# Piyush Waradkar | piyushwaradkar@gmail.com

# **Technical Design Document**

## **Data Ingestion Workflow**

#### 1. Proprietary Restaurant Dataset

- The prototype reads a proprietary restaurant dataset from an Excel file (Sample\_Ingredients\_File.xlsx).
- This file includes restaurant information, menu items, categories, prices, addresses, ratings, etc.
- Data is cached in a pickle file (data\_cache.pkl) to avoid repeated loading from Excel.
   This speeds up subsequent runs.

#### 2. Unstructured External Documents

- Wikipedia Connector: Fetches summaries from Wikipedia by calling the Wikipedia API.
- Google Search Connector: Uses the googlesearch library to retrieve URLs and then scrapes content from those URLs using requests and BeautifulSoup.
- For each external source, large text blocks are chunked into 500-character segments.
   This ensures that only relevant excerpts are passed to the LLM and prevents excessively large context blocks.

#### **Vectorization and Indexing**

## 1. Embedding Model

 A SentenceTransformer model (all-MiniLM-L6-v2) encodes each piece of text (both from the proprietary dataset and external sources if we choose to index them) into high-dimensional embeddings.

## 2. Vector Database (FAISS)

- The embeddings are stored in a FAISS index (faiss\_index.index).
- A separate NumPy array file (embeddings.npy) stores the raw embeddings on disk.

## 3. BM25 for Complementary Retrieval

 In addition to vector retrieval, BM250kapi is used to ensure that classic lexical matching complements the vector-based similarity. This is known as hybrid search, combining semantic similarity with token-based ranking.

## **Retrieval Pipeline**

## 1. Query Embedding & Vector Similarity

• When a user makes a query, the system encodes the query into an embedding and searches the FAISS index for the top-k candidates.

- The same query is also tokenized and used to compute BM25 scores on the entire corpus.
- The top results from both methods are combined and deduplicated, ensuring the final top-k results are a balanced mix of semantic and lexical matches.

#### 2. External Content Fetch & Chunking

- The pipeline concurrently fetches external sources (Wikipedia, Google-search-based web content).
- These external texts are chunked into 500-character pieces. Each chunk is stored with the metadata of its source, URL, and title. This makes the final "retrieval set" more fine-grained.

#### 3. Factual Consistency

- The system merges:
  - 1. Internal data from the restaurant dataset.
  - 2. External data from Wikipedia or other websites.
- A final "prompt" is built by concatenating these different blocks of text. The user query is appended, along with instructions on how to reference sources.

## **LLM Integration & Prompt Engineering**

## 1. System/Instruction Prompt

- The code builds a prompt that includes each chunk of text labelled with a bracketed reference.
- The user query is added afterward, along with guidelines:
  - Cite sources with reference numbers.
  - Distinguish facts from opinions.
  - Compare prices to category averages.
  - Include relevant menu items from the dataset and external context.
  - Handle conflicting information gracefully.

## 2. Chain-of-Thought vs. Final Output

- Although minimal chain-of-thought is shown, the code constructs a single prompt that provides the relevant context to the LLM.
- The LLM sees enough textual context to produce a factual, well-grounded answer, citing references.
- o If the external or internal data is lacking or contradictory, the LLM's final instructions prompt it to offer disclaimers or clarifications rather than guess.

## **References & Source Tracking**

#### 1. Internal References

o For proprietary data, references include item\_id and relevant textual snippets so the user can see which row of the dataset the LLM used.

#### 2. External References

- For each external chunk, the system keeps track of the source (Wikipedia, web), the URL, and any metadata like title.
- o In the final response, the user can optionally view the references to see where the LLM got its information.

## **Scalability & Future Extensions**

#### 1. Larger Datasets

- o The code uses FAISS for vector search, which easily scales to millions of documents.
- o The BM25 approach can also handle large corpora with efficient inverted indices.

#### 2. Real-Time Ingestion

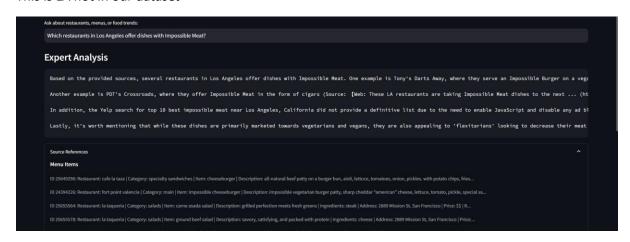
- New data can be vectorized and added to the FAISS index incrementally (FAISS supports adding vectors to an IndexFlatL2, although for very large scale, advanced indexing structures might be needed).
- For external real-time news or websites, a queue-based approach with periodic reindexing can be adopted.

## 3. Caching Embeddings & Connectors

- The code already includes caching for embeddings and data with local pickle and NumPy files.
- Future updates could incorporate more advanced caching or versioning strategies for frequently updated datasets.

