Standard Policy

Publication Date: October 01, 2025
Applicability: All employees, contractors, and affiliates.
Status: For Internal Use Only

A. Scope and Summary

- This Standard establishes the framework and standards for the development, deployment, and use of all Artificial Intelligence (AI), Machine Learning (ML), Generative AI, and Agentic AI systems across the Firm.
- The primary objective is to ensure that the Firm's use of AI is responsible, trustworthy, secure, and compliant with HIPAA and all applicable healthcare, privacy, and data protection laws. This includes strict adherence to the Health Insurance Portability and Accountability Act (HIPAA) for systems processing Protected Health Information (PHI).
- This Standard is designed to align with external standards including the NIST AI Risk Management Framework (RMF), the EU AI Act, and the HIPAA Security and Privacy Rules, by establishing a risk-based approach to governance, management, and control.
- All AI systems that handle PHI must ensure confidentiality, integrity, and availability of health data through encryption, access control, auditability, and human-in-the-loop verification, as outlined in the Firm's Trustworthy AI in Healthcare Claims Processing Framework.

B. Core Principles (The "Govern" Function)

All AI systems at the Firm must be designed, developed, and deployed in alignment with the NIST RMF and HIPAA Privacy & Security principles of trustworthy AI:

- **Accountability and Transparency:**
  All AI systems must have a clear Business Sponsor and Primary Accountable Owner. Decision-making processes involving AI—especially those impacting patient care or claim reimbursement—must be explainable and logged for audit purposes. All AI actions must be traceable to either the human operator or the autonomous agent identity.
- **Fairness and Bias Mitigation:**
  The Firm prohibits AI systems from profiling individuals based on race, ethnicity, or medical conditions. AI must not use or infer sensitive attributes in decision-making. Controls are in place to identify and mitigate data or model bias during the design and training stages.
- **Reliability and Safety:**
  AI systems must be reliable, clinically safe, and robust. A "Human-in-the-Loop"

(HITL) is mandated for high-risk or healthcare-impacting use cases. Model and data monitoring must detect and mitigate drift, false positives in fraud detection, or errors that could affect claim outcomes.

- **Privacy and Security:**
  AI systems handling PHI must comply with HIPAA's Privacy, Security, and Breach Notification Rules. All PHI must be encrypted in transit and at rest using approved cryptographic standards. Access is restricted via Role-Based Access Control (RBAC) and Multi-Factor Authentication (MFA). Systems must implement Zero Trust Architecture (ZTA) and agent sandboxing to ensure isolation and minimize data leakage.

C. Governance and Roles

A multi-layered governance structure ensures alignment with both HIPAA and firmwide AI standards:

- **AI Risk and Controls Council:**
  Responsible for approving this AI Standard, assessing high-risk healthcare AI use cases (including PHI processing), and ensuring HIPAA compliance controls (encryption, masking, access logging) are implemented.
- **AI Risk Management Group:**
  Maintains a comprehensive AI inventory including model versioning, ownership, HIPAA compliance status, and validation logs. Performs periodic reviews for drift, bias, and compliance with PHI handling standards.
- **AI Risk Assessment Group:**
  Reviews and approves low-risk healthcare use cases (Category 3), ensuring compliance with HIPAA's minimum necessary rule and verifying that only de-identified or masked data is used in test environments.
- **AI Metrics and Incidents Group:**
  Tracks AI-related Key Risk Indicators (KRIs) including "Risk of non-compliance with HIPAA," model drift events, or data leakage incidents. Reviews root-cause analyses and ensures incident closure through documented mitigation.
- **Business Sponsors:**
  Serve as accountable owners of AI systems, ensuring HIPAA-mandated controls are applied, including consent management, PHI encryption, secure audit logging, and breach notification preparedness.
- **AI Users:**
  All employees and contractors must complete HIPAA and AI compliance training before using AI tools. They must use AI strictly within authorized scope and report any PHI exposure, hallucination, or data misuse immediately.

