

ReviewPal: A Restaurant Review Generator

Final Report

Word Count: 1994 Penalty: 0%

Tian Wang 1005982769
Yining Wang 1005728134
Yiping Wang 1005776144
Shiming Zhang 1011821129

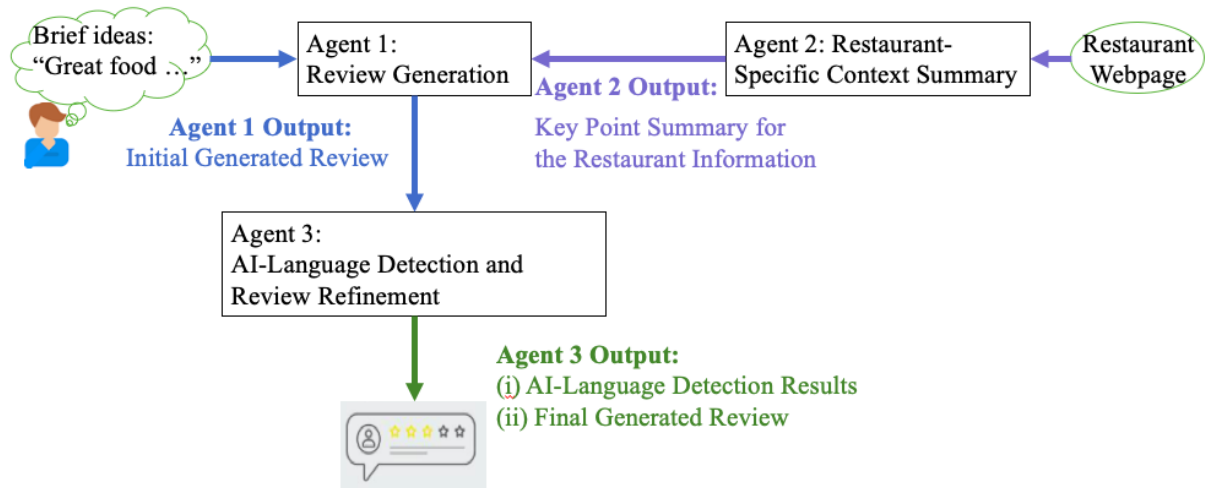
I. Permissions

	Tian Wang	Yining Wang	Yiping Wang	Shiming Zhang
Post Video	Yes	Yes	Yes	Yes
Post Final Report	Yes	Yes	Yes	Yes
Post Source Code	Yes	Yes	Yes	Yes

II. Introduction

Online reviews play an important role in guiding customer decisions, especially in the restaurant industry. Platforms like Google and Yelp enable users to share their experiences and have their opinions seen by more people. However, people usually have only a few quick thoughts to share, resulting in short and vague comments like “great food” or “poor service”, which don’t fully communicate the user’s experience or provide valuable insights to other customers. On the other hand, platforms offer rewards such as bonus points or exclusive events to encourage high-quality reviews, but drafting such reviews remains challenging for many people. Therefore, we want to develop a multi-agent system that leverages Large Language Models (LLMs) to help users turn quick thoughts into detailed reviews effortlessly.

III. Illustration



IV. Background & Related Work

Juuti et al. proposed a refined neural machine translation (NMT) model that generates context-specific restaurant reviews with greater authenticity and credibility [1]. Their NMT models were structured with stacked recurrent neural networks (RNNs) for review generation, whereas our approach will utilize LLMs. This methodological difference would emphasize the novelty of our project.

Peng et al. found that directly applying the LLMs for product review generation resulted in responses that lacked personalization. They explored the use of LLMs to generate product reviews with prompts based on user’s historical behaviors [2]. Their approach shares

similarities with our project in its use of LLMs to generate customized reviews. However, their approach was designed for product reviews, while our proposed system focuses on restaurant reviews. Moreover, our approach employs a multi-agent system, with one agent performing RAG to improve relevance and accuracy in review generation.

