

## Technical Appendix

**Foodie. AI - Team Members:** ChengHsin Chang (cc5211), Yu-Heng Chi (yc4548),  
Tirada Nilphanaphan (tn2559), Sarah Pradita (syp2132)

### 1. Model Choices & Architecture

- Embedding Model
  - Model: Sentence-Transformer (sentence-bert or mpnet-base-v2)
  - Rationale: State-of-the-art for semantic similarity; fine-tuned on NLI and STS tasks; captures nuanced relationships like "rich tonkotsu broth"  $\approx$  "deep, savory pork flavor"
  - Input: Restaurant reviews, menu descriptions, user queries
  - Output: Dense 384-768 dimensional vectors (model-dependent)
- Vector Search Infrastructure
  - Solution: FAISS (Facebook AI Similarity Search)
  - Indexing: IVF-Flat or HNSW for fast approximate nearest-neighbor retrieval
  - Query Latency: <50ms for top-K retrieval (typical K=50)
  - Scalability: Supports millions of restaurant embeddings with efficient memory usage

Language Model (LLM) for AI Insights

- Models: Mistral 7B
- Role: Post-processing of top-K results to generate:
  - \* Match score with rationale (e.g., "79.1% match")
  - \* Review-backed explanations grounded in real reviewer language
  - \* Dish-level matching (specific menu items aligned with user craving)
  - \* Trade-offs and context (wait times, crowd levels)
- Prompting Strategy: Few-shot examples showing desired insight structure; chain-of-thought reasoning to ensure evidence-based explanations

### 2. Data Sources & Integration

- Primary Data Source: Kaggle NYC Restaurants Dataset
- Source: <https://www.kaggle.com/datasets/beridzeg45/nyc-restaurants>
- Content:
  - \* Restaurant metadata (names, addresses, coordinates, cuisines, neighborhoods, ZIP codes)
  - \* Menu descriptions and dish categories
  - \* Aggregated Google reviews (raw review text)
  - \* Star ratings and review counts
- Volume: ~10,000 unique restaurants in Manhattan with 50,000+ reviews

- Refresh Cadence: Weekly ingest of new reviews
  - Secondary Data Source: NYC Department of Health Restaurant Grades
- Source: <https://data.cityofnewyork.us/Health/Restaurant-Grades/gra9-xbjk/>