Foodie. AI - Your NYC Ultimate Guide for Food Adventure

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

- **Background**

Foodie. AI - Your NYC Ultimate Guide for Food Adventure

Manhattan's restaurant scene offers limitless choices, which is exciting but overwhelming for food lovers. With thousands of places, deciding can be challenging. Popular review sites often highlight only the trendiest spots, leaving out hidden gems and important details.

Google Maps can also feel random. It often suggests places that do not meet your needs or are not in your location. Searching for a quiet coffee shop with Wi-Fi or an authentic Indian restaurant near Midtown usually means wading through reviews and piecing together information. You lose personalization.

That's why we created Foodie.AI: to make dining in Manhattan more personal. Foodie.AI combines advanced AI with official NYC restaurant inspection data. Other platforms don't offer this feature. You get smart, tailored recommendations based on your tastes, dietary needs, priorities, and location. Whether you're a local looking for hidden gems or a visitor wanting a safe meal, Foodie.AI helps you explore with confidence.

- **Problem Statement**

Manhattan's restaurant scene is exciting, but it can quickly become overwhelming. With so many choices, it's hard for locals and visitors to pick a place with confidence.

Pain Points:

· **Proximity-Biased Recommendations**: Popular review sites often highlight only the most talked-about spots, so you might miss out on hidden gems or overlook important details. Even Google Maps can feel random, often suggesting places that don't fit your needs, ranking primarily by proximity rather than actual preference fit.

· **Generic Filter**: Finding a quiet coffee shop with Wi-Fi or authentic Indian curry near Midtown requires sifting through many reviews across platforms.

· **Limited Dish-Level Intent Matching**: Most platforms only let you search by broad cuisine types. You can't look for something specific, like "spicy tonkotsu ramen in a quiet place" or "authentic cacio e pepe in a cozy spot." To find these details, you must read through many reviews.

· **Lack of Trustworthy:** Popular platforms emphasize ad placement and star ratings, often missing the nuanced quality signals embedded in 100% real user reviews. Users struggle to distinguish between a well-marketed mediocre place and a genuinely excellent hidden gem.

Why Foodie. AI?

Foodie. AI was born to transform the dining experience in Manhattan with a level of personalization unmatched by other tools. By uniquely blending cutting-edge AI embeddings with real Google reviews and restaurant menu data—grounded in trust and intent matching rather than popularity and proximity alone— Foodie. AI empowers you to discover not only delicious food but also restaurants that genuinely match your tastes and context.

No more disappointment—just smart, dynamic recommendations based on your exact cravings, location, and the authentic voices of real diners. Whether you're a local hoping to uncover hidden gems in your neighborhood or a visitor seeking a memorable, confident meal, Foodie. AI puts exploration and trust back on the menu.

- **Solution**

Core Value Proposition

Foodie. AI is a restaurant search platform that eliminates decision fatigue by ranking restaurants on three key dimensions:

1. **Menu Match**: Does this restaurant serve the specific dish or food style you're craving?
2. **Experiential Match**: Do real reviews suggest you will genuinely enjoy dining here, given your preferences?
3. **Location Fit**: Is the restaurant realistically accessible from where you are, or where you want to be?

By grounding every recommendation in real user Google reviews instead of ad placement or generic popularity, Foodie. AI delivers fewer but higher-confidence options, helping users feel more certain when trying something new.

User Experience Flow

Step 1: Natural Language Food Input

Users can describe the food they are craving with natural language, including preferred flavors, meal types, specific proteins, or dining context. Our system interprets flexible, conversational queries without requiring predefined filters or menu navigation.
Example: "I want to eat spicy Thai meat curry in Midtown"

Step 2: Location Detection

Users can either enter their ZIP code manually or allow our GPS-based location service to automatically detect their current position for more accurate, localized recommendations.

Step 3: Intelligent Restaurant Search

Foodie. AI processes the user's food description, location, and context. Leveraging semantic embeddings and our curated restaurant database (built from Google reviews and menus), the system identifies and ranks the top K restaurants that best match the user's request across the three core dimensions.

Step 4: Rich Restaurant Results Display
Our website presents a structured, user-friendly results page that includes:

- Restaurant name and key details (address, distance, neighborhood)
- Highlights of dishes that match the user's query
- Summary of what real reviewers liked or disliked powered by AI
- Menu overview and restaurant description
- Match score and AI-generated explanations of why each restaurant is a good fit

This provides users with a seamless, transparent, and confidence-building browsing experience.

