Finetuning Large Language Models to Capture the Specific Consumer Preference

Xiangwen Deng
dengxw22@mails.tsinghua.edu.cn

Abstract

This paper explores the potential of using Large Language models (LLMs) to capture specific consumer preferences. In particular, we fine-tuned an LLM using unstructured comment data from consumers of a particular product and then test whether the LLM can be simulateted to the distribution of this consumer's specific preferences (e.g., "taste, service, environment") for this product. Our results demonstrate that the fine-tuned LLM is 5% better at predicting the score mean value and 35% better at predicting the std and variance value compared to the untrained LLM. In addition, our experiments also demonstrate that the LLM does not simply memorize reviews, but understands consumer tendencies and can make complex combinatorial inferences. This provides meaningful results for the future use of LLMs to understand consumer tendencies, which is difficult to do with traditional machine learning and manual questionnaires.

1. Introduction

Large Language Models (LLMs), represented by ChatGPT and GPT4, are now playing an increasingly important role in people's lives. The difference between LLMs and general AI is that LLMs are pre-trained on large-scale data thus showing emergent intelligence [XCG+23, WTB+22] and can understand human rules and language. Therefore many researchers [SMR23, TJ23] demonstrate that LLMs can do role-playing, i.e., simulate specific consumers. The latest research, [BIN23, ZWV24, GS24, LCKS24] has studied the application of LLMs in marketing, proving that the LLM has some common sense and has the potential to simulate consumers. However, all of the above studies use GPT4 directly to simulate the potential of random humans, that is, to conduct generalized random surveys. It lacks the ability to model the preferences of specific particular consumer groups over specific products.

Our research aims to fine-tune an LLM by training it so as to simulate specific consumer group preferences and distributions for specific products. Specifically, we have the following questions: 1) Is it possible to simulate the consumer distribution better with the trained LLM, compared to not training the LLM, etc.? 2) Is it possible to train an LLM with only unstructured data, such as consumer reviews? 3) Is it possible to manipulate the LLM to answer some questions that are difficult to solve by traditional survey methods or traditional machine learning methods(like LDA [JWY+19])?

Our approach firstly collects 1000 real consumer reviews from multiple stores on "dazongdianping", each review will contain some consumer scores: "taste, environment, service". We use the 800 review data in the format of dialog text through the LoRA [HSW+21] LLM training technique, fine-tune the training of a local ChatGLM [DQL+21] model. The question was then repeated 200 times to get a response containing a score and a description of the text. The experimental results demonstrate that both the consumer score features and the text data are closer to the distribution of real consumers compared to the untrained model.

Our contributions are listed below:

- We are the first to explore methods for consumer-specific simulateting of LLMs for specific products, rather than previous work focusing on generalized basic product attribute insights.
- We provide a method for fine-tuning the LLM using unstructured data (UGC), which is effective in improving the simulateting accuracy of the data and is not limited to the use of structured questionnaire data.
- We demonstrate that this UGC fine-tuning LLM truly understands consumer tendencies, rather than just reciting comments. Therefore, it has the potential to replace traditional manual questionnaires.

2. Related Work

2.1. LLMs for Role-playing

Researchers started to treat the LLM as an intelligent body. For example, [CWX+24, XCG+23] illustrated that an LLM-based agent has 4 parts including perception, reasoning, action, and tools., which can realize complex tasks like humans. Therefore an important direction is how to do role-play [SMR23]. [POC+23] can realize a community of multiple intelligences through prompt, and [WPQ+23] can also realize role-play imitation through fine training. [TJ23] points out the potential of using UGC data to train LLM, and [XCG+23] illustrates a promising approach to the training and simulation of demographic populations.

2.2. LLMs for Consumer Simulation

Recently the ability of LLMs has attracted the attention of many marketing researchers, who particularly focused on exploring the difference between LLM's preference tendencies and real consumers. [BIN23] Using GPT3.5 responds to sets of survey questions in ways that are consistent with economic theory and well-documented patterns of consumer behavior.[ZWV24] evaluates LLMs, like GPT-4, for their ability to mimic consumer-revealed preferences in risk-influenced decision-making.[GS24] Uses GPT3.5 to reveal the preferences of LLMs in simulateting consumers' historical date investment rewards and explores the impact of different languages. It also proposes the use of COT-based [WWS+22] mechanism to better utilize LLM. [LCKS24] uses GPT4 to explore perceptual analysis based on brand similarity metrics and product attribute ratings, and demonstrates that

LLMs and humans have comparable effects. The significant difference between us and the above studies is that we fine-tune the local LLM model to better simulate consumer propensities, rather than just using Openai's GPT interactions. Thus it is possible to better simulate consumer groups for specific products, or new products that GPT has not seen before. Additionally, we use unstructured data of UGC consumer comments to simulate insights rather than structured data like questionnaires.

