RACI matrix)
- c) Establish success criteria and acceptance testing requirements
- d) Plan for testing in sandbox/staging environment before production
- e) Identify pre-deployment security controls that must be implemented to minimize risk
- f) Define rollback procedures in case of implementation issues

2) Threat Modeling

AI threat modeling must be conducted before production deployment to identify vulnerabilities specific to machine learning systems. Traditional threat models focus on infrastructure and application-layer risks; AI threat modeling extends this to cover behavioral vulnerabilities, adversarial inputs, and the expanded attack surface that AI creates in clinical environments.

Healthcare Attack Surface Analysis

Map how AI vulnerabilities manifest across each integration point in your environment:

- **EHR Integrations:** Patient data exfiltration via prompt manipulation, unauthorized access through AI-assisted queries, PHI leakage in model responses
- **Clinical Decision Support:** Model poisoning affecting treatment recommendations, adversarial inputs leading to misdiagnosis, liability exposure from AI-influenced clinical errors
- **Patient-Facing Chatbots:** Prompt injection for unauthorized data access, social engineering through AI manipulation, HIPAA violations via conversational exploitation
- **Ambient Documentation:** Sensitive conversation capture and leakage, injection attacks via spoken commands, unauthorized data aggregation

OWASP Top 10 for LLM Assessment

Assess the AI solution against the [OWASP Top 10](#) for LLM Applications, documenting applicable threats and required controls for each:

- **LLMo1 — Prompt Injection:** Evaluate exposure to user prompts that alter model behavior or output in unintended ways; implement input/output filtering, privilege enforcement, content segregation, and adversarial testing
- **LLMo2 — Sensitive Information Disclosure:** Assess risk of the model exposing PHI, proprietary algorithms, or confidential details through its output; implement data sanitization, access controls, and federated learning where appropriate
- **LLMo3 — Supply Chain:** Evaluate integrity of training data, pre-trained models, and deployment platforms from third-party sources; maintain SBOMs, apply scanning, and implement strict monitoring and auditing

- **LLMo4 — Data and Model Poisoning:** Assess whether pre-training, fine-tuning, or embedding data could be manipulated to introduce vulnerabilities, backdoors, or biases; track data origins, vet data vendors, validate outputs against trusted sources, and implement strict sandboxing
- **LLMo5 — Improper Output Handling:** Evaluate validation and sanitization of model outputs before they pass downstream to clinical systems; adopt zero-trust approach, implement context-aware output encoding, and use parameterized queries
- **LLMo6 — Excessive Agency:** Assess whether the AI system has been granted more autonomy than necessary — in healthcare, AI should recommend, not execute, medication changes or order modifications; implement least-privilege access and require confirmation for sensitive operations
- **LLMo7 — System Prompt Leakage:** Evaluate whether system prompts or instructions contain sensitive information that could be extracted; separate sensitive data from system prompts and implement guardrails and security controls outside the model
- **LLMo8 — Vector and Embedding Weaknesses:** For RAG-based systems, assess how vectors and embeddings are generated, stored, and retrieved; implement fine-grained access controls and robust data validation pipelines
- **LLMo9 — Misinformation:** Evaluate the risk of the model producing false or misleading clinical information that appears credible; implement RAG, cross-check outputs, automatically validate key outputs, and communicate known limitations to users
- **LLMo10 — Unbounded Consumption:** Assess risk of excessive or uncontrolled inferences leading to denial of service, economic loss, or service degradation; implement rate limiting, resource allocation management, timeouts, and throttling

Note: the assessment above is version 2.0 of the OWASP. Organizations should verify they are using the current version at <https://genai.owasp.org/llm-top-10/>.

While the OWASP Top 10 for LLM Application provides useful guidance for generative AI systems, organizations should adapt these threat modeling approaches to the specific AI architecture used, including predictive models, imaging system, and other machine learning applications.

Agentic AI Threat Assessment

If the AI solution includes autonomous or semi-autonomous agent capabilities, conduct additional threat modeling that treats AI agents as a new category of insider:

- **Identity Management:** Verify every AI agent has a unique identity with documented capabilities and boundaries
- **Credential Management:** Confirm AI credentials are managed with the same rigor as privileged service accounts
- **Behavioral Baselines:** Establish and document normal operating patterns for each AI agent to enable anomaly detection
- **Rogue Detection:** Define monitoring for deviation from expected behavior, unexpected data access, or privilege escalation
- **EHR Access:** An AI agent with EHR access and excessive permissions is functionally equivalent to a compromised insider with clinical system privileges — scope and constrain accordingly

Threat Model Documentation

- Document all identified threats, their likelihood or exploitability and potential impact, and the controls required to mitigate each
- Map threats to specific integration points, data flows, and user interactions
- Establish risk acceptance criteria for residual threats that cannot be fully mitigated
- Include threat model findings in go-live approval criteria

3) Technical Integration & Testing

a. Sandbox/Staging Environment Testing

- i. Deploy AI solution in non-production test environment
- ii. Validate integration with EHR and other clinical systems (API, HL7, FHIR interfaces)
- iii. Test data flows and transformations
- iv. Verify security controls and access permissions
- v. Conduct user acceptance testing (UAT) with representative workflows
- vi. Validate AI model performance against expected baselines
- vii. Test for adverse interactions with existing systems
- viii. Validate rollback and recovery procedures

b. Security Validation

- i. Verify encryption implementation (data at rest and in transit)
- ii. Validate authentication and authorization controls
- iii. Test audit logging and monitoring capabilities
- iv. Confirm data isolation and access controls
- v. Conduct vulnerability scan of integrated solution
- vi. Conduct penetration test based on risk level
- vii. Validate any other specific security controls identified as required pre-deployment
- viii. Validate controls identified in AI threat model are implemented and functioning

c. AI-Specific Security Testing

- i. Conduct adversarial testing against prompt injection attack vectors identified in threat model
- ii. Test input validation and output filtering controls
- iii. Verify that AI agent permissions enforce least-privilege access
- iv. Test behavioral monitoring and alerting capabilities
- v. Validate that AI outputs are treated as untrusted and validated before integration with clinical systems (e.g., clinical notes validated before EHR integration)
- vi. Confirm content segregation between system instructions and user inputs

d. Clinical Validation & Safety Testing

- i. Confirm AI outputs meet clinical accuracy requirements
- ii. Validate decision support recommendations with clinical experts
- iii. Test edge cases and failure modes
- iv. Verify appropriate handling of missing or incomplete data

- v. Confirm human override capabilities function correctly
 - vi. Validate alert fatigue mitigation strategies
 - vii. Test for overreliance scenarios by verifying that workflows enforce human verification for critical clinical decisions, ensuring diagnostic AI augments rather than replaces clinical judgment
 - viii. Evaluate model performance under local data distributions to identify potential dataset shifts or performance degradation compared to vendor-reported results.
- e. **Verification & Validation Activities**
- i. Document that AI system performs as intended for specified use cases
 - ii. Verify that system meets contractual performance requirements
 - iii. Validate compliance with regulatory and privacy requirements (FDA, HIPAA, etc.)
 - iv. Confirm bias mitigation controls are functioning
 - v. Test fail-safe mechanisms and degradation scenarios
 - vi. Document all activities for regulatory and compliance purposes
 - vii. Validate threat model controls are in place and functioning as designed
2. **Privacy & Consent Management**
- a. **Privacy Impact Assessment**
- i. Conduct or update Privacy Impact Assessment (PIA) for AI solution
 - ii. Document data flows and PHI/PII processing activities
 - iii. Assess privacy risks specific to AI processing, including risks identified in threat model (sensitive information disclosure, data aggregation, conversational exploitation)
 - iv. Implement privacy controls appropriate to risk
- b. **Patient Consent & Disclosure**
- i. Determine disclosure requirements based on risk tolerance and organizational policy
 - ii. Update Notice of Privacy Practices (NPP) if AI use requires disclosure
 - iii. Implement patient consent processes where required
 - iv. Develop patient-facing communications about AI use
 - v. Consider risk-based disclosure (high-risk AI decisions require greater transparency)
3. **User Training & Change Management**
- a. **Role-Specific Training Programs**
- i. **Clinical Users:** AI-assisted clinical workflows, interpreting AI outputs, override procedures, documentation requirements
 - ii. **IT Staff:** System administration, troubleshooting, integration management, update procedures
 - iii. **Security Team:** Monitoring AI-specific threats (prompt injection, data exfiltration, model manipulation), AI incident response procedures, audit log review, agentic AI behavioral anomaly detection
 - iv. **Compliance Staff:** Regulatory requirements, documentation obligations, reporting procedures
 - v. **Business Owners:** Performance monitoring, escalation procedures, governance oversight

- b. **Training Content & Delivery**
 - i. Hands-on training in test environment
 - ii. Workflow integration and change management
 - iii. Understanding AI limitations and appropriate use
 - iv. Recognizing and reporting AI errors, bias, or unexpected outputs
 - v. AI-specific security awareness: prompt injection risks, social engineering via AI, and appropriate data handling
 - vi. Privacy and security considerations
 - vii. Documentation of training completion
- c. **Competency Assessment**
 - i. Establish competency requirements by role
 - ii. Conduct skills assessment before granting production access
 - iii. Implement ongoing competency validation
 - iv. Refresher training requirements (e.g., annually or after major updates)
- 4. **Documentation & Asset Registration**
 - a. Register AI system in organizational inventory/asset management system
 - b. Document intended use, approved workflows, and user populations
 - c. Record integration points and data flows
 - d. Maintain configuration baselines and settings documentation
 - e. Document business owner, technical owner, and vendor contacts
 - f. Establish change control procedures for future modifications
 - g. Include AI threat model and risk assessment documentation in system record
 - h. Document AI agent identities, permissions, and behavioral baselines (if applicable)
- 5. **AI Incident Response Preparedness**
 - a. Establish AI-specific incident response capabilities that account for the unique characteristics of AI failures and attacks
 - b. Define behavioral alerts, user reporting channels, and output anomaly thresholds; establish criteria for assessing harm, blast radius, and severity
 - c. Document graduated response with escalating controls: rate limiting → input/output restriction → capability reduction → full shutdown; ensure clinical continuity plans exist for each escalation level
 - d. Define procedures for investigating and collecting prompts, outputs, data lineage, logs, and network traffic; establish processes for identifying attack vectors, data exposure, and downstream clinical impact
 - e. Document remediation procedures for patching guardrails, rotating credentials, blocking attack patterns; define criteria for when model retraining or system redesign is required
 - f. Plan for recovery, reporting, and restoration with heightened monitoring; define board, regulatory, and industry reporting obligations
 - g. Conduct tabletop exercises for AI-specific incident scenarios before production deployment
- 6. **Phased Production Rollout**
 - a. Consider phased deployment approach (pilot users, single department, then broader)

- b. Implement enhanced monitoring during initial rollout period, including behavioral baselines established during threat modeling
- c. Collect user feedback and address issues before full deployment
- d. Validate performance metrics in production environment
- e. Confirm audit logging is capturing required information
- f. Monitor for threats identified in AI threat model during initial production period

AI-Specific Considerations:

- AI systems may behave differently in production than in testing due to data distribution differences
- Plan for "learning period" if AI adapts over time, with enhanced monitoring
- Ensure users understand AI is a tool to augment, not replace, professional judgment
- Training must address how to recognize and respond to AI errors or unexpected outputs
- Consider whether AI indicators should be visible to patients (e.g., chatbot disclosures)
- AI vulnerabilities emerge from behavior, not just code. Continuous behavioral monitoring and periodic adversarial testing (red teaming) must continue post-deployment
- AI security is an ongoing program that must evolve as AI capabilities and threats evolve; organizations that succeed will build AI security into their DNA rather than bolting it on as an afterthought

Deliverables:

- Implementation Project Plan
- AI Threat Model and Risk Assessment
- [OWASP Top 10 for LLM](#) Assessment Results
- Integration Test Results
- Security Validation Report (including AI-specific security testing)
- Validation and Verification Documentation Package
- Privacy Impact Assessment
- AI Incident Response Playbook
- Training Materials and Completion Records
- System Documentation and Operational Procedures
- Go-Live Approval and Sign-off

Key Takeaway: Successful AI implementation requires rigorous threat modeling, testing, validation, and training before production use. Traditional application security is insufficient; therefore, organizations must model threats specific to AI behavior, including prompt injection, data poisoning, excessive agency, and the expanded attack surface AI creates in clinical environments. Organizations must verify that AI systems perform as expected in their specific environment, that controls mitigate identified threats, and that users understand both capabilities and limitations of the AI solution.

Phase 4: Ongoing Monitoring & Performance Management

Phase 4 represents the longest and most resource-intensive period of the AI lifecycle, spanning from deployment through end-of-life. Unlike traditional software requiring only periodic patches and incident response, AI systems demand continuous monitoring due to their dynamic nature. AI models drift as input data changes, performance degrades gradually without systematic measurement, and frequent vendor updates involving model retraining introduce risks that standard change management cannot address. Security configurations may reset to default settings during updates, bias may emerge across different populations, and new AI-specific vulnerabilities like prompt injection attacks require constant vigilance.

Effective monitoring requires sustainable processes balancing comprehensiveness with feasibility through automation, including dashboards tracking performance indicators, alerting systems flagging anomalies, and validation tools detecting drift, while maintaining human oversight. Monitoring should be risk-based with clear escalation paths. Phase 4 encompasses managing vendor updates through test environment deployment, validation of security and performance, change management approval, and post-deployment monitoring. Periodic reassessments, typically annually or at contract renewal, revisit vendor evaluation, update risk classifications, and validate continued appropriateness throughout the operational lifecycle.

For additional information on verification and validation during this phase, see [Appendix I](#).

Key Activities:

1. Performance Monitoring & Metrics

a. AI-Specific Performance Indicators

- i. Model accuracy, precision, recall (as applicable to use case)
- ii. False positive and false negative rates
- iii. User override frequency and patterns
- iv. Decision confidence score distributions
- v. Processing time and system responsiveness
- vi. User adoption and utilization rates
- vii. Clinical Outcome Correlation: Compare real-world performance metrics against validation benchmarks provided during procurement or implementation

b. Model Drift Detection

- i. Monitor for degradation in model performance over time
- ii. Track changes in input data distributions
- iii. Identify concept drift (relationships between inputs and outputs change)
- iv. Establish thresholds for acceptable performance variation
- v. Trigger revalidation when drift exceeds thresholds

c. Bias & Fairness Monitoring

- i. Track AI performance across demographic subgroups (age, gender, race, ethnicity, socioeconomic status)
- ii. Monitor for discriminatory outcomes or disparate impact

- iii. Investigate anomalies or performance disparities
 - iv. Report bias concerns through governance channels
2. **Security & Compliance Monitoring**
- a. **Security Controls Validation**
 - i. Regular review of access logs and authentication events
 - ii. Monitoring for unusual AI system behavior or anomalies
 - iii. Scanning for vulnerabilities in AI components
 - iv. Validation that security configurations remain intact
 - v. Assessment of new AI-specific threat intelligence
 - vi. Monitor for AI-specific attack patterns such as prompt injection attempts, adversarial inputs, abnormal interference activity or agent behavior abnormalities
 - b. **Compliance Auditing**
 - i. Periodic audits of AI system against contractual requirements
 - ii. Validation of Business Associate Agreements (BAAs, compliance and PHI handling)
 - iii. Review of patient consent and disclosure practices
 - iv. Regulatory compliance verification (FDA, HIPAA, state laws)
 - v. Documentation of audit findings and corrective actions
3. **Vendor Performance Management**
- a. **SLA Monitoring & Reporting**
 - i. Track vendor performance against SLA commitments
 - ii. Document incidents of non-compliance
 - iii. Conduct regular vendor performance reviews
 - iv. Escalate persistent issues through governance channels
 - b. **Vendor Communication & Coordination**
 - i. Regular vendor check-ins and status meetings
 - ii. Review of vendor-provided performance reports
 - iii. Discussion of emerging risks or issues
 - iv. Coordination on planned updates or changes
 - v. Sharing of relevant threat intelligence
4. **Update & Patch Management**
- a. **Update Notification & Assessment**
 - i. Receive and review vendor update notifications and release notes
 - ii. Assess criticality and potential impact of updates
 - iii. Evaluate changes to AI model, algorithms, or configurations
 - iv. Determine testing requirements based on update scope
 - v. Submit updates through organizational change management process
 - b. **Testing & Validation of Updates**
 - i. Deploy updates to sandbox/test environment first
 - ii. Conduct functional testing of updated capabilities
 - iii. Validate integration with other systems remains intact
 - iv. Verify security settings were not reset to defaults

- v. Test AI model performance after update
- vi. Review vendor's validation and verification documentation for the update
- vii. Obtain business owner and IT approval before production deployment
- c. **Production Deployment & Post-Update Validation**
 - i. Schedule production deployment during maintenance windows
 - ii. Implement update according to change management procedures
 - iii. Immediately verify critical security and privacy settings
 - iv. Conduct post-deployment validation testing
 - v. Monitor closely for unexpected behavior during window period between update and setting revalidation
 - vi. Document any configuration changes required to restore desired state
- 5. **Algorithmic Transparency & Governance**
 - a. Maintain documentation of AI model architecture and training
 - b. Update records when vendor makes model changes
 - i. Maintain version history for AI models including training datasets, parameter configurations, and deployment dates.
 - c. Ensure explainability remains appropriate for use case
 - d. Review AI decision-making processes periodically
 - e. Validate continued appropriateness of AI solution for intended use
- 6. **Incident & Issue Management**
 - a. Establish reporting mechanisms for AI-related concerns (accuracy issues, bias, errors, security concerns)
 - b. Investigate reported incidents according to severity
 - c. Coordinate with vendor on issue resolution
 - d. Document incidents and corrective actions
 - e. Track trends and patterns in AI issues
 - f. Report material incidents to governance committee
- 7. **Periodic Reassessment & Revalidation**
 - a. Conduct comprehensive reassessment at defined intervals (e.g., annually or at contract renewal)
 - b. Revalidate AI system performance and safety
 - c. Review and update risk classification if use cases have changed
 - d. Reassess vendor security posture and compliance
 - e. Update GRC questionnaire to address new risks or requirements
 - f. Document continued appropriateness of AI solution

AI-Specific Considerations:

- Updates to AI systems occur more frequently than traditional software and may have greater impact
- Organizations should automate monitoring where possible to scale with frequency of AI changes
- Model drift is a unique AI risk that requires continuous attention
- Performance degradation may be gradual and difficult to detect without systematic monitoring

- Vendors may reset security configurations during updates—vigilance is essential
- Consider whether to implement AI-specific monitoring tools or platforms

Deliverables:

- Performance Monitoring Dashboards
- Monthly/Quarterly Performance Reports
- Vendor Performance Scorecards
- Audit Reports and Findings
- Update Testing Documentation
- Incident Reports and Corrective Action Plans
- Annual Reassessment Reports

Key Takeaway: AI systems require more intensive ongoing monitoring than traditional software due to model drift, frequent updates, and evolving risks. Organizations must establish sustainable monitoring processes with appropriate automation and clear escalation paths for performance degradation or security concerns.

