



THE ERDŐS INSTITUTE  
Revolutionary Collaborations in  
Academia and Industry



# Flavor Finder

Review-Based Chat Client

by

Xue Xiao, Michael Shteyn, William Porteous,  
Andres Martinez, Zhihan Li, and Daniel Colón Amill

# Problem: How do we simplify ordering food?



Choosing from an unfamiliar menu is challenging and time consuming for diners

Diners want personalized guidance, but servers have limited time

Scrolling through hundreds of online reviews doesn't simplify the decision



# Solution: Automatic, Diner-Informed Recommendation System



**Target:** first-time diners, eaters with dietary restrictions, and enthusiasts wanting to maximize their restaurant experience

# Flavor-Finder: User Perspective

*What's  
something good  
to eat at Apteka,  
Pittsburgh?*



# Flavor-Finder: User Perspective

*What's something good to eat at Apteka, Pittsburgh?*



*The Szparagi, a black-pepper sauteed asparagus, is the most popular dish according to reviewers. Reviewers also liked...*

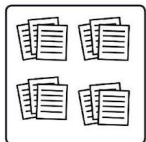


# Flavor-Finder: Backend (RAG)

 = HuggingFace Model

## Step 1: Process Data

Reviews



Filter with  
NER-Model



Reviews with  
Food Items



Text  
Embedding  
Model



Vector Database



 Pinecone

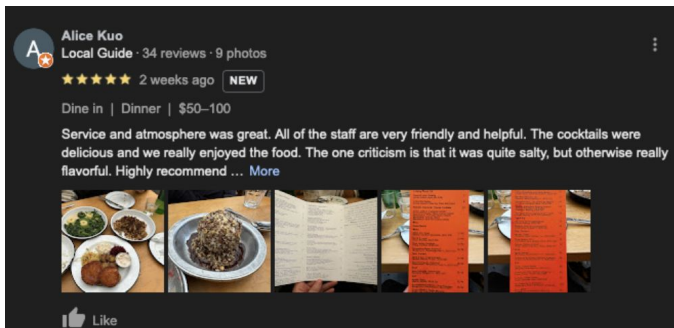
*What's  
something good  
to eat at Apteka,  
Pittsburgh?*



# Dataset

Google Local Data (2021), Pennsylvania  
T. Zhang & J. Li, UCSD

~13M reviews of ~190K businesses



name	rating	text	gmap_id
Judges Corner	4	The smothered pork chops are banging Yams ar...	0x882d7fc7b02c7a39:0xedc083a71e5be96d
Avocado From Mexico	5	I used to go here when I was a student at Linc...	0x89c64cb230711c3f:0x567be5ee6fea1ac6
Avocado From Mexico	5	Favorite place for tacos Their guacamole is v...	0x89c64cb230711c3f:0x567be5ee6fea1ac6
Avocado From Mexico	5	Quick tasty good value especially onTaco T...	0x89c64cb230711c3f:0x567be5ee6fea1ac6
Avocado From Mexico	5	Delicious beef tacos	0x89c64cb230711c3f:0x567be5ee6fea1ac6
...	...	...	...

## Reviews:

- ID of the reviewer
- Name of the reviewer
- Time of the review
- Rating of the business
- Text of the review
- Pictures of the review
- Business response to the review
- ID of the business