D. AI Risk Management Framework ("Map, Measure, Manage")

The Firm adopts a risk-based categorization methodology aligned with both the **EU AI Act** and **HIPAA** to ensure proportional oversight and controls for healthcare-related systems.

Category	**Risk Level**	**Description & Examples**	**Approval Requirement**
Category 1	**High-Risk / HIPAA-Protected Use**	Includes AI systems managing or processing PHI, such as healthcare claim adjudication, fraud detection, or medical code validation. Prohibited uses include patient profiling or discriminatory automated decisions.	Requires Enterprise Risk Committee (ERC) and AI Risk and Controls Council approval, with HIPAA security validation.
Category 2	**Significant-Risk**	Generative or Agentic AI used for healthcare administration, claims processing, or fraud identification involving non-anonymized data.	Requires AI Risk and Controls Council review and PHI protection confirmation.
Category 3	**Low-Risk / Limited-Ris k**	Systems that use de-identified data or anonymized claim sets in a controlled environment (e.g., sandboxed testing).	Requires AI Risk Assessment Group confirmation.
Category 4	**Minimal-Ris k**	Non-PHI AI tools for internal use, analysis, or reporting.	Registration in AI Model Inventory.

E. Standard Control Requirements

1. General Controls (All Categories)

- **AI Inventory:** All AI systems must maintain metadata including ownership, PHI use, encryption status, and HIPAA compliance review date.
- **Ongoing Monitoring:** AI systems must be monitored for model drift, anomalies, and non-compliance with HIPAA data retention or deletion rules.
- **Change Management:** All model updates must undergo Secure SDLC review and HIPAA impact assessment before production deployment.

2. Generative AI Controls (Categories 1, 2, 3)

- **Data Masking & Classification:** All PHI must be classified and masked in testing environments.
- **Encryption:** Data-at-Rest and Data-in-Transit encryption using AES-256 and TLS 1.3 must be enforced.

- **Human-in-the-Loop (HITL):** Required for final decision-making in healthcare or claims adjudication workflows.
- **Explainability:** Generative outputs must include rationale for claim approvals/denials for auditing and patient clarity.

3. Agentic AI Systems Controls (Categories 1 & 2)

Agentic AI systems, including healthcare claims processing agents, require HIPAA-specific safeguards:

- **Identity and Authentication:** Each agent must have a unique identity. All agent actions must be logged and attributable.
- **Access Control:** RBAC ensures agents only access the minimal data required to complete assigned healthcare tasks.
- **Guardrails and Sandboxing:** AI input/output filters must block prompt injections and PHI exposure.
- **Audit Logging:** Every claim adjudication or PHI interaction must be time stamped and retained for seven years per HIPAA.
- **Incident Management:** AI-specific incident response procedures must handle PHI-related anomalies and model mispredictions.

F. Reporting, Escalations, and Training

Incident Reporting:
All AI users must report potential PHI breaches, hallucinations, or misjudged claim outputs immediately. Reports must be escalated to the Privacy Officer and AI Risk and Controls Council.

Training and Awareness:
All employees must complete:

- Annual HIPAA Certification Training
- Firmwide Cyber Essentials and AI Security Training
- Specialized module on “Trustworthy AI for Healthcare Claims” before system access

Audits and Compliance Checks:
Quarterly audits validate HIPAA adherence, encryption standards, and AI performance KPIs (e.g., reducing claim processing time from 15 days to 2 days, improving fraud detection from 95% to 97%).

G. Continuous Improvement and Optimization

The Firm’s AI environment operates in a continuous HIPAA-compliant lifecycle:

1. **AI Readiness:** Establish policy and accountability for PHI handling.
2. **AI Pilot:** Mask PHI, test securely in sandboxed environments.
3. **AI Integration:** Deploy with full guardrails, access controls, and encryption.
4. **AI Optimization:** Monitor drift, retrain models, and update the AI Risk Register.

All improvements must be documented under Traceability & Artifact Management to ensure auditability and compliance.