Phase 5: Incident Response & Recovery

Despite rigorous due diligence, implementation testing, and ongoing monitoring, AI incidents should be anticipated despite rigorous controls. Whether triggered by security breaches, model failures, bias events, adversarial attacks, or vendor issues, AI incidents present unique challenges that traditional IT incident response procedures cannot address. AI failures can be subtle and difficult to detect, may manifest as gradual degradation rather than catastrophic failure, and often require specialized forensic analysis. Unlike conventional software where rollback is straightforward, AI incidents may involve corrupted training data, accumulated model drift, or emergent behaviors that cannot be easily reversed. The complexity of AI systems and heavy dependencies on vendor expertise mean effective incident response requires carefully orchestrated coordination between HCOs and AI vendors.

Phase 5 establishes frameworks, procedures, and coordination mechanisms to detect, respond to, and recover from AI-specific incidents while protecting patient safety, protecting organizational data, and maintaining regulatory compliance. Organizations must prepare for diverse AI incident scenarios, from security breaches affecting training data to bias events producing discriminatory outputs to model hallucinations generating erroneous recommendations, each requiring tailored response strategies. By establishing clear incident response plans, vendor coordination protocols, and recovery procedures before incidents occur, HCOs can significantly reduce response times, limit damage, and restore operations more safely than organizations improvising during crisis.

Key Activities:

1. **Incident Response Planning & Preparation**
 - a. **AI-Specific Incident Response Plan**
 - i. Develop or enhance incident response plans to address AI-specific scenarios:
 1. Security breaches affecting AI systems or training data

2. Model performance failures or unexpected degradation
3. Bias events or discriminatory AI outputs
4. Data poisoning or adversarial attacks
5. Model hallucinations or erroneous outputs in production
6. Privacy breaches involving AI-processed PHI
- ii. Define severity levels and escalation criteria for AI incidents
- iii. Identify AI incident response team members and roles
- iv. Establish communication protocols and notification chains
- v. Define SLA response notification in alignment with regulations

b. Vendor Coordination Protocols

- i. Contractually require vendor participation in incident response
- ii. Establish joint incident response procedures and communication channels
- iii. Define vendor notification requirements and timeframes (e.g., 2-24 hours depending on severity)
- iv. Agree on information sharing during incident investigation
- v. Clarify vendor responsibilities for forensics, remediation, and recovery support
- vi. Conduct joint tabletop exercises to test incident response coordination

2. AI Incident Detection & Classification

a. Detection Mechanisms

- i. Monitor for AI performance anomalies and degradation
- ii. Detect unusual AI system behavior or outputs
- iii. Track security alerts related to AI infrastructure
- iv. Monitor user reports of AI errors or concerns
- v. Review audit logs for suspicious AI system access
- vi. Assess alerts from AI-specific security tools

b. Incident Classification

- i. Determine incident type (security breach, model failure, bias event, data issue)
- ii. Assess severity based on:
 1. Patient safety impact (actual or potential harm)
 2. PHI/PII exposure risk
 3. Operational disruption/Business Impact Analysis
 4. Regulatory implications
 5. Reputational impact
- iii. Assign incident priority and escalation level
- iv. Determine if regulatory notification is required ([21 CFR Part 803](#) for med device-related adverse reporting, HHS OCR for HIPAA breach, relevant state and federal reporting requirements, etc.)

3. Containment & Mitigation

a. Immediate Actions

- i. Isolate affected AI systems if necessary to prevent further harm

- ii. Implement safe mode or fallback procedures for critical functions
 - iii. Suspend AI system if risk of harm outweighs benefit of continued operation
 - iv. Notify vendor immediately per contractual requirements
 - v. Preserve evidence for forensic investigation
 - vi. Implement workarounds to maintain essential operations
- b. Coordinated Response Activities**
- i. Engage vendor in joint incident response
 - ii. Conduct forensic investigation to determine root cause
 - iii. Assess scope and impact of incident
 - iv. Identify affected patients, users, or data
 - v. Coordinate on containment strategies with vendor
 - vi. Implement temporary mitigations while permanent fixes are developed
- 4. Recovery & Restoration**
- a. System Recovery Procedures**
- i. Develop recovery plan in coordination with vendor
 - ii. Define Recover Time Objective (RTO) and Recovery Point Objective (RPO)
 - iii. Restore AI systems from known-good backups or model versions
 - iv. Conduct rollback to previous model version if update caused incident
 - v. Implement vendor-provided patches or fixes
 - vi. Validate security and privacy controls after restoration
 - vii. Verify AI model performance and accuracy post-recovery
- b. Model Rollback & Revalidation**
- i. Maintain documented rollback procedures for AI models
 - ii. Revert to previously validated model version when necessary
 - iii. Conduct abbreviated revalidation of rolled-back model in current environment
 - iv. Document justification for rollback decision
 - v. Plan for long-term resolution (may require different vendor or solution)
- c. Data Recovery & Integrity Validation**
- i. Assess any data corruption or loss during incident
 - ii. Restore data from secure backups if necessary
 - iii. Validate data integrity after restoration
 - iv. Verify training data and model parameters remain uncorrupted
 - v. Ensure audit trails and logs are preserved for investigation
- 5. Post-Incident Activities**
- a. Documentation & Reporting**
- i. Complete incident report documenting timeline, actions taken, and outcomes
 - ii. Conduct root cause analysis with vendor participation
 - iii. Report to regulatory authorities if required (FDA, OCR, state agencies)
 - iv. Notify affected individuals according to regulatory requirements and organizational policy
 - v. Document lessons learned and improvement opportunities

- vi. Update incident response procedures based on experience
 - b. **Corrective & Preventive Actions (CAPA)**
 - i. Identify systemic issues revealed by incident
 - ii. Develop corrective actions to address root causes
 - iii. Implement preventive measures to reduce likelihood of recurrence
 - iv. Update controls, configurations, or procedures as needed
 - v. Require vendor to implement corrections in their environment
 - vi. Validate effectiveness of corrective actions
 - vii. Track Corrective and Preventive Action Plans (CAPA) completion and close-out
 - c. **Return to Normal Operations**
 - i. Define criteria for "all-clear" and return to full production
 - ii. Obtain third-party attestation of system security if warranted by incident severity
 - iii. Conduct final validation before resuming normal operations
 - iv. Determine appropriate timeline for re-establishing vendor connectivity after breach
 - v. Implement enhanced monitoring during recovery period
 - vi. Communicate resolution to stakeholders and users
6. **Continuous Improvement & Preparedness**
- a. Conduct regular tabletop exercises simulating AI-specific incidents including vendors
 - b. Update incident response playbooks based on evolving AI threats
 - c. Train incident response team on AI-specific scenarios
 - d. Review vendor incident response participation and capability
 - e. Assess adequacy of incident response tools and resources for AI systems
 - f. Incorporate lessons learned into policy and procedure updates

AI-Specific Considerations:

- AI model failures may be subtle and difficult to detect compared to traditional system failures
- Rollback of AI models is more complex than traditional software rollback
- Determining "normal operation" for AI systems requires revalidation of model performance
- Bias events and other model failures may require different response than security incidents (e.g. regulatory reporting, communication strategy)
- Vendor dependencies are higher for AI incidents due to complexity of model and training data
- Recovery time objectives (RTO) may be longer for AI systems requiring revalidation

Deliverables:

- AI-Specific Incident Response Plan and Playbooks
- Vendor Incident Response Coordination Agreements
- Incident Reports and Documentation
- Tabletop exercise standard operating procedures
- Root Cause Analysis Reports

- Corrective and Preventive Action (CAPA) Plans
- Regulatory Notifications (as required)
- Lessons Learned Documentation

Key Takeaway: Ensure that third-party AI providers are actively engaged in the HCO's cybersecurity incident response, disaster recovery, and QA/VV processes, particularly for AI-enabled services or models that influence clinical or operational outcomes.

Key Recommendations:

1. Integrated Response and Recovery Coordination

- Establish contractual obligations for third-party AI vendors to participate in coordinated incident response and recovery activities
- Vendors must notify covered entities of security incidents, degradations, or AI model anomalies affecting integrity, performance, or safety within HDO's defined per incident response Service Level Agreement (SLA)
- Include third-party vendors in tabletop exercises, model rollback scenarios, and cyberattack simulations involving AI-dependent processes
- Clearly define requirements for return to "normal" business after an event

2. Resilience and Rollback of AI Models

- Require vendors to provide documented rollback and recovery procedures for AI models, including reversion to a known-good baseline or safe mode operation in the event of model corruption, adversarial attacks, or hallucination propagation
- Recovery plans should define how to restore model output accuracy, data lineage integrity, and system interoperability
- AI vendors should participate in the recovery efforts. This requirement should be defined in the contract as well as in the recovery plan

3. Policy Integration

- Incorporate AI-specific validation requirements and incident response expectations into the organization's broader cybersecurity and AI governance policies
- Procedures should clarify how third-party vendors are evaluated, monitored, and held accountable during initial onboarding, ongoing reassessment and contract termination

4. Assurance of Continuous Improvement

- Require vendors to conduct periodic reassessments and revalidation of AI models following:
 - Model updates, retraining events, or architecture changes
 - Emerging threat intelligence relevant to AI exploitation
 - Regulatory or clinical performance issues

Phase 6: End-of-Life & Transition Management

All AI systems eventually reach end-of-life through planned vendor discontinuation, organizational decisions, technological obsolescence, or unplanned events like vendor failure or third-party model deprecation. Unlike traditional software with predictable support cycles, AI systems face unique EOL challenges requiring specialized planning. AI models may depend on external services deprecated without the third-party vendor control, as when fourth-party vendors sunset specific models forcing urgent migrations. Organizational data may be embedded in model weights requiring specialized destruction beyond standard data deletion. Simply replacing one AI model with another may not maintain equivalent performance without comprehensive revalidation, and replacement systems may require substantially different workflows disrupting clinical operations.

Proactive planning prevents patient care disruption, data loss, and compliance gaps. Phase 6 establishes frameworks to manage planned and unplanned AI discontinuation while protecting data integrity, maintaining continuity of care, and meeting regulatory obligations. Effective EOL management begins during initial contracting by establishing vendor notification requirements, data extraction rights, and secure destruction procedures. Organizations must prepare to rapidly decide whether to replace discontinued systems, conduct expedited vendor evaluations, extract and migrate operational data, validate replacement system equivalence, retrain users, and ensure secure data destruction from all vendor environments including production systems, backups, and AI models.

Key Activities:

1. End-of-Life Planning & Notification

a. Proactive EOL Planning

- i. Negotiate EOL notification requirements during initial contracting (12-18 months advance notice minimum)
- ii. Incorporate AI system lifecycle into organizational technology roadmap
- iii. Monitor vendor announcements for EOL indicators
- iv. Track AI system age and vendor support commitment duration
- v. Plan for EOL of underlying models not controlled by vendor (e.g., OpenAI model deprecations)

b. EOL Notification & Impact Assessment

- i. Receive formal EOL notification from vendor including:
 1. Last date of full support
 2. Last date of security patching
 3. Recommended migration path or replacement options
 4. Data export and transition support availability
- ii. Assess operational, clinical, cybersecurity, safety and regulatory impact
- iii. Determine urgency of transition based on criticality of AI function
- iv. Evaluate risk of continuing to use unsupported AI system

2. Transition Decision & Strategy

a. Replace vs. Discontinue Decision

- i. Evaluate whether AI solution remains necessary for operations
- ii. Assess alternative solutions (different vendor, internally developed, non-AI approach)

- iii. Consider cost-benefit of migration versus accepting risks of continued use
 - iv. Determine if gaps in security or functionality require immediate replacement
 - v. Document decision rationale and obtain governance approval
 - b. **Replacement Solution Selection** (if applicable)
 - i. Conduct expedited vendor evaluation for replacement solution
 - ii. Prioritize solutions with similar functionality to minimize workflow disruption
 - iii. Ensure new vendor contracts meet or exceed previous standards
 - iv. Negotiate transition support from outgoing vendor
 - v. Plan for migration timeline that minimizes operational disruption
3. **Data Management & Migration**
- a. **Data Inventory & Classification**
 - i. Identify all data associated with AI system:
 - 1. Training datasets and fine-tuning data
 - 2. Operational data and transaction logs
 - 3. Audit trails and override histories
 - 4. Clinical records and decision documentation
 - 5. User interaction data and prompts
 - ii. Classify data by retention requirements and regulatory obligations
 - iii. Determine what data must be migrated versus archived versus destroyed
 - b. **Data Extraction & Export**
 - i. Require vendor to provide data in interoperable, non-proprietary formats
 - ii. Extract all organizational data from vendor systems
 - iii. Obtain vendor tooling and support for data export
 - iv. Validate completeness and accuracy of exported data
 - v. Document data export process and results
 - c. **Data Migration** (if replacing system)
 - i. Map data elements from legacy to replacement system
 - ii. Transform data formats as needed for compatibility
 - iii. Import data into replacement system with validation
 - iv. Conduct data reconciliation to ensure no loss or corruption
 - v. Validate that migrated data remains accessible and usable
 - d. **Data Archival & Retention**
 - i. Archive data according to organizational retention policies and legal requirements
 - ii. Ensure archived data remains accessible for regulatory audits or legal discovery
 - iii. Implement long-term storage strategy (on-premises, cloud, hybrid)
 - iv. Document archival location and access procedures
4. **Secure Decommissioning**
- a. **System Shutdown Procedures**
 - i. Develop decommissioning plan in coordination with IT and vendor
 - ii. Notify users of shutdown timeline and transition plan

- iii. Implement phased shutdown if appropriate (pilot reversal approach/staged decommissioning)
 - iv. Disable user access systematically
 - v. Disconnect AI system from production networks and integrations
 - vi. Document final system state and configuration
- b. Data Destruction & Vendor Cleanup**
- i. Require vendor to securely destroy all organizational data per [NIST 800-88](#) or equivalent
 - ii. Obtain certification of data destruction from vendor
 - iii. Verify data removal from vendor systems including:
 - 1. Production databases
 - 2. Backup systems
 - 3. Training datasets
 - 4. Logs and caches
 - iv. Confirm no residual PHI/PII remains in vendor environment
 - v. Validate destruction of any AI models containing embedded organizational data
- c. Infrastructure Cleanup**
- i. Sanitize or destroy hardware used for AI system (if on-premises)
 - ii. Remove AI system software and configurations from servers
 - iii. Clean up cloud resources and terminate vendor access
 - iv. Remove access credentials, API keys, and integration tokens
 - v. Update firewall rules and network configurations
- 5. Replacement System Onboarding** (if applicable)
- a. Accelerated Implementation**
- i. Follow Phase 3 implementation procedures with appropriate adaptations for urgency
 - ii. Prioritize critical workflows for initial deployment
 - iii. Leverage lessons learned from previous implementation
 - iv. Conduct equivalence testing comparing legacy and replacement AI outputs
 - v. Validate replacement system performance meets or exceeds legacy system
- b. User Transition & Training**
- i. Provide training on replacement system with emphasis on workflow changes
 - ii. Support users during transition period with additional help desk resources
 - iii. Collect feedback and address usability issues promptly
 - iv. Monitor user adoption and satisfaction
 - v. Document known differences from legacy system
- 6. Regulatory & Compliance Considerations**
- a. Documentation & Audit Trail**
- i. Document complete End of Life (EOL) process for regulatory compliance
 - ii. Maintain records demonstrating how EOL risks were managed
 - iii. Update risk management files and security documentation
 - iv. Preserve final configuration and performance records
 - v. Document validation of replacement system (if applicable)

Key Takeaway: End-of-life planning for AI systems must begin during initial contracting and address the unique challenge of model dependencies beyond vendor control. Organizations need clear procedures for data extraction, secure destruction, and validation of replacement systems to ensure continuity of care and regulatory compliance.

Conclusion and How to Use This Guide

Healthcare's rapid AI adoption demands a fundamental shift in managing third-party technology risk. Traditional vendor risk practices fail to address AI systems that learn, drift, and rely on opaque supply chains. This Guide provides a structured, lifecycle-based framework for healthcare organizations to mitigate risks, ensuring AI delivers value without compromising patient safety, data privacy, or operational continuity.

The Process: A Seven-Phase Lifecycle

The continuous AI third-party risk management lifecycle is organized into seven interdependent phases:

- **Phase 0 — AI Use Case Justification & Strategic Assessment:** Define the problem, confirm AI suitability, classify the use case by safety impact (Low, Medium, High, or Critical), and establish accountability to ensure strategic alignment and an understood risk profile.
- **Phase 1 — Due Diligence & Vendor Evaluation:** Extend standard vendor assessment (financial, cybersecurity) with AI-specific Governance, Risk, and Compliance (GRC) assessments. Key areas include data lineage/bias, model transparency/explainability, security controls, supply chain dependencies, and ethical practices. Assessment rigor must scale based on the risk classification from Phase 0, and vendors must provide Quality Assurance, Verification, and Validation (QA/VV) documentation.
- **Phase 2 — Contract Negotiation & Legal Protections:** Standard agreements are insufficient for AI. This phase establishes shared responsibility through AI-specific contract clauses covering data ownership, training restrictions, change approval, performance obligations, incident response, and end-of-life support. Enhanced Business Associate Agreement (BAA) provisions address the unique risks of processing Protected Health Information (PHI) with AI.
- **Phase 3 — Implementation, Integration & Training:** Covers the highest-risk transition to production. Conduct AI-specific threat modeling (e.g., OWASP Top 10 for LLMs), rigorous sandbox/clinical validation, and security testing. Ensure role-specific training is completed for all staff and utilize phased production rollout with enhanced monitoring.
- **Phase 4 — Ongoing Monitoring & Performance Management:** This is the longest phase, requiring continuous attention to model drift, bias, performance degradation, and security integrity, especially after vendor updates. It encompasses security/compliance auditing, patch validation, and periodic reassessment. Quality assurance and re-validation are essential after every system change.
- **Phase 5 — Incident Response & Recovery:** Prepares for inevitable AI incidents (model degradation, breaches, bias events). Establish detection, classification, containment, vendor coordination protocols, model rollback/revalidation, and post-incident corrective actions. AI incidents are uniquely gradual and difficult to detect, requiring specialized forensics and vendor expertise.

- **Phase 6 — End-of-Life & Transition Management:** Addresses planned and unplanned AI discontinuation. Proactive planning ensures continuity of care, secure data destruction, and regulatory compliance, managing unique challenges like embedded organizational data and replacement system revalidation.