- **Technical Approach**

1. Preprocessing Restaurant Data

1.1 Data Collection

We pull raw text descriptions of restaurants—menus, tags, reviews, categories, and other contextual information. For each restaurant, we maintain metadata including ZIP code, coordinates, and aggregated review summaries.

1.2 Review Embedding

We use **Sentence-Transformer** (a state-of-the-art embedding model fine-tuned on semantic similarity tasks) to convert restaurant reviews into dense vector representations. Each review or aggregate restaurant description is transformed into a high-dimensional vector that captures semantic meaning—including cuisine type, dish quality, atmosphere, service tone, and other nuanced signals that humans extract when reading reviews.

Storage Structure:
| Index | Zipcode | Restaurant_ID | Embedded_Vector | Review_Text | Metadata |

2. User Input Processing

2.1 User Provides Description
The user enters text describing what they want.
Example: "spicy ramen near Times Square"

2.2 Convert User Input into Embedding
We embed the user's query using the **same Sentence-Transformer model**, placing it in the same vector space as the restaurant embeddings.
Example embedding: [0.656, 0.875, 0.964, ...]

2.3 Extract User Location

We extract the user's location context (ZIP code, e.g., "10018" or GPS coordinates) to enable geographic filtering.

3. Intelligent Filtering

3.1 Geographic Filtering

We filter restaurant embeddings by user location, selecting only restaurants in matching or nearby ZIP codes. This reduces the search space to geographically relevant options, ensuring users don't receive recommendations across the entire city for a casual lunch.

Filtering Logic:

- Exact ZIP code match (highest priority)
- Neighboring ZIP codes within a user-specified radius
- Optional: all of Manhattan (for exploratory searches)

4. Vector Similarity Search (FAISS)

We store all restaurant embeddings in a **FAISS (Facebook AI Similarity Search) index**, which enables fast approximate nearest-neighbor search over large numbers of vectors.

4.1 Cosine Similarity Ranking

After geographic filtering, we compute cosine similarity between the user's embedding and each candidate restaurant embedding. Cosine similarity ranges from -1 to 1, with values closer to 1 indicating high semantic similarity.

Similarity Score Formula:

4.2 Retrieve Top-K Candidates

We return the indices of the K most relevant restaurants based on similarity scores.
Example: top-K returns indices [364, 1024, 1036, 1124, ...]

5. AI-Powered Insight Generation

5.1 Fetch Restaurant Data

Using the top-K indices, we retrieve full restaurant descriptions, menus, aggregated reviews, and metadata.

5.2 LLM-Generated AI Insights

We send the user's original query along with top-K restaurant details to a large language model (Mistral AI). The LLM generates concise, grounded AI insights displayed directly in the results that include:

- **Match Score with Rationale**: A percentage match score (e.g., "79.1% match") paired with a brief, evidence-based explanation of why this restaurant fits the user's query
- **Review-Backed Reasoning**: Insights are grounded in actual review language and signals, e.g., "*LAN LARB SOHO is highly rated (4.7) and praised for its authentic Thai curries, including green curry, which suggests they can handle spicy crab meat curry well. The review highlights excellent flavor and portion size, making it a top match.*"
- **Dish-Level Matching**: Reference specific menu items and dishes that align with the user's craving

· **Key Attributes Highlighting**: Surface relevant review themes (ambiance, service quality, cuisine authenticity) that match user intent

· **Trade-Offs and Context**: Where relevant, mention practical considerations (e.g., wait times, crowd levels, price range)

These AI insights are **displayed prominently in the results UI**, providing users with immediate, transparent reasoning for why each recommendation appeared—eliminating guesswork and building confidence in their choice.

6. Final Output with AI Insights

6.1 Present Ranked Recommendations with AI-Generated Explanations
The system outputs an easy-to-read, high-confidence result list featuring:

· **Ranked List** of 3–5 top restaurants (not dozens)

· **Match Score**: Visual indicator (e.g., "79.1% match") showing semantic and contextual fit

· **AI Insight Card**: AI-generated explanation grounded in real reviews, showing exactly why this restaurant matches the user's query

· **Key Details**: Restaurant name, cuisine type, address, distance, neighborhood, health inspection grade

· **Review Highlights**: Key dishes, review themes, and attributes most relevant to the user's search

· **Link to Full Details**: Option to view complete menu, all reviews, and booking information

This multi-layered output transforms restaurant discovery from browsing-heavy to insight-driven, allowing users to make confident decisions in seconds rather than minutes of searching.

- **How Foodie. AI Solves the Pain Points**

From Proximity-Based to Intent-Matched Ranking

Traditional Google Maps Approach:

"Show me all restaurants within 0.5 miles, sorted by star rating and popularity."
Result: Same high-traffic chains every time; hidden gems buried below.

