Agentic AI Chatbot

Overview

This project implements a unified multi-agent conversational system that centralizes access to organizational knowledge, analytical data, and product information. A Master Agent routes user queries to specialized sub-agents, including the Help, API, Review, and Specbook Agents, allowing employees to interact naturally with one interface while drawing from both structured APIs and unstructured documentation. The system uses an LLM and a Vector Database to process more than 6,000 help files and an extensive product library, with a Data Processor converting raw documents into embedding for fast, context-aware retrieval.

For analytical and operational tasks, the API Agent connects directly to internal and third-party APIs to provide real-time financial, sales, and customer insights, while the Review Agent interprets vendor surveys and summarizes feedback. Together, these agents form an orchestrated, modular architecture that delivers accurate responses, reduces information-search friction, and significantly improves decision-making efficiency across the organization.

Client Details

Name: Confidential | **Industry:** Software | **Location:** USA

Technologies

Python, FastAPI, Docker, LangChain, LangGraph, Streamlit, Gemini Models, Quadrant Vector Database, GCP VM , GitHub and GitHub Actions.

Agentic AI Chatbot

Project Description

Modern enterprises often maintain thousands of help articles, process documents, and FAQs, making it challenging for employees to quickly locate reliable information. In this case, more than 6,000 help topics existed, yet users still struggled to find the correct steps for ticket creation, troubleshooting, and routine operations. This revealed a pressing need for a natural language–driven assistant that could turn static documentation into easily accessible, conversational support.

Beyond documentation, the organization depended on multiple dashboards and APIs that exposed financial metrics, customer records, sales data, and accounting information. However, accessing and interpreting this data required technical proficiency, which restricted its usefulness. To unlock its full value, the company needed a fact-driven analytical Chatbot capable of querying these APIs directly and delivering accurate, real-time insights to users in simple, natural language.

Additional needs surfaced for specialized AI agents; one to analyze vendor survey data and distill qualitative feedback, and another to navigate a vast spec-book library containing over 8,000 products. These agents were to be designed to give users instant access to product details, usage information, and vendor performance insights, eliminating the need for manual searches or complex data interpretation.

To unify these capabilities, a master–slave orchestration model was implemented, allowing a central agent to route each query to the most relevant domain-specific agent. This approach delivered a single intelligent interface that simplified access to documentation, analytics, vendor insights, and product information, ultimately improving operational efficiency and decision-making across the organization.

Project Phases

1. Data Preprocessing (Training Chatbot)

The data preprocessing phase involved preparing multiple heterogeneous data sources to train the suite of specialized Chatbots. For the Help Agent, over 6,000 help-center topics spanning HTML files, DAT files, JavaScript content and other formats were ingested and processed into structured embeddings before being stored in the vector database. In parallel, the API Agent followed a different approach. Instead of loading local data, it consumed real-time API endpoints defined dynamically through a configuration-driven JSON file, enabling live analytical responses. The Vendor Review Agent relied on a streamlined pipeline optimized specifically for PDF ingestion, processing survey documents and storing their embeddings separately to support performance summarization. Finally, the Specbook Agent used a dynamic workflow that fetched product PDFs through real-time APIs, processed them on demand, and automatically updated the vector store

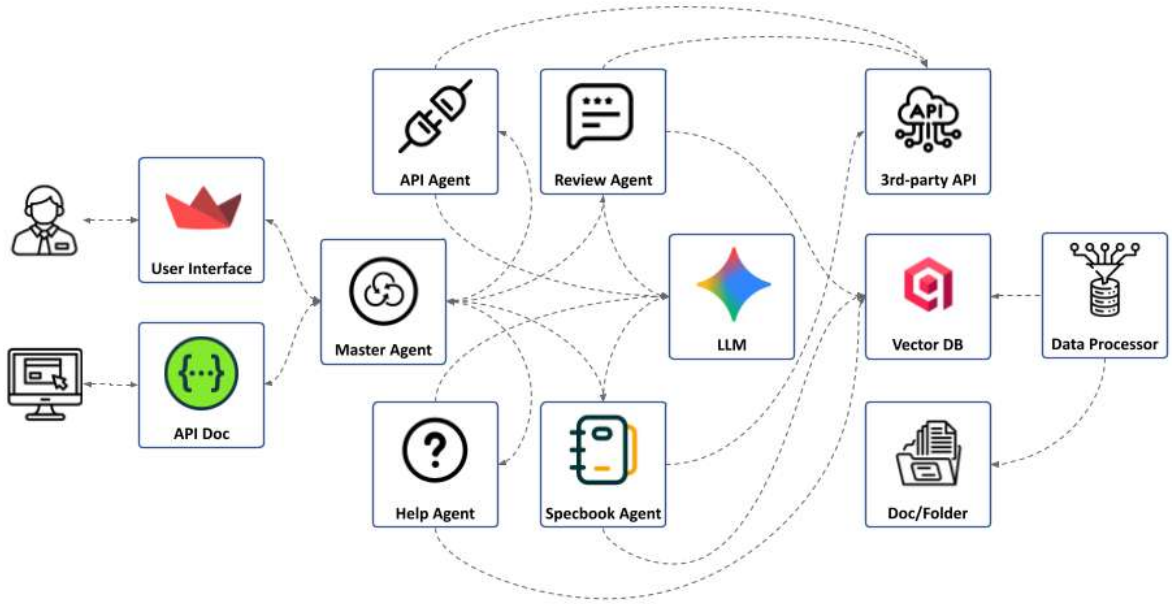
Agentic AI Chatbot

whenever product documents were added, modified, or removed. This modular preprocessing approach ensured that each agent was trained using the most appropriate method for its data domain.

