Yelpy: A Generative AI Framework for Grounded, Sentiment-Controlled Review Synthesis

Ribence Kadel [1] **Austin Robinson** [1] **Fisher Ramsey** [1] **Jorge Castaneda** [1]

Abstract

In the restaurant industry, online review platforms are crucial for business reputation and consumer discovery. This serves to benefit those with an established supportive clientele, however newer establishments lacking customers and reviews often have a tough time gaining an online presence for visibility. We present Yelpy, a conceptual framework utilizing Large Language Models (LLMs) for effortless generation of artificial, highly engaged customer reviews. By using a prompt grounding technique, our system conditions the generative process on several inputs such as the restaurant's menu, name, and cuisine type. Allowing for the creation of specific tailored reviews referencing actual menu items. Additionally, the system allows for control over the reviews sentiment (positive/negative) and quantity. Giving the capability for boosting one's visibility and "review bombing" competitors. This paper goes into depth on the system architecture, the prompt methodology, and the dual use implications of this tech.

1. Introduction

In our modern society's digital age, online reviews on platforms like Yelp, Google, Tripadvisor, etc. have become a form of social credibility for establishments. How highly and the amount of ratings have a significant impact on consumer behavior and the discovery of a restaurant, playing a major role in a business's financial success. Meaning a lack of existing reviews causes a business to remain invisible to the platform and potential customers, creating a wall for new owners to overcome organically.

This obstacle has created a black market for fake reviews (Luca & Zervas, 2016), however these reviews are often generic and detectable putting the business at risk for platform penalties. We introduce a more personalized and conceptual approach: could generative AI create artificial reviews that are indistinguishable from those of real customers?

To answer this we created Yelpy, a tool designed to explore this question. Yelpy is a web app that allows a user to create batches of natural sounding reviews. Unlike a generic text generator, Yelpy conditions the LLM's output on specific parameters related to said restaurant. Allowing synthetic reviews to appear more natural by referencing specific attributes related to the business. Simulating a consumer who genuinely visited the establishment.

Additionally the system allows control over two additional parameters:

- **Sentiment:** A toggle allowing the user to generate "Positive" or "Negative" reviews.
- **Quantity:** A counter that allows the user to generate a batch of 1-10 distinct reviews at once

This sentiment toggle is a vital component of our application. A new restaurant could not only utilize the "Positive" feature to establish a solid online presence, but maliciously use the "Negative" setting to "review bomb" rivals. Generating realistically sounding reviews tailored to their actual menu to create reputational damage.

2. Related Work

Our idea aligns with several key research areas:

Sentiment Analysis: Classifying text based on its polarity (positive, negative, neutral) is a well established idea (Pang & Lee, 2008; Truong, 2025). Our work inverts this by using sentiment as an input used for generation rather than a label placed on the output.

Generative Language Models: The innovation of RNNs (Sutskever et al., 2014) to Transformers (Vaswani et al., 2017) and large-scale pre-training (Brown et al., 2020; Ope-

[1] Department of Computer Science, Conceptual University. Correspondence to: Ribence Kadel <kadel@conceptual.edu>.

Proceedings of the 39^{th} International Conference on Machine Learning, Baltimore, Maryland, USA, PMLR 162, 2022.

nAI, 2023) has created models capable of producing human-like text. We leverage these advancements.

Grounded Text Generation: Research has been done on grounding text generation on inputs such as tables or documents (Lewis et al., 2020; Wiseman et al., 2017). Yelpy employs a simplistic but effective form of this with grounding the generation in specific restaurant text details.

Synthetic Data & Misinformation: As LLMs get stronger, their potential for creating misinformation from news articles to reviews has become a major concern (Gehrmann et al., 2019; Borji, 2023; Associated Press, 2024). Solutions have been developed for detecting generated text, however this has quickly become a game of cat and mouse (Ippolito et al., 2020). Yelpy serves as one of these tools capable of misinformation.

3. Methodology: The ”Yelpy” System

Yelpy is designed as a full-stack application using Next.js (with Turbopack) as seen in the `package.json`. The system's divided in two with a user-facing frontend and a serverless API backend.

3.1. System Architecture and UI

The frontend, defined by `app/page.tsx`, is formatted with a two-column interface (Figure 1, conceptual). The left column being the `InputForm` component, and the right column displaying the generated reviews.