V. Data and Data Processing

It is difficult to gather enough raw user inputs, such as quick thoughts, in a consistent and evaluable format. Thus, we first collected real, high-quality reviews and then used a separate LLM to reverse-engineer them into brief opinions that simulate the test inputs.

We collected 100 authentic restaurant reviews from 20 different restaurants, with an equal split between Google and Yelp (50 reviews from each platform). To ensure high-quality data, we selected the top five reviews for each restaurant based on the "Yelp sort" or "Google relevance" filters, which prioritize reviews deemed most relevant and useful by the platforms.

There were 72 positive (rating > 3) and 28 negative (rating ≤ 3) reviews. This imbalance happened because most chosen restaurants had high ratings, and online platforms tend to feature more positive reviews. Despite the inherent imbalance, we ensured that at least one negative review was included for each restaurant to make the final dataset representative and capable of testing the system's ability to generate both positive and negative reviews.

We used GPT-4 to transform original reviews into simulated user inputs by following specific criteria: preserving the original sentiment, avoiding introducing new ideas, and simplifying key details. Initially, we tested the reverse-engineering process on a subset of 10 examples, refining the prompt iteratively until it successfully converted all examples into valid test inputs. Once finalized, this prompt was applied to process the remaining reviews. The average review length was reduced from 130 words to 34 words.

One example of an original review:

Chubby's in Toronto serves up delicious and authentic Jamaican flavors in a vibrant, Caribbean-inspired setting. The curry goat is a standout dish, packed with bold spices. The service is excellent, with attentive staff who offered up great recommendations. The laid-back yet lively atmosphere makes it a perfect spot for a flavorful and fun dining experience. Highly recommend for anyone craving quality Jamaican cuisine!

Its processed result:

Delicious authentic Jamaican flavors. Standout curry goat with bold spices. Excellent service with great recommendations. Vibrant, laid-back atmosphere. Highly recommend for quality Jamaican cuisine!

VI. Architecture and Software

Our system architecture is shown in the Illustration Section. The proposed multi-agent system takes the user's brief opinions along with the URL of the restaurant webpage as the input and generates a high-quality review as output. The system consists of three agents.

- Agent 1: It receives the user's idea, along with the restaurant-specific context summary provided by Agent 2, to generate a review. It is guided by a structured prompt that emphasizes expanding naturally on the user's input while using the restaurant context as a supporting element to augment the input without overshadowing it. The prompt also instructs the agent to avoid introductory statements or overly formal language, as well as emphasizes the importance of consistency, simplicity, and relevance.
- Agent 2: It processes the content of the restaurant's webpage, which our system automatically extracts from the provided restaurant URL using a code script to ensure a fully automated workflow. We prompted Agent 2 to analyze the extracted content, identify restaurant-specific information such as popular dishes, recent events, or other relevant details, and organize them into a key point summary to support the review generation process. The summary format is flexible, allowing Agent 2 to decide which key points from the webpage are most relevant for creating high-quality reviews.
- Agent3: Although Agent 1 is instructed to generate realistic and authentic reviews, its outputs still include excessive AI-like language. Hence, we added Agent 3 for review refinement.

We initially had Agent 3 directly refine Agent 1's review, but the results were unsatisfactory. To address this, we designed a prompt that employs the Chain-of-Thought (CoT) technique to better guide Agent 3's refinement process. This approach first instructs Agent 3 to detect and list overly AI-like phrases in Agent 1's output,

and then focus on refining these parts to ensure the final review adopts a more human-like tone.

VII. Comparison Method

We defined four criteria to evaluate the generated reviews. Each criterion was rated on a scale of 1 to 5 by a single group member to ensure consistent scoring. To be considered as successful, an output needs to score at least 3 in all criteria. The entire system needs to achieve a success rate of over 90% to be considered successful.

- **Consistency with user input** evaluates whether the review reflects the user's original thoughts and sentiment, as well as includes all the key points the user wants to emphasize.
- **Contextual relevance** evaluates whether the review contains relevant details about the restaurant, as accurate information enhances the credibility of the review.
- **Human-like tone** evaluates the review's naturalness and fluency. A human-like tone makes the review more readable and trustworthy, which can increase its likelihood of being recognized as authentic by the platform.
- **Overall quality** evaluates the review's overall coherence, meaningfulness, and usefulness.

