

Case Study

Engineering AI-Powered Legal Document Discovery Across 85M+ Enterprise Records

Client Overview

A large UK enterprise needed to transform how its legal teams discovered historical property and legal documents across a complex legacy archive.

Legal and property teams depended on a vast archive of historical documentation distributed across legacy cloud storage. With inconsistent metadata, duplicate records and hundreds of millions of files spanning multiple formats, finding relevant documents was complex and time-consuming.

Exponentia.ai engineered a metadata-first, AI-enriched document discovery platform designed not simply to answer questions, but to provide reliable, explainable and scalable retrieval across the enterprise archive.

The Challenge?

- The archive presented a combination of data, retrieval and operational challenges:

~174.6M	~285.5M	520.3 TB
Duplicate records identified	Inventory records across eight storage accounts	Archive estate

- The estate included PDFs, Office documents, emails, ZIP archives, CAD/DWG files and non-text assets.
- Legal teams needed the ability to identify the right historical document, understand why it matched their search and retrieve the original source file.
- A semantic-search-only approach was insufficient because documents needed to remain discoverable even when their contents could not be completely extracted.

- The existing solution incurred high amount of cost for file storage on cloud. Also, the search method was manual and takes up to several weeks to search for relevant files.

Solution Delivered

Exponentia.ai engineered an enterprise document intelligence platform around four principles.



Inventory Before Intelligence

Source paths, metadata, file types and identifiers were standardized and linked to original documents.



Selective Processing

Duplicate detection, file classification and format-specific extraction allowed processing approaches to be matched to each document type.



Hybrid Retrieval

Three complementary search strategies supported discovery:

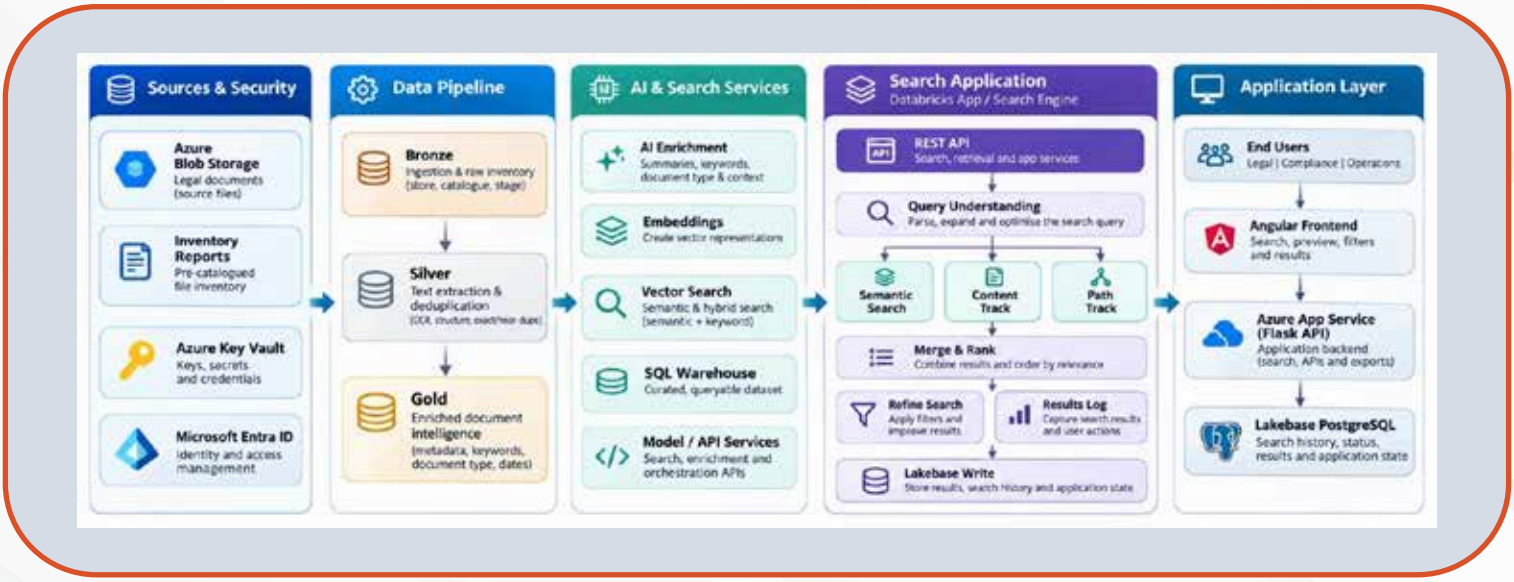
- Path and metadata
- Extracted and enriched content
- Semantic retrieval through Databricks Vector Search



Search to Evidence

Deterministic ranking and match explanations helped users understand results. Large searches and exports ran asynchronously, while selected original files could be packaged and delivered through SharePoint.

The architecture combines enterprise storage, Medallion data engineering, AI enrichment, hybrid retrieval, ranked search, application workflows and original-document export.



The Impact

37.16M
Files
text-extracted

~36.75M
Documents
AI-enriched

3
Retrieval strategies across path, content and semantics

55–95 seconds
Representative primary-search performance

21–52 seconds
Representative refine-search performance

10,000
Ranked rows available to the user interface while full result sets remain preserved

Example user journey

- A representative search for “Victoria Hall” returned 889,426 files in approximately 55 seconds.
- A refinement for “fire safety” reduced the result set to 24,451 files in approximately 27 seconds.
- Users could then select original evidence and trigger a controlled export workflow through Databricks and SharePoint.

Outcome

The significance of the solution extends beyond search.

Reliable enterprise AI depends on the engineering surrounding the model: metadata, pipelines, identity, processing, ranking, application workflows, security, observability and operational ownership.

By bringing these components together, Exponentia.ai helped turn fragmented enterprise content into a production-ready intelligence and discovery capability.



Interested in a customized solution designed to meet your unique requirements and achieve comparable results?

At Exponentia.ai, we partner with businesses to address complex data challenges and build trust in enterprise data. Our approach helps organizations streamline data management, improve visibility, and empower teams with reliable insights for better decision-making.

Engage with Us

www.exponentia.ai

engage@exponentia.ai

US UK UAE IND SGP