

Task: Create N8N + Supabase RAG chatbot

Introduction

n8n is a workflow automation platform that gives technical teams the flexibility of code with the speed of no-code. Supabase is an open source Firebase alternative. It provides Hosted Postgres Database, Authentication and Authorisation, Auto-generated APIs, Database and Edge functions, File storage, vector embeddings toolkit and a dashboard. Retrieval-augmented generation is a technique that enables generative artificial intelligence models to retrieve and incorporate new information. It modifies interactions with a large language model so that the model responds to user queries with reference to a specified set of documents, using this information to supplement information from its pre-existing training data.

Architecture:

There are 2 main sub-workflows set up in N8N.

The first workflow gets a specified file from a connected Google Drive, extracts relevant text from it, and inserts it into a supabase vector store, after creating embeddings using Google Gemini models.

The second workflow retrieves the vector embeddings from Supabase, uses it to augment the query to the LLM, which is Gemini, and outputs the result of the LLM.

Instructions

The prerequisites to running this n8n workflow is to have accounts created on n8n, Google Cloud Platform and supabase. Next, you need to set up credentials for Google Cloud Platform for using Gemini, and to hook up Supabase, n8n has excellent documentation online to do the same. Finally, you need to set up the vector store in Supabase itself, Supabase also has excellent documentation for that.

Finally, import the json file of the workflow into n8n and run the workflow.

Work done by:

Aseem Athale

athaleaseem@gmail.com