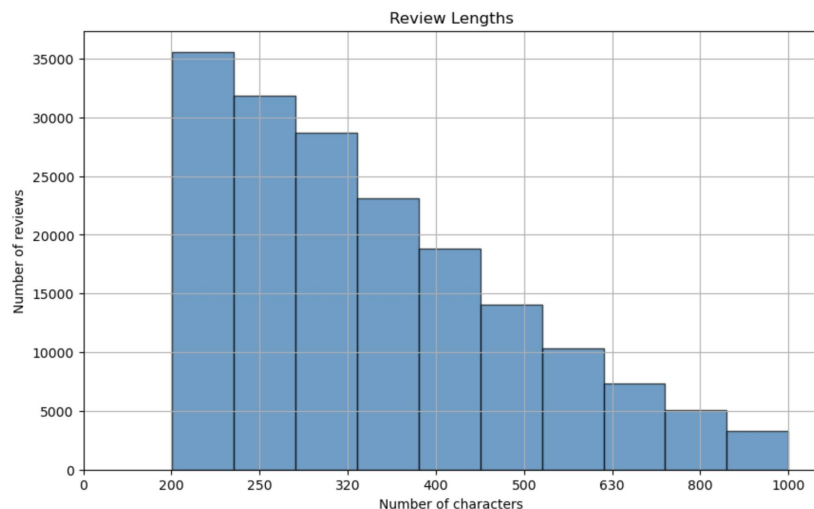
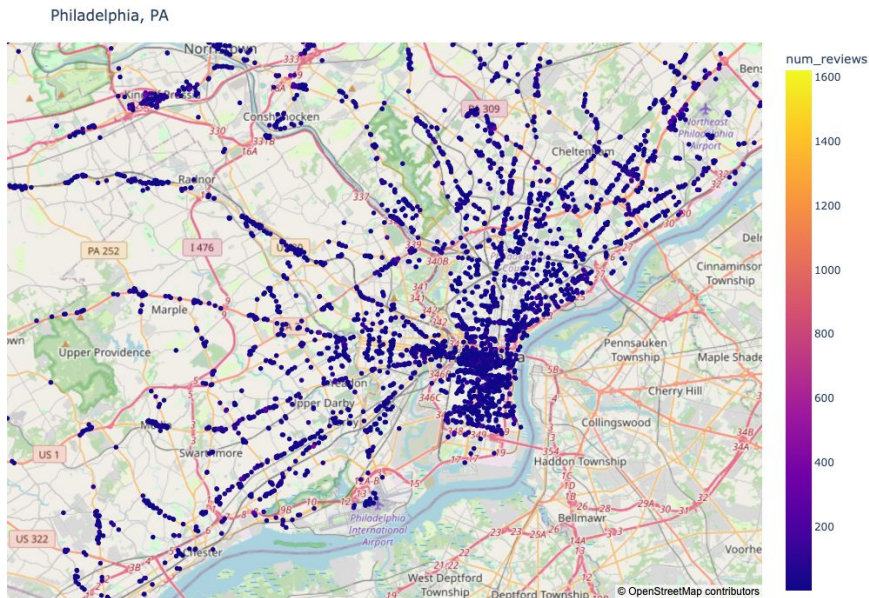
## Metadata:

- Name of the business
- Address of the business
- ID of the business
- Latitude
- Longitude
- Category
- Average rating
- Number of reviews
- URL of the business
- ... and more

# Data Cleaning

We filtered restaurant reviews with:

- Ratings of 4 or 5 stars
- Lengths between 200 and 1000 characters
- Food item mentions



~800k reviews of ~20k restaurants

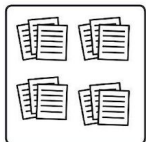


# Flavor-Finder: Backend (RAG)

 = HuggingFace Model

## Step 1: Process Data

Reviews



Filter with  
NER-Model



Reviews with  
Food Items



Text  
Embedding  
Model



Vector Database



 Pinecone

*What's  
something good  
to eat at Apteka,  
Pittsburgh?*



# Flavor-Finder: Backend (Data)

## Step 2: Put The R in RAG

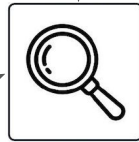
What's something good to eat at Apteka, Pittsburgh?