Foodie. AI Approach:

"Show me restaurants where real reviewers mention spicy food, quiet atmosphere, and good vegetarian options, within my preferred area."
Result: Fewer options, but each one is a strong fit for your context.

Semantic Understanding Over Categorical Filters

By embedding entire reviews and descriptions, Foodie. AI understands nuanced, conversational queries that would break traditional category-based systems. The embedding space captures relationships like:

- "Rich tonkotsu broth" ≈ "deep, savory pork flavor"
- "Cozy date-night spot" ≈ "intimate ambiance, attentive service"
- "Great for solo diners" ≈ "friendly bartenders, seat at the counter"

This semantic richness means users can express genuine intent, not squint at predefined checkbox filters.

Building Trust Through Transparency

Every recommendation is grounded in 100% real user reviews. The LLM-generated explanation shows users exactly which review signals drove the ranking, helping them understand why a restaurant might work for them. This transparency converts skeptical users into confident explorers.

- **Intended Users**

Our target users include:

- **NYC Residents**: Locals seeking to uncover hidden gems in their neighborhood or explore new cuisines without the browsing fatigue of traditional review platforms.
- **Tourists and Visitors**: People new to the city who want curated, trustworthy recommendations that feel authentic and tailored to their specific cravings and context.
- **Food Enthusiasts**: Anyone who wants a smarter, faster way to discover places to eat and feels more confident when choices are grounded in real user voices, not popularity hype.

By allowing users to type natural language requests such as *"I want to eat spicy Thai crab curry in the East Village"* or *"cozy solo dinner with good natural wine in Nolita,"* Foodie. AI makes restaurant searching more intuitive and personalized than existing tools.

- **Data Sources and Integration**

Foodie. AI leverages two authoritative, publicly available datasets to build a comprehensive and trustworthy restaurant knowledge base:

Primary Data Source 1: Kaggle NYC Restaurants Dataset

Source: https://www.kaggle.com/datasets/beridzeg45/nyc-restaurants

This dataset provides:

- **Restaurant Metadata**: Names, addresses, coordinates, cuisine types, neighborhoods, and ZIP codes
- **Menu Information**: Aggregated menu descriptions and dish categories
- **Review Text**: Cleaned Google reviews and user feedback

- **Rating and Aggregates**: Star ratings, review counts, and sentiment indicators

We use this dataset as our primary source for:

- Building initial restaurant-to-embedding mappings via Sentence-Transformer
- Extracting menu and dish-level language for semantic matching
- Populating the FAISS index with review embeddings
- Establishing baseline restaurant metadata for the results display

Secondary Data Source 2: NYC Department of Health Restaurant Grades

Source: https://data.cityofnewyork.us/Health/Restaurant-Grades/gra9-xbjk/about_data

This dataset provides:

- **Restaurant Metadata**: Names, addresses, coordinates, cuisine types, neighborhoods, and ZIP codes
- **Menu Information**: Aggregated menu descriptions and dish categories
- **Geographic Coverage**: All restaurants inspected across NYC's five boroughs

We use this dataset to:

- Extracting menu and dish-level language for semantic matching
- Combining NYC restaurant data with Kaggle data

Data Integration Pipeline

Step 1: Data Merging

Join the Kaggle dataset with the DOH dataset using restaurant name fuzzy matching and geographic proximity. This creates a unified restaurant record that includes both AI-discoverable signals (reviews, menus) and official compliance data.

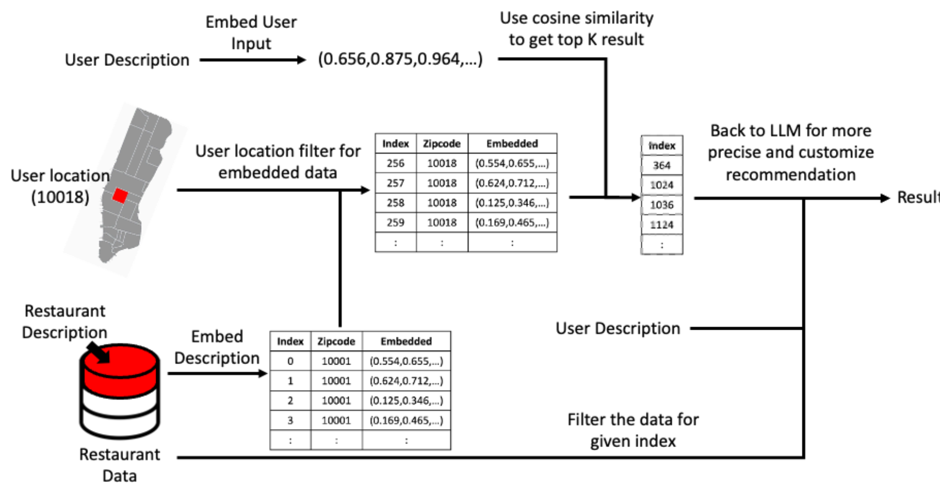
Step 2: Data Cleaning and Standardization

