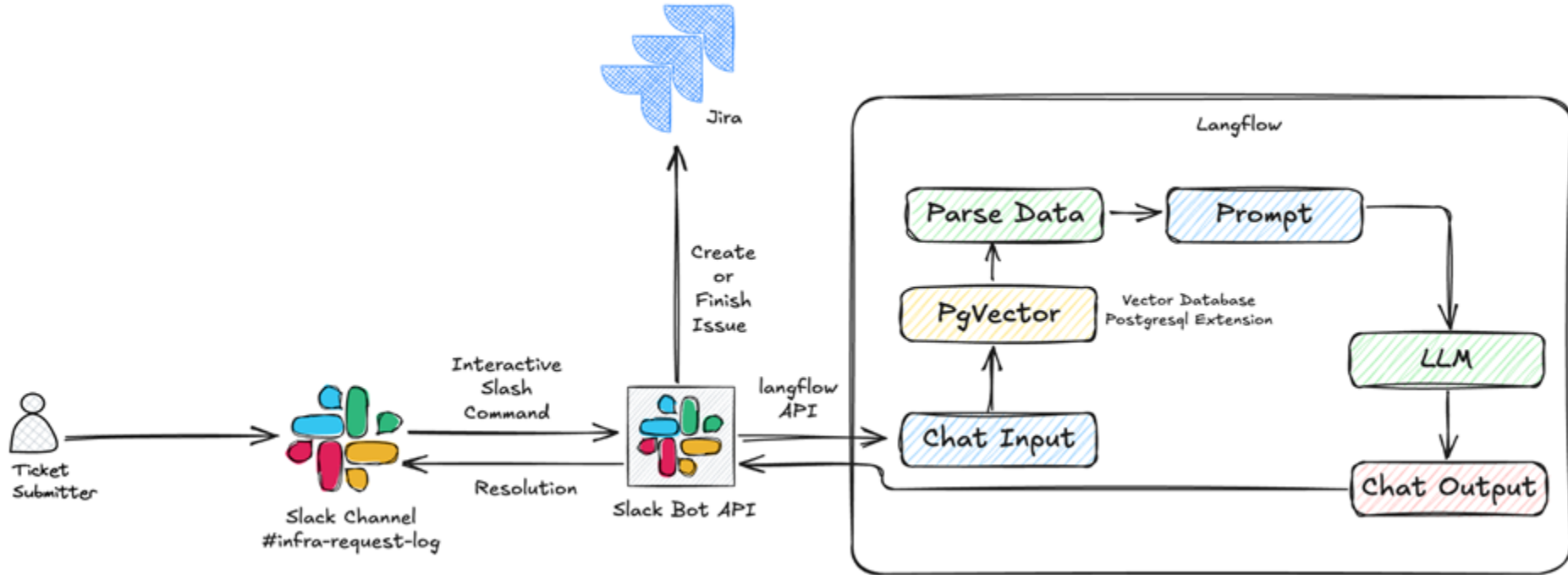# Ticketing System using slack, LLM & Langflow

The Goals

...

