GWC DATA.AI
SOLUTION MATTERS

Next-Gen Knowledge Engine on Azure

Transforming Enterprise Knowledge Access with AI-Powered Retrieval

Azure OpenAI | RAG Architecture | 70% Faster Retrieval

Problem Statement

A leading manufacturing enterprise managed over **80,000+ documents** in multiple formats—PDFs, Word, Excel, and image-based records. However, retrieving accurate, context-aware information was time-consuming and inconsistent.

Ineffective Search
Keyword-based search incapable of semantic understanding

High Manual Effort
Time-consuming reviews leading to delays and inconsistency

Fragmented Data
Lacking centralized access and integration across sources

Business Impact

- Employees spent excessive time locating the right information
- Decision-making was delayed and inconsistent
- Productivity and efficiency were negatively impacted

Current State & Operational Gaps

Analysis revealed critical inefficiencies in the existing knowledge management system

- Document repositories fragmented across multiple sources
- Keyword-based search with no contextual understanding
- Manual data ingestion and content extraction
- Lack of unified access and centralized retrieval system
- Limited visibility into usage, access, and performance metric
- Security compliance risks due to non-uniform governance

The Solution

Azure-Powered RAG Knowledge Retrieval Framework

We deployed a Retrieval-Augmented Generation (RAG) framework leveraging Azure OpenAI, Azure AI Search, and Azure-native services to transform enterprise knowledge management. The solution delivers precise, context-driven answers while ensuring data security, compliance, and scalability.

Strategic Requirements

Knowledge Accessibility
- Build a secure, scalable, AI-driven RAG system on Microsoft Azure
- Enable context-aware, natural language querying across all document types

Automation & Intelligence
- Automate document ingestion, embedding generation and vector indexing
- Deliver accurate, contextually relevant answers through generative AI

Security & Compliance
- Ensure enterprise-grade security using Azure AD, Private Link, and Key Vault
- Maintain compliance via Microsoft Defender for Cloud and Azure Policy

Scalability & Monitoring
- Deploy modular, containerized microservices for scalability
- Enable real-time system monitoring and proactive alerting through Azure Monitor

Architecture Overview

A comprehensive multi-layer architecture built on Azure-native services

Data Ingestion Layer
Automated text and metadata extraction using Unstructured.io for PDFs, Word, Excel, and image files

Embedding Pipeline
Azure OpenAI generates vector embeddings capturing semantic meaning

Vector Indexing & Retrieval
Azure AI Search powers contextual search and vector-based retrieval

Generative Response Layer
Azure OpenAI GPT synthesizes human- readable, contextually relevant responses

Application Layer
FastAPI backend containerized with Docker, deployed on AKS. React JS frontend with Azure AD integration

Security & Monitoring
Azure Key Vault, Private Link, Defender for Cloud ensure secure operations with real-time monitoring

Core Solution Highlights

- Automated Data Ingestion: Parsed PDFs, Word, Excel, images with Unstructured.io
- AI-Generated Answers: GPT-based synthesis from retrieved enterprise documents
- Secure Enterprise Access: Azure AD-based authentication and SSO
- Contextual Semantic Search: Azure AI Search powered by OpenAI embeddings
- Scalable Infrastructure: Modular architecture via Docker and AKS
- Continuous Monitoring: Azure Monitor dashboards for real-time visibility

Technology Stack

Enterprise-grade technologies and Azure-native services

Core Technologies

Azure OpenAI
Embedding generation and generative response

LangChain
RAG orchestration and prompt flow management

Python (FastAPI)
Backend API and workflow management

React JS
User interface for natural language interaction

Docker & AKS
Scalable microservice deployment

Unstructured.io
Document parsing and text extraction

Azure Services Integrated

Compute & Containers	Azure Kubernetes Service (AKS), Azure Container Registry	Host and manage containerized services
AI & Cognitive	Azure OpenAI, Azure AI Search	Enable embeddings, semantic retrieval, and generative AI
Storage & Data	Azure Storage Account, Azure Cosmos DB, Azure SQL Database	Store data, vector indices, and metadata securely
Security & Compliance	Azure Key Vault, Microsoft Defender for Cloud, Azure Policy	Protect secrets and enforce compliance
Networking & Access	Azure Private Link, Azure AD	Secure communication and SSO-based authentication
Monitoring & Management	Azure Monitor, Azure Backup	System observability and disaster recovery
Integration & Messaging	Azure Service Bus, API Management	Event management and API orchestration

Business Impact

Measurable improvements across key operational metrics

Metric	Before	After
Information Retrieval Time	Manual keyword search	→ 70% faster retrieval
Document Review Time	Time-intensive	→ 50% reduction
Search Accuracy	Context-agnostic	→ 60% improvement
Monitoring & Insights	Limited	→ Real-time via Azure Monitor
Security & Compliance	Fragmented controls	→ Centralized Azure-native governance

Quantitative Outcomes

Significant, measurable improvements in enterprise knowledge management

70%
Faster Information Retrieval
compared to traditional keyword searches

50%
Reduction in Manual Review Time
for technical and business teams

60%
Improvement in Accuracy
and relevance of search results

Additional Key Achievements

- Real-time performance visibility through Azure Monitor dashboards
- Enhanced data security and compliance with Azure-native governance and access controls
- Scalable foundation supporting further AI-driven automation initiatives across the enterprise