The Guide is a practical and scalable resource to be applied in the following ways:

- **Start with Governance:** Establish an AI governance body to formalize accountability, approval processes, and risk criteria, using the sample Governance Policy ([Appendix B](#)).
- **Assess Current State:** Inventory existing AI-enabled systems (including embedded features) using discovery techniques ([Appendix C](#)) and retroactively evaluate them against the Phase 1 framework.
- **Scale Controls:** Implement controls proportionate to organizational size: Small organizations use baseline requirements; mid-sized implement enhanced controls; and large organizations deploy advanced risk stratification and extensive validation.
- **Integrate AI Activities:** Enhance existing vendor risk management programs, contract templates, and incident response plans with AI-specific activities, mapping responsibilities using the RACI matrix ([Appendix D](#)).
- **Leverage Tools and Vendors:** Utilize the ready-to-adopt templates and checklists in the appendices, and share the Guide with AI vendors to enforce expectations for transparency and collaborative lifecycle management. The appendices provide templates, tools, and reference materials:
 - [Appendix A](#): Phase 0 AI Use Case Justification Template and Risk Level Definitions.
 - [Appendix B](#): Governance Policy for AI Third-Party Risk.
 - [Appendix C](#): Inventory Management (Discovery techniques and catalog of AI systems).
 - [Appendix D](#): RACI Matrix (Roles and responsibilities across the seven phases).
 - [Appendix E](#): Sample Commercial Contract Language (Templates for seventeen critical areas).
 - [Appendix F](#): Sample BAA Contract Language (AI-specific PHI provisions).
 - [Appendix G](#): AI Vendor Assessment Questions for Procurement and GRC (69 questions for risk scrutiny).
 - [Appendix H](#): Training Completion Checklist and Curriculum (Twelve-module training).
 - [Appendix I](#): Quality Assurance/Verification/Validation with AI Third-Party Providers (QA/VV framework for Phases 1 and 4).
 - [Appendix J](#): References (Foundational guidance from HSCC, FDA, NIST, etc.).

Appendix A. Phase 0 AI Use Case Justification Template and Risk Level Definitions

1. Purpose and When to Use

This template helps healthcare organizations justify, document and approve AI use cases before procurement, pilot, or activation. It captures intended use, expected benefits, risks, and required controls to support consistent risk-based governance and auditability.

2. Recommendations for Using This Template (Practical Guidance)

Complete the template before: signing a contract, enabling an embedded AI feature, starting a pilot, connecting PHI/PII, or deploying a model into production.

a. Always document these fields (all risk levels)

- i. Problem the solution intends to solve
- ii. Intended use and prohibited uses (so the tool isn't used "off-label")
- iii. Human oversight (who reviews what aspect or output and when review is conducted)
- iv. Data types (PHI/PII vs de-identified data), data flows, integrations and locations for storing the data, and whether any data is used for training purposes
- v. Risk Classification/Categorization and 1–2 sentence justification
- vi. Stakeholders
- vii. Decisions, conditions/controls, sign-off

b. Don't confuse internal risk tier with regulatory status

- i. A tool can be risk assessed internally even if it isn't marketed as a medical device. Use this tiering to drive controls; regulatory classification can be assessed separately. Just because a product is approved via a regulatory body does not ensure that the risk it imposes to the organization is mitigated.

c. Use "highest applicable risk"

- i. When determining risk classification, if any part of the workflow is High/Critical (e.g., a triage recommendation that changes patient routing), classify the *use case* at the higher tier.

3. Risk Classification Definitions (Impact-Based)

a. Low Impact (Minimal or no safety/financial impact)

- i. AI failure would not affect patient care, safety, or significant financial outcomes
- ii. Errors are easily detected and corrected by users
- iii. Examples: word prediction tools, email autocomplete, meeting scheduling assistants

b. Medium Impact (Moderate safety/financial impact with strong human oversight)

- i. AI supports decision-making but humans retain full control and review all outputs
- ii. Errors could affect individual patients or transactions but are caught through validation; human override and correction mechanisms are required
- iii. Examples: clinical decision support where the clinician makes the final decision, coding assistance tools, supply chain forecasting

c. High Impact (Significant safety/financial impact with limited human intervention)

- i. AI substantially influences important clinical or financial decisions
- ii. Errors could affect multiple patients, department/service-line operations, and/or significant financial outcomes
- iii. Examples: automated medication dosing recommendations, AI-driven triage systems, fraud detection with automatic claim denials, predictive models influencing treatment planning

d. Critical Impact (Life-threatening safety risk or enterprise-critical operational/financial impact)

- i. AI may result in autonomous or near-autonomous decisions affecting patient life/safety or organization-wide operations
- b. Failure could result in serious harm, death, or catastrophic financial/operational consequences
- c. Examples: autonomous diagnostic AI without physician review, AI-controlled medical devices (ventilators, insulin pumps), fully automated clinical pathways, enterprise-wide operational AI (staffing/capacity), AI controlling critical infrastructure

4. Quick Classification (Fast Triage)

Use these questions to quickly evaluate risk tier:

1. How much financial risk would the organization bear should the tool fail?
2. Could incorrect output from the tool delay care, misroute patients, or change clinical decisions/misdiagnosis?
3. Does the tool generate clinical content (notes, instructions) that could be copied into the medical record?
4. Does the tool influence access to or priority for services (triage, scheduling priority, eligibility, billing)?
5. What level of validation, human override, or review is part of the process?
6. Does the tool process and/or store PHI/PII externally or involve third parties/subprocessors?

7. Is the model updated frequently or is the tool opaque (“black box”) with limited explainability?

Scoring Guidance

1. If Finance response is moderate, results are fully overseen/validated, tool does not externally store PHI/PII and model is rarely updated → likely Medium+.
2. If Finance response is moderate to high, results have “limited human intervention”, tool stores/processes PHI/PII → likely High.
3. If Finance response is high, results are “autonomous diagnosis/treatment without physician/clinical review”, tool stores/processes PHI/PII and involves third parties or model is updated frequently or unexplainable → Critical.

5. Alternative Scoring Method (0–3 each)

Score each dimension 0 (none) to 3 (high). Total the score for all dimensions to obtain the risk classification. If scoring and the definitions conflict, use the higher tier. If the tool meets any of the critical triggers listed below, the categorization should immediately be listed as critical.

Dimensions

- Patient safety impact if erroneous response is provided by tool
- Influence on clinical decision-making
- Degree of autonomy (human oversight vs automated)
- Sensitivity of data and exposure risk
- Equity and bias risk (differential impact across groups)

Recommended mapping to Risk Classification

0–4 = Low; 5–8 = Medium; 9–12 = High; 13–15 = Critical

Critical triggers: autonomous clinical actions without human approval; direct diagnosis /treatment decisions; failure could cause severe harm/death.

6. Purpose and When to Use

This template helps healthcare organizations justify, document and approve AI use cases before procurement, pilot, or activation. It captures intended use, expected benefits, risks, and required controls to support consistent risk-based governance and auditability.

7. Recommendations for Using This Template (Practical Guidance)

Complete the template before: signing a contract, enabling an embedded AI feature, starting a pilot, connecting PHI/PII, or deploying a model into production.

a. Always document these fields (all risk levels)

- i. Problem the solution intends to solve
- ii. Intended use and prohibited uses (so the tool isn't used "off-label")
- iii. Human oversight (who reviews what aspect or output and when review is conducted)
- iv. Data types (PHI/PII vs de-identified data), data flows, integrations and locations for storing the data, and whether any data is used for training purposes
- v. Risk Classification/Categorization and 1–2 sentence justification
- vi. Stakeholders
- vii. Decisions, conditions/controls, sign-off

b. Don't confuse internal risk tier with regulatory status

- i. A tool can be risk assessed internally even if it isn't marketed as a medical device. Use this tiering to drive controls; regulatory classification can be assessed separately. Just because a product is approved via a regulatory body does not ensure that the risk it imposes to the organization is mitigated.

c. Use "highest applicable risk"

- i. When determining risk classification, if any part of the workflow is High/Critical (e.g., a triage recommendation that changes patient routing), classify the *use case* at the higher tier.

8. Risk Classification Definitions (Impact-Based)

a. Low Impact (Minimal or no safety/financial impact)

- i. AI failure would not affect patient care, safety, or significant financial outcomes

- ii. Errors are easily detected and corrected by users
 - iii. Examples: word prediction tools, email autocomplete, meeting scheduling assistants
- b. Medium Impact (Moderate safety/financial impact with strong human oversight)**
- i. AI supports decision-making but humans retain full control and review all outputs
 - ii. Errors could affect individual patients or transactions but are caught through validation; human override and correction mechanisms are required
 - iii. Examples: clinical decision support where the clinician makes the final decision, coding assistance tools, supply chain forecasting
- c. High Impact (Significant safety/financial impact with limited human intervention)**
- i. AI substantially influences important clinical or financial decisions
 - ii. Errors could affect multiple patients, department/service-line operations, and/or significant financial outcomes
 - iii. Examples: automated medication dosing recommendations, AI-driven triage systems, fraud detection with automatic claim denials, predictive models influencing treatment planning
- d. Critical Impact (Life-threatening safety risk or enterprise-critical operational/financial impact)**
- i. AI may result in autonomous or near-autonomous decisions affecting patient life/safety or organization-wide operations
 - ii. Failure could result in serious harm, death, or catastrophic financial/operational consequences
 - iii. Examples: autonomous diagnostic AI without physician review, AI-controlled medical devices (ventilators, insulin pumps), fully automated clinical pathways, enterprise-wide operational AI (staffing/capacity), AI controlling critical infrastructure

9. Quick Classification (Fast Triage)

Use these questions to quickly evaluate risk tier:

- a. How much financial risk would the organization bear should the tool fail?
- b. Could incorrect output from the tool delay care, misroute patients, or change clinical decisions/misdiagnosis?
- c. Does the tool generate clinical content (notes, instructions) that could be copied into the medical record?
- d. Does the tool influence access to or priority for services (triage, scheduling priority, eligibility, billing)?
- e. What level of validation, human override, or review is part of the process?
- f. Does the tool process and/or store PHI/PII externally or involve third parties/subprocessors?

- g. Is the model updated frequently or is the tool opaque (“black box”) with limited explainability?

Scoring Guidance

- a. If Finance response is moderate, results are fully overseen/validated, tool does not externally store PHI/PII and model is rarely updated → likely Medium+.
- b. If Finance response is moderate to high, results have “limited human intervention”, tool stores/processes PHI/PII → likely High.
- c. If Finance response is high, results are “autonomous diagnosis/treatment without physician/clinical review”, tool stores/processes PHI/PII and involves third parties or model is updated frequently or unexplainable → Critical.

10. Alternative Scoring Method (0–3 each)

Score each dimension 0 (none) to 3 (high). Total the score for all dimensions to obtain the risk classification. If scoring and the definitions conflict, use the higher tier. If the tool meets any of the critical triggers listed below, the categorization should immediately be listed as critical.

Dimensions

- Patient safety impact if erroneous response is provided by tool
- Influence on clinical decision-making
- Degree of autonomy (human oversight vs automated)
- Sensitivity of data and exposure risk
- Equity and bias risk (differential impact across groups)

Recommended mapping to Risk Classification

0–4 = Low; 5–8 = Medium; 9–12 = High; 13–15 = Critical

Critical triggers: autonomous clinical actions without human approval; direct diagnosis /treatment decisions; failure could cause severe harm/death.

Phase 0 AI Use Case Justification – Checklist Form

Directions: Users please complete sections 1 – 5. Section 6 should be completed by the AI governance team upon review.

1. Requestor:

Date: ____ / ____ / ____

Requestor (name, role, org): _____

Reviewer / Owner (AI governance): _____

Business unit: _____ Site(s): _____

2. Use Case Summary

Use case title:

Business Problem:

AI capability type (check one):

Predictive model NLP / LLM GenAI content Computer vision

Rules + ML Other: _____

Proposed solution (if known):

Vendor: _____ Product: _____

Model name/version (if known): _____

Deployment (check one):

SaaS On-prem Hybrid Embedded in existing tool API / integration

Workflow description and context (where used, by whom, how often):

Intended use (what it will do):

Out of scope / prohibited uses (what it will NOT do):

Required Integrations (what other systems both internal and external must it be connected):

3. Expected Benefits (measurable where possible)

Primary benefit(s):

Target users:

Patients Clinicians Admin Revenue cycle Other: _____

Expected impact (metrics):

Time saved Cost reduction Throughput Quality Safety Other: _____

Baseline metric: _____ Target metric: _____ Timeframe: _____

4. Data and Access

Data inputs (check all that apply):

PHI PII De-identified Claims Imaging Notes Labs
 Scheduling Other: _____

Data sources (check all that apply):

EHR HRIS CRM Data warehouse Patient portal Other: _____

Data outputs (check all that apply):

Recommendation Summary Classification Generated text Alert Other: _____

Data sharing (check all that apply):

- No data leaves organization
- Data sent to vendor for processing
- Data used for training/fine-tuning or model improvement (requires explicit approval)

5. Initial Risk Classification

Overall risk level (select one): Low Medium High Critical

(See definitions in Section 3)

Brief justification (1–2 sentences):

6. Required Controls and Approvals (based on risk) – To be completed by AI Governance Team

Minimum controls (all levels):

- Intended use and prohibited uses documented
- Vendor identifies AI components and third parties
- Security and privacy review
- User training and acceptable use communicated
- Human oversight plan defined (who reviews, what they verify, and when)
- Downtime plan

Additional controls (Medium):

- Validation plan for the use case
- Monitoring plan (quality/drift, oversight, and incident triggers)
- Contract language for AI updates + incident notification

Additional controls (High):

- Formal evaluation results for intended use
- Bias assessment + fairness monitoring plan
- Fail-safe/fallback procedures tested
- Expanded vendor due diligence (full questionnaire)

Additional controls (Critical):

- Clinical safety case documented
- Independent validation / third-party assessment considered
- Regulatory pathway confirmed (as applicable)
- Ongoing monitoring with alerting + escalation playbooks

Approvals required (check all that apply):

- Security Privacy Compliance Legal Clinical safety
- IT/Architecture Procurement Data governance Other: _____

7. Decision

Decision: Approve Approve with conditions Pilot only Reject Defer pending info

8. Next Steps and Ownership

Next Steps

Owners:

- | | |
|--|-------|
| <input type="checkbox"/> Vendor Evaluation/Third Party Risk Assessment | _____ |
| <input type="checkbox"/> Security Risk Assessment/GRC Risk Assessment | _____ |
| <input type="checkbox"/> Reference Checks | _____ |
| <input type="checkbox"/> BAA | _____ |
| <input type="checkbox"/> Contract Requirements | _____ |
| <input type="checkbox"/> Service Level Agreement | _____ |
| <input type="checkbox"/> Data Processing Agreement | _____ |
| <input type="checkbox"/> Implementation Plan | _____ |
| <input type="checkbox"/> User training Plan | _____ |
| <input type="checkbox"/> Human oversight plan | _____ |
| <input type="checkbox"/> Vendor/Tool Monitoring Plan | _____ |
| <input type="checkbox"/> Validation/ Patching/Update Plan | _____ |
| <input type="checkbox"/> Downtime plan | _____ |
| <input type="checkbox"/> System Documentation and Operational Procedures | _____ |

[] Other _____

9. Sign-off

Requestor: _____ Date: _____

AI governance owner: _____ Date: _____

Clinical owner (if applicable): _____ Date: _____

Appendix B: Governance Policy for AI Third-Party Risk

The following policy is developed as a baseline for organizations of all sizes and complexities to adopt and implement as a foundation to AI Governance. Most organizations, due to their size, structure and risk tolerance, will have various ways to implement this policy and so instead of providing overly prescriptive methodologies, the authors instead have called out several areas where examples of ways to implement these generalized policy statements may be illustrative.

		Policy Title:	Artificial Intelligence
Effective Date:		Policy Number:	
Review Date:		Section:	
Revised Date:		Oversight Level:	
Administrative Responsibility:	VP Compliance and VP Information Security		

1. Purpose

- 1.1. To provide guidelines and security controls to enforce responsible use of Artificial Intelligence (“AI”) in all projects, products and applications.
- 1.2. To foster public trust, protect patient privacy and security, support business outcomes, ensure ethical and legal compliance for transparent, accountable and responsible implementation of AI technology.

2. Scope

- 2.1. “ORGANIZATION NAME” its subsidiaries, any other entity or organization in which “ORGANIZATON” or an “ORGANIZATION” subsidiary owns a direct or indirect equity interest of 50% or more, provided that organization has agreed to adopt “ORGNIZATION” policies.

3. Definitions

- 3.1. **AI** means a machine-based system that can, for a given set of human-defined objectives, make predictions and recommendations influencing real or virtual environments. AI systems use machine and human-based inputs such as patterns and structures learned from existing data, deep learning, neural networks, and Machine Learning techniques to, among other actions, (1) perceive real and virtual environments; (2) abstract such perceptions into models through analysis in an automated manner; (3) create new, original content, such as images, text, or music; (4) produce

content autonomously that closely resembles human-created output; (5) produce natural language texts based on a given input, such as a prompt, a keyword, or a query; and/or (6) use model inference to formulate options for information or action.

- 3.2. **Generative AI (GAI)** means a technology that can create new content in response to prompts, including but not limited to text, speech, and images.
- 3.3. **Algorithmic AI (AAI)** means a technology that analyzes data with machine-learning algorithms and can make decisions or predictions based on the data.
- 3.4. **Machine Learning** means a set of techniques that can be used to train AI algorithms to improve performance at a task based on data.
- 3.5. **Natural Language Processing** is a subfield of AI that enables computers to understand, process, interpret and generate human language. Natural Language Processing systems can perform tasks such as text classification, sentiment analysis, and translation, using techniques from computational linguistics to process and analyze natural language data.

4. Policy

4.1. Accountability and Governance/Oversight.

4.1.1. A governing body shall exist to oversee this policy. [*Call Out: No prescribed body or group of individuals is identified as the size, scope and make-up of this body is dependent on the organization. Additionally, the mechanism this group uses to categorize and prioritize AI tools will depend on the organization's tolerance for risk. Creating a categorization system such as high/medium/low, or one based on functionality, sensitivity and patient impact are roles of the governing body and should align to the organization's risk tolerance. Finally, there is no defined criteria for approval of AI technology as each organization's minimum requirements and approval process will be different. Evaluation criteria should be defined by the governing body and may include items such as compliance with HIPAA or NIST or other items seen in this policy.*].

4.1.2. A mechanism for human oversight shall exist for all AI-driven tools.

4.1.3. Before using AI outputs, users must engage in an independent review, including editing as needed to improve clarity; correcting grammatical and other errors; and identifying potential vulnerabilities and opportunities for improvement.

4.1.4. An inventory of all systems and applications that include AI will be maintained by the I.T. Applications Department. This inventory should include solutions with embedded features that can be activated at a later time as well as periodic discoveries to find shadow AI usage.

4.2. Business Associate Agreement.

4.2.1. A BAA, or Business Associate Agreement, is a contract required by HIPAA (Health Insurance Portability and Accountability Act) when a business associate handles protected health information (PHI) on behalf of a covered entity. In the context of AI, a BAA is crucial when AI tools are used, or could be used, with patient data to ensure HIPAA compliance and protect sensitive information.

4.3. Security.

4.3.1. All software/tools/applications with embedded AI should have the AI disabled until the AI is reviewed by this governance committee.

4.3.2. All AI vendors and tools/software purchased or procured that contain any AI tools will require review, verification, validation and approval from the Organization's Information Security Department. [*Call Out: Organizations may also require evaluations of AI vendors supply chains and disclosure of third-party services/models depending on their risk tolerance - see recommended GRC questions and purchasing terms/conditions.*]

4.3.3. The Organization's Information Security Department may restrict or limit the use of AI capabilities if they present risks that cannot be effectively mitigated.