Speed up ticket resolution with AI
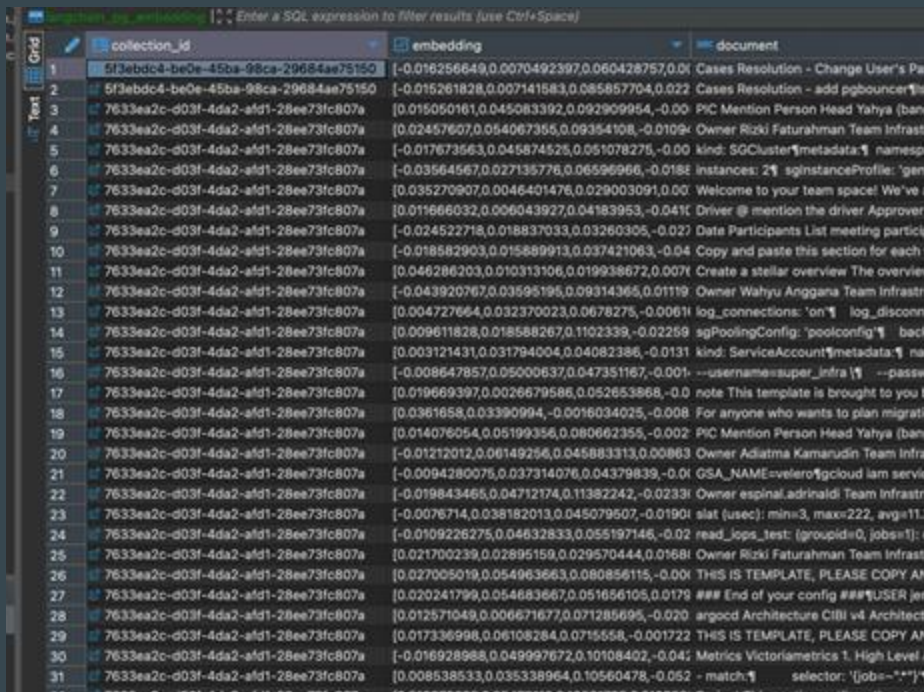
Reduce team workload and boost focus

Improve response times and efficiency

Automate repetitive tasks using RAG

Leverage past solutions for faster fixes

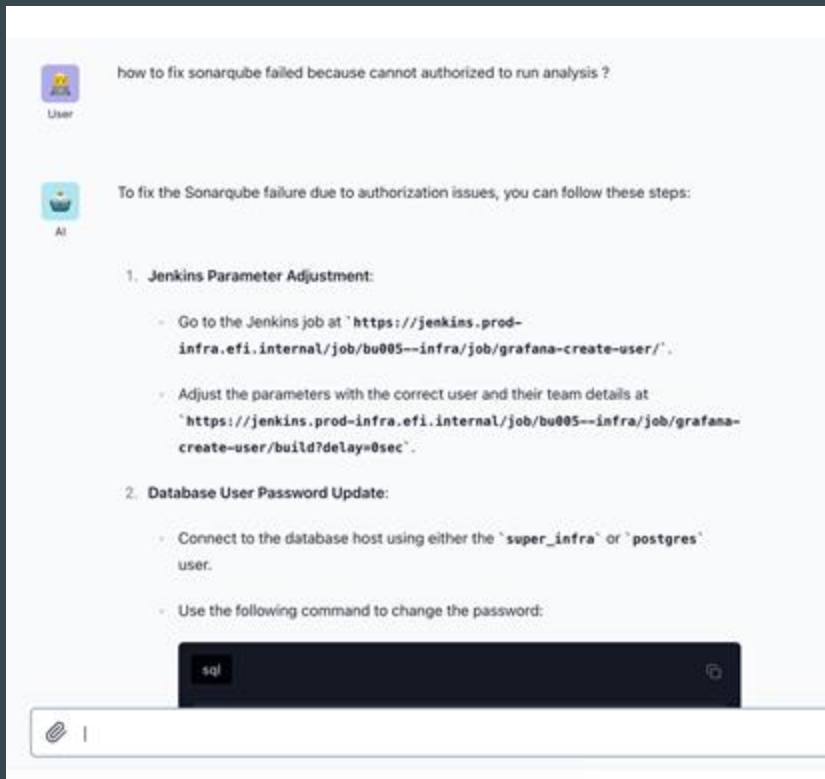Scale support for growing ticket volumes

Vector Store on PgVector

- The PGVector vector store is queried with an embedding of the input message to find similar request tickets in the vector store.
- Similar tickets are retrieved based on their proximity to the current query in the embedding space, leveraging semantic similarity.

The Results

- The prompt is fed to the OpenAI language model (gpt-4o) with a low temperature setting (0.1) to encourage more deterministic and precise responses.
- The language model generates a response based on the input query and the context derived from the retrieval process.
- Finally, the generated response is displayed in the chat playground to answer the original query.