Embed User Query



Relevant Reviews



Search Vector-Database To **Retrieve** Reviews Relevant to Query

Reviews



Filter with NER-Model



Reviews with Food Items



Text Embedding Model



Vector Database

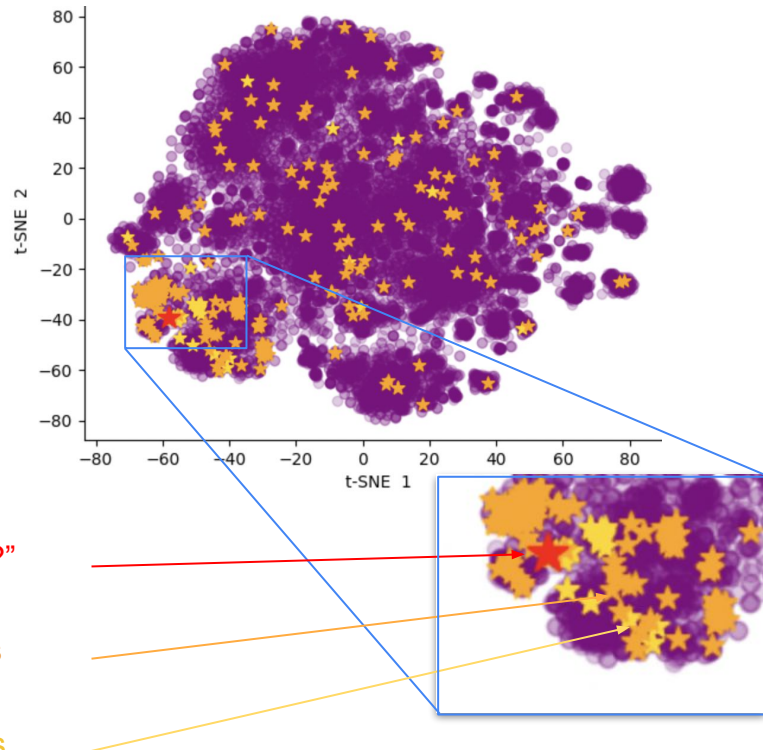
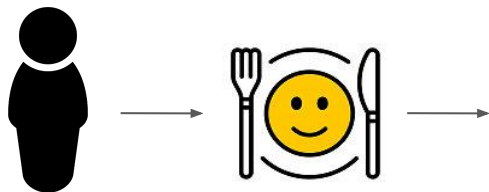


Pinecone

# Relevant Reviews are Selected by Similarity to Query

t-SNE: Mapping User Query In Vector Database  
(t-distribution Stochastic Neighbor Embedding)

*What's a good dish at Noodlehead?*



**“What’s a good dish at Noodlehead?”**

Reviews of Thai Restaurants

Reviews of noodle restaurants

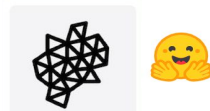
# Flavor-Finder: Backend (RAG)

## Step 2: Put The R in RAG

What's something good to eat at Apteka, Pittsburgh?