4.3.3.1. This may include, but is not limited to, managing data on premises so that sensitive data does not leave the network.

4.3.4. The Organization's security protocols shall be in place to prevent unauthorized access and manipulation of AI systems.

4.3.5. AI vendors must provide a documented Quality Assurance and Verification/Validation (QA/VV) plan that is tailored to the AI system's intended use, safety classification, and criticality in the healthcare setting.

4.4. Data Handling.

4.4.1. Users shall only input approved data into approved AI systems, tools, products, and projects.

4.4.1.1. Workforce members must never load Organization's internal, confidential, sensitive or patient-related data, or any Organization-provided credentials, into a publicly available AI tool, (i.e., tool residing outside of Organization's secure IT environment), including meeting note taker applications or any home/personal use AI applications.

4.4.2. Sensitive or confidential information, including but not limited to PHI/PII shall not be input into AI tools or software containing AI tools without permission from the Organization's Health Care Corporate Privacy Officer and applicable confidentiality agreements and BAAs.

4.4.3. In the event sensitive or confidential information, including PHI/PII, is permitted to be inputted into AI tools or software, only the minimum necessary of the

information and/or data needed to perform the intended AI function should be used.
Whenever possible, de-identified information/data should be utilized instead of PHI/PII.

4.4.4. AI tools, systems, products, and projects will not be used on public-facing applications without organizational approval.

4.5. Ethical Considerations [*Call out: Organizations may want to formalize an Ethical Review Board or other escalation path for products the governance committee or GRC deems to be at high risk or does not meet the organization's minimum requirements.*]

4.5.1. Users/systems shall not use AI tools, products, or applications to create text, audio or visual content for purposes of committing fraud or to misrepresent an individual's identity.

4.5.2. All AI models shall be trained and tested for biases as reasonable and necessary by the providing vendor for the prevention of discrimination.

4.5.3. Users shall review output of AI tools, products, or applications to ensure it meets Organization's standards for ethics, equity and appropriateness.

4.5.4. Users shall not use AI tools, products or solutions to distort, impair, trick or otherwise interfere with the ability of an individual to make autonomous and informed choices or decisions.

4.5.5. Users shall not maliciously prompt or alter the AI tool, product or solution or otherwise engage in any unauthorized modifications that could compromise the integrity of the AI output.

4.6. Medical Record Documentation. In no event may a provider or clinician utilize any AI technology in the analysis or creation of patient medical record documentation without reviewing the record for accuracy and completeness and ultimately signing the record within the time period required.

4.7. Fairness and Transparency.

4.7.1. AI applications, tools, and products shall maintain transparency in their purpose, goals and objectives.

4.7.2. For AI outputs, metadata shall be kept for traceability and shall provide context and references from where the information was pulled, to allow for verification and validation where applicable.

4.7.3. Systems using AI should be designed to provide consistent and similar quality of service for all users.

4.8. Assessment, Monitoring and Auditing. Minimally, at contract renewal or at Organization-defined vendor performance evaluation timeframes, products and applications with AI embedded within will be reviewed to ensure ethical use and adherence with this policy and applicable laws and regulations.

4.8.1. For AI Solutions

4.8.1.1. Ongoing monitoring may include additional items based on the risk profile of the product, as deemed by the governance committee. *[Call Out: These may include KPIs and Performance Metrics that monitor the effectiveness, fairness or impact of the AI tool, and/or items provided by the vendor including AI Model drift, revalidation of the model, notification of procedures for vendor updates, and notification of breaches from vendors. The size and scope of reporting should be scalable for the size of your organization.]*

4.8.2. For Generative AI solutions

4.8.2.1. Safeguards will be added for detecting and flagging synthetic outputs. *[Call Out: For example, GAI will not be used in medical or employment decisions without human review. Organizations can add items that align with their risk tolerance.]*

4.9. Disclosure. When AI is used, especially in real-time communications, its use shall be disclosed.

4.10. Consent. If any AI tool/software/application is used in decision making, it shall be disclosed in the notice of privacy practices and/or consent forms.

4.11. Education/Training. All workforce members will receive education related to this policy. Personnel involved in AI projects, products, and applications, directly or indirectly, shall undergo specific training as part of AI project onboarding. *[Call out: Role-Specific training may or may not be identified in this policy, dependent on the type of AI implemented in your organization and risk tolerance. The specific details may belong in a separate policy within your organization dedicated to competencies, so please check with HR and compliance prior to adding details in this policy.]*

4.12. Incident Reporting.

4.12.1. AI-related concerns or incidents related to material deviation in the accuracy of outputs, biased or discriminatory outputs, or outputs that deviate from expectations shall be reported through the Organization's incident reporting system.

4.12.2. AI-related concerns or incidents related to suspected/actual inadvertent disclosure of proprietary company data, Protected Health Information (PHI), or Personal Health Information (PII) shall be immediately reported to the applicable Subsidiary(ies) Corporate Compliance Officer.

4.13. Sanctions. Violations of this policy may result in disciplinary action, up to and including termination of employment or contract, as well as potential financial and legal penalties. Suspected violations of this policy shall be immediately reported to the individual's immediate supervisor, HR and Compliance and Information Security as applicable.

5. Procedure

5.1. Prior to entering into any agreement/contract for the purchase or use of any technology intended to be used in connection with the performance of services for, or on behalf of the Organization, the business owner or contract requester will indicate in the contract management software/workflow whether the technology uses or includes AI.

5.1.1. If the technology uses or includes AI, the technology will not be purchased or implemented until it has undergone evaluation by, and received approval from, the Organization's Information Security Department, the Information Technology Department, and the Compliance Department.

5.1.1.1. A Security Risk Assessment will be required from a vendor, attesting that the AI output(s) does not result in biased or discriminatory outcomes/results. Once approved, the AI tool, solution or product will be subject to periodic audit during the lifespan of AI usage.

5.1.2. If technology requires FDA authorization or clearance as a medical device, the Clinical Engineering Department will:

5.1.2.1. Verify the AI technology has the required FDA authorization. *[Call Out: Depending on the risk tolerance of your organization, you may want to also ensure clinical validation of results/outputs prior to implementing in clinical practice]*

5.1.2.2. Perform rigorous testing procedures prior to implementing in clinical practice.

5.1.2.3. Ensure the vendor has provided applicable training and documentation (labeling) for healthcare providers on the appropriate use and limitations of AI-enabled medical devices is provided.

5.2. If the AI technology will be used in the delivery of patient care, billing/coding of items or services provided by the Organization, for employment related purposes, or for other uses involving PHI, PII, PCI, Business Sensitive information, the applicable department leader(s)/committee(s) must review and approve the AI technology for its intended use, and, if approved, implement appropriate protocols to govern its use in a manner that does not jeopardize patient safety or quality and complies with all applicable laws, regulations, and Organizational policies.

5.2.1. The applicable leader(s)/committee(s) will also consider, as part of its analysis, how the AI technology addresses and attempts to limit or avoid the potential for bias, including identifying specific employees, patients or other populations the AI is utilized for; risks associated with the potential bias; and how those risks will be mitigated. Consideration will include the size of the data set used to develop the AI within the tool/equipment, and the source of the data.

5.3. If after approval, the AI technology or its intended use within the Organization changes, the revised technology and/or intended use should be re-evaluated.

6. Modifications to Policy.

6.1. AI and the laws and regulations governing AI are rapidly evolving, and this policy may be amended from time to time to reflect the evolving landscape.

7. References

- 7.1. Privacy and Security Violation Corrective Action Policy
- 7.2. o6 Business Associate and Data Use Agreements Policy
- 7.3. HIPAA Uses & Disclosures of PHI Policy
- 7.4. Contract Management and Authorization Policy
- 7.5. Standards of Conduct Policy
- 7.6. Disclosure of Misconduct Policy
- 7.7. Acceptable Use of Technology Resources Policy
- 7.8. FDA’s Quality System Regulation (21 CFR 820) requirements for manufacturers, which applies to all AI enabled medical devices.
- 7.9. Link to FDA List of all AI/ML Approved Medical Devices: <https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-and-machine-learning-aiml-enabled-medical-devices>

8. Appendix : None

Approvals:

Insert Committee Name: Insert Committee Date

Insert Signer's Name	Date
Insert Signer's Title	

Appendix C: Inventory Management

There are two steps in developing an Inventory/Asset management system for tracking AI within an organization. First is finding all the existing systems, solutions, and devices that are already deployed that contain AI. Second is creating a process for capturing and documenting all newly approved systems that are implemented.

To find existing systems within an organization's environment, there are a few tools and techniques that can be used to identify possible AI technologies. First is a traffic analysis on the network. By monitoring network flows for API calls to known AI services such as Cloud AI SDKs such as aws-sdk, google-ai or azure you can identify possible use. Additionally, tracking bandwidth spikes could indicate large model downloads. If an organization wants to limit the use of cloud AI by end users, it could use proxy content categories for AI and deny those by default.

Another technique to identify existing AI in the environment is to use DNS monitoring. By querying AI service domains, one might identify organizational uses of AI. One example is looking for queries to machine learning frameworks such as tensorflow. Similar to DNS monitoring, process monitoring can also be used to look for resource intensive processes utilizing high CPU/GPU resources.

If an organization has a software inventory, that can be used to identify software packages with known AI built in such as Python packages (tensorflow) or Cloud AI SDKs (AWS-sdk, Google-AI, Azure). For organizations with medical device inventories, the FDA keeps current a list of approved medical devices with AI embedded (<https://www.fda.gov/medical-devices/software-medical-device-samd/artificial-intelligence-enabled-medical-devices>) this list can be compared to the organization's inventory to identify medical devices with embedded AI.

Finally, organizations can review any exceptions granted previously for specific AI use cases. By using these various techniques, a relatively complete inventory of systems, solutions, and devices can be gathered of existing AI already deployed within an organization's environment. All discovery and monitoring should be consistent with organizational policy and be deployed in alignment with privacy and other laws. Please ensure that the appropriate departments are involved in planning, deployment, and monitoring these tools.

Below are examples of AI-enabled third-party systems by category with example solutions provided for each category. This is not an exhaustive list nor an endorsement for any solution mentioned. Sample Categorization may aid in inventory management goals, but each organization should identify their categories in a way to aid in risk management of the systems.

Clinical Care Device Categories:

- AI Diagnostic Imaging: Aidoc for stroke detection, iCAD for mammography screening, Zebra Medical Vision for radiology findings
- AI Pathology: PathAI digital pathology systems, Paige.AI for cancer detection
- AI Cardiology: Caption Health ultrasound guidance, Viz.ai for large vessel occlusion detection
- AI Ophthalmology: IDx-DR for diabetic retinopathy screening (autonomous diagnosis)

- AI Sepsis/Deterioration Prediction: Epic Sepsis Model, TREWS (Targeted Real-time Early Warning System)
- AI Medication Management: Automated dosing recommendations, predictive models influencing treatment planning
- Closed-Loop Insulin Delivery: Medtronic MiniMed 780G, Tandem Control-IQ, Omnipod 5 (AI-driven automated insulin adjustment)
- AI-Controlled Ventilators: INTELLiVENT-ASV adaptive support ventilation, Puritan Bennett 980 with automated weaning protocols
- Autonomous Robotic Surgery Systems: Da Vinci with autonomous tissue recognition and movement assistance
- AI Anesthesia Delivery: Sedasys (automated propofol delivery - withdrawn but example of category)
- AI Cardiac Monitoring with Intervention: Implantable cardioverter-defibrillators (ICDs) with AI algorithms, automated external defibrillators (AEDs) with AI rhythm analysis
- AI Radiation Therapy: Ethos adaptive radiotherapy with AI-driven real-time treatment adjustments
- AI Critical Care Automation: Closed-loop hemodynamic management systems, AI-driven code blue response systems
- Enterprise Infrastructure AI: Organization-wide bed management, emergency department flow optimization affecting patient placement, AI-driven staffing systems affecting care coverage

Operational and Administrative Categories:

- Medical Record AI: Tools to aid in documenting notes, patient scheduling, ambient listening, creating patient letters, or any other AI tools integrated into the electronic medical record system. Examples include Heidi, Abirdge, DAX, and Freed.
- Medical Coding and Billing AI: Tools to aid in billing and coding of claims, deal with insurance denials, or other processes used in the revenue cycle processes. Examples include XpertDox, RapidClaims.AI, and Claimocity
- Enterprise Infrastructure AI: other general AI used by operational departments such as Claude, ChatGPT or other function specific AI to manage day-to-day activities such as contracts., Email filtering, security, network monitoring, virtual assistant, etc.

Data and Analytics Categories

- Population Health Analytics AI Platforms: Use predictive risk stratification, automated care management, social determinates of health and Point-of-Care insights. Most platforms such as Navina, Health Catalyst, and Persivia have AI engines included in their platforms.
- Data Mining and Pattern Recognition: Dataiku, Alteryx, Databricks, and DataRobot.
- Research and Development AI: DataRobot, Sigli, and Scientist.com

Appendix D: RACI Matrix

This RACI matrix defines roles and responsibilities for key activities within the AI third-party risk management lifecycle. Organizations should adapt this matrix to their specific structure, size, and governance model.

RACI Key

R = Responsible (performs the work) A = Accountable (ultimately answerable for completion) C = Consulted (provides input) I = Informed (kept updated on progress)

Activity	Responsible (R)	Accountable (A)	Consulted (C)	Informed (I)
<i>PHASE 0: USE CASE JUSTIFICATION & STRATEGIC ASSESSMENT</i>				
Use Case Justification & Problem Definition	Business Owner	CIO/CISO	Clinical Leadership, Privacy Officer, Security Team, Compliance	Senior Management
Initial Risk Classification	Security Team, Business Owner	CISO	Clinical Leadership, Compliance, Privacy Officer	Senior Management
<i>PHASE 1: DUE DILIGENCE & VENDOR EVALUATION</i>				
Vendor Vetting & GRC Assessment	Security & Compliance Teams	CISO	Business Owner, Privacy Officer, Clinical Leadership, Legal	Senior Management
Security Risk Assessment (SRA)	Information Security Team	CISO	IT Team, Business Owner	Compliance, Senior Management
AI-Specific Risk Evaluation (Model, Bias, Transparency)	Security Team, Compliance	CISO, Chief Compliance Officer	Clinical Leadership, Data Science Team (if available), Business Owner	Senior Management
<i>PHASE 2: CONTRACT NEGOTIATION & LEGAL PROTECTIONS</i>				
Contract Negotiation	Legal Team, Procurement	General Counsel	CISO, Compliance, Business Owner, Privacy Officer	Senior Management
BAA Negotiation & AI Clauses	Privacy Officer, Legal Team	Chief Privacy Officer	CISO, Compliance, Business Owner	Senior Management

SLA Definition & Performance Metrics	Business Owner, IT Team	CIO	Security, Vendor Management, Clinical Leadership	Senior Management
<i>PHASE 3: IMPLEMENTATION, INTEGRATION & TRAINING</i>				
Technical Implementation & Integration	IT Team	CIO	Security, Business Owner, Clinical Leadership, Vendor	Senior Management, Compliance
Sandbox/Staging Testing	IT Team, Business Owner	CIO, Business Owner	Security, Clinical Leadership, Quality/Clinical Engineering	Compliance
Verification & Validation (V&V)	Business Owner, Quality/Clinical Engineering	Business Owner	IT, Security, Clinical Leadership, Compliance	Senior Management
Privacy Impact Assessment	Privacy Officer	Chief Privacy Officer	Security, Compliance, Business Owner, Legal	Senior Management
User Training & Change Management	Business Owner, Training/Learning Team	Business Owner	IT, Security, Clinical Leadership, HR	Compliance
Production Deployment	IT Team	CIO	Security, Business Owner, Clinical Leadership	Senior Management, Compliance
<i>PHASE 4: ONGOING MONITORING & PERFORMANCE MANAGEMENT</i>				
Performance Monitoring (Operational)	Business Owner, IT Operations	Business Owner	Security, Compliance	Senior Management
Security Monitoring & Threat Detection	Security Operations Team	CISO	Business Owner, IT Operations	Compliance, Senior Management
Model Drift & Bias Monitoring	Business Owner, Data Science/Analytics Team	Business Owner	Clinical Leadership, Quality, Compliance	Senior Management, CISO

Vendor Performance Management	Vendor Management/Procurement	CPO (Chief Procurement Officer)	Business Owner, IT, Security, Compliance	Senior Management
Update/Patch Testing & Validation	IT Team, Business Owner	CIO, Business Owner	Security, Clinical Leadership (if clinical system), Vendor	Compliance
Configuration Re-validation After Updates	IT Team, Security Team	CISO, CIO	Business Owner	Compliance
Periodic Reassessment (Annual or at Renewal)	Security & Compliance Teams	CISO, Chief Compliance Officer	Business Owner, IT, Privacy Officer	Senior Management
<i>PHASE 5: INCIDENT RESPONSE & RECOVERY</i>				
Incident Detection & Classification	Security Operations/IT Operations	CISO	Business Owner, Compliance, Privacy Officer	Senior Management, Legal
Incident Response & Vendor Coordination	Incident Response Team, Vendor	CISO, General Counsel	Public Relations, Clinical Leadership, Business Owner, Regulatory Affairs	Senior Management, Board (if major incident)
Model Rollback & System Recovery	IT Team, Vendor	CIO, Business Owner	Security, Clinical Leadership, Quality/Clinical Engineering	Senior Management, Compliance
Post-Incident Analysis & CAPA	Security Team, Business Owner, Vendor	CISO, Business Owner	Compliance, Legal, Clinical Leadership, Quality	Senior Management
Regulatory Notification (FDA, OCR, etc.)	Regulatory Affairs, Compliance	General Counsel, Chief Compliance Officer	CISO, Privacy Officer, Clinical Leadership	Senior Management, Board
<i>PHASE 6: END-OF-LIFE & TRANSITION MANAGEMENT</i>				
EOL Planning & Risk Assessment	Business Owner, IT Team	CIO, Business Owner	Security, Compliance, Vendor Management	Senior Management

Data Extraction & Migration	IT Team	CIO	Business Owner, Security, Privacy Officer, Vendor	Compliance, Senior Management
Data Migration Validation	Business Owner, IT Team	Business Owner	Clinical Leadership, Quality, Privacy Officer	Compliance, Senior Management
Secure Decommissioning & Data Destruction	IT Team, Security Team	CISO, CIO	Vendor, Business Owner, Privacy Officer	Compliance, Senior Management
Replacement System Implementation (if applicable)	IT Team, Business Owner	CIO, Business Owner	Security, Clinical Leadership, Compliance, New Vendor	Senior Management
<i>CONTINUOUS ACTIVITIES (ALL PHASES)</i>				
Inventory/Asset Management	IT Asset Management	CIO	Security, Business Owners	Compliance
Policy & Procedure Updates	Compliance, Privacy Officer, Security Team	Chief Compliance Officer	Legal, Business Leadership, Clinical Leadership	Senior Management, Board
Governance Oversight & Reporting	Compliance, CISO	Chief Compliance Officer, CISO	Business Owners, Legal, Privacy Officer	Senior Management, Board
Documentation & Audit Trail Maintenance	Business Owner, Compliance	Chief Compliance Officer	IT, Security, Privacy Officer	Internal Audit
Threat Modeling	Vendor	Vendor, Business Owners, IT Security	Legal, IT Security	CIO, Chief Compliance Office, CISO. Privacy Officer

Appendix E: Sample Commercial Contract Language

These clauses apply to commercial contracts involving AI use, such as license agreements, evaluation agreements, or Software-as-a-Service (SaaS) contracts. Organizations should incorporate these provisions into master services agreements, statements of work, or AI-specific contract addendums.