H. Key Control Indicators (HIPAA-Aligned Framework)

Pilot Stage

Description	HIPAA Control Area	AI Lifecycle Stage
Maintain inventory of all AI agents, models, and systems, capturing versioning, ownership, risk profiles, and approval scope.	Administrative Safeguards, Security Management Process/Asset Inventory	AI Infrastructure Setup
Assign Primary/Secondary Owners, Business Sponsors, and define clear escalation paths for every AI model to ensure accountability.	Administrative Safeguards, Assigned Security Responsibility	AI Infrastructure Setup
Use a formal mechanism to classify all data (sensitive, financial, health) ensuring appropriate handling controls are applied.	Administrative Safeguards, Information Access Management	AI Data Pipeline & Labeling
Explicitly prohibit AI use with sensitive data for discriminatory decision-making, social scoring, or human profiling purposes.	Administrative Safeguards, Sanction Policy/Information Access	AI Enterprise Governance
Establish unique identities for all AI agents and use mechanisms (e.g., tokens) to attribute all actions to the originator.	Technical Safeguards, Access Control/Audit Control	AI Deployment
Implement deterministic and non-deterministic guardrails to detect and block threats like prompt injection and sensitive data exfiltration.	Technical Safeguards, Security Management Process/Risk Analysis	AI Deployment

Define and track AI-specific KPIs, observability metrics, and explainability methods to mitigate "black box" risks.	Administrative Safeguards, Security Management Process/Evaluation	AI Monitoring
Define AI-specific Key Performance Indicators (KPIs) to establish baselines and measure the efficiency of model infrastructure and setup.	Administrative Safeguards, Quality/Efficiency Metric	AI Infrastructure Setup
Define AI-specific Key Performance Indicators (KPIs) to measure the effectiveness and accuracy targets of model optimization efforts.	Administrative Safeguards, Quality/Efficiency Metric	AI Validation & Testing
Document the current process (e.g., claims processing) including time, resources, and cost to benchmark AI automation ROI.	Administrative Safeguards, Risk Management/Contingency Plan	AI Infrastructure Setup
Explicitly prohibit AI use with sensitive data for discriminatory decision-making, social scoring, or human profiling purposes.	Administrative Safeguards, Information Access Management	AI Development
Define Key Control Indicators (KCIs) to measure organizational readiness, policy completion, and governance structure status.	Administrative Safeguards, Security Management Process/Policy Review	AI Enterprise Governance
Define Key Control Indicators (KCIs) to validate data classification, access controls, and policy compliance within the AI pilot phase.	Technical Safeguards, Access Control/Data Integrity	AI Data Pipeline & Labeling (Specifically "AI Pilot" sub-stage)
Evaluate the performance of the Adjudication Agent in accurately checking and applying complex medical or financial codes.	Administrative Safeguards, Evaluation/Quality Metric	AI Validation & Testing
How are unique, manageable identities established for all AI agents, and what mechanisms (e.g., OAuth on-behalf-of tokens) used to attribute all actions to either the originating user or the autonomous agent?	Administrative Safeguards, Workforce Security/Assigned Security Responsibility	AI Development

