# Document Metadata

author: jfraze@mycomp.org

title: AI and PHI Boundary Controls

nist\_function: Manage

priority\_phase: Must

last\_reviewed: 2025-06-24

status: draft

## **Purpose**

To prevent inappropriate access, use, storage, or leakage of Protected Health Information (PHI) in generative AI systems used in hospital environments.

## **Scope**

Applies to all GenAI systems that interface with, infer from, or generate content related to PHI, including embedded LLMs, documentation tools, summarizers, and chatbots.

## **Policy Statement**

All GenAI models must implement strict PHI boundaries at the input, processing, and output stages. PHI exposure through training data, prompt memory, or inferred responses must be minimized and monitored.

## **Roles and Responsibilities**

• IT Security: Enforce access controls and prompt PHI boundary inspection.

• Model Developers: Implement redaction, token blocking, and context filters.

• Privacy Officers: Validate that PHI use complies with HIPAA, HITECH, and ONC interoperability rules.

## **Implementation Phases**

### **Must Do**

• Enforce opt-out for PHI storage in hosted/inference interfaces.

• Apply prompt filtering and output redaction for PHI signatures (MANAGE-2.2).

• Block context carryover across sessions unless explicitly authorized (SP 800-218A §3.1.2).

### **Should Do**

• Conduct PHI leakage red-teaming and inference attacks.

• Maintain PHI-aware audit logs, version-controlled prompts, and patient traceability.

• Require vendors to attest to zero data retention in APIs unless contracted otherwise.

### **Recommended**

• Use automated PHI detection classifiers on logs and outputs.

• Integrate with DLP (Data Loss Prevention) systems for real-time alerting.

• Apply differential privacy techniques for GenAI models trained on mixed datasets.