- Normalize restaurant names and addresses
- Handle duplicates and aliases (e.g., "Foo Bar" vs "FooBar")
- Standardize cuisine tags and neighborhood classifications
- Validate coordinate data and ZIP codes

Step 3: Review Preprocessing

- Remove spam and low-quality reviews
- Tokenize and clean review text for embedding
- Extract key phrases (dish names, attributes like "spicy," "quiet," "good for groups")

Step 4: Embedding and Indexing



· Embed all reviews and restaurant descriptions using Sentence-Transformer

· Store embeddings + metadata in FAISS index for efficient retrieval

· Maintain auxiliary lookup tables linking FAISS indices back to restaurant IDs, health grades, and source metadata

Step 5: Continuous Updates

· Weekly or monthly refresh cycles to ingest new reviews from the Kaggle dataset

· Quarterly sync with NYC DOH inspection data to keep health grades current

· Incremental FAISS index updates to add new restaurants without full reindexing

- **MVP Features**

1. Personalized Restaurant Search

Food lovers can type free-text cravings such as *"I want spicy Thai crab curry near the East Village."* Foodie. AI understands their intent using natural language embeddings and semantic matching to recommend the most relevant restaurants based on cuisine, style, mood, and real review signals.

2. Location-Based Filtering

Users can search for restaurants near their current location by clicking the 'Location Set' button or in a specific area (e.g., Midtown, East Village, around Central Park). The system uses geographic metadata to find nearby spots that match preferences.

3. Restaurant Details Display with AI Insights

Each recommendation includes:

· Restaurant name, cuisine type, and star rating

· Address, distance, and neighborhood

· **Match Score**: Percentage-based semantic fit indicator (e.g., "79.1% match")

· **AI Insight**: LLM-generated explanation showing why this restaurant matches the user's query, grounded in real review evidence

· Highlights of dishes matching the user's query

· Summary of review sentiment and key themes

· Menu overview and hours

The AI Insight card is the centerpiece of the results display, providing users with immediate, transparent reasoning for the recommendation.

4. **Smart Ranking and AI-Powered Explanation**

Foodie. AI uses AI-powered semantic similarity scoring combined with review-based signals and health inspection data to rank restaurants according to food preference, location, and trustworthiness. For each recommendation, the system generates an AI Insight—a concise, evidence-based explanation that shows users exactly why the restaurant matches their query. This combination of intelligent ranking and transparent AI reasoning ensures every user gets the best possible match and understands why.

5. Simple and Intuitive User Interface

A clean, responsive web interface invites users to type what they're craving and instantly see top-matching restaurants with clear reasoning, turning every meal search into a confident culinary discovery.

- **Future Improvements and Roadmap**

Short-Term Enhancements (Q1–Q2 2026)

1. Richer Multi-Signal Ranking
Incorporate additional structured signals into the ranking function:

· Price level and budget constraints (extracted from review mentions and menu data)

· Peak-time crowding and wait-time estimates (derived from review frequency and temporal patterns)

· Temporal context (brunch vs. late-night, weekday vs. weekend recommendations)

· Special dietary accommodations (vegan, gluten-free, kosher, halal) detected via review mining

Implement a learn-to-rank model that combines embedding similarity with these features, trained on aggregated user engagement data (clicks, saves, bookings) to continuously improve ranking quality.

2. Aspect-Based Sentiment Analysis
Move beyond aggregate review scores to extract sentiment along specific dimensions:

- **Taste and Flavor Quality**: How reviewers describe dishes, flavor profiles, and food quality
- **Service and Hospitality**: Speed, attentiveness, and friendliness signals
- **Ambiance and Noise Level**: Atmosphere, crowd density, and suitability for different occasions
- **Value and Price-to-Quality Ratio**: Whether reviewers feel the restaurant is worth the cost

3. **Personalization and User Profiles**
Build user embeddings over time by tracking:

- Past restaurant choices and feedback (likes, saves, skips, written reviews)
- Explicit preferences (cuisine types, price range, dietary restrictions, occasion type)
- Behavioral signals (search history, time-of-day patterns, neighborhood focus)

Re-weight similarity scores for returning users so recommendations become increasingly personalized over time.