VIII. Quantitative Results

Our system achieved a generation success rate of 96%, which met our pre-set objective (>90%). The table below summarizes the average score for each criterion.

	Consistency with User Input	Use of Restaurant-Specific Information	Human-Like Tone	Overall Quality
Average Score	4.87	4.87	3.89	4.57
Number of failed results (rating < 3)	0	0	4	0

The system performed well (i.e., achieving an average score > 4.5) in all criteria except for Human-Like Tone. This was the area where the system struggled the most, with all four failed generations attributed to the use of excessive AI language. Common issues include overly formal or generic phrasing and long, complicated clauses, which we've tried to

address through refining our prompt design. However, there is still room for future improvement to make the generated reviews more natural and engaging.

IX. Qualitative Results

In this section, we present an example of the user input and the final generated review, along with an explanation of how well our system works. The intermediate outputs from each agent are provided in Appendix A for reference.

User Input:

Delicious authentic Jamaican flavors. Standout curry goat with bold spices. Excellent service with great recommendations. Vibrant, laid-back atmosphere. Highly recommend for quality Jamaican cuisine!

(Indeed, it also comes with [the restaurant URL](#). We chose to use the restaurant’s TripAdvisor webpage because its content is easily accessible and extractable, whereas platforms like Yelp or Google Reviews require an API key for data access.)

Final Generated Review:

Chubby's Jamaican Kitchen offered a delicious taste that transported me right to Jamaica. The curry goat was definitely the star—with spices hitting just the right notes and the meat melting into the curry. What I loved was how friendly and helpful the staff were—they knew their stuff and made great suggestions. The place had a relaxing island feel with colorful decor, and you could hear gentle island music softly playing, setting the perfect vibe. For anyone craving authentic Jamaican food, Chubby's is a must. I left with a happy heart and a fuller belly!

Explanation: The generated review effectively captures all the key ideas from the user input, including the delicious Jamaican flavors, standout curry goat, excellent service, and vibrant atmosphere. The sentiment is positive, which remains consistent with the original input. Information retrieved about the restaurant's tropical-themed decor and relaxing island atmosphere is seamlessly incorporated to augment the user’s experience. The review uses simple and conversational language, which makes it easy to read and relatable. Overall, the output is natural, coherent, and provides meaningful insights that would be valuable to anyone considering a visit to Chubby's Jamaican Kitchen.

X. Discussion and Learnings

The system effectively leverages RAG to boost the review generation process. By combining user input with relevant retrieved information, the system generates reviews that are both detailed and contextually enriched. For example, when a user mentioned “lamb,” the system identified the specific dish (e.g., Braised Lamb Shoulder) from the retrieved data and expanded on its description, which added depth and relevance to the review. This capability ensures that even when user input is brief or vague, the generated review remains informative and engaging.

Furthermore, the system properly handles inconsistencies between user input and retrieved information. It always prioritizes the user inputs, with retrieved information only selectively used without altering the user’s intent. An example in Appendix B shows how the system retained the user’s negative sentiment about food quality and expanded on it with specific details, instead of focusing on the retrieved data that mentioned the restaurant’s good food quality and positive reputation.

Lastly, the system effectively utilizes LLM capabilities to derive meaningful advice and insights based on the user’s experience. For example, the system could suggest making reservations when a user mentioned long waiting times or recommend using public transit if limited parking was noted. These insights add value to the generated reviews, making them not only reflective of user sentiment but also practically useful to future customers.

One key takeaway is that, while LLMs could be effectively and creatively applied to natural language generation tasks to save human effort, overcoming AI-like language remains a big challenge. Ensuring the generated results maintain a personal, human-centered, and stylistic touch are the most difficult parts. In future projects, we would prioritize collecting sufficient real inputs from actual users, as time permits, since some of the AI-like language in the generated results appeared to stem from the synthetic test inputs we used.