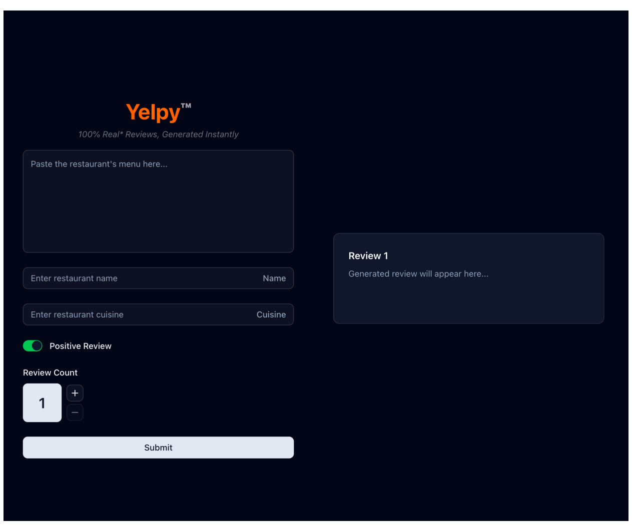

Figure 1. UI of the Yelpy system, showing the input form on the left and the generated review output on the right.

The `InputForm` (from `components/input-form.tsx`) is the main user interface, responsible for collecting five key parameters:

- **Menu Text:** A large `InputGroupTextarea` where the user pastes the restaurant's menu.
- **Restaurant Name:** An `InputGroupInput`.
- **Cuisine:** An `InputGroupInput`.
- **Sentiment:** A `Switch` component (from `components/ui/switch.tsx`) that toggles a React state `isPositive`.
- **Review Count:** A `CounterButton` component (from `components/counter-button.tsx`) that allows the user to select a value between 1 and 10.

After a user submits, the `handleSubmit` function establishes a processing state sending a `POST` request to the backend API.

3.2. Core Generation Engine

The backend logic is contained within `app/api/generate-reviews/route.ts`. This route uses OpenAIs API SDK (`openai` v6.8.1) to interface with the `gpt-4.1-mini` model.

The API route performs the following steps:

- **Parses Request:** It de-structures the `menuText`, `restaurantName`, `cuisine`, `isPositive`, and `reviewCount` from the incoming JSON request body.
- **Sets Parameters:** It translates the `isPositive` boolean into a `tone` string (”positive” or ”negative”) and clamps the `reviewCount` to a safe maximum of 10.
- **Constructs Prompt:** It dynamically constructs a detailed prompt, which is the core of our methodology (Gao et al., 2023).
- **Calls LLM:** It sends the prompt to the OpenAI API, specifically requesting a JSON object as output.
- **Returns Response:** It parses the LLM's JSON output and returns the array of review strings to the frontend.

3.3. Prompt Engineering for Grounded Generation

Our artificial reviews quality is based entirely on how we engineer our prompt (Gao et al., 2023). The template as found in `route.ts`, is as follows:

```
Using this menu:

${menuText}

Write ${count} distinct, ${tone}, natural-sounding
customer reviews for this restaurant:

Name: ${restaurantName}
Cuisine: ${cuisine}

Return ONLY valid JSON in this exact shape...:
{
  "reviews": [
    "review 1 text",
    "review 2 text"
  ]
}
```

This formation works to achieve several goals:

- **Groundedness:** By providing an entire menu, the LLM is able to mention specific items that are accurate, making the reviews appear more authentic.
- **Control:** The `$tone` and `$count` give us direct control over the sentiment and quantity of our output.
- **Structure:** The explicit JSON output instruction simplifies parsing on the backend.

This context grounding is highly effective and requires no model fine-tuning, demonstrating the powerful capabilities of modern LLMs.

