Word List Storage

- Stores list of words used by the web scraper to search for WeChat articles.

Soguo Web Scraper

- Scrapes Sogou, a Chinese search engine, for WeChat articles using the terms in the word list as keywords.

Vector Embedding Model

- Assigns a vector to user queries and sections of the scraped articles to store in ChromaDB.

Front End UI

- User adds or deletes words from the wordlist.

- User submits queries to the LLM which will generate a response.

- General documentation can be viewed from within the UI.

Python Back-End

- Facilitates the transport of information throughout numerous components.

AYA LLM

- Generates a response to user query using solely the articles stored in the ChromaDB that most closely match the vector of the query.

- Handles the translation between English and Chinese.

ChromaDB

- Vector database that stores our scraped articles along with associated metadata.