1: Scope of Use and Permitted Applications

Clearly define what the AI solution can be used for, including specific use cases, user populations, data types, and deployment environments. Distinguish between evaluation/pilot use and production deployment, with separate approval requirements and restrictions for each phase. Specify any prohibited uses or applications that exceed the scope of the agreement.

Example language: "Customer may use the AI Solution solely for [specific clinical/operational use case] within [specified departments/facilities]. The AI Solution shall not be used for [prohibited applications]. Production deployment requires Vendor's written approval following successful pilot validation."

2: License Restrictions and Access Controls

Prevent unauthorized use, copying, modification, or resale of the AI product. Define authorized users, permitted access locations (on-premises, cloud, remote), and any geographic or jurisdictional restrictions. Specify whether the license is perpetual or term-based, exclusive or non-exclusive, and whether sublicensing is permitted.

Example language: "License is non-exclusive, non-transferable, and limited to Customer's employees and credentialed staff at [specified locations]. Customer shall not reverse engineer, decompile, or attempt to extract the AI model. Access from outside [jurisdiction] requires prior written approval."

3: Updates, Patches, and Change Management

Define the cadence for updates and patches (e.g., quarterly planned updates, emergency security patches within 72 hours). Require vendor to provide comprehensive release notes, change documentation, and validation evidence at least [30 days] prior to planned updates. Establish testing requirements, including mandatory deployment to sandbox/staging environment before production release. Require vendor to obtain Customer approval before deploying updates that modify AI model architecture, change default security configurations, alter data processing methods, or affect system integrations. Specify rollback procedures and vendor support obligations if updates cause performance degradation or operational issues.

Example language: "Vendor shall provide detailed release notes including all functional changes, security updates, and AI model modifications at least 30 days prior to deployment. All updates must be deployed to Customer's test environment for validation before production release. Customer reserves the right to delay or reject updates that do not meet validation criteria. Vendor shall support rollback to previous version if update causes system instability or performance degradation."

(Cross-reference: See Quality Assurance/Verification/Validation section for detailed testing requirements)

4: Data Ownership and Control

Organization retains full ownership of all input data, output data, derived insights, and any fine-tuning or customization data generated through use of the AI system. Vendor shall have no rights to organizational data except as strictly necessary to provide contracted services. Vendor cannot use, disclose, sell, or license organizational data to third parties without explicit written consent. Specify ownership of any custom AI models developed using organizational data.

Example language: "Customer retains all right, title, and interest in Customer Data, including all inputs, outputs, annotations, feedback, and derived insights generated through use of the AI Solution. Vendor shall not use Customer Data for any purpose other than providing Services under this Agreement. Any custom models trained on Customer Data shall be owned by Customer, with Vendor retaining only a license to operate such models on Customer's behalf."

5: Confidentiality and Intellectual Property Protection

Establish mutual confidentiality obligations protecting trade secrets, proprietary algorithms, model architectures, training methodologies, performance metrics, and test results. Define what constitutes confidential information, permitted disclosures (e.g., to subcontractors under NDA, to regulators as required by law), and duration of confidentiality obligations (typically surviving contract termination).

Example language: "Each party agrees to maintain in confidence all Confidential Information of the other party, including but not limited to AI model architecture, training data, algorithms, security assessments, and performance benchmarks. Confidential Information may not be disclosed except to employees and contractors with need-to-know and confidentiality obligations. Obligations survive contract termination for [5 years]."

6: Prohibition on AI Training Without Consent

Explicitly prohibit vendor from using organizational data to train, improve, fine-tune, or validate AI models—whether vendor's own models or third-party models—without prior written consent specifying the exact purpose, data elements, and safeguards. Prohibit use of organizational data to train general-purpose or multi-tenant AI models. Require vendor to disclose if organizational data will be used for any model development activities.

Example language: "Vendor shall not use Customer Data to train, retrain, fine-tune, validate, or improve any AI models, whether proprietary or third-party, general-purpose or specialized, without Customer's prior written consent specifying the exact data elements, purpose, de-identification methods, and time period for such use. Vendor represents that Customer Data will not be co-mingled with other customers' data or used to improve Vendor's general AI capabilities."

7: Security and Compliance Requirements

Require vendor to implement and maintain security controls commensurate with the AI system's risk classification and the sensitivity of data processed. Specify minimum security requirements including: encryption standards (e.g., AES-256 for data at rest, TLS 1.3 for data in transit); multi-factor authentication; access logging and monitoring; vulnerability management; penetration testing frequency; and security incident response capabilities. Require adherence to applicable frameworks (e.g., NIST Cybersecurity Framework, HITRUST, SOC 2 Type II). Mandate

compliance with all applicable laws and regulations including HIPAA, HITECH, FDA regulations, state privacy laws, and organizational policies.

Example language: "Vendor shall maintain security controls meeting or exceeding [HITRUST CSF / SOC 2 Type II] standards, including: (a) AES-256 encryption for data at rest and TLS 1.3 for data in transit; (b) multi-factor authentication for all user access; (c) comprehensive audit logging; (d) annual penetration testing by qualified third party; (e) vulnerability scanning and patching within [timeframe]. Vendor shall comply with HIPAA, HITECH, FDA regulations applicable to [device classification], and all Customer security policies provided in Exhibit [X]."

(Cross-reference: Security requirements should be informed by GRC assessment findings)

8: Termination and Data Return/Destruction

Establish clear termination rights for both parties, including termination for convenience (with notice period), termination for cause (breach, insolvency, regulatory issues), and termination upon end-of-life or discontinuation of AI service. Upon termination, require vendor to: return all organizational data in usable, non-proprietary formats within [30 days]; provide data export tools and support at no additional charge; securely destroy all data retained in vendor systems, including backups, caches, and logs; and provide certification of data destruction meeting NIST 800-88 standards or equivalent. Specify transition assistance obligations to facilitate migration to replacement vendor or system.

Example language: "Either party may terminate this Agreement [for convenience with 90 days' notice / for cause upon 30 days' notice of uncured breach]. Upon termination, Vendor shall: (a) within 30 days, return all Customer Data in [CSV/JSON/FHIR] format suitable for import to alternative systems; (b) provide extraction tools and technical support at no additional charge; (c) securely destroy all Customer Data in Vendor's systems per NIST 800-88 Guidelines; (d) provide written certification of destruction signed by Vendor's Chief Security Officer; (e) provide up to [40 hours] of transition assistance to facilitate migration to replacement system."

9: Performance Standards and Support Obligations

Define Service Level Agreements (SLAs) specifying system uptime/availability targets (e.g., 99.9% uptime), response times for support requests by severity level, resolution timeframes for critical issues, and AI model performance baselines (accuracy, precision, recall, or other metrics relevant to the use case). Establish penalties or service credits for SLA breaches. Require vendor to provide comprehensive documentation, user training, ongoing technical support, and access to subject matter experts.

Example language: "Vendor guarantees: (a) 99.9% system uptime measured monthly, excluding scheduled maintenance; (b) AI model accuracy of no less than [X%] as measured by [specified methodology]; (c) support response within [1 hour for critical / 4 hours for high / 24 hours for medium / 72 hours for low priority issues]; (d) resolution within [timeframes by severity]. For each 0.1% shortfall in uptime SLA, Customer receives [5%] service credit. Vendor shall provide 24/7 technical support, comprehensive user documentation, training materials, and quarterly business reviews."

10: Limitation of Liability and Harm Notification

Carefully negotiate limitation of liability provisions to ensure adequate protection for both parties while not inappropriately capping damages for serious breaches or AI-related patient harm. Establish clear obligations for

vendor to notify Customer and regulatory authorities (FDA, other applicable agencies) of any near-harm events, adverse events, serious injuries, or deaths associated with AI system use, in accordance with applicable regulations including the Safe Medical Devices Act of 1990. Prohibit vendor from imposing contractual limitations that would prevent required regulatory reporting by either party.

Example language: "Vendor's aggregate liability under this Agreement shall not exceed [amount], except that such limitation shall not apply to: (a) breaches of confidentiality or data security; (b) intellectual property infringement; (c) gross negligence or willful misconduct; (d) violations of HIPAA or other privacy laws. Vendor shall immediately notify Customer and applicable regulatory authorities (FDA, etc.) of any adverse events, serious injuries, near-harm events, or deaths potentially associated with AI Solution, in compliance with Safe Medical Devices Act and other applicable regulations. Nothing in this Agreement shall limit either party's obligation to report safety events to regulatory authorities."

11: PHI Processing Limitations and Model Modification Restrictions

For AI systems processing Protected Health Information, establish clear boundaries on how PHI may be accessed, used, transformed, or disclosed by the AI system. Require vendor to notify Customer and obtain written approval before implementing any changes to AI model architecture, training data sources, or data processing methods that would alter how PHI is used, accessed, or potentially exposed. Any such changes require contract amendment and potentially new privacy impact assessment and security risk assessment.

Example language: "Vendor shall process PHI solely as specified in Exhibit [X] and shall not modify AI model's PHI processing methods without Customer's prior written approval. Any changes to: (a) PHI data elements accessed or processed; (b) AI model architecture or training that affects PHI; (c) third-party AI services processing PHI; (d) data retention, storage, or destruction methods for PHI; shall require contract amendment, updated privacy impact assessment, and Customer governance approval before implementation."

12: Audit Rights and Compliance Verification

Grant Customer (and Customer's authorized auditors, regulators, and certification bodies) the right to audit vendor's compliance with contractual obligations, security controls, data handling practices, and AI model governance. Specify audit scope, frequency (e.g., annually or upon reasonable notice), vendor's cooperation obligations, and requirements for remediation of audit findings. For high-risk AI systems, consider requiring annual independent security audits or certifications (SOC 2, HITRUST, ISO 27001) at vendor's expense.

Example language: "Customer may, upon [30 days' notice], audit Vendor's compliance with this Agreement, including inspection of facilities, systems, security controls, and data handling practices related to the AI Solution. Vendor shall cooperate fully, provide access to relevant personnel and documentation, and remediate any deficiencies within [30 days]. Customer may engage qualified third-party auditors under confidentiality obligations. For critical-impact AI systems, Vendor shall obtain annual SOC 2 Type II certification and provide reports to Customer."

13: Mutual Indemnification

Establish mutual indemnification obligations protecting each party from third-party claims arising from the other party's breach of contract, negligence, or misconduct. For AI systems, specifically address indemnification for: intellectual property infringement claims related to AI models or training data; claims arising from AI-generated

outputs or decisions; data breaches or privacy violations; regulatory enforcement actions; and patient harm allegedly caused by AI system failures or errors (subject to appropriate limitations and carve-outs).

Example language: "Each party shall indemnify, defend, and hold harmless the other party from third-party claims arising from: (a) indemnifying party's breach of this Agreement; (b) negligence or willful misconduct; (c) violation of applicable laws. Vendor shall indemnify Customer for claims arising from: (i) AI model intellectual property infringement; (ii) unauthorized use of training data; (iii) AI-generated outputs (subject to Customer's proper use). Customer shall indemnify Vendor for claims arising from Customer's misuse of AI Solution contrary to documentation and training. [Specify procedures, notice requirements, cooperation obligations, and any caps]."

14: Coordinated Incident Response and Recovery

Require vendor to actively participate in Customer's incident response procedures for security incidents, AI system failures, or performance degradation affecting the AI solution. Establish vendor's obligations to: provide timely notification of incidents (within [2-24 hours] depending on severity); cooperate in incident investigation and forensic analysis; provide technical expertise and system access necessary for response; assist in containment, eradication, and recovery activities; support communications with regulators and affected individuals; and conduct post-incident root cause analysis and implement corrective actions.

Example language: "Vendor shall immediately notify Customer of any security incidents, AI model failures, performance anomalies, or other events affecting the AI Solution (within 2 hours for Critical incidents, 24 hours for others). Vendor shall: (a) participate in joint incident response per Customer's incident response plan; (b) provide technical expertise, system logs, and forensic support; (c) assist in containment and remediation at no additional charge; (d) support regulatory notifications and communications; (e) conduct root cause analysis and implement preventive measures; (f) provide post-incident reporting within [timeframe]."

(Cross-reference: See Response and Recovery Planning section for detailed incident response requirements)

15: Post-Incident Recovery and Return-to-Normal Criteria

Define clear criteria and procedures for determining when systems can return to normal operation following a security incident, AI failure, or significant performance degradation. Specify what documentation or attestations are required from vendor to validate that threats have been eliminated, vulnerabilities remediated, and system integrity restored. For significant incidents, require independent third-party attestation that vendor systems are secure before re-establishing connectivity. Establish reasonable timelines for recovery and resumption of service, with escalation procedures if timelines are not met.

Example language: "Following any Critical or High severity incident, vendor must provide: (a) detailed incident report documenting root cause, scope of impact, and remediation actions; (b) evidence that all vulnerabilities have been patched and threats eliminated; (c) updated security risk assessment; (d) for breaches involving potential compromise of vendor infrastructure, independent third-party penetration test report confirming no persistent threats; (e) certification by vendor's CISO that systems are secure for reconnection. Customer reserves the right to maintain isolation of affected systems until satisfied that risks are adequately mitigated. Target recovery timelines: [specify by severity and system criticality]."

(Cross-reference: See Response and Recovery Planning section)

16: Model Drift Monitoring and Revalidation

Require vendor to implement continuous or periodic monitoring for AI model drift (degradation in performance due to changes in input data distributions or concept drift). Establish thresholds for acceptable performance variation and require vendor to notify Customer when drift exceeds thresholds. Mandate periodic reassessment and revalidation of AI models following: scheduled retraining events or model updates; significant changes to model architecture; emergence of new threat intelligence relevant to AI exploitation; regulatory or clinical performance issues identified by Customer or reported by other users; and at minimum annually for Critical and High-impact systems.

Example language: "Vendor shall monitor AI model performance continuously for drift, using metrics including [specify: accuracy, precision, recall, etc.]. Vendor shall notify Customer within [48 hours] if performance degrades by more than [X%] from baseline. Vendor shall revalidate AI model following: (a) any model update or retraining; (b) architecture changes; (c) newly identified AI vulnerabilities or exploits; (d) reported clinical performance issues; (e) at minimum annually. Revalidation shall include [specify: test data set validation, bias assessment, security testing] and be documented in validation report provided to Customer."

(Cross-reference: See Response and Recovery Planning section and Ongoing Monitoring section)

17: Quality Assurance and Verification/Validation (QA/VV) Documentation

Require vendor to provide comprehensive Quality Assurance and Validation documentation tailored to the AI system's intended use, safety classification (Low/Medium/High/Critical impact), and criticality in the healthcare setting. QA/VV documentation must include: system description and intended use statement; risk classification and hazard analysis; verification and validation test plans and results; model performance baselines and acceptance criteria; training data characterization and bias assessment; version control and traceability to development changes; clinical validation evidence (for clinical AI); and regulatory compliance documentation (FDA submissions, CE marking, etc. as applicable).

Example language: "Vendor shall provide QA/VV documentation including: (a) intended use statement and risk classification; (b) verification test results demonstrating AI model performs as specified; (c) validation evidence confirming AI model meets user needs and intended use; (d) performance baselines for [accuracy/precision/recall/other relevant metrics]; (e) training data description, sources, and bias assessment methodology and results; (f) security testing results including adversarial robustness; (g) clinical validation study results (for clinical AI); (h) version history and change log; (i) regulatory submissions and clearances (FDA 510(k), De Novo, PMA as applicable). Documentation shall be updated with each model version and provided to Customer within [timeframe]."

Business Associate Agreement (BAA) - AI-Specific Provisions

For AI systems that access, process, store, or transmit Protected Health Information (PHI), HCOs must ensure that BAAs include provisions addressing AI-specific risks and requirements. These provisions should be incorporated into standard BAAs as amendments or addendums or may warrant separate AI-focused BAAs for high-risk clinical AI systems. See [Appendix F](#) for Sample BAA Contract Language.

Appendix F: Sample BAA Contract Language

The following provisions establish baseline protections that all AI-related BAAs should include:

Prohibition on AI Training with PHI: Business Associate shall not use, disclose, or permit access to any Covered Entity PHI for the purpose of training, retraining, fine-tuning, validating, or improving any artificial intelligence models, whether proprietary, third-party, general-purpose, or specialized, without Covered Entity's prior written consent. Any approved use of PHI for AI training shall be governed by a separate written agreement specifying exact data elements, de-identification methods, purpose limitations, duration, and destruction requirements. Business Associate represents that Covered Entity PHI will not be co-mingled with data from other sources for training multi-tenant or general-purpose AI models.

Permitted Uses and Disclosures: Business Associate may access, use, and disclose PHI solely for the specific AI-enabled services described in Exhibit [X] and only to the extent necessary to perform such services. Permitted uses and disclosures shall align with HIPAA's Privacy Rule minimum necessary standard. Business Associate shall not use or disclose PHI for any purpose other than those explicitly authorized in this Agreement or required by law. Any expansion of AI system functionality that requires access to additional PHI data elements or new uses of existing PHI requires Covered Entity's prior written approval and contract amendment.

AI Purpose Limitation: Business Associate's AI Solution shall process PHI only for the clinical, operational, or administrative purposes specified in this Agreement. Business Associate shall not use PHI to develop AI capabilities beyond the contracted scope, sell or license AI models trained on PHI, provide AI services to other parties using PHI-derived insights, or use PHI for marketing, advertising, or commercial purposes unrelated to services provided to Covered Entity.

Minimum Necessary Standard: Business Associate's AI Solution shall be designed and configured to access and process only the minimum necessary PHI required to accomplish the specified purpose. Business Associate shall implement technical controls to limit AI model access to PHI data elements, restrict AI processing to defined patient populations or time periods, enforce role-based access controls for AI system users, and regularly review and optimize PHI access patterns to ensure minimum necessary standard compliance.

Compliance with Laws and Regulations: Business Associate warrants that its AI Solution and all AI-related services comply with HIPAA Privacy Rule, Security Rule, and Breach Notification Rule; HITECH Act requirements; applicable state privacy laws; FDA regulations (if AI is a medical device); Covered Entity's privacy and security policies provided in Exhibit [X]; and all other applicable federal, state, and local laws and regulations.

1: Permitted Uses and Disclosures of PHI for AI Processing

Specifically enumerate the permitted uses and disclosures of PHI by the AI system, including: AI model inference and prediction activities; clinical decision support recommendations; patient risk stratification and identification; automated documentation or coding assistance; quality measurement and reporting; population health analytics; or other specified functions. Prohibit use or disclosure for any purpose not explicitly listed.