Embed User Query

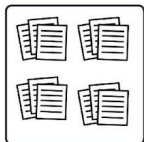


Relevant Reviews



Search Vector-Database To **Retrieve** Reviews Relevant to Query

Reviews



Filter with NER-Model



Reviews with Food Items



Text Embedding Model

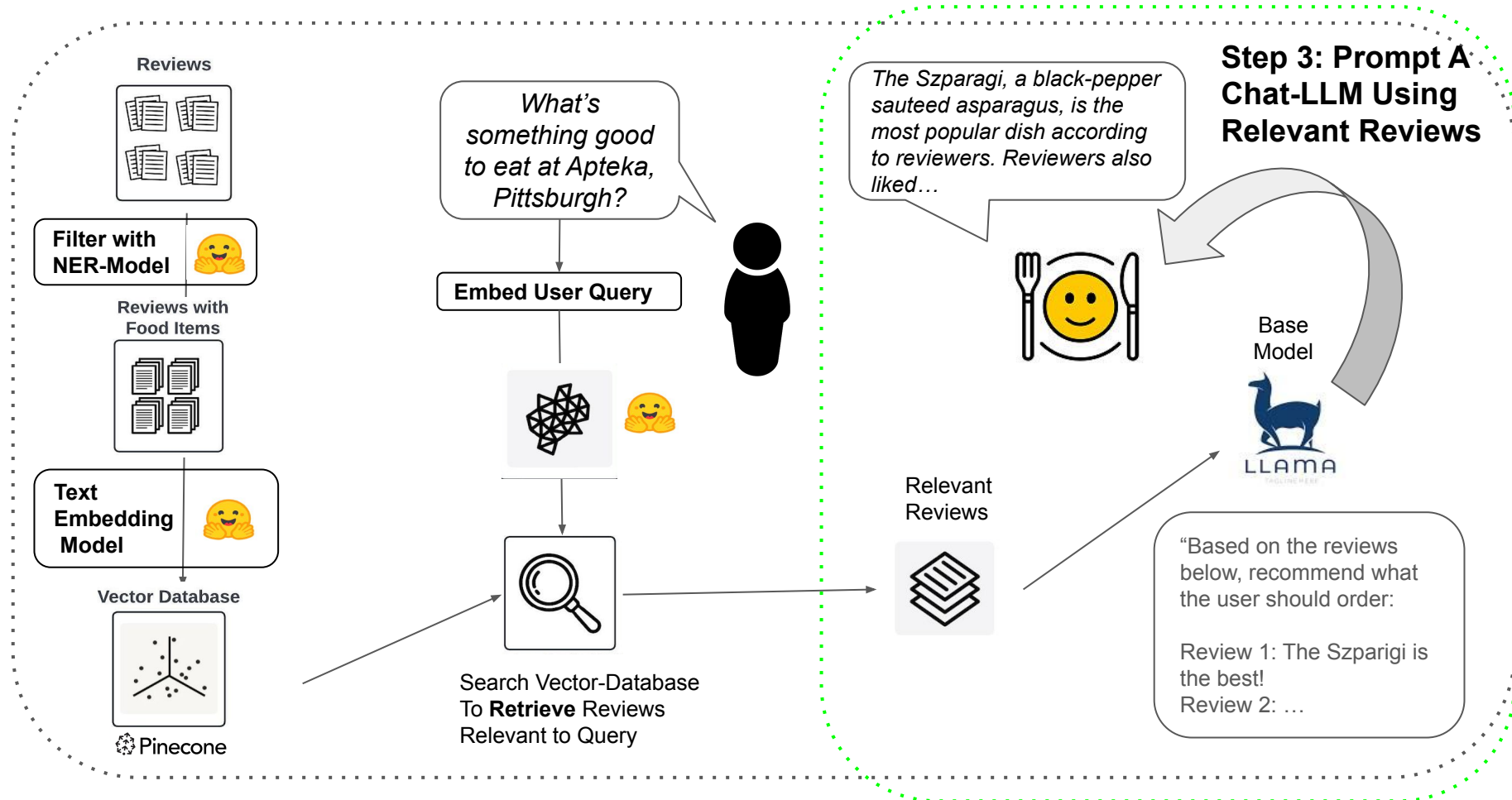


Vector Database



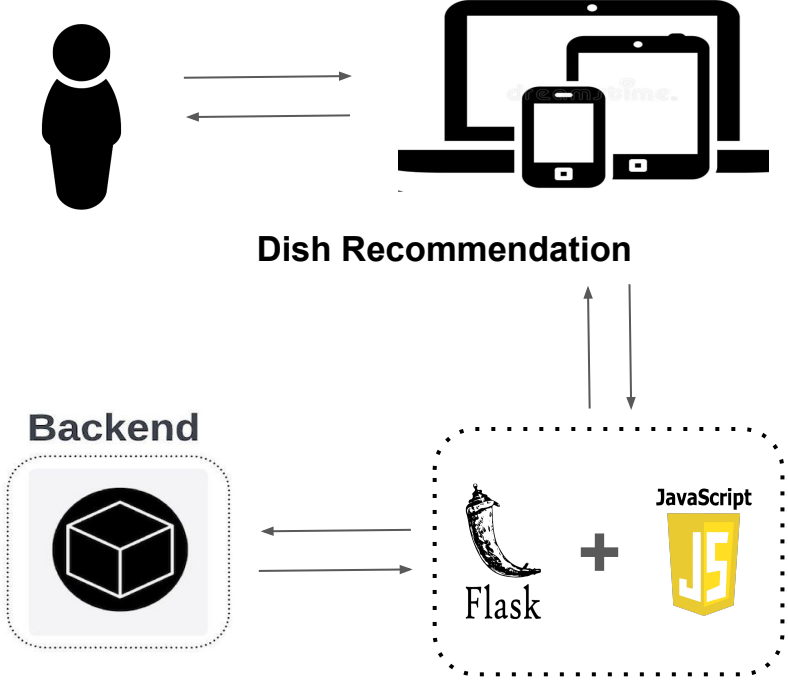
Pinecone

# Flavor-Finder: Front End



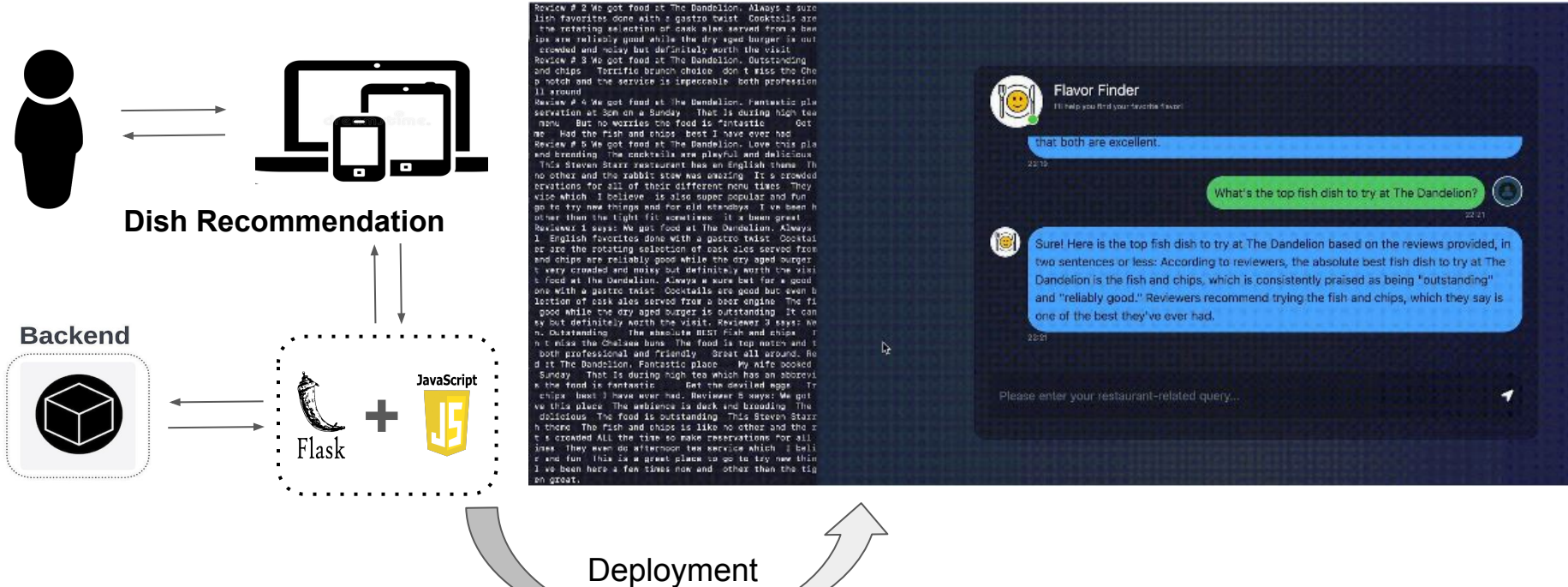
# Chatbot API: Frontend

Query: For the best vegetarian plate, what should I order at Noodlehead?



# Chatbot API: Frontend

Query: For the best vegetarian plate, what should I order at Noodlehead?



# Performance Criteria

Hold out set: test with custom untrained prompts

Wrong restaurant	0
Correct restaurant only	1
Correct restaurant and off-menu recommendation	2
Correct restaurant and incorrect menu item recommendation	3
Correct restaurant and correct menu item recommendation	4



Tell me, what's the winning seafood dish at Cooper's Seafood House?

Based on the reviews provided, the winning seafood dish at **Cooper's Seafood House** is the **Maryland Crab Bisque**.



**COOPER'S**  
Seafood House  
EST. 1948

Menu

## STARTERS

Maryland Crab Bisque

The one and only!

