PROJECT CASE STUDY

Enterprise Knowledge AI Chatbot

Vector-Powered RAG Integration for Zoho Cliq

AI / ENTERPRISE | 2024

01

Project Overview

An enterprise-grade AI chatbot that ingests organisational knowledge, stores it in a vector database, and surfaces answers via RAG directly inside Zoho Cliq.

Project Overview

YEAR	CATEGORY	STATUS
2024	AI / Enterprise	Production Deployed

Executive Summary

The Enterprise Knowledge AI Chatbot is a production system that transforms scattered organisational documents into an intelligent, conversational knowledge base. By combining vector database technology with Retrieval-Augmented Generation (RAG), the system allows employees to ask natural-language questions and receive accurate, context-aware answers sourced directly from internal documentation — all without leaving their Zoho Cliq workspace.

The architecture decouples data ingestion from retrieval and response generation, enabling the knowledge base to expand continuously while maintaining low-latency, high-accuracy responses. Integration with Zoho Cliq means zero context switching for employees: they chat with the bot as they would any colleague.

Technology Stack

Python | Vector Database | RAG | Zoho Cliq API | LLM | Embeddings

RAG
ARCHITECTURE

Vector
STORAGE

Zoho
INTEGRATION

02

The Problem

Enterprise knowledge is trapped in documents, scattered across teams, and difficult to surface when employees need it most.

The Problem

CORE CHALLENGE

Organisations accumulate vast amounts of internal knowledge — policy documents, technical manuals, process guides, FAQs, and project documentation — spread across file systems, cloud storage, and wiki platforms. Employees waste hours searching for information that already exists but is effectively invisible. Existing search tools return keyword matches, not answers.

Key Pain Points

1

Information Silos

Critical organisational knowledge is fragmented across Confluence, Google Drive, PDFs, Slack threads, and email archives. No single access point exists.

2

Keyword Search Falls Short

Traditional search returns document lists, not answers. Employees must open multiple files and manually scan for relevant information.

3

Context Switching Overhead

Employees work in Zoho Cliq but must leave the platform to search wikis, documents, and file shares for every knowledge query.

4

Knowledge Loss

When experienced employees leave, institutional knowledge walks out the door. There is no systematic capture or retrieval mechanism.

03

The Solution

Vector-powered RAG pipeline that ingests documents, retrieves relevant context, and delivers answers inside Zoho Cliq.

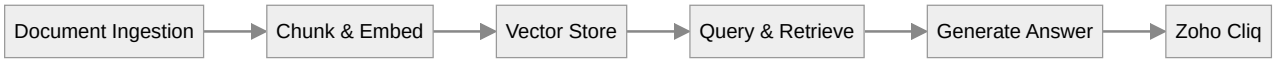
The Solution

APPROACH

Built an end-to-end RAG pipeline that ingests organisational documents, chunks and embeds them into a vector database, retrieves semantically relevant context at query time, and augments an LLM prompt to generate accurate, sourced answers. The entire system is exposed as a chatbot inside Zoho Cliq, enabling employees to query organisational knowledge conversationally without leaving their workspace.

System Architecture

The pipeline follows the canonical RAG pattern with three distinct phases: ingestion, retrieval, and generation. Each phase is independently scalable and instrumented for monitoring.



Pipeline Components

Table 1 RAG Pipeline Stage Breakdown

Stage	Technology	Function
Ingestion	Python, Document Parsers	Extracts text from PDFs, Word docs, wikis, and web pages. Handles multiple formats, image-based PDFs via OCR, and incremental updates.
Chunking & Embedding	Sentence Transformers	Documents are split into semantically coherent chunks and converted to high-dimensional vector embeddings using an open-source embedding model.
Vector Storage	Vector Database	Embeddings are indexed in a vector database optimised for fast approximate nearest-neighbour search at scale.
Retrieval	Semantic Search	User queries are embedded and matched against the vector index to retrieve the top-k most semantically relevant document chunks.

Generation	LLM via API	Retrieved chunks are injected into a structured prompt. The LLM synthesises a concise, accurate answer with citations to source documents.
Delivery	Zoho Cliq API	Answers are formatted and delivered back to the user within the Zoho Cliq conversation thread, maintaining conversational context.

04

Key Features

Enterprise-grade RAG capabilities designed for accuracy, scale, and seamless workspace integration.

Key Features & Capabilities

1

Multi-Format Document Ingestion

Supports PDF, Word, Markdown, HTML, and plain text sources. OCR layer handles scanned documents. Incremental ingestion ensures new documents are indexed without full re-processing.

2

Semantic Vector Search

Document chunks are embedded using a high-quality sentence transformer model and stored in a purpose-built vector database. Retrieval is based on semantic meaning, not keyword overlap, enabling natural-language queries.

3

Context-Aware Answer Generation

Retrieved chunks are ranked by relevance and injected into a carefully structured prompt. The LLM generates answers grounded in retrieved context, with source citations for traceability.

4

Native Zoho Cliq Integration

The bot operates as a native Zoho Cliq user. Employees @mention the bot or DM it directly. Responses appear inline with full formatting, citations, and conversational memory.

5

Conversational Memory

The bot maintains session context, allowing follow-up questions and clarifications without requiring users to restate the full query context each time.

6

Monitoring & Feedback Loop

Every query, retrieval, and response is logged. Admin dashboard tracks usage patterns, response quality, and identifies knowledge gaps for targeted document updates.

05

Results & Impact

Measurable improvements in knowledge accessibility, employee productivity, and information retrieval accuracy.

Results & Impact

Operational Outcomes

Deployment of the Enterprise Knowledge AI Chatbot fundamentally changed how employees access organisational information. The shift from manual document hunting to conversational retrieval delivered immediate productivity gains and improved decision-making speed.

Instant
ANSWER RETRIEVAL

100%
SOURCE CITATIONS

Zoho
NATIVE INTEGRATION

Key Achievements

Table 2 Operational Impact Summary

Dimension	Before	After
Information Access	Manual search across multiple platforms	Single conversational query in Cliq
Search Quality	Keyword-based, often irrelevant results	Semantic retrieval, contextually accurate
Answer Format	Document links requiring manual reading	Concise, synthesised answers with citations
Context Switching	Leave Cliq to search wikis and drives	Query and receive answers entirely in Cliq
Knowledge Capture	Tacit, at risk of loss on employee departure	Systematically indexed and retrievable
Scalability	Limited by manual curation effort	Automated ingestion scales with document volume

Conclusion

The Enterprise Knowledge AI Chatbot demonstrates how RAG architecture, when paired with a vector database and native chat platform integration, can transform an organisation's relationship with its own knowledge. By meeting employees where they already work — inside Zoho Cliq — and delivering accurate, cited answers in seconds, the system eliminates friction from information retrieval and preserves institutional knowledge at scale.