2. Chatting with a Bot

In this phase, users engage with the trained Chatbot through a conversational interface, where they can ask questions and retrieve information from the vector store. The Chatbot provides accurate, context-specific responses based on the data it has been trained on, which includes details about the procedures, real-time API data, products, vendor reviews, and many more. By leveraging the knowledge from the training phase, the Chatbot offers precise answers to user queries, making it easier for users to access the information they need without having to navigate through the website manually or to any other data specific to the system. This phase enhances user experience by delivering quick, relevant responses to a wide range of questions.

Architecture



Agentic AI Chatbot

The system provides a unified platform for training, managing, and interacting with multiple specialized AI Chatbots designed to support various operational needs within an organization. Through a single interface, users can access a conversational assistant that understands natural language and retrieves information from help documentation, real-time APIs, vendor review data, and product libraries. The system integrates several domain-specific agents, such as the Help Agent, API Agent, Review Agent, and Specbook Agent; each responsible for processing and responding to queries related to its specialized dataset.

To prepare each agent, the system supports multiple data ingestion workflows. Help documentation consisting of more than 6,000 topics is processed through a structured pipeline that handles diverse file types, including HTML, DAT, JavaScript, and other documents. These files are cleaned, vectorized, and stored in a vector database for fast retrieval. For real-time operational data, the API Agent is configured through JSON-based endpoint definitions, enabling it to interact directly with internal and third-party APIs instead of relying on static files. Vendor review files and product PDFs follow their own ingestion flows—survey data is processed and indexed from PDFs, while product information is dynamically downloaded, processed, and automatically updated in the vector store whenever new items are added or modified.

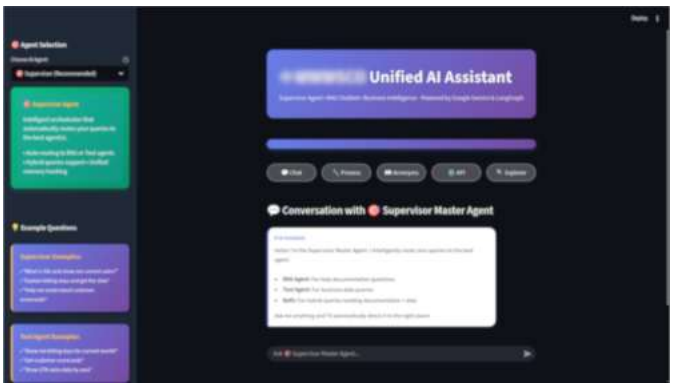
Once training is complete, users interact with the system through a conversational interface that automatically routes their queries to the appropriate agent using a central Master Agent. Employees can ask procedural questions from the help library, request real-time sales or financial metrics, analyze vendor performance summaries, or explore product specifications from the Specbook—all within the same chat window. The LLM coordinates responses across agents to deliver accurate, context-aware answers without requiring users to manually navigate documents, dashboards, or APIs.

This approach ensures that organizational knowledge, analytical data, and product information become immediately accessible in a natural, user-friendly format, allowing teams to work more efficiently and make faster, well-informed decisions.

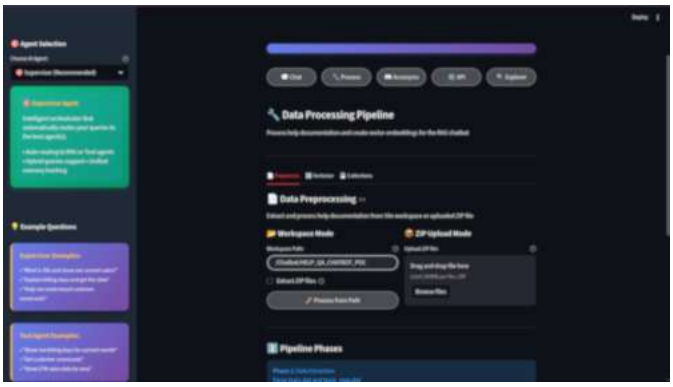
Agentic AI Chatbot

Screenshots

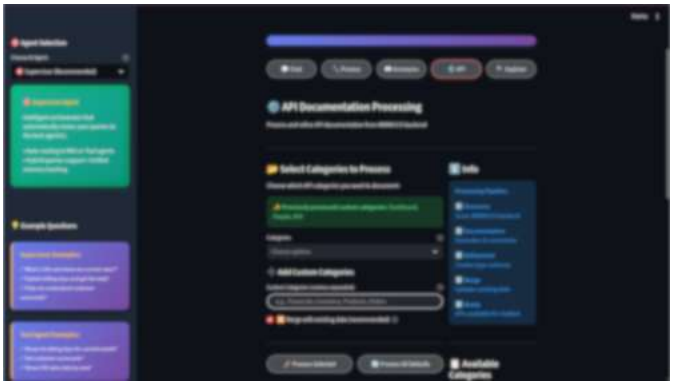
User Interface



Data Training



API Data Training



Agentic AI Chatbot

Processing and Results Training

Chatting with the Bot