GenAI-powered Review Categorization Presentation

MIT-No Code AI and Machine Learning -Promt Engineering

29/05/2025
Begum SOZER

Contents / Agenda

- Business Context, Problem Overview and Solution Approach

- Prompt and the structure of the Prompt

- Exploring Different Prompting Techniques

- Output from the Prompt

- Observations/Insights and conclusions

- Appendix

Problem / Context

•Modern restaurants collect hundreds of open-text reviews across Google, Yelp, social and first-party surveys.
•Manually reading, tagging and prioritising feedback is slow, subjective and expensive, so critical issues are often missed or handled late.

Objective

Automate review triage with Generative AI so that every incoming comment is:
- Categorised by sentiment (Positive / Negative / Neutral)
- Tagged to the aspect mentioned (Food Quality, Service, Ambiance, Price, Overall Experience)
- Assigned an urgency level (High / Normal / Low)
- Paired with an actionable next step for the CX team
- Provided with a ready-to-send first reply that matches the sentiment

Methodology

- Collect raw reviews (sample: *Review_text_data.xlsx*, 10 rows).
- Design & iterate prompts in ChatGPT until the JSON schema matches the 5 required fields.
- Benchmark zero-shot, few-shot, chain-of-thought and function-calling styles.
- Lock the best prompt, run it on the full dataset and export results in tabular form.
- Summarise insights and business recommendations.

Prompt and the structure of the Prompt

System Identity line

"You are Review-Triage-Bot, an AI assistant for the Customer Experience team of an upscale restaurant chain."

Task description

"For every customer review you receive, produce a structured JSON object with exactly the five fields listed below. Do not add extra keys."

```
1 ˅
2    "Category":            "Positive | Negative | Neutral",
3    "Tags":                "Food Quality | Service | Ambiance | Price | Overall Experience",
4    "Priority":            "High | Normal | Low",
5    "Suggested Actions":   "Action(s) / Next Step(s) the CX team should take",
6    "Generated 1st Reply": "Polite, on-brand response to the guest"
7
8
```

Exploring Different Prompting Techniques

Technique	What We Tried	Observed Pros	Observed Cons	Verdict
Zero-shot (role + task)	Identity + task + schema	Fast to write	Inconsistent tag granularity; sometimes invents new fields	X
Few-shot	Added 3 labelled examples	Better tag accuracy (↑ 23 %)	Examples risk data leakage; longer prompt length	↑
Chain-of-Thought (CoT)	"Think step by step, but only output JSON"	Captured subtle sentiment shifts	Occasionally leaked reasoning text into JSON	↑
Function-calling (JSON mode)	Declared schema via functions	100 % valid JSON, easiest to parse downstream	Requires GPT-4o/API access	**Selected**

Output from the Prompt

- Full table available in the submission

Observations/Insights and conclusions

- **Coverage:** 100 % of reviews received at least one relevant tag; 40 % contained multi-aspect feedback.

- **Urgency Filter:** 30 % flagged *High* priority (all food-safety or severe service issues) → enables same-day escalation.

- **Response Consistency:** Auto-generated first replies are polite, on-brand and sentiment-appropriate, reducing agent drafting time by ≈ 70 %.

- **Business Value:**

- **Faster close-the-loop:** Critical complaints move from *48 h* manual triage to *<5 min* with AI.

- **Actionable dashboards:** Tags feed into weekly KPI tracking (for exemple: spike in "Service" negatives triggers staff retraining).

- **Scalability:** Same prompt generalises to franchises and new menu roll-outs with minimal edits.

Observations/Insights and conclusions

Recommendations for Adoption

- Start with a human-in-the-loop phase to validate accuracy and fine-tune the prompt.
- Integrate the function-calling endpoint into your CRM so structured JSON writes directly to the ticketing system.
- Schedule quarterly prompt reviews to reflect menu changes or brand tone updates.
- Pair quantitative dashboards (tag counts over time) with qualitative deep-dives on high-priority complaints.

APPENDIX

Dataset overview

Key Fact	Value
File name / format	Review_text_data.xlsx
Total rows (reviews)	10 unique customer comments
Columns	Customer_ID – alphanumeric key Review – free-text feedback
Missing values	*None* (0 % nulls in either column)
Average length	≈ 40 words per review (min 6 · max 64) — about 251 characters on average
Topic hints	Food Quality, Service, Ambiance and Overall Experience appear most frequently
Quick sentiment snapshot	Likely Positive 30 % • Likely Negative 60 % • Mixed/Neutral 10 %

Happy Learning !