

Company Knowledge Base

SearchSphere AI — Knowledge Management Platform

Overview

SearchSphere AI is an intelligent knowledge-management platform designed to unify enterprise documents, logs, and datasets under a single AI-powered search layer.

It leverages **Elastic Search AI** and **Google Cloud Vertex AI (Gemini)** to deliver contextual, conversational answers across organizational knowledge sources in real time.

Contents of This Document

1. **AI Search Architecture** – Description of hybrid retrieval using BM25 + dense vector similarity for balanced precision and recall.
 2. **Data Ingestion Workflows** – Step-by-step pipeline from raw PDFs, CSVs, and HTML pages → chunking → embedding → Elastic index storage.
 3. **FinOps Reporting Examples** – Sample cost-optimization analyses powered by SearchSphere AI, highlighting automated cloud-spend insights.
 4. **Vertex AI and Elastic Hybrid Use Cases** – Demonstrations of combining Vertex AI's text-embedding-005 model with Elastic's KNN search for RAG-based QA and document summarization.
-

Key Highlights

- Enables **natural-language querying** of technical or financial data.
- Delivers **cited, source-linked answers** for transparency.
- Designed for **multi-tenant security** and access control.

- Deployable on **Google Cloud Run**, integrating seamlessly with existing data pipelines.
-

Purpose

This document serves as an **internal reference** for evaluating the hybrid AI search capabilities of SearchSphere AI and acts as a **test corpus** for Retrieval-Augmented Generation (RAG) chat experiments.

Created by: SearchSphere AI Team

Version: 1.0

Created on: October 2025