Medium-Term Extensions (Q3–Q4 2026)

1. **Dish-Level and Menu-Item Embeddings**
Move from restaurant-level to dish-level semantic matching:

- Embed individual menu items from the Kaggle dataset and link them to specific review passages
- Capture dish-specific sentiment (e.g., "ramen quality," "tiramisu richness," "falafel crispiness")
- Build a secondary FAISS index for dishes, enabling queries like "best tonkotsu ramen in Manhattan"
- Allow users to search for specific dishes and receive restaurants that excel at that item

2. **Enhanced Explainability and Transparency (Beyond MVP)**
Build on the existing AI Insight layer with deeper transparency options:

- **Expandable Review Evidence**: Allow users to expand AI insights to view actual review excerpts that support recommendations
- **Aspect-Based Breakdown**: Show AI-generated breakdowns by attribute (taste, service, ambiance, cleanliness)
- **Trade-Off Transparency**: Surface potential considerations like wait times or crowd levels
- **Confidence Indicators**: Show the system's confidence level based on review volume, recency, and consistency

3. **Temporal and Contextual Recommendations**

· **Time-of-Day Awareness**: Recommend different restaurants for breakfast, lunch, dinner, or late-night

· **Weather Adaptation**: Suggest cozy indoor spots on rainy days or outdoor seating on nice days

· **Event-Based Context**: Brunch spots for weekends, quick lunch for weekdays, date-night ambiance

· **Seasonal Menus**: Track and highlight restaurants with fresh, seasonal offerings

4. **Integration with Reservation and Booking Systems**

Partner with platforms like Resy and OpenTable to enable:

· Real-time availability checking and display in Foodie.AI results

· Seamless direct booking from recommendations

· Historical booking data to validate popularity signals

· Conversion tracking from recommendation to actual reservation

Long-Term Vision (2027+)

Scalability Across Cities and Regions

Extend beyond Manhattan to:

· Other NYC boroughs (Brooklyn, Queens, Bronx, Staten Island)

· Other major US cities (San Francisco, Los Angeles, Chicago)

· International expansion by partnering with local health authorities and review platforms

Multi-Modal Embeddings

Enhance embeddings to incorporate:

· **Food Photography**: Use vision models (e.g., CLIP) to understand dish appearance and presentation

· **Menu Images**: OCR to extract menu text from photos and embed alongside text embeddings

· **Audio/Video Reviews**: Extract signals from YouTube or TikTok food reviews

Community and Social Features

Build community engagement:

· User-generated reviews and ratings specific to Foodie.AI

· "Food Neighborhoods" or curated collections based on collaborative filtering

· Social discovery showing what friends and influencers are recommending

· Leaderboards recognizing trusted reviewers

- Integration with Instagram and TikTok to surface recommendations

Advanced AI and Recommendation Techniques
Implement cutting-edge approaches:

- **Fine-Tuned Embedding Models**: Train custom models on Foodie. AI's user interactions
- **Graph-Based Recommendations**: Use graph neural networks for pattern-based suggestions
- **Conversational AI**: Enable multi-turn dialogue for iterative search refinement
- **Predictive Personalization**: Predict user preferences and offer proactive suggestions

A/B Testing and Continuous Optimization
Implement systematic experimentation:

- Test different embedding models and FAISS configurations
- Optimize ranking strategies based on user engagement
- Experiment with LLM providers and prompting strategies
- Track conversion rates, user satisfaction, and retention metrics

Data Quality and Feedback Loop
Establish quality maintenance mechanisms:

- User feedback integration for rating recommendations
- Automated data validation for freshness and accuracy
- Active learning to identify and improve weak recommendations
- Community moderation to flag low-quality or spam reviews

- **Conclusion**

Foodie.AI transforms restaurant discovery in Manhattan by replacing proximity-based, generic recommendations with intent-matched, trust-centered results grounded in real user voices. By combining semantic embeddings, efficient vector search, intelligent filtering, and LLM-powered explanations, Foodie.AI will help food lovers, locals and visitors alike, discover restaurants with confidence.

The platform's core strength lies in its simplicity for users combined with technical depth: conversational input, transparent reasoning, and high-confidence recommendations. As we scale features like personalization, multi-signal ranking, and dish-level matching, Foodie.AI will become the go-to tool for anyone seeking not just a meal, but a discovery that genuinely aligns with their tastes and context.