CASE STUDY

UNLOCKING HANDWRITTEN CLINICAL DATA FOR AGENTIC AI TESTING WITH UDA-REDACT

Client Overview

A U.S. Fortune 100 health insurance and healthcare services organization operates an AI-powered assistance platform for its member-support teams.

The platform helps member advocates handle complex healthcare inquiries by summarizing conversations in real time, retrieving relevant benefit and eligibility information, and presenting compliance guidance during member interactions. Advocates remain responsible for reviewing the information and making final decisions.

To improve the platform, the organization needed realistic clinical documents for testing its AI agents and the retrieval systems supporting them.

Challenge: Sensitive Handwritten Data Could Not Be Used for AI Testing

Many of the organization's clinical documents contained notes and other information written by physicians. These documents provided valuable test data because they represented the conditions encountered in real healthcare workflows, including handwriting, abbreviations, mixed typed and handwritten content, incomplete sentences, corrections, and variations in document structure.

However, the documents also contained protected health information (PHI), including member names, dates of birth, addresses, identification numbers, contact information, record numbers, and other sensitive details.

The organization could not place unredacted documents into non-production testing environments or ingest them into the vector store used by its AI platform. Manually redacting large document collections was slow, inconsistent, and difficult to validate.

Replacing the documents with entirely synthetic text was also insufficient. Synthetic documents did not fully represent the handwriting, clinical terminology, layout variation, and document complexity the AI agents would encounter in real workflows.

The organization needed a way to preserve the useful characteristics of authentic clinical documents while reliably removing sensitive member information.

The Turning Point: Creating a Privacy-Safe Clinical Document Corpus

The organization adopted UDA-Redact to de-identify clinical documents containing both typed and handwritten information.

UDA-Redact processes documents inside the organization's controlled environment and identifies sensitive information wherever it appears. This includes PHI contained in typed fields, handwritten physician notes, annotations, and other unstructured areas of a document.

Sensitive values are removed while the approved clinical context, handwriting characteristics, terminology, and overall document structure are preserved. This transforms restricted production documents into privacy-safe testing assets that retain the complexity required for meaningful AI evaluation.

Following validation, the redacted documents can be ingested into the organization's controlled vector store and used in agentic test workflows.

Results: From Restricted Documents to a Reusable AI Evaluation Corpus

With UDA-Redact, the organization was able to:

- Convert sensitive clinical documents into a reusable, privacy-safe testing corpus.
- Preserve realistic physician handwriting, abbreviations, annotations, and mixed-content layouts.
- Remove member-identifying information before documents entered AI development and testing environments.
- Store approved redacted documents in the vector store supporting the AI platform.
- Test document ingestion, text extraction, chunking, embedding, indexing, and semantic retrieval using realistic clinical content.
- Evaluate whether AI agents retrieved the correct document and supporting passage for a given test question.
- Build grounded test cases with known source documents, expected facts, and expected retrieval results.
- Test how agents interpreted and summarized information contained in handwritten physician notes.
- Validate agent behavior when handwriting was incomplete, ambiguous, corrected, or difficult to interpret.
- Determine whether agents cited or surfaced the appropriate source information rather than generating unsupported answers.
- Test appropriate abstention and escalation behavior when the retrieved evidence was missing, insufficient, or unclear.
- Evaluate compliance guidance and privacy controls without exposing real member identities.
- Reproduce failed retrieval and reasoning scenarios for investigation and regression testing.
- Reuse the same validated corpus across model, prompt, embedding, and vector-index updates.
- Give development and quality teams access to realistic clinical documents without providing access to raw PHI.

How the Redacted Data Supports the AI Platform

The privacy-safe document corpus enables the organization to test the complete agentic workflow:

1. A representative member-support question is submitted to the AI agent.
2. The agent searches the vector store for relevant clinical documentation.
3. The retrieval system identifies the most relevant document chunks.
4. The agent interprets the retrieved typed and handwritten content.
5. The agent summarizes the relevant information and presents it to the member advocate.
6. The system applies compliance guidance and determines whether human review or escalation is required.
7. The response is compared with the expected document, facts, and outcome.

This allows the organization to evaluate more than whether the agent produced a plausible answer. Testing can determine whether the agent found the correct evidence, interpreted it accurately, remained grounded in approved source material, and escalated appropriately when the evidence was uncertain.

Conclusion and Next Steps

The organization currently uses UDA-Redact to transform sensitive clinical documents containing physician handwriting into safe, reusable assets for agentic AI testing.

The redacted data can be stored in the organization's vector store and used to evaluate retrieval, reasoning, summarization, grounding, compliance, and human-escalation behavior. This enables more realistic testing than fully synthetic clinical text while preventing raw member information from entering AI development environments.

Future priorities include:

- Expanding the redacted corpus to additional clinical-document types and handwriting styles.
- Building larger golden datasets with defined questions, expected source documents, and expected answers.
- Measuring retrieval accuracy across different chunking, embedding, and indexing strategies.
- Testing combinations of handwritten notes, typed fields, scanned documents, and supporting attachments.
- Automating agent regression testing whenever models, prompts, retrieval logic, or vector indexes change.
- Extending the same privacy-safe data foundation to additional member-support and clinical-document workflows.