Production Stage

Description	HIPAA Control Area	AI Lifecycle Stage
Maintain inventory of all AI agents, models, and systems, capturing versioning, ownership, risk profiles, and approval scope.	Administrative Safeguards, Security Management Process/Asset Inventory	AI Infrastructure Setup
Assign Primary/Secondary Owners, Business Sponsors, and define clear escalation paths for every AI model to ensure accountability.	Administrative Safeguards, Assigned Security Responsibility	AI Infrastructure Setup
Use a formal mechanism to classify all data (sensitive, financial, health) ensuring appropriate handling controls are applied.	Administrative Safeguards, Information Access Management	AI Data Pipeline & Labeling
Explicitly prohibit AI use with sensitive data for discriminatory decision-making, social scoring, or human profiling purposes.	Administrative Safeguards, Sanction Policy/Information Access	AI Enterprise Governance
Establish unique identities for all AI agents and use mechanisms (e.g., tokens) to attribute all actions to the originator.	Technical Safeguards, Person/Entity Authentication	AI Deployment
Define, enforce, and continuously monitor secure baseline configurations for all AI platforms, runtimes, and infrastructure.	Technical Safeguards, Integrity Controls/Secure Infrastructure	AI Infrastructure Setup
Enforce strong authentication mechanisms (e.g., MFA) for all access to AI systems and their components to mitigate impersonation.	Technical Safeguards, Person/Entity Authentication	AI Deployment
Implement and enforce Role-Based Access Control (RBAC) based on the principle of least privilege for all identities.	Technical Safeguards, Access Control	AI Deployment
Enforce strong, cryptographic protocols (e.g., TLS 1.3) to encrypt all data in transit (API calls, inference requests).	Technical Safeguards, Transmission Security	AI Deployment

Implement deterministic and non-deterministic guardrails to detect and block threats like prompt injection and sensitive data exfiltration.	Technical Safeguards, Integrity Controls/Audit Controls	AI Deployment
Continuously monitor datasets and predictions to track drift, anomalies, bias, and performance degradation for timely response.	Administrative Safeguards, Information System Activity Review	AI Monitoring
Implement robust Human-in-the-Loop controls for high-risk decisions, including user consent, review, override, and escalation mechanisms.	Administrative Safeguards, Information Access Management/Contingency Plan	AI Monitoring
Define and track AI-specific KPIs, observability metrics, and explainability methods to mitigate "black box" risks.	Administrative Safeguards, Information System Activity Review/Evaluation	AI Monitoring
Identify, document, and ensure compliance with all applicable data privacy and regulatory requirements (e.g., GDPR, EU AI Act).	Administrative Safeguards, Security Management Process/Evaluation	AI Enterprise Governance
Establish due diligence and legal contract protections to manage third-party AI risk against data leakage and unauthorized access.	Administrative Safeguards, Business Associate Contracts	AI Infrastructure Setup
Define AI-specific Key Performance Indicators (KPIs) to establish baselines and measure the efficiency of model infrastructure and setup.	Administrative Safeguards, Quality/Efficiency Metric	AI Infrastructure Setup
Define AI-specific Key Performance Indicators (KPIs) to measure the effectiveness and accuracy targets of model optimization efforts.	Administrative Safeguards, Quality/Efficiency Metric	AI Validation & Testing
Document the current process (e.g., claims processing) including time, resources, and cost to benchmark AI automation ROI.	Administrative Safeguards, Risk Management/Contingency Plan	AI Infrastructure Setup
Define the key ethical and legal concept (e.g., PHI Trust) that serves as a mandate for data handling policies.	Administrative Safeguards, Policy/Security Management Process	AI Enterprise Governance
Explicitly prohibit AI use with sensitive data for discriminatory decision-making, social scoring, or human profiling purposes.	Administrative Safeguards, Policy/Sanction Policy	AI Development

Define Key Control Indicators (KCIs) to validate data classification, access controls, and policy compliance within the AI pilot phase.	Technical Safeguards, Integrity Controls	AI Data Pipeline & Labeling
Define Key Control Indicators (KCIs) to measure security, identity, and access controls during the final integration of the AI model.	Technical Safeguards, Integrity Controls/Audit Controls	AI Deployment
Continuously monitor datasets and model predictions to detect drift, anomalies, bias, and performance degradation.	Administrative Safeguards, Information System Activity Review	AI Monitoring
Evaluate the performance of the Adjudication Agent in accurately checking and applying complex medical or financial codes.	Administrative Safeguards, Evaluation/Quality Metric	AI Validation & Testing
How are unique, manageable identities established for all AI agents, and what mechanisms (e.g., OAuth on-behalf-of tokens) used to attribute all actions to either the originating user or the autonomous agent?	Administrative Safeguards, Workforce Security/Contingency Plan	AI Development