2: Prohibition on AI Training with PHI Without Explicit Authorization

Categorically prohibit Business Associate from using PHI to train general-purpose AI models, large language models, or any AI systems not exclusively dedicated to Covered Entity's use. Require separate written authorization specifying data governance, de-identification standards, model ownership, and disposition of PHI-derived models if any AI training using PHI is contemplated.

3: De-Identification and Limited Data Set Restrictions

If Business Associate requests use of de-identified data or limited data sets for AI development or research, require: compliance with HIPAA de-identification standards (expert determination or safe harbor method); documentation of de-identification methodology and validation; prohibition on re-identification attempts; restrictions on use of limited data sets consistent with data use agreements; and requirement that any AI models developed using de-identified or limited data sets shall not be used to re-identify individuals.

4: AI-Specific Safeguards

Require Business Associate to implement technical and administrative safeguards specific to AI risks, including: encryption of PHI during AI model training, inference, and storage; access controls limiting AI model access to authorized data elements; audit logging of all AI system access to and processing of PHI; protection against adversarial attacks, model inversion, or membership inference that could expose PHI; safeguards against unintended disclosure of PHI in AI-generated outputs (e.g., hallucinations containing PHI); and monitoring for anomalous AI behavior that could indicate security compromise.

(Cross-reference: Safeguards should be informed by GRC assessment and security risk assessment findings)

5: Breach Notification for AI-Related Incidents

Establish expedited breach notification timelines for AI-specific incidents, recognizing that AI breaches may be more difficult to detect and scope. Require Business Associate to notify Covered Entity within [2 days] of discovery of: unauthorized access to PHI through AI system compromise; data exfiltration potentially including PHI used in AI training or inference; model theft or unauthorized copying that may contain embedded PHI; adversarial attacks that caused PHI disclosure; or any other security incident affecting AI system integrity or PHI confidentiality. Require comprehensive breach reports including scope assessment methodology, potentially affected individuals, and remediation actions.

6: Subcontractor and Third-Party AI Service Flowdown

Require Business Associate to ensure that any subcontractors, cloud service providers, AI platform vendors, or other third parties that create, receive, maintain, or transmit PHI on behalf of Business Associate enter into written agreements imposing the same restrictions and conditions that apply to Business Associate under the BAA. Specifically address AI platform providers (e.g., AWS, Azure, Google Cloud AI services, OpenAI, Anthropic, etc.) and require Business Associate to obtain BAAs from all third-party AI service providers before PHI is processed through such services.

7: Individual Access and Amendment Rights

Establish procedures for Covered Entity or individuals to: access PHI that has been processed by the AI system; obtain explanations of AI-generated recommendations or decisions affecting their care; request amendment or correction of PHI processed by AI; and obtain documentation of how PHI was used in AI decision-making. Require

Business Associate to provide reasonable assistance in facilitating these rights, including providing AI decision audit trails and model explanation capabilities where technically feasible.

8: Data Return or Destruction Including AI Model Disposal

Upon termination of the agreement, require Business Associate to: return or destroy all PHI in any form, including structured data, unstructured data, and any PHI embedded in AI model weights, parameters, or training artifacts; securely dispose of AI models that were trained using PHI or that may contain embedded PHI; provide certification that all PHI has been returned or destroyed per NIST 800-88 Guidelines; and address disposition of de-identified data sets or AI models that Business Associate claims no longer contain PHI (require expert determination or similar validation).

9: Audit and Oversight of AI Data Flows

Grant Covered Entity the right to audit Business Associate's AI systems and data flows to verify: PHI is accessed and processed only as authorized; minimum necessary standard is enforced by AI system design; security safeguards are functioning as designed; AI training prohibitions are being respected; and subcontractor BAAs are in place for all third-party AI services. Require Business Associate to provide documentation, system logs, audit trails, and access to technical personnel necessary to conduct audits. For Critical and High-impact AI systems, consider requiring periodic independent security audits focused on PHI protection.

Appendix G: AI Vendor Assessment Questions for Procurement and GRC

This appendix provides comprehensive question sets for evaluating AI vendors and solutions during the procurement process and Governance, Risk, and Compliance (GRC) assessments. These questions are designed to help HCOs:

- Justify the need for AI solutions before entering procurement
- Conduct thorough vendor due diligence aligned with Phase 1 activities
- Assess AI-specific risks that standard IT vendor assessments do not address
- Scale assessment rigor based on AI system risk classification (Low/Medium/High/Critical)
- Ensure consistent evaluation across all AI procurement activities

Section 1: Pre-Procurement Use Case Justification Questions

These questions should be incorporated into the procurement intake process when staff, faculty, or physicians request AI tools or products. Answers to these questions inform the initial risk classification and governance approval process before formal vendor evaluation begins.

- 1. What is the specific problem the AI is addressing, and how does it align with our organization's strategic objectives?**
 - a. Describe the current state problem or opportunity
 - b. Explain how AI solutions support organizational goals
 - c. Quantify expected benefits (clinical outcomes, efficiency gains, cost savings, etc.)
- 2. Is there documentation on how the AI system supports clinical, operational, or administrative outcomes?**
 - a. Provide evidence of effectiveness (peer-reviewed studies, case studies, vendor-provided data)
 - b. Identify key performance indicators that will measure success
- 3. Have you identified alternative methods to accomplish this process/task without AI? If so, why is AI preferred given its added complexity and risk?**
 - a. List non-AI alternatives considered
 - b. Justify why AI approach is superior or necessary
 - c. Explain why benefits outweigh risks and complexity
- 4. How does the vendor recommend the AI be used (what is the ideal workflow)?**
 - a. Describe vendor's recommended implementation and use patterns

- b. Identify critical workflow touchpoints and user interactions
5. **How does the recommended AI workflow match the current workflow used by the department/organization?**
- a. Assess degree of workflow disruption or change required
 - b. Identify gaps between current state and AI-enabled workflow
 - c. Estimate change management effort and user training requirements
6. **What is the preliminary risk classification of this AI solution? (Low/Medium/High/Critical)**
- a. Assess patient safety impact
 - b. Evaluate financial and operational criticality
 - c. Determine PHI/PII processing requirements
 - d. Identify regulatory classification (FDA device, etc.)
7. **Which stakeholders must be engaged for this AI procurement?**
- a. Business owner and clinical champion
 - b. Privacy, security, legal, compliance teams
 - c. Governance committee or AI review board

Section 2: Standard GRC Questions for AI Vendors

During a normal GRC assessment process, these general questions should be asked of all vendors, but responses should be scrutinized especially carefully for AI vendors due to the unique data processing, security, and compliance risks AI systems introduce.

1. **Does the AI service access, process, or store Protected Health Information (PHI)?**
 - a. If yes, what type of AI model is used (classical ML, generative AI, LLM, etc.)?
 - b. How is PHI utilized in the AI system (training, inference, both)?
2. **Can you provide a signed Business Associate Agreement (BAA)?**
 - a. Does the BAA address AI-specific data processing?
 - b. Are all subcontractors and third-party AI services covered by BAAs?
3. **Is any PHI used to train the AI model?**
 - a. If yes, how is patient anonymity preserved?
 - b. What de-identification methods are used?
 - c. Is training data segregated from production data?

4. **Do you use de-identified or anonymized data in any part of the system? How is de-identification validated?**
 - a. Describe de-identification methodology (safe harbor, expert determination)
 - b. Provide validation/certification of de-identification process
 - c. Confirm prohibition on re-identification attempts
5. **Are you compliant with HIPAA, HITECH, and applicable state privacy laws?**
 - a. Provide evidence of compliance (certifications, audits, attestations)
 - b. Describe compliance program and oversight
 - c. Identify state-specific privacy law compliance (California CMIA, etc.)
6. **Who has access to our organization's data, including subcontractors?**
 - a. List all personnel, contractors, and third parties with data access
 - b. Describe access controls and monitoring
 - c. Provide background check and training requirements for personnel
7. **What data residency requirements exist and how are they enforced?**
 - a. Specify where data is stored geographically
 - b. Describe controls preventing data from leaving authorized jurisdictions
 - c. Address any offshore processing or storage
8. **What mechanisms exist for data subject rights (access, deletion, portability)?**
 - a. Describe processes for handling individual rights requests
 - b. Specify response timeframes
 - c. Explain how rights are exercised for AI-processed data
9. **How is data anonymization/de-identification implemented and validated?**
 - a. Detail anonymization techniques and validation procedures
 - b. Provide expert determination or certification documentation
10. **Is there a retention period for all customer data within the AI solution?**
 - a. Specify retention periods by data type
 - b. Does retention policy include prompt logs, interaction data, and training data?
 - c. Describe data destruction procedures at end of retention
11. **If an external LLM SaaS is used (OpenAI, Anthropic, etc.), is there a BAA with all LLM vendors?**

- a. Identify all third-party AI services used
 - b. Confirm BAAs are in place before PHI is processed
 - c. Describe data flows to third-party AI services
12. **What security frameworks do you follow?** (e.g., NIST 800-53, HITRUST, SOC 2)
- a. Provide certifications or audit reports (SOC 2 Type II, HITRUST CSF, ISO 27001)
 - b. Describe alignment with healthcare security standards
13. **Is there a privacy policy for the AI system?**
- a. Provide current privacy policy
 - b. Confirm policy addresses AI-specific data processing
14. **Do you implement Responsible AI principles?**
- a. Describe responsible AI framework (fairness, transparency, accountability)
 - b. Provide evidence of responsible AI practices
15. **Do you conduct regular penetration tests and vulnerability assessments?**
- a. Specify frequency (annually, quarterly, continuous)
 - b. Provide most recent penetration test summary
 - c. Describe vulnerability remediation processes and timelines
16. **Is multi-factor authentication (MFA) required for system access?**
- a. Confirm MFA is enforced for all user access
 - b. Describe authentication methods supported
17. **What is your incident response plan in the event of a cybersecurity breach?**
- a. Provide incident response plan overview
 - b. Specify notification timelines and procedures
 - c. Describe coordination with customers during incidents
18. **Do you provide audit logs and system activity tracking?**
- a. Describe logging capabilities and retention periods
 - b. Confirm availability of logs to customers
 - c. Specify what activities are logged (access, changes, AI decisions, etc.)
19. **Is data obscured or restricted in chat/prompt logs?**
- a. If not, how is access to chat or prompt logs secured from unauthorized access?
 - b. Describe controls preventing PHI exposure in logs

20. **Does the solution integrate with our EHR or clinical systems?**
 - a. If yes, how? (API, HL7, FHIR, other standards)
 - b. Describe integration architecture and data flows
 - c. Provide integration documentation and support
21. **What onboarding, training, and support services are included?**
 - a. Detail training programs and materials provided
 - b. Describe ongoing support model (24/7, business hours, tiered support)
 - c. Specify included versus additional-cost services
22. **Are user permissions role-based and configurable?**
 - a. Describe role-based access control (RBAC) capabilities
 - b. Confirm granular permission management
 - c. Explain how roles align with healthcare workflows
23. **What warranties or guarantees are provided about system performance or compliance?**
 - a. Specify performance guarantees and SLAs
 - b. Describe compliance warranties (HIPAA, FDA, etc.)
 - c. Identify limitations or exclusions
24. **What is your data retention and destruction policy after contract termination?**
 - a. Describe data return procedures and formats
 - b. Specify destruction methods and timelines
 - c. Provide certification of destruction
25. **Do we retain ownership of data and insights derived from our organization's use?**
 - a. Confirm customer data ownership
 - b. Clarify ownership of AI-generated insights
 - c. Address any vendor claims to derived data

Section 3: AI-Specific Additional Questions for GRC Assessment

When AI tools are assessed by GRC teams, these additional questions are necessary to understand the scope, intent, and impact of the AI solution. Organizations should adapt these questions based on their risk tolerance, technical sophistication, and the specific AI system being evaluated.

1. **Describe your process for identifying all AI tools and applications currently in use within your organization.**

- a. How do you maintain an inventory of AI components and models?
 - b. What metadata do you track? (tool purpose, data types accessed, model type, version, last update)
2. **Do you provide customers with an inventory of AI components in your solution?**
 - a. Can you identify which features use AI and which do not?
 - b. How are AI model versions tracked and communicated to customers?
3. **How do you assign ownership and accountability for AI systems?**
 - a. Who is responsible for model governance and oversight?
 - b. Who owns the underlying training data?
4. **What is your process for periodic inventory updates?**
 - a. How often is the AI inventory reviewed and updated?
 - b. How are customers notified of new AI components or changes?
5. **Can you describe how your AI model was trained, tested, and validated?**
 - a. Provide detailed model development lifecycle documentation
 - b. Describe training, validation, and test dataset characteristics
 - c. Explain validation methodology and performance metrics
6. **What data is used to train the AI model?**
 - a. Identify data sources (public datasets, proprietary data, customer data, synthetic data)
 - b. Describe data provenance and licensing
 - c. Confirm no unauthorized use of copyrighted or proprietary data
7. **How is training data quality assessed and maintained?**
 - a. Describe data quality assurance processes
 - b. Explain data cleaning, preprocessing, and augmentation methods
 - c. Identify mechanisms for detecting and correcting data quality issues
8. **Can you trace data flow from source to AI model output?**
 - a. Provide data lineage documentation
 - b. Describe data transformations and processing steps
 - c. Explain how outputs are generated from inputs
9. **What controls prevent data poisoning or adversarial inputs?**
 - a. Describe input validation and sanitization
 - b. Explain protections against adversarial attacks

- c. Detail monitoring for anomalous inputs or behaviors
10. **Is PHI or PII redacted prior to AI model training or processing?**
- a. Describe redaction or de-identification methods
 - b. Confirm PHI/PII is not used in training without explicit authorization
 - c. Explain how de-identification is validated
11. **How is synthetic data generation governed if used?**
- a. Describe synthetic data generation methods
 - b. Explain validation that synthetic data doesn't contain real patient information
 - c. Address potential for synthetic data to reveal sensitive information
12. **What steps have been taken to assess and mitigate bias in the AI model?** (e.g., race, gender, age, socioeconomic status)
- a. Provide bias assessment methodology and results
 - b. Describe fairness metrics used (demographic parity, equal opportunity, etc.)
 - c. Explain bias mitigation strategies implemented
 - d. Identify any known limitations or disparate performance across subgroups
13. **Do clinicians or users have visibility into how AI outputs are generated (explainability)?**
- a. Describe explainability features available to users
 - b. Explain level of transparency appropriate to use case criticality
 - c. Provide examples of explanations generated by the system
14. **Are there mechanisms for ensuring ongoing accuracy of data and model outputs?**
- a. Describe continuous quality monitoring
 - b. Explain feedback loops for error detection and correction
 - c. Detail retraining or model update processes
15. **What happens when AI systems fail or produce erroneous outputs?**
- a. Describe fail-safe mechanisms and degradation handling
 - b. Explain user notification of AI failures or low-confidence outputs
 - c. Detail fallback procedures when AI cannot provide reliable recommendations
16. **What is the safety impact if the AI product fails?** Select impact level: Low / Medium / High / Critical
- a. **Low:** No safety or financial impact (e.g., word prediction, autocomplete)

- b. **Medium:** Some safety/financial impact with human oversight (e.g., clinical decision support with clinician having final decision)
 - c. **High:** Significant safety/financial impact, limited human intervention (e.g., automated medication dosing, AI-driven triage)
 - d. **Critical:** Life-threatening or enterprise-critical impact, autonomous decisions (e.g., AI-controlled medical devices, fully automated clinical pathways)
17. **Has an impact analysis been conducted on patient or user safety?**
- a. Provide hazard analysis and risk assessment documentation
 - b. Describe failure mode and effects analysis (FMEA)
 - c. Identify risk mitigation strategies for identified hazards
18. **Are any AI clinical or diagnostic decisions made without human involvement?**
- a. Specify which decisions are fully automated
 - b. Describe human oversight mechanisms where they exist
 - c. Explain circumstances under which human review can be bypassed
19. **How is the AI infrastructure hardened against attacks?**
- a. Describe security architecture and defensive measures
 - b. Explain network segmentation and access controls
 - c. Detail security monitoring and intrusion detection
20. **What controls prevent model theft or reverse engineering?**
- a. Describe model protection mechanisms
 - b. Explain access controls for model parameters and weights
 - c. Detail intellectual property protections
21. **How are prompt injections and adversarial attacks mitigated?**
- a. Describe input validation and filtering for prompts (if applicable)
 - b. Explain adversarial robustness testing
 - c. Detail defenses against model manipulation attacks
22. **What monitoring exists for abnormal AI behavior?**
- a. Describe anomaly detection for AI outputs
 - b. Explain alerting for unusual model performance
 - c. Detail investigation procedures for AI anomalies
23. **What AI security standards or frameworks do you follow?**

- a. Identify standards used (NIST AI RMF, MITRE ATLAS, OWASP ML Top 10, etc.)
 - b. Describe alignment with AI security best practices
 - c. Provide evidence of framework implementation
24. **Does the AI service use any multitenant SaaS generative AI solutions?** (e.g., ChatGPT, Hugging Face, OpenAI, Anthropic, Azure OpenAI)
- a. If yes, identify all third-party AI services used
 - b. Confirm BAAs are in place with all third-party AI vendors
 - c. Describe data isolation and tenant separation mechanisms
25. **Are all AI applications validated for safety and impact in a sandbox environment before production deployment?**
- a. Describe pre-production testing environments
 - b. Explain validation procedures and acceptance criteria
 - c. Detail production deployment gates and approval processes
26. **If chatbots are utilized, do users have clear indicators or notifications that they are interacting with AI?**
- a. Describe AI disclosure mechanisms
 - b. Explain how users can distinguish AI from human interactions
 - c. Detail any regulatory compliance for AI disclosure (FTC, state laws)
27. **How are AI systems periodically or continuously monitored and evaluated for risks to accuracy, safety, or privacy?**
- a. Describe continuous monitoring approach and metrics
 - b. Explain periodic evaluation schedule and methodology
 - c. Detail triggers for revalidation or intervention
28. **Do you have an internal AI governance framework?**
- a. Describe governance structure and oversight bodies
 - b. Identify who is accountable for AI risk management
 - c. Explain escalation paths for AI-related issues
29. **Who owns model oversight, and who owns the underlying data?**
- a. Clarify organizational accountability for AI systems
 - b. Describe data stewardship and ownership policies