Cup ..... 6.50



# A Clear Upgrade From Base Model

## Average Recommendation Scores

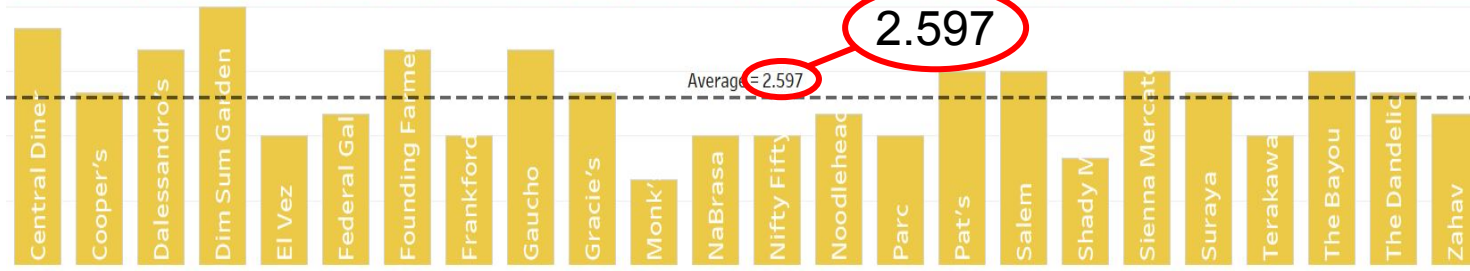
Pennsylvania local restaurants with >100 reviews with dish mentions