4. Limitations

- **Hallucinated Details:** Large Language Models (LLMs) can generate convincing but inaccurate content (Ji et al., 2023), such as referencing menu items that don't exist. Yelpy has no built-in fact-checking, so hallucinated details can reduce authenticity.
- **Tone Control Granularity:** The `isPositive` boolean switch in `input-form.tsx` translates to only "positive" or "negative" tone strings in `route.ts`, preventing generation of nuanced mixed-sentiment reviews like realistic 3-star critiques.
- **Output Format Fragility:** Despite prompting for strict JSON, the LLM occasionally returns malformed or extra-text responses, complicating automated parsing.
- **Bias and Stereotypes:** The LLM's training biases may surface in generated reviews (Bommasani et al., 2021), and our prompt in `route.ts` requesting "natural-sounding" text provides no mechanism to filter or detect such stereotypes.
- **Lack of Iterative Feedback:** Yelpy does not include mechanisms for refining outputs or flagging inaccuracies, which could improve review quality.

5. Ethical Considerations

- **Dual-Use Potential:** The `isPositive` toggle in `input-form.tsx` explicitly enables both constructive bootstrapping (positive reviews) and malicious review bombing (negative reviews) through the same interface, making Yelpy a textbook dual-use tool.
- **Deception and Consumer Trust:** Synthetic reviews generated by Yelpy may mislead users into trusting inauthentic experiences. This undermines consumer trust in platforms that rely on honest user feedback (Associated Press, 2024).
- **Legal Risks:** Tools like Yelpy may violate the FTC's final rule announced in August 2024, which explicitly prohibits AI-generated fake reviews and testimonials used for deceptive marketing purposes (Federal Trade Commission, 2024).
- **Economic Harm:** Negative synthetic reviews can cause measurable financial damage to restaurants, especially small businesses that rely on online reputation. Research shows that one additional Yelp star can increase restaurant revenue by 5-9%, demonstrating the significant economic stakes of review manipulation (Luca & Zervas, 2016).
- **Platform Integrity:** Platforms like Yelp and Google are investing heavily in fake review detection. Proliferation of tools like Yelpy risks polluting these ecosystems and triggering stricter moderation policies (Gu & Spann, 2025).

6. Future Work

- **Fine-Grained Control:** Expand control beyond sentiment polarity to allow star ratings, tone/style, or emphasis areas (e.g., service vs. food).
- **Reviewer Personas:** Add persona selection UI components to `input-form.tsx` (e.g., "foodie", "critic", "casual diner") and incorporate these into the prompt template in `route.ts` to simulate diverse writing voices beyond the current style variation rules.
- **Multilingual and Cultural Support:** Expand `input-form.tsx` to include language/region selectors and update the prompt in `route.ts` with

cultural review pattern instructions, while extending `names.ts` with culturally appropriate name pools.

- **Evaluation Pipeline:** Implement a scoring system in `route.ts` that evaluates generated reviews against authenticity metrics before returning to the frontend, possibly adding confidence scores to the `Review` interface in `types.ts`.

- **Detection and Safeguards:** Modify `route.ts` to inject cryptographic watermarks into the LLM output text before parsing and assignment of random names from `names.ts`, allowing platforms to trace AI-generated content.

7. Competition

Most existing tools address fake reviews through detection rather than generation. Platforms typically rely on classification models, behavioral analysis, and anomaly detection to identify suspicious activity (Gehrmann et al., 2019; Ippolito et al., 2020). These systems do not provide mechanisms for automatically creating review text.

Yelpy fills this gap by focusing on controlled generation. The system uses:

- **A structured Next.js form** that collects inputs such as restaurant name, cuisine, menu text, sentiment, and number of reviews.

- **A backend prompt builder** that converts these fields into a grounded prompt containing specific restaurant details.

- **Batch LLM generation** that produces multiple distinct reviews in one call.

This design enables Yelpy to create realistic, high-variation reviews in bulk. Unlike detection-focused systems, Yelpy shows how targeted prompt engineering and grounding can be used to generate content that closely resembles natural user reviews.

8. Cross-Platform Applicability

Yelpy's architecture is modular and can be adapted to other review platforms with minimal changes. The core pipeline remains the same:

1. Collect structured metadata in the frontend.

2. Serialize it to JSON.

3. Inject it into a domain-specific prompt template.

4. Generate multiple grounded reviews in a single backend request.

Only the input fields and template wording need modification:

- **Amazon:** replace menu text with product specifications, features, and target star rating (Schermerhorn, 2024).

- **TripAdvisor:** collect hotel or attraction details such as amenities, room type, or location (Pan & Gang, 2025).

- **Google Reviews:** combine product, service, and location inputs depending on the entity type.