Additionally, we would improve the system's usability by making each agent's output visible to users and enabling user interaction at any step of the process. By accepting user feedback dynamically, the system could further refine the generated reviews to make them sound more authentic and aligned with the user’s expectations better.

XI. Individual Contributions

Tian Wang

- Collected 25 Yelp reviews.
- Designed the reverse-engineering prompt.
- Labeled 50 outputs after reverse-engineering.
- Modified Agent 1's prompt and designed Agent 3 based on the two-agent one.
- Generated the final outputs of the system.
- Evaluated the outputs on Consistency criterion.

Yining Wang

- Collected 25 Google reviews.
- Designed Agent 1's prompt.
- Developed the single-agent pipeline.
- Tested the final pipeline design for better generation results.
- Evaluated the outputs on AI-tone criterion.

Yiping Wang

- Collected 25 Yelp reviews
- Designed Agent 2's prompt.
- Developed the two-agent pipeline based on the single-agent one.
- Evaluated the outputs on Contextual-Relevance criterion
- Organized and analyzed the evaluation results.

Shiming Zhang:

- Collected 25 Google reviews.
- Designed the reverse-engineering prompt.
- Labeled 50 outputs after reverse-engineering.
- Evaluated the outputs on Overall-Quality criterion.

XII. Reference

- [1] M. Juuti, B. Sun, T. Mori, and N. Asokan, "Stay on-topic: Generating context-specific fake restaurant reviews," in *European Symposium on Research in Computer Security*, Cham: Springer International Publishing, 2018, pp. 132-151.
- [2] Q. Peng, H. Liu, H. Xu, Q. Yang, M. Shao, and W. Wang, "LLM: Harnessing Large Language Models for Personalized Review Generation," *arXiv preprint* arXiv:2407.07487, 2024.

Appendix A

1. Complete output from Agent 2:

*****Chubby's Jamaican Kitchen Summary*****

- *****Cuisine Style:*****
 - *Caribbean*
 - *Jamaican*
 - *Vegetarian friendly*
 - *Offers gluten-free options*

- *****Restaurant Features:*****
 - *Located at 104 Portland St, Toronto, Ontario M5V 2N2, Canada.*
 - *Price Range: \$6.00 to \$25.00*
 - *Meals offered: Lunch, Dinner, Brunch, Drinks*
 - *Features: Takeout, Reservations, Outdoor Seating, Serves Alcohol, Full Bar, Wine and Beer, Accepts Credit Cards, Table Service*
 - *Atmosphere is lively with tropical-themed decor, reminiscent of a Caribbean island setting.*
 - *The venue includes seating on two levels in a refurbished old house with wooden beam ceilings and indoor plants.*

- *****Popular Dishes/Drinks:*****
 - *Pepper shrimp praised for its perfect kick and flavorful sauce*
 - *Jerk chicken described as tender and flavorful*
 - *Turkey and vegetable patties well-liked for appetizers*
 - *Saltfish fritters and oxtail, though oxtail reviews are mixed*
 - *Goat curry and rice, oxtail, and jerk pork frequently highlighted*
 - *Popular drinks include rum punch, Red Stripe, Jamaican mule, and ginger beer*

- *****Customer Ratings:*****
 - *Overall rating of 4.5 out of 5, based on 135 reviews, placing it 126 out of 9,203 Restaurants*

2. Complete output from Agent 1:

Chubby's Jamaican Kitchen delivers a delightful wave of authentic Jamaican flavors, closely resembling a culinary journey to the Caribbean. The standout dish, without a doubt, is the curry goat. Its savory richness laced with bold spices creates a transcendent experience that I find myself vividly reminiscing about. I appreciate how the tender meat seamlessly melds into the fragrant curry sauce, creating a dish as heartwarming as it is flavorful. The service complemented the meal beautifully, each staff member brought joy with knowledgeable recommendations that truly made a difference in the dining experience. The ambiance exudes a laid-back tropical vibe, further immersing you in a Jerk-fueled feast, accompanied by vibrant decor and soothing island tunes playing low in the background. Chubby's Jamaican Kitchen is a high recommendation for anyone longing for excellent Jamaican cuisine in an atmosphere both warm and engaging.