3. Method

We outline our main methodology. It is mainly divided into 3 parts: data collection, model training, and question evaluation.

First, we collect 1000 real consumer reviews from a website called "dazongdianping", and each review contains not only textual comments but also consumer scores on "taste, environment, and service" dimensions.

Then, We randomly sample 800 reviews, merge their text data and scoring data, and then construct training data pairs using the format of dialog text. We use the training data to fine-tune some of the parameters of the ChatGLM through a LoRA technique.

Finally, we measure the training results by first randomly asking the LLM 200 questions, and then randomly sampling the remaining 200 comments to compare with the LLM results. The metrics include the mean-variance of the scores and the distribution of PCA texts. The experimental results demonstrate that the distribution of scores of consumers of the trained LLM is closer to the scores of real consumers compared to the pre-training.

3.1. Data and Training

Our data source is "dazongdianping(大众点评)"[1], which is one of the largest local lifestyle platforms in China, and it is also famous for its ability to rate products in multiple dimensions. We first crawled "dazongdianping" to collect data from 1000 real consumer reviews, and each review will contain some consumer scores: "taste, environment, service" (shown in Fig. 7). The products we choose include a cake store, a hotpot store, and a barbecue store. Each restaurant crawled 1,000 pieces of data. All of them are stores that LLM has no contact with and no specific knowledge.

[LCKS24, EEA+23] have been proved that scoring directly using the LLM is not credible. [LCKS24] proposed an effective method, specifically, they converted the direct scoring task into a continuation task by counting the number of keywords in the output text without direct scoring but by counting the proportional scoring, which allows estimating some common-sense product attribute perceptions, such as car brands. But the previous method of statistical frequency in [LCKS24] is also not suitable for the task because the LLM has no knowledge of the specific product and we are using unstructured data like consumer reviews rather than a questionnaire. So we were inspired by the COT-based [WWS+22] mechanism, although direct scoring by the LLM does not work, LLM can solve the problem in a step-by-step manner by analyzing the data textual evaluations first and then scoring them. Therefore, we recombined the data of the LLM and added

[1] 大众点评：https://www.dianping.com/

the scores of “taste, environment, service” to the end of the corresponding statements in the reviews. This resulted in the final training data pair format (shown in Fig. 7).

Since “dazongdianping” is a Chinese website, most of the comments are in Chinese, considering the performance of the model on Chinese data, we use the leading open-source Chinese LLM as our base model, ChatGLM [DQL+21], and we train the LLM by the LoRA technique [HSW+21]. We use LoRA to train the parameters of the LoRA part of the LLM. We run a total of 10000 epochs, and the model converges gradually, taking about 3 hours on a single RTX 3090 GPU.

Table 3.1: Statistical Measures for Hotpot, Cake, and Barbecue Categories. Where % refers to the value of deviation from human_GT, the closer to 0 % the better.

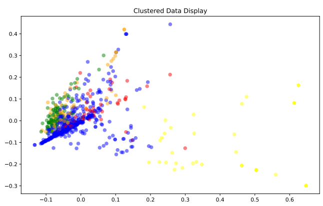
Category	**Measure**	**Metric**	**llm_trained vs. human_GT (%)**	**llm vs. human_GT (%)**
Hotpot	Taste	Mean	7.76%	-10.84%
		Std	-66.12%	-100%
		Variance	-88.53%	-100%
	Environment	Mean	6.83%	-22.88%
		Std	-63.94%	-85.76%
		Variance	-87.00%	-97.96%
	Service	Mean	4.89%	-2.57%
		Std	-63.60%	-92.38%
		Variance	-86.77%	-99.42%
Cake	Taste	Mean	-4.45%	-6.52%
		Std	-61.73%	-65.56%
		Variance	-85.37%	-88.13%
	Environment	Mean	-3.11%	-10.32%
		Std	-46.71%	-67.73%
		Variance	-71.56%	-89.58%
	Service	Mean	-5.84%	4.08%
		Std	-38.09%	-84.77%
		Variance	-61.70%	-97.67%
Barbecue	Taste	Mean	1.05%	1.13%
		Std	-69.78%	-100%
		Variance	-90.86%	-100%
	Environment	Mean	-7.52%	-10.13%
		Std	-54.06%	-93.59%
		Variance	-78.89%	-99.59%
	Service	Mean	-2.42%	11.35%
		Std	-45.05%	-93.26%
		Variance	-69.80%	-99.55%