Complete Flows using Langflow

How It Works

- Chat Input:
  - A query is received from the chat playground, containing a request or question from the user.
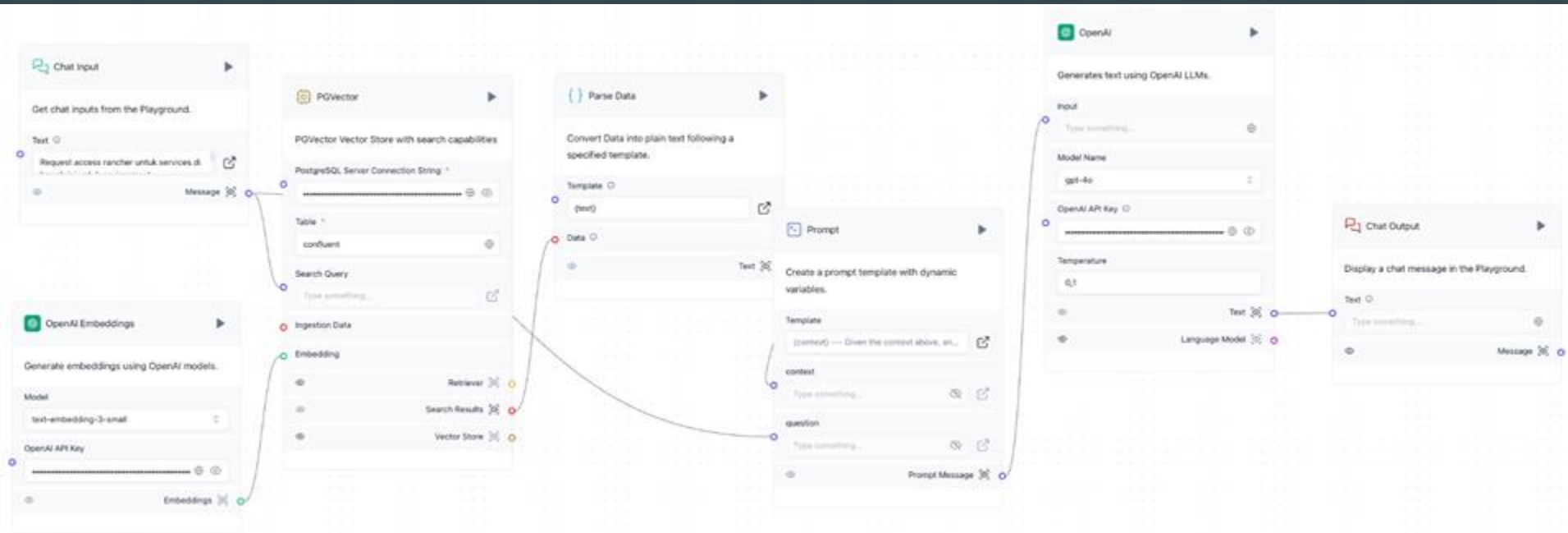- Retrieve Similar Entries:
  - The PGVector vector store is queried with an embedding of the input message to find similar request tickets in the vector store.
  - Similar tickets are retrieved based on their proximity to the current query in the embedding space, leveraging semantic similarity.
- Parse Data:
  - The retrieved data is parsed into plain text, which is formatted using a predefined template.
- Prompt Creation:
  - A prompt is created that incorporates the retrieved context, providing additional information to the model.
  - This prompt includes relevant prior cases and resolutions to guide the language model's response.
- Language Model (OpenAI):
  - The prompt is fed to the OpenAI language model (gpt-4o) with a low temperature setting (0.1) to encourage more deterministic and precise responses.
  - The language model generates a response based on the input query and the context derived from the retrieval process.
- Chat Output:
  - Finally, the generated response is displayed in the chat playground to answer the original query.

Thanks!

Wahyu Anggana

Cloud Engineer

https://www.linkedin.com/in/wahyuanggana/
https://github.com/k0wl0n
Telegram: @kowlon