- c. Explain how customer data rights are protected
30. **What is the AI model development lifecycle and approval process?**
- a. Describe development methodology and quality gates
 - b. Explain approval requirements before model deployment
 - c. Detail documentation and traceability requirements
31. **How are model versions controlled, and how are rollback procedures defined?**
- a. Describe version control and change management for AI models
 - b. Explain rollback procedures and testing
 - c. Detail customer notification and approval for version changes
32. **What risk assessments have been completed for the AI system?** (e.g., failure modes, harm analysis, security risk assessment)
- a. Provide risk assessment documentation
 - b. Describe identified risks and mitigation strategies
 - c. Explain risk acceptance decisions and residual risks
33. **Do audit logs exist for all AI system processes or outputs rendered by AI systems?**
- a. Confirm comprehensive logging of AI decisions and recommendations
 - b. Describe log retention periods and accessibility
 - c. Explain how logs support traceability and accountability
34. **What is your approach to monitoring AI model drift or degradation over time?**
- a. Describe drift detection methods and metrics
 - b. Explain alerting thresholds and response procedures
 - c. Detail retraining or recalibration processes
35. **Can the AI system provide meaningful explanations for its decisions?**
- a. Describe explainability capabilities (attention maps, SHAP values, counterfactuals, etc.)
 - b. Explain how explanations are tailored to user sophistication
 - c. Provide examples of decision explanations
36. **What documentation exists for model architecture and training?**
- a. Provide model cards, technical documentation, or equivalent
 - b. Describe training datasets, hyperparameters, and performance characteristics
 - c. Explain updates to documentation when models change

37. **How are AI decision boundaries and confidence levels communicated to users?**
 - a. Describe confidence scoring and uncertainty quantification
 - b. Explain how users are informed of AI limitations
 - c. Detail recommendations for appropriate use within validated boundaries
38. **How are AI model vulnerabilities identified and remediated?**
 - a. Describe vulnerability scanning and assessment for AI components
 - b. Explain patch management for AI-specific vulnerabilities
 - c. Detail coordination with security research community
39. **What channels exist for contesting or appealing decisions made by AI systems?**
 - a. Describe human review and appeal processes
 - b. Explain how contested decisions are investigated
 - c. Detail correction procedures when AI errors are identified
40. **How does the AI system comply with FDA regulations for medical devices** (if applicable)?
 - a. Provide FDA clearance, granting, or approval documentation (510(k), De Novo, PMA)
 - b. Describe intended use and indications for use
 - c. Explain how the AI system meets FDA's regulations for medical devices (if applicable)?
 - d. Explain how the AI systems meet other non-required FDA regulations (e.g. Good Machine Learning Practices)
41. **Does the AI system also comply with international regulations** (e.g., EU AI Act, MDR, IVDR)?
 - a. Identify international regulatory clearances or certifications
 - b. Describe compliance approach for jurisdictions where deployed
 - c. Explain how system adapts to varying regulatory requirements
42. **What clinical validation has been performed for diagnostic or clinical AI?**
 - a. Provide clinical study results and publications
 - b. Describe validation methodology and performance metrics
 - c. Explain generalizability to target populations
43. **How are HIPAA requirements met in AI processing workflows?**
 - a. Describe HIPAA compliance program
 - b. Explain technical and administrative safeguards
 - c. Detail breach response and notification procedures

44. **What quality management system governs AI in clinical settings?**
 - a. Describe QMS framework (ISO 13485, FDA 21 CFR Part 820, etc.)
 - b. Explain quality assurance and control procedures
 - c. Detail continuous improvement and CAPA processes
45. **How will the system adapt to evolving AI governance requirements?**
 - a. Describe monitoring of regulatory landscape
 - b. Explain product update strategy for regulatory changes
 - c. Detail customer communication about compliance updates
46. **What impact assessments have been conducted for high-risk AI applications?**
 - a. Provide Data Protection Impact Assessment (DPIA) or Algorithm Impact Assessment (AIA)
 - b. Describe human rights impact assessment (if applicable)
 - c. Explain how assessment findings inform risk mitigation
47. **What reporting mechanisms exist for AI incidents or failures?**
 - a. Describe internal incident tracking and escalation
 - b. Explain external reporting to regulators (FDA MedWatch, etc.)
 - c. Detail transparency with customers about incidents
48. **How is the AI system maintained over time?** (patches, upgrades, model updates)
 - a. Describe update frequency and types (security patches, feature updates, model retraining)
 - b. Explain customer notification and approval processes
 - c. Detail testing and validation before updates are deployed
49. **What SLAs do you offer for system uptime and support?**
 - a. Specify uptime guarantees (e.g., 99.9%)
 - b. Describe support tiers and response times
 - c. Explain escalation procedures and business continuity plans
50. **How is staff trained to work with and supervise AI systems?**
 - a. Describe training programs provided to customers
 - b. Explain competency assessment and certification
 - c. Detail ongoing training for system updates
51. **Do you use any third-party models, APIs, or data services?**
 - a. Identify all third-party AI components and dependencies

- b. Describe integration with external AI services (cloud ML platforms, LLMs, etc.)
 - c. Explain vendor selection and vetting processes
52. **Are your third-party providers also under BAAs and compliant with healthcare regulations?**
- a. Confirm BAAs with all subcontractors handling PHI
 - b. Describe flowdown of compliance requirements
 - c. Provide evidence of subcontractor compliance
53. **How do you evaluate and monitor your own vendors for risk?**
- a. Describe vendor risk assessment processes
 - b. Explain ongoing monitoring and performance management
 - c. Detail remediation procedures for vendor issues
54. **Have you undergone HIC-SCRiM or similar supply chain risk assessments?**
- a. Provide supply chain risk assessment results
 - b. Describe supply chain security practices
 - c. Explain transparency into multi-tier supply chain
55. **How are AI service providers assessed and monitored?**
- a. Describe due diligence for AI-specific vendors
 - b. Explain technical and security assessments of AI dependencies
 - c. Detail continuous monitoring of third-party AI performance
56. **What data sharing agreements govern AI vendor relationships?**
- a. Describe contractual protections with AI subcontractors
 - b. Explain data use limitations and restrictions
 - c. Detail audit rights over third-party AI services
57. **How is vendor AI model performance and security validated?**
- a. Describe independent validation of third-party models
 - b. Explain security assessments of integrated AI components
 - c. Detail ongoing performance monitoring
58. **What contingency plans exist for AI service disruptions?**
- a. Describe fallback procedures if third-party AI services fail
 - b. Explain alternative vendors or approaches identified
 - c. Detail business continuity planning for AI dependencies

59. **Is there a designated point of contact for the third-party vendor's AI governance and incident investigation?**
- Identify AI governance contact person and role
 - Provide escalation chain for AI-related issues
 - Describe availability and responsiveness commitments
60. **How are AI ethics principles implemented and monitored?**
- Describe ethical AI framework (fairness, accountability, transparency, etc.)
 - Explain governance mechanisms for ethical oversight
 - Detail metrics and monitoring for ethical compliance
61. **What are the intended and prohibited uses of your AI application in a healthcare setting?**
- Specify validated use cases and clinical indications
 - Identify explicitly prohibited or off-label uses
 - Explain enforcement of appropriate use
62. **How is human oversight maintained in AI-influenced decision-making?**
- Describe human-in-the-loop design features
 - Explain override capabilities and procedures
 - Detail documentation of human review and decisions
63. **Have you established an AI ethics board or review process internally?**
- Describe ethics board composition and charter
 - Explain review processes for new AI applications
 - Detail decision-making authority and escalation
64. **Do you disclose whether generative or large language models (LLMs) are in use?**
- Confirm transparency about use of generative AI
 - Describe disclosure to end users and customers
 - Explain any risks specific to generative AI in your application
65. **Do you offer indemnification for harm caused by AI outputs or errors?**
- Describe indemnification provisions offered
 - Explain scope and limitations of indemnification
 - Detail insurance coverage for AI-related liability
66. **What warranties or guarantees are provided about AI system performance or compliance?**

- a. Specify performance warranties and acceptance criteria
- b. Describe compliance warranties (regulatory, privacy, security)
- c. Explain warranty limitations and exclusions

67. What is your data retention and destruction policy after contract termination?

- a. Describe data return procedures, formats, and timelines
- b. Explain secure destruction methods (NIST 800-88 compliance)
- c. Detail certification of destruction provided

68. How will liability be handled if the AI makes a harmful clinical recommendation?

- a. Describe liability allocation framework
- b. Explain shared responsibility model with customers
- c. Detail liability caps and carve-outs

69. Do we retain ownership of data and insights derived from our organization's use?

- a. Confirm customer ownership of all data and outputs
- b. Clarify ownership of custom models or configurations
- c. Explain any vendor rights to derived data or insights

Appendix H: Training Completion Checklist and Curriculum

This training program provides healthcare professionals across all organizational roles with the essential knowledge and practical skills needed to identify, assess, and mitigate AI-specific risks throughout the vendor relationship lifecycle.

The training checklist and a sample curriculum is provided below. Both assume learners have foundational understanding of what artificial intelligence is, how AI can be incorporated into third-party applications and medical devices, and the general categories of risks AI introduces to HCOs. The training focuses on practical governance frameworks, processes, tools, and techniques that enable HCOs of any size to implement effective third-party AI risk management programs. Whether you are a clinician evaluating AI tools, a procurement officer negotiating contracts, an IT professional implementing AI systems, a security analyst monitoring AI performance, a compliance officer ensuring regulatory adherence, or an executive overseeing AI strategy, this training equips you with role-appropriate knowledge to fulfill your responsibilities in protecting patients, organizational data, and operational continuity.

Primary Objective: Enable healthcare professionals to understand and implement comprehensive third-party AI risk management practices aligned with the Health Industry Third-Party AI Risk and Supply Chain Transparency Guide.

Specific Goals:

- Understand the AI third-party risk management lifecycle from use case justification through end-of-life
- Recognize unique AI risks requiring specialized assessment, contracting, and monitoring approaches
- Apply risk-based governance principles to scale oversight appropriately by AI impact classification
- Implement practical tools including GRC assessments, contract provisions, and monitoring frameworks
- Establish shared responsibility models with AI vendors throughout the system lifecycle
- Develop organizational capabilities to start or enhance AI risk management programs

Primary Audiences:

- Executive Leadership (C-suite, Board members): Strategic oversight, resource allocation, governance accountability
- Clinical Leadership (CMO, CNO, Department Chairs): Clinical validation, patient safety, use case evaluation
- Information Technology (CIO, IT Directors, System Administrators): Technical integration, testing, ongoing operations
- Information Security (CISO, Security Analysts): Risk assessment, security controls, incident response
- Privacy & Compliance (CPO, Compliance Officers): HIPAA/regulatory compliance, privacy impact assessments, BAAs
- Legal & Contracting (General Counsel, Procurement): Contract negotiation, legal protections, vendor agreements

- Vendor Management (Procurement, Third-Party Risk): Vendor evaluation, performance management, relationship oversight
- Business Owners (Department Leaders): Use case justification, implementation oversight, user training

Training Path Customization:

- **All Roles:** Complete Core Modules 1-3
- **Governance & Executive Roles:** Add Modules 4, 11
- **Procurement & Legal Roles:** Add Modules 5, 6, 7
- **Technical & Security Roles:** Add Modules 8, 9, 10
- **Clinical & Business Owner Roles:** Add Modules 4, 8, 12

Learning Objectives:

Upon completion of this training program, participants will be able to:

1. **Explain** the seven-phase AI third-party risk management lifecycle and key activities in each phase
2. **Classify** AI systems by safety impact level (Low/Medium/High/Critical) and apply risk-based governance
3. **Identify** AI-specific risks in vendor assessments including model bias, drift, supply chain dependencies, and training data issues
4. **Evaluate** AI vendors using structured GRC questionnaires addressing data lineage, security controls, governance frameworks, and regulatory compliance
5. **Negotiate** AI-specific contract provisions and enhanced BAAs protecting organizational interests
6. **Implement** testing, validation, and user training protocols for safe AI deployment
7. **Monitor** AI system performance, security, and compliance throughout operational lifecycle
8. **Respond** to AI-specific incidents using coordinated vendor response procedures
9. **Manage** AI system end-of-life transitions including data extraction, secure destruction, and replacement validation
10. **Develop** organizational implementation roadmaps for starting or enhancing AI risk management programs

- Training Completion Checklist

Use this checklist to track training program implementation and individual participant completion:

Organizational Readiness

- Training coordinator and instructors identified
- Training schedule established with dates and logistics
- Training materials customized to organizational context
- Participants identified and enrolled across all relevant roles
- Pre-training communications sent to participants
- Training venue/technology platform prepared
- Templates and tools assembled and tested
- Pre-training reading materials distributed

Core Training Modules (Check modules required for your role)

- Module 1: Foundations of Third-Party AI Risk (All roles - Required)
- Module 2: AI Risk Classification and Shared Responsibility (All roles - Required)
- Module 3: AI Risk Management Lifecycle Overview (All roles - Required)
- Module 4: Phase 0 - Use Case Justification (Governance, Clinical, Business Owners)
- Module 5: Phase 1 - Vendor Due Diligence and GRC Assessment (Procurement, Security, Compliance)
- Module 6: Phase 2 - Contract Negotiation (Legal, Procurement)
- Module 7: BAAs and AI-Specific Protections (Legal, Privacy, Compliance)
- Module 8: Phase 3 - Implementation and Training (IT, Security, Business Owners, Clinical)
- Module 9: Phase 4 - Ongoing Monitoring (IT, Security, Business Owners)
- Module 10: Phase 5 - Incident Response (Security, IT, Clinical, Compliance)
- Module 11: Phase 6 - End-of-Life Management (IT, Procurement, Business Owners)
- Module 12: Implementation Roadmap (All roles - Recommended)

Practical Exercises and Deliverables

- Use case justification document completed
- AI risk classification exercise completed with 100% accuracy

- GRC questionnaire reviewed and customized
- Vendor scorecard completed for practice scenario
- Contract clause checklist reviewed
- BAA amendment template reviewed and customized
- Implementation project plan template completed
- V&V checklist developed
- Performance monitoring dashboard designed
- Update validation protocol developed
- Incident response playbook created or updated
- EOL transition plan template completed
- Organizational implementation roadmap developed (capstone project)

Assessments

- Module knowledge check quizzes passed (80% or higher)
- Use case classification competency demonstrated
- Vendor evaluation scoring exercise completed successfully
- Contract review exercise passed
- Implementation readiness checklist validated
- Tabletop exercise participated and debriefed
- Final capstone implementation roadmap submitted

Certification

- All required modules attended
- All assessments passed
- All practical exercises completed
- Capstone project submitted and approved
- Certificate of completion received
- Added to organizational AI governance resource roster

Post-Training Application

- Participated in post-training implementation support activities
- Applied training to at least one real AI procurement or management scenario
- Shared lessons learned with peer learning community
- Provided feedback on training effectiveness
- Scheduled for annual refresher training

Organizational Implementation (Leadership/Governance Committee)

- Training completion rates tracked and reported
- Implementation roadmaps consolidated into organizational plan
- Quick wins identified and executed
- Resources allocated for ongoing AI risk management program
- Governance committee established or enhanced
- Policies and procedures updated based on training
- Progress metrics established and monitored
- Executive leadership briefed on training outcomes and implementation plan

Training Implementation Steps

Step 1: Pre-Training Preparation (2-4 weeks before training)

Activities:

- Identify training participants across all relevant organizational roles
- Customize training content to organizational context and existing AI deployments
- Gather example AI systems from your organization for case studies
- Review current AI governance policies, contracts, and processes
- Prepare training materials, templates, and tools
- Schedule training sessions with appropriate duration and spacing
- Communicate training objectives and expectations to participants
- Assign pre-training reading: Guide executive summary and relevant sections

Responsible: Training coordinator, AI governance lead, HR/Learning & Development

Step 2: Core Training Delivery (2-3 days, can be split across multiple weeks)

Day 1: Foundations and Governance (Modules 1-4)

- Morning: Modules 1-2 (AI risk fundamentals and classification)
- Afternoon: Modules 3-4 (Lifecycle overview and use case justification)

Day 2: Procurement and Contracting (Modules 5-7)

- Morning: Module 5 (Vendor due diligence and GRC assessment)
- Afternoon: Modules 6-7 (Contract negotiation and BAA provisions)

Day 3: Implementation and Operations (Modules 8-12)

- Morning: Modules 8-9 (Implementation and ongoing monitoring)
- Afternoon: Modules 10-12 (Incident response, EOL, and implementation roadmap)

Delivery Options:

- In-person intensive: 3 consecutive days
- Virtual: 6 half-day sessions over 2-3 weeks
- Hybrid: Combination of self-paced online modules and live workshops
- Modular: Role-specific tracks delivered separately

Responsible: Instructors, facilitators, subject matter experts

Step 3: Hands-On Application and Practice (Concurrent with or immediately following training)

Activities:

- Apply templates and tools to real organizational AI systems
- Conduct actual GRC assessment of current or prospective AI vendor
- Review and enhance existing AI contracts using new clause templates
- Develop or update organizational AI governance policy
- Create monitoring dashboards for existing AI systems
- Conduct tabletop exercise for AI incident response
- Begin organizational implementation roadmap development

Responsible: Training participants, functional teams, working groups

Step 4: Assessment and Certification (Within 2 weeks of training completion)

Assessment Components:

- Module knowledge check quizzes (passing score: 80%)
- Practical exercises and template completion
- Final capstone project: Organizational implementation roadmap
- Peer review and instructor feedback

Certification Requirements:

- Attendance at all required modules for role
- Passing scores on all knowledge assessments
- Completion of practical exercises
- Submission of capstone implementation roadmap

Certificate: "Health Industry Third-Party AI Risk Management Professional"

Responsible: Training coordinator, instructors

Step 5: Post-Training Implementation Support (Ongoing, 3-6 months)

Activities:

- Monthly implementation office hours for Q&A
- Quarterly check-ins on implementation progress
- Peer learning community for sharing experiences
- Access to updated templates and tools as guide evolves
- Refresher training on specific topics as needed
- Advanced training for specialized roles

Responsible: AI governance committee, training coordinator, subject matter experts

Step 6: Continuous Learning and Recertification (Annual)

Activities:

- Annual refresher training on guide updates
- New AI technology and threat landscape updates

- Regulatory changes and compliance requirements
- Lessons learned sharing from incidents and implementations
- Recertification assessment

Responsible: AI governance committee, training coordinator

- Training Curriculum

Proposed AI Training Curriculum can be seen in the modules below. Each module can be adapted to the needs of the organization.

Module 1: Foundations of Third-Party AI Risk in Healthcare (60 minutes)

Format: Instructor-led presentation with case studies

Content:

- Why traditional third-party risk management is insufficient for AI systems
- Unique characteristics of AI: model drift, training data dependencies, explainability challenges
- AI supply chain complexity: multi-tier dependencies, open-source components, cloud AI services
- Patient safety implications of AI failures, bias events, and security breaches
- Regulatory landscape: FDA medical device requirements, HIPAA considerations, emerging AI regulations
- Overview of the seven-phase AI risk management lifecycle

Learning Activities:

- Case study analysis: Real-world AI incident scenarios and lessons learned
- Group discussion: AI systems currently in use at your organization

Module 2: AI Risk Classification and Shared Responsibility (45 minutes)

Format: Instructor-led with interactive exercises

Content:

- Four-tier AI safety impact classification system (Low/Medium/High/Critical)
- Specific examples of AI systems in each category
- Risk-based governance: Scaling oversight to AI criticality
- Shared responsibility framework between healthcare organizations and vendors

- Roles and responsibilities across organizational functions (RACI matrix overview)
- When to engage governance committees and executive leadership

Learning Activities:

- Classification exercise: Categorize 10 AI use cases by safety impact
- Role-play: RACI responsibilities for AI procurement scenario

Module 3: The AI Risk Management Lifecycle Overview (90 minutes)

Format: Instructor-led presentation with workflow diagrams

Content:

- **Phase 0:** Use case justification and strategic assessment
- **Phase 1:** Due diligence and vendor evaluation
- **Phase 2:** Contract negotiation and legal protections
- **Phase 3:** Implementation, integration, and training
- **Phase 4:** Ongoing monitoring and performance management
- **Phase 5:** Incident response and recovery
- **Phase 6:** End-of-life and transition management
- Continuous activities: Inventory management, documentation, policy updates

Learning Activities:

- Lifecycle mapping exercise: Plot your organization's current AI procurement process against the framework
- Gap identification: Where does your process need enhancement?