Because the JSON schema, API call structure, and batch-generation logic remain identical, Yelpy can be adapted to new review ecosystems simply by changing the form fields and the prompt template used to ground the generated text.

9. Conclusion

Yelpy shows that with the right prompt structure and a few well-chosen inputs, an LLM can generate restaurant reviews that feel specific, varied, and grounded in real menu details. By combining simple user controls like sentiment and review count with a consistent JSON output format, the system reliably produces batches of readable, tailored reviews without extra model training. At the same time, the project highlights how easily these tools can be misused to shape online ratings in misleading ways. Moving forward, adding stronger guardrails, better sentiment control, and ways to check or label generated text will be important for keeping systems like Yelpy both useful and responsible.

References

Associated Press. Fake ai-generated reviews are all over the internet, research shows. *Fast Company*, Dec 2024. URL https://www.fastcompany.com/91252048/fake-ai-generated-reviews-internet-research.

Bommasani, R. et al. On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*, 2021.

Borji, A. Synthetic data generation: A survey. *arXiv preprint arXiv:2301.01381*, 2023.

Brown, T. B., Mann, B., Ryder, N., Subbiah, M., Kaplan, J., Dhariwal, P., Neelakantan, A., Shyam, P., Sastry, G., Askell, A., et al. Language models are few-shot learners. *Advances in neural information processing systems*, 33: 1877–1901, 2020.

Federal Trade Commission. Federal trade commission announces final rule banning fake reviews and testimonials. FTC Press Release, Aug 2024. URL https://www.ftc.gov/news-events/news/press-releases/2024/08/federal-trade-commission-announces-final-rule-banning-fake-reviews-testimonials.

Gao, L. et al. Prompting: Better ways of using language models for nlp tasks. In *EMNLP Tutorial*, 2023.

Gehrmann, S., Strobelt, H., and Rush, A. M. Gltr: Statistical detection and visualization of generated text. *arXiv preprint arXiv:1906.04043*, 2019.

Gu, W. K. M. and Spann, M. The impact of generative artificial intelligence on consumer reviews: Insights from a quasi-experiment. *SSRN Electronic Journal*, 2025. doi: 10.2139/ssrn.5328200.

Ippolito, D. et al. Automatic detection of generated text is easiest when humans are fooled. In *ACL*, 2020.

Ji, S., Pan, S., Cambria, E., Kong, L., and Zhao, J. Survey of hallucination in natural language generation. *ACM Computing Surveys*, 2023.

Lewis, P., Perez, E., Piktus, A., Petroni, F., Karpukhin, V., Nogueira, R., He, H., Genc, H., Kiela, D., and Riedel, S. Retrieval-augmented generation for knowledge-intensive nlp tasks. *Advances in Neural Information Processing Systems*, 33:9459–9474, 2020.

Luca, M. and Zervas, G. Fake it till you make it: Reputation, competition, and yelp review fraud. *Management Science*, 62(12):3412–3427, 2016.

OpenAI. Gpt-4 technical report. Technical report, OpenAI, 2023.

Pan, M. and Gang, V. Travel faqs made smarter: Ai-driven answers from real-world discussion. Tripadvisor Tech Blog (Medium), Apr 2025. URL https://medium.com/tripadvisor/travel-faqs-made-smarter-ai-driven-answers-from-r

Pang, B. and Lee, L. Opinion mining and sentiment analysis. *Foundations and Trends® in Information Retrieval*, 2(1–2):1–135, 2008.

Schermerhorn, V. How amazon continues to improve the customer reviews experience with generative ai. Amazon Newsroom, 2024. URL https://www.aboutamazon.com/news/amazon-ai/amazon-improves-customer-reviews-with-generative-

Sutskever, I., Vinyals, O., and Le, Q. V. Sequence to sequence learning with neural networks. In *Advances in neural information processing systems*, volume 27, 2014.

Truong, V. Examining the sentiment and emotional differences in product and service reviews: The moderating role of culture. *arXiv preprint arXiv:2507.21057*, 2025.

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, Ł., and Polosukhin, I. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

Wiseman, S., Shieber, S. M., and Rush, A. M. Challenges in data-to-document generation. In *Proceedings of the 2017 Conference on Empirical Methods in Natural Language Processing*, pp. 2253–2263, 2017.