# **Outputs:**

## This is LA not in our dataset



## This is San Franciso, note that SF is also recognised



## This is in Chicago

```
Expert Analysis

To find restaurants near you that serve gluten-free pizza, I recommend using the following resources:

1. Download the Find Me Gluten Free app (Source: [Web: Gluten-Free Pizza in Chicago - 2025 (https://www.findmeglutenfree.com/us/il/chicago/pizza)], last updated January 2025.

2. Check out the Top 10 Best Gluten Free Pizza options in Chicago, Il according to Yelp as of 2025 (Source: [Web: TOP 10 BEST Gluten Free Pizza in Chicago, IL - Updated 2025 - Yelp

3. Explore gluten-free pizza options on Slice, which offers a list of gluten-free pizza places in Chicago (Source: [Web: Find Gluten Free Pizza Places in Chicago, IL - Updated 2025 - Yelp

3. Explore gluten-free pizza options on Slice, which offers a list of gluten-free pizza places in Chicago (Source: [Web: Find Gluten Free Pizza Places in Chicago - Slice (https://sl

3737 N Western Ave, Chicago, IL 60618

- 1084 N Observes Place, Chicago, IL 60657

- 2389 N Sheffield Ave, Chicago, IL 60657

- 2389 Sith St., Chicago, IL 60643

- 3477 N Broadowy St., Chicago, IL 60640

- 10314 S Halsted St., Chicago, IL 60630

- 5358 N Herodewy, Chicago, IL 60630

- 5358 N Herodewy, Chicago, IL 60630

- 5051 S Arber Ave, Chicago, IL 60640

- 1081 S N Harlem Ave, Chicago, IL 60640

- 1081 S N Harlem Ave, Chicago, IL 60640

- 1081 S N Harlem Ave, Chicago, IL 60640

- 1081 S N Harlem Ave, Chicago, IL 60640

- 10
```

## This is San Francisco





Ask about restaurants, menus, or food trends:

What is the history of sushi, and which restaurants in SF are known for it?

Expert Analysis

The history of sushi can be traced back to Japan, where it originated as a method of preserving fish using fermented rice. However, the introduction of sushi to the United States is Fast forward to modern times, San Francisco is home to several restaurants known for their sushi offerings. One such establishment is cha-ya san francisco. On their menu, they offer While cha-ya san francisco is a standout in San Francisco's sushi scene, it's important to acknowledge that conflicting information exists regarding the history of sushi in the U.S.

Source References

Menu Items

10 243922519: Restaurant: cha-ya san francisco | Category: sushi | Item: medium sized rolls (6 pcs) | Description: vegan, | Ingredients: carrots | Address: 762 Valencia St, San Francisco | Price: \$5 | Rating: 4.0 | 10....

10 243922519: Restaurant: cha-ya san francisco | Category: sushi | Item: medium sized rolls (6 pcs) | Description: vegan, | Ingredients: rori | Address: 762 Valencia St, San Francisco | Price: \$5 | Rating: 4.0 | 10....

10 243922519: Restaurant: cha-ya san francisco | Category: sushi | Item: small rolls (6 pcs) | Description: vegan, | Ingredients: rori | Address: 762 Valencia St, San Francisco | Price: \$5 | Rating: 4.0 | 10....

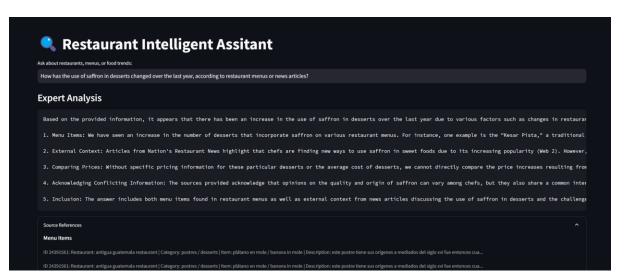
10 243922519: Restaurant: cha-ya san francisco | Category: sushi | Item: small rolls (6 pcs) | Description: vegan, | Ingredients: rori | Address: 762 Valencia St, San Francisco | Price: \$5 | Rating: 4.0 | 10....

10 243922519: Restaurant: cha-ya san francisco | Category: sushi | Item: small rolls (6 pcs) | Description: vegan, | Ingredients: rori | Address: 762 Valencia St, San Francisco | Price: \$5 | Rating: 4.0 | 10....

10 243922519: Restaurant: cha-ya san francisco | Category: sushi | Item: small rolls (6 pcs) | Description: vegan, | Ingredients: rori | Address: 762 Valencia St, San Francisco | Price: \$5 | Rating: 4.0 | 10....

Prices St | Rating: 4.0 | 10....







Restaurants in the Bay Area that use saffron as an ingredient in their dishes include:

1. Blue Plate (ID: 26085136) - This restaurant, located at 3218 Mission St, San Francisco, offers a pan roasted rockfish entree with soupy rice in tomato-saffron broth, green garlic

2. Deccan House (ID: 24393265) - This restaurant located at 1142 Valencia St, San Francisco serves a Deccan special biryani in a large tray suitable for 16-20 guests. The dish is see

. Espertento (ID: 26083625) - This restaurant at 3295 22nd St, San Francisco offers albondigas, a meatball stew with saffron. Additionally, they have a paella de carnes y mariscos,

In terms of opinions, the ratings provided by Yelp users offer insight into customers' experiences with these restaurants and their dishes containing saffron. However, it is essentia

Course Deference

ID 26085136: Restaurant: blue plate | Category: entrees !!!\* | Item: pan roasted rockfish | Description: soupy rice in tomato-saffron broth, green garlic, fava beans, escarole, black olive vinaigrette | Ingredient.