3. Complete output from Agent 3:

****CoT****

1. *"delivers a delightful wave of authentic Jamaican flavors"*
2. *"closely resembling a culinary journey to the Caribbean"*
3. *"transcendent experience"*
4. *"vividly reminiscing"*
5. *"seamlessly melds into"*
6. *"heartwarming as it is flavorful"*
7. *"The service complemented the meal beautifully"*
8. *"knowledgeable recommendations that truly made a difference"*
9. *"ambiance exudes a laid-back tropical vibe"*
10. *"further immersing you in a Jerk-fueled feast"*
11. *"soothing island tunes playing low in the background"*
12. *"a high recommendation"*
13. *"atmosphere both warm and engaging"*

****Agent 3 Rewrite****

Chubby's Jamaican Kitchen offered a delicious taste that transported me right to Jamaica. The curry goat was definitely the star—with spices hitting just the right notes and the meat melting into the curry. What I loved was how friendly and helpful the staff were—they knew their stuff and made great suggestions. The place had a relaxing island feel with colorful decor, and you could hear gentle island music softly playing, setting the perfect vibe. For anyone craving authentic Jamaican food, Chubby's is a must. I left with a happy heart and a fuller belly!

Appendix B

1. An example of negative user input:

*Great welcoming and cozy vibe. Beautiful interior. Top patio added to the experience. Service was lacking; waitress seemed unhappy. **Food arrived cold:** oily kale salad with cold chicken, burnt pizza crust, and cold steak even after sending it back. Potential is there, but major issues affected the experience. Rating: 3/5. Hope they improve.*

2. **Agent 2 Output** (According to the extracted information from the restaurant's webpage, Trattoria Nervosa is mostly positively rated. However, as highlighted in red (food quality and overall restaurant feedback), the retrieved data reveals some contradictions with the user's input) :

****Trattoria Nervosa Summary****

- ****Cuisine Style:****

- *Italian*
- *Offers Pizza, Vegetarian friendly options*
- *Vegan and Gluten-Free options available*

- ****Popular Dishes:****

- *Kale Salad: Notable for being highly recommended even by those who don't usually enjoy kale*

- *Pasta: Includes varieties like pasta with a red sauce and shrimp, Mafalde ai Funghi with mushrooms and cream sauce, spicy shrimp pasta, and Risotto Nervosa with beef short rib*

- *Pizza: Known for prosciutto and arugula toppings, considered by some to be the best pizza ever*

- *Branzino: Grilled fish dish with accompanying sides*

- *Desserts: Highly praised based on observed satisfaction from nearby patrons*

- ****Notable Features:****

- *Located in Yorkville, Toronto*

- *Recognized with a Travelers' Choice Award*
- *Offers a cozy and friendly neighborhood dining experience*
- *Operates on a first-come, first-served basis*
- *Provides delivery, takeout, reservations, outdoor seating, alcohol service, and free WiFi*
- *****Recent Events & Observations:*****
 - *Generally described as a place **with good food quality**, but with some negative feedback on service, particularly concerning the behavior of the hostess and handling of customer issues*
 - *Pricing is seen as reasonable by some and expensive by others, particularly with comments on portion sizes in takeout*
- ***Positive***

3. The final generated review (the final generated review remains negative, with the description of the food quality unchanged from the user's input rather than reflecting the retrieved information):

*Trattoria Nervosa feels like a quaint corner of southern Italy. The rooftop patio's open-air setup was a treat. But the visit took a nosedive with service; our waitress seemed far from happy, giving things an awkward start. **The food didn't help.** The kale salad, sadly chilled by clunky chicken pieces, missed the mark. The pizza could have passed for a charcoal sculpture, and even after a rerun, the steak was still as chilly. It's frustrating since this spot has been treasured in Yorkville since 1996. Such a charming venue should reflect better on the plate. Hopefully, they can step up soon. For now, a 3 out of 5 – waiting for the next visit to be more delightful.*