Module 4: Phase 0 - Use Case Justification and Strategic Assessment (60 minutes)

Format: Workshop format with templates

Content:

- Pre-procurement gate-keeping: Preventing unnecessary AI adoption
- Problem definition and solution assessment
- Use case documentation requirements

- Risk classification methodology
- Stakeholder identification and engagement
- Alternatives analysis: AI vs. non-AI approaches
- Business case development: ROI and total cost of ownership

Learning Activities:

- Template completion: Use case justification document for sample AI project
- Peer review: Evaluate business cases developed by other participants

Module 5: Phase 1 - Vendor Due Diligence and GRC Assessment (120 minutes)

Format: Workshop with GRC questionnaire tools

Content:

- Standard vs. AI-specific vendor assessments
- Categories of AI-specific risks to assess:
 - Data lineage, quality, and bias
 - Model transparency and explainability
 - Security controls and AI-specific vulnerabilities
 - AI governance and risk management frameworks
 - Third-party supply chain dependencies
 - Regulatory compliance and validation
 - Ethical and responsible AI practices
- Tiered GRC questionnaire approach (Essential/Enhanced/Comprehensive)
- Vendor response evaluation and scoring
- Security risk assessment requirements
- Reference checks and validation
- Gap analysis and risk acceptance decisions

Learning Activities:

- GRC assessment practice: Evaluate mock vendor responses to AI questionnaire
- Red flags identification: Spot concerning vendor responses

Module 6: Phase 2 - Contract Negotiation and Legal Protections (90 minutes)

Format: Instructor-led with contract clause review

Content:

- Why standard software contracts fail for AI systems
- 17 essential AI commercial contract clauses:
 - Data ownership and AI training restrictions
 - Update management and change control
 - Security and compliance requirements
 - Performance standards and quality assurance
 - Model drift monitoring and revalidation
 - Incident response coordination
 - Data return and destruction
 - End-of-life notification and support
- Enhanced BAA provisions for AI processing PHI
- Negotiation strategies and priorities
- Non-negotiable terms vs. acceptable compromises
- Contract monitoring and enforcement

Learning Activities:

- Contract review exercise: Identify missing AI protections in sample agreement
- Negotiation simulation: Vendor pushback on key AI provisions

Module 7: Contracting Deep Dive - BAAs and AI-Specific Protections (60 minutes)

Format: Legal compliance workshop

Content:

- HIPAA requirements for AI processing PHI
- Standard BAA limitations for AI systems
- Nine essential AI-specific BAA provisions:
 - Prohibition on AI training with PHI
 - Permitted uses and AI purpose limitations
 - De-identification and limited data set restrictions
 - AI-specific safeguards requirements

- o Expedited breach notification for AI incidents
- o Subcontractor flowdown to AI service providers
- o Individual access and amendment rights
- o Data destruction including embedded PHI in models
- o Audit rights for AI data flows
- When to require separate AI-focused BAAs
- Privacy officer and legal counsel collaboration

Learning Activities:

- BAA gap analysis: Review your organization's standard BAA against AI requirements
- Case study: PHI exposure through AI training data

Module 8: Phase 3 - Implementation, Integration, and Training (120 minutes)

Format: Technical workshop with validation protocols

Content:

- Pre-implementation planning and project management
- Sandbox/staging environment testing requirements
- Security validation protocols
- Clinical validation and safety testing
- Verification and Validation (V&V) documentation
- Privacy impact assessments for AI
- Patient consent and disclosure considerations
- Role-specific user training programs
- Competency assessment before production access
- Phased production rollout strategies
- Documentation and asset registration
- Go-live approval criteria

Learning Activities:

- Plan validation activities for sample AI system
- Training curriculum design: Develop role-specific training for AI tool

Module 9: Phase 4 - Ongoing Monitoring and Performance Management (90 minutes)

Format: Technical workshop with monitoring tools

Content:

- AI-specific performance indicators: Accuracy, precision, recall, false positive/negative rates
- Model drift detection and thresholds
- Bias and fairness monitoring across demographic groups
- Security controls validation
- Compliance auditing requirements
- Vendor performance management and SLA tracking
- Update and patch management process
- Critical importance of post-update configuration validation
- Vendor update testing in sandbox environments
- Algorithmic transparency maintenance
- Incident and issue reporting mechanisms
- Periodic reassessment and revalidation (annual/at renewal)

Learning Activities:

- Monitoring dashboard design: Identify key metrics for sample AI system
- Update validation protocol: Develop testing checklist for AI model updates

Module 10: Phase 5 - Incident Response and Recovery (90 minutes)

Format: Tabletop exercise and planning workshop

Content:

- AI-specific incident scenarios:
 - Security breaches affecting AI systems
 - Model performance failures and degradation
 - Bias events and discriminatory outputs
 - Adversarial attacks and data poisoning
 - Model hallucinations and erroneous outputs
 - Privacy breaches involving AI-processed PHI

- Detection mechanisms for AI incidents
- Incident classification and severity assessment
- Vendor coordination protocols and communication
- Containment and mitigation strategies
- Model rollback and revalidation procedures
- Data recovery and integrity validation
- Post-incident documentation and CAPA
- Regulatory notification requirements
- Return to normal operations criteria

Learning Activities:

- Tabletop exercise: Simulated AI incident response (model failure scenario)
- Vendor coordination planning: Develop communication protocols

Module 11: Phase 6 - End-of-Life and Transition Management (60 minutes)

Format: Instructor-led with case studies

Content:

- Planned vs. unplanned EOL scenarios
- EOL notification requirements in contracts (12-18 months)
- Impact assessment and transition decisions
- Replace vs. discontinue evaluation
- Data inventory, extraction, and migration
- Secure decommissioning procedures
- Vendor data destruction certification
- Replacement system validation and user transition
- Regulatory compliance during transitions
- Post-transition review and lessons learned

Learning Activities:

- EOL planning exercise: Develop transition plan for sample AI system

- Case study: Unplanned EOL scenario (vendor model deprecation)

Module 12: Implementation Roadmap - Getting Started (90 minutes)

Format: Strategic planning workshop

Content:

- Assessing organizational AI maturity and readiness
- Securing executive buy-in and sponsorship
- Current state inventory: Discovering existing AI systems
- Rapid risk assessment of current AI deployments
- Immediate risk mitigation for high-priority gaps
- Designing future state governance framework
- Incremental implementation approach: Start small, scale up
- Measuring progress and demonstrating value
- Common implementation challenges and solutions
- Building sustainable AI risk management programs
- Continuous improvement and adaptation

Learning Activities:

- Organizational assessment: Evaluate your AI governance maturity
- Implementation roadmap development: Create 12-month action plan for your organization
- Priority setting: Identify quick wins vs. long-term initiatives

Appendix I: Quality assurance/verification/validation with AI third-party providers

Ensuring the reliability, safety, and security of third-party AI solutions requires a structured Quality Assurance, Verification, and Validation (QA/VV) framework. HCOs frequently integrate third-party AI components into their systems, yet these dependencies introduce risks when vendor-driven updates, patches, or configuration resets occur. For example, major software and AI vendors re-set certain custom configurations back to default configurations during major updates, including AI-related features, workflows, etc. These resets often undo critical security, privacy, or performance settings that organizations had deliberately customized to align with healthcare workflows and regulatory requirements. As a result, operational teams may spend significant time reconfiguring systems and validating that clinical and cybersecurity safeguards are intact. Such disruptions can not only delay care delivery but also introduce unrecognized safety or compliance risks if resets are not promptly detected. Updates and upgrades to the software containing AI features, as well as changes to the AI model itself, may occur more frequently than traditional software environments. Organizations should create as many automated processes as possible, including review, testing, and revalidation processes, so that support of these technologies can scale.

To ensure effective response, recovery, and quality assurance processes are implemented in collaboration with third-party AI vendors, the following quality assurance and validation and verification documentation requirements should be embedded into organizational policies, contracts, and evaluation frameworks. These requirements align with the NIST AI Risk Management Framework (AI RMF) and the FDA's Computer Software Assurance (CSA) guidance.

1. During Phase 1: Vendor Evaluation and Due Diligence

Purpose: To demonstrate that the AI system has been developed and validated using appropriate methodologies and that risks have been proactively identified and mitigated.

Compliance Framework alignment:

- NIST AI RMF: MEASURE (M1) – Understand and document intended purpose, risk, and impact.
- For FDA regulated devices/systems: Testing activities and supporting documentation must be risk-based and commensurate with the AI system's role in clinical or operational decision-making within the organization.

Minimum Acceptable Evidence:

- System description and intended use
- Risk classification and criticality assessment
- Verification and validation test plans and results
- Model performance baselines and thresholds

- Version history and traceability to development changes

Key Takeaway: AI incidents require coordinated response between HCO and vendors, with specific attention to model recovery, performance revalidation, and regulatory reporting. Organizations must prepare for AI-specific incident scenarios and ensure vendors are contractually obligated to provide timely support during response and recovery.

2. During Phase 4: Ongoing Monitoring and Performance

In addition to the vendor vetting completed during the evaluation stage, quality assurance, verification and validation must also be completed after each update, upgrade, or change to the system.

Purpose: To demonstrate that the update, upgrade or change to the AI system has been reviewed and tested prior to deployment using appropriate methodologies and that risks have been proactively identified and mitigated.

Compliance Framework alignment:

- NIST AI RMF: MEASURE (M1) – Understand and document intended purpose, risk, and impact.
- NIST AI RMF: MEASURE (M2) – AI systems are evaluated for trustworthy characteristics.
- NIST AI RMF: MEASURE (M3) – Mechanisms for tracking identified AI risks over time are in place.
- NIST AI RMP: MEASURE (M4) – Feedback about efficacy of measurement is gathered and assessed.
- For FDA regulated devices/systems: Testing activities and supporting documentation must be risk-based and commensurate with the AI system’s role in clinical or operational decision-making within the organization.

Review of Software Notes and Change Logs

For all changes, updates, and upgrades to the software as well as model changes made by the owner of the model, the third-party vendor should notify customers. As part of the communication, a change log or software notes should be provided detailing the changes. This change log should be used to develop installation plans and testing scenarios using the organization's standard change management process. To ensure these software notes are provided in a timely manner every time an update or upgrade is available, there should be terms and conditions requiring the documentation. In addition to the change log, the vendor should provide information on the type of testing they have conducted to validate the changes. Depending on the type of software, function the software performs, and risk tolerance of the organization, this testing may be sufficient for the organization to sign off on the change with little to no testing. If the software complexity, function, or risk tolerance is not low, then additional testing and validation is required.

Testing in Staging or Test System

For critical systems, it is best practice to have the solution available in a staging or test system so that changes, updates, and upgrades can be tested in a non-production environment. This allows testing in a low-risk manner where interfaces, data transfers, and complex interactions with other connected systems can be mimicked without the risk of impacting day-to-day operations. There is always a cost to stand up a duplicate test system in both

hardware and resources, and so the choice of whether to implement a test system depends on the risk tolerance of the organization and the complexity and risk of the solution being implemented. The amount of testing done by the vendor versus done by the organization themselves may also play a role in the decision to stand up a test system. Once testing is successful in the test system, the planning and deployment of the solution to the production system can be planned and implemented.

Re-validation of Settings Upon Update

One of the most overlooked challenges in third-party AI system management is the reversion of configurations back to vendor defaults during updates. Every update must be treated as a potential point of failure until proven otherwise. Organizations must implement a process to:

- Confirm whether updates alter pre-established configurations.
- Determine whether critical security, privacy, or model performance settings have been reset.
- Assess the downstream impact on clinical performance, user workflows, and security posture.

Validation testing should include functional testing, cybersecurity impact assessment, and model verification activities to confirm that the system continues to operate within its validated parameters. If an update modifies or resets configurations (e.g., privacy settings defaulting to broader data access, or security settings weakening authentication), the consequences may include data leakage, compromised AI output integrity, or regulatory noncompliance. Therefore, re-validation of settings after every update is non-negotiable. Because of this re-validation step, there will be a time between the update and when settings are reverted to the known “good” settings. Organizations should have a plan for this window of time and how to handle issues that arise during this time.

Shared Responsibility and Vendor Transparency

AI must be managed as a shared responsibility between vendors and HCOs. AI-enabled systems in healthcare environments cannot be managed under a “black box” model where vendors push changes without accountability. Shared responsibility requires a clear delineation of obligations between HCOs and third-party vendors, supported by contractual language, technical controls, and operational processes.

Vendors must provide proactive, detailed communication of all upcoming changes, including security patches, model updates, algorithmic adjustments, and potential configuration resets. This communication should occur with sufficient lead time to allow internal security, quality, and clinical teams to prepare, test, and validate changes before they are deployed into production. Silent updates or inadequate release notes create unacceptable risk in regulated environments where unverified modifications can directly affect patient safety and regulatory compliance.

HCOs should establish vendor transparency requirements as part of supplier agreements, including:

- Advance notice periods for all updates, with a minimum threshold (e.g., 30 days for non-critical updates, 72 hours for urgent security patches).
- Full release documentation describing not only the technical content of the update but also its impact on default settings, integrations, and data handling practices.
- Explicit disclosure of known limitations or risks, including any loss of functionality or security safeguards introduced by the update.

- Availability of validation evidence, showing that the vendor has tested updates in conditions that approximate the healthcare environment, with emphasis on security, robustness, and safety.
- Point-of-contact accountability, requiring vendors to designate a responsible individual or team who can address questions, provide clarification, and support issue resolution during the update process.

Responsibility must also extend to incident management. If an update introduces an unintended failure mode, such as resetting privacy controls, breaking an integration, or degrading AI performance, the vendor must commit to timely remediation, root cause analysis, and support for corrective and preventive actions (CAPA). Vendors should not treat updates as a unilateral activity but as part of a collaborative lifecycle with their healthcare customers.

Ultimately, transparency is not optional in regulated environments. Vendors who fail to provide sufficient detail or who deliver undocumented changes increase the likelihood of system downtime, degraded AI performance, and regulatory noncompliance. Shared responsibility ensures that HCOs retain visibility and control, while vendors remain accountable for the safety and reliability of their products.

QA/VV for Third-Party AI Systems

In summary, quality assurance and verification/validation should have the following minimum requirements to ensure success:

- Require detailed vendor release notes: Vendors must provide comprehensive documentation describing updated content, configuration changes, and potential risks before deployment.
- Mandate change management processes: All updates must be reviewed by operational, security, and clinical teams through a formal change management pathway.
- Demand validation evidence: Vendors should supply verification and validation documentation aligned with the intended use, safety classification, and criticality of the AI system.
- Validation documentation must cover:
 - Baseline functional verification and performance thresholds.
 - Security testing results (e.g., adversarial robustness, input manipulation, model inversion).
 - Post-deployment monitoring and drift detection mechanisms.
 - Documentation of human-in-the-loop or override capabilities where applicable.
- Vendor-supported revalidation: AI vendors must actively participate in on-site or operational revalidation, particularly following model updates, integrations, or data migrations.
- Data recovery and retraining support: Vendors should assist in restoring model accuracy, ensuring clinical reliability, and maintaining regulatory compliance in case of data corruption, drift, or loss.

Appendix J: References

HSCC Guidance

- [1] Health Sector Coordinating Council Cybersecurity Working Group (HSCC-CWG). *AI Third-Party and Supply Chain Transparency Guide*. Washington, DC: HSCC, 2025. Available at:
- [2] HSCC-CWG. *AI Cyber Governance Framework Implementation Guide*. Washington, DC: HSCC, 2025.
- [3] HSCC-CWG. *AI Cyber Operations & Defense Playbook*. Washington, DC: HSCC, 2025.
- [4] HSCC-CWG. *AI Education & Enablement Guide*. Washington, DC: HSCC, 2025.

FDA Guidance & Publications

- [6] U.S. Food and Drug Administration (FDA). *Artificial Intelligence and Machine Learning (AI/ML) Software as a Medical Device (SaMD) Action Plan*. Silver Spring, MD: FDA, 2021.
- [7] FDA. *Good Machine Learning Practice (GMLP) for Medical Device Development: Guiding Principles*. FDA/Health Canada/MHRA Joint Document, 2021.
- [8] FDA. *Proposed Regulatory Framework for Modifications to Artificial Intelligence/Machine Learning-Based Software as a Medical Device (SaMD)*. Discussion Paper, 2019.
- [9] FDA. *Predetermined Change Control Plan (PCCP) for AI/ML-Based Software as a Medical Device*. Draft Guidance, 2023.
- [10] FDA. *Cybersecurity in Medical Devices: Quality System Considerations and Content of Premarket Submissions*. Final Guidance, 2023.

NIST Frameworks & Standards

- [11] National Institute of Standards and Technology (NIST). *AI Risk Management Framework (AI RMF) 1.0*. Gaithersburg, MD: NIST, 2023.
- [12] NIST. *Cybersecurity Framework (CSF) 2.0*. Gaithersburg, MD: NIST, 2024.
- [13] NIST. *Special Publication 800-53 Rev. 5: Security and Privacy Controls for Information Systems and Organizations*. Gaithersburg, MD: NIST, 2020.
- [14] NIST. *Special Publication 800-218: Secure Software Development Framework (SSDF)*. Gaithersburg, MD: NIST, 2022.
- [15] NIST. *Special Publication 1270: Towards a Standard for Identifying and Managing Bias in Artificial Intelligence*. Gaithersburg, MD: NIST, 2022.

MITRE & Adversarial AI Resources

- [16] MITRE. *ATT&CK for Enterprise Framework*. Bedford, MA: MITRE, 2023. Available at: <https://attack.mitre.org>
- [17] MITRE. *Adversarial Threat Landscape for Artificial-Intelligence Systems (ATLAS)*. Bedford, MA: MITRE, 2023. Available at: <https://atlas.mitre.org>

International Standards & Guidance

[18] International Medical Device Regulators Forum (IMDRF). *Machine Learning-enabled Medical Devices: Key Terms and Definitions*. IMDRF/SaMD WG, 2023.

[19] International Medical Device Regulators Forum (IMDRF). *Proposed Framework for AI/ML-enabled Medical Devices: Regulatory Considerations*. IMDRF/SaMD WG, 2024.

[20] International Organization for Standardization (ISO). *ISO/IEC 23894:2023 Information technology – Artificial intelligence – Guidance on risk management*. Geneva: ISO, 2023.

[21] ISO/IEC JTC 1/SC 42. *Artificial Intelligence Standards Roadmap*. Geneva: ISO, 2022.