Flavor Finder



Base Model



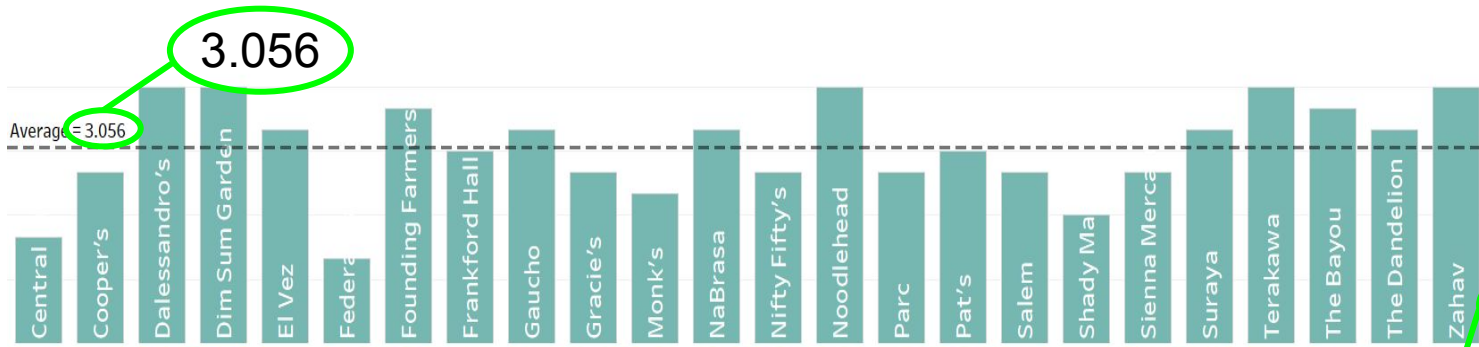
# Comparison with GPT-4o

## Comparative Scoring Against OpenAI Products

Pennsylvania local restaurants with >100 reviews with dish mentions



Flavor Finder



3.056

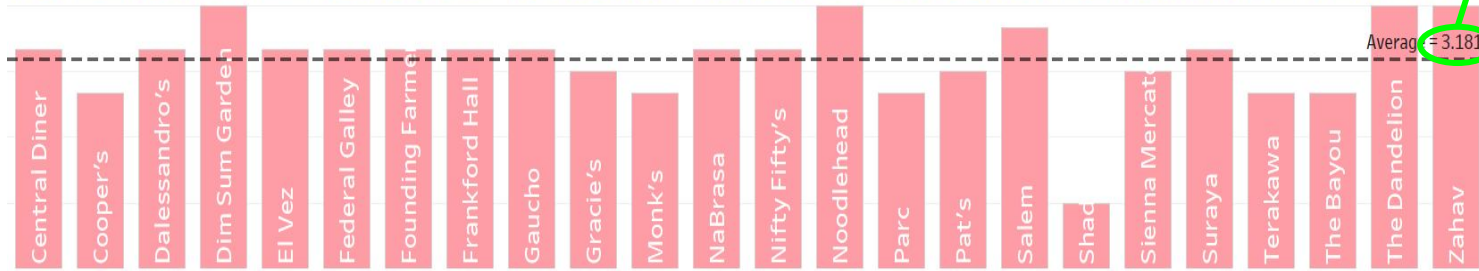
Average = 3.056

3.181

Average = 3.181

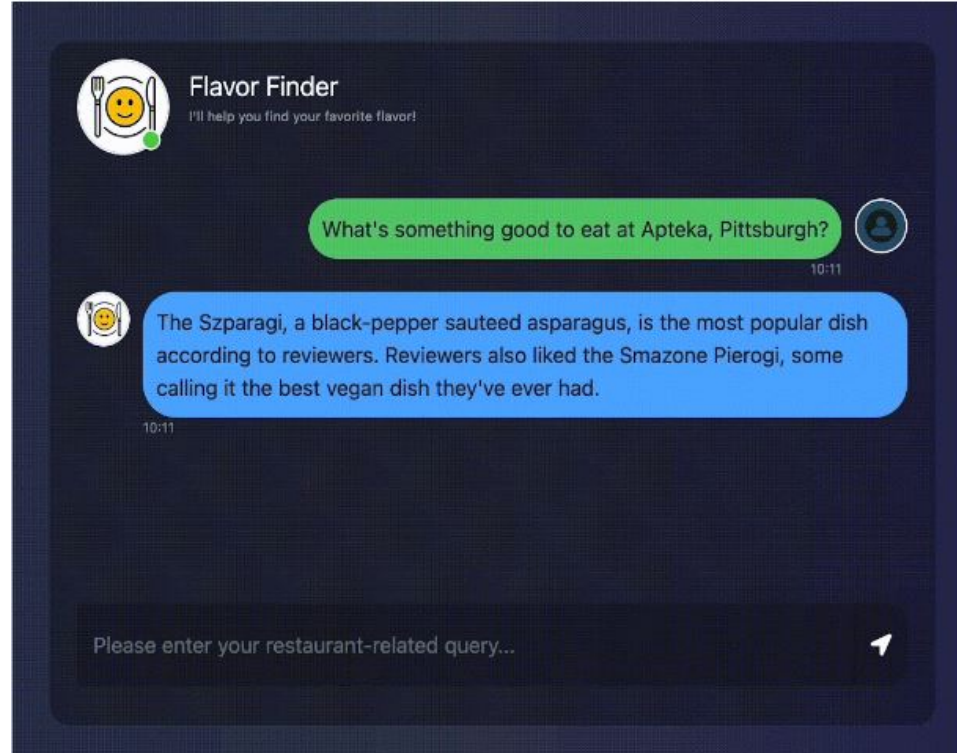


ChatGPT 4o



# In Conclusion

1. Built & deployed chat client which can inform menu-recommendations
2. Using RAG, we overcame limited pre-training data of open-source LLMs
3. Observed better performance than base Llama - can we beat OpenAI products?



# Next Steps

- **Near-term:**

- Deploy online with Amazon SageMaker
- Finish integrating live web-scraping of reviews and menus near the user location

- **Longer-Term:**

- Speed-up the retrieval system (important for live-user)
- Integrate menu “star-based” review system
- Handle negative review information
- Accept more diverse query formats



Google Places API



Amazon SageMaker



# Thanks To



**Hugging Face**



LangChain



PyTorch



Pinecone



Flask



LLAMA  
TAGLINE HERE

Special Thank You To Soheyl Anbouhi, Steve Gubkin, Alec Clott,  
Roman Holowinsky, & Everyone

@



THE ERDŐS INSTITUTE

Revolutionary Collaborations in  
Academia and Industry