3.2. Evaluation and Results

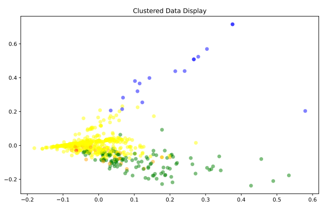
After completing the training of the model, we asked the model 200 questions using the same prompts as during training, and then used the test set (i.e., 200 random reviews that the model had not seen) as a comparison metric. This is because consumer reviews are randomly distributed and groups of consumers of the same product should have similar characteristics. Therefore, if the model is trained effectively, it should be able to match the distributional characteristics of 200 new consumer comments.

The metrics tested include the numerical distribution of the mean-variance of the scoring values and the distribution of PCA features of the output text. Specifically, we first use TfidfVectorizer to transform the text data into TF-IDF [SY73] feature vectors. The original high-dimensional TF-IDF feature vectors are then downscaled to two dimensions using PCA [AW10] in order to visualize and display the clustering results on a two-dimensional plane.

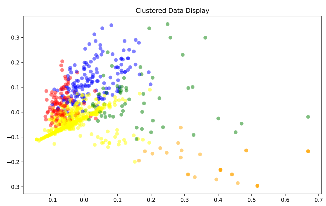
The results show that our trained LLM model captures consumer preference better than the untrained macro model. As shown in Tab. (3.1), Our results demonstrate that the fine-tuned LLM is 5% better at predicting the score mean value and 35% better at predicting the std and variance value compared to the untrained LLM. As shown in Fig. (3.1), the results indicate that the PCA visualization is also closer to the distribution of the real values, compared to the almost random distribution of the original LLM model.



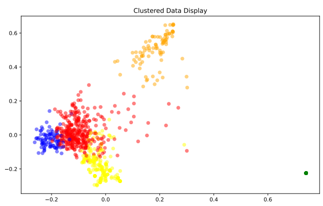
(a) human_GT BBQ



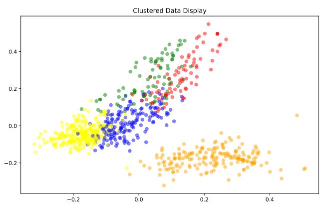
(b) human_GT Cake



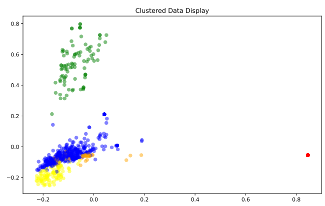
(c) human_GT Hotpot



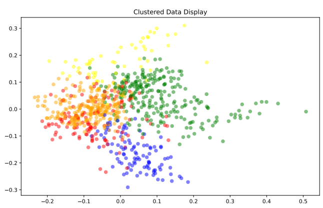
(d) Trained BBQ



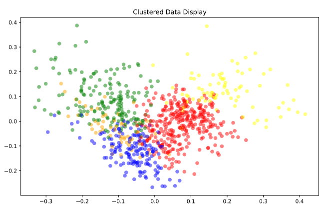
(e) Trained Cake



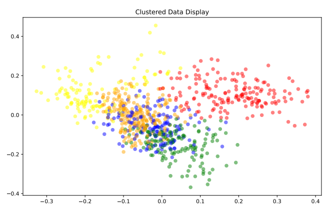
(f) Trained Hotpot



(g) LLM BBQ



(h) LLM Cake



(i) LLM Hotpot

Figure 3.1: Comparison of PCA Text analysis for Different Categories. The colors are randomly assigned, which indicate different categories of PCA.

4. Discussion

In this section, we wish to show the advantages of the LLM approach in marketing research [BIN23, ZWV24, GS24, LCKS24]. We show this in two ways: 1) Comparison with traditional ML methods. 2) Comparison with manual questionnaires.

Among the traditional ML topic analysis methods, LDA [JWY+19] is a very common method that is widely used. However, LDA is based on the principle of bag-of-words statistics, which cannot recognize complex semantics and can only output some words (e.g., "delicious" , "good service"). In contrast, the LLM approach not only outputs words but also recognizes complex semantics (“the service is not good because there are too many queues”).

In addition, we demonstrate the ability of the trained model to answer some “combinatorial analysis” type of questions. One of the difficulties in traditional manual questionnaires is the inability to simulate consumer bias, which makes it difficult [LCKS24] to get high-quality consumers to answer truthfully when facing some complex questions. As shown in Fig. (7), summarizing the question “Why is the service bad?” The LLM's answers show that it recognizes the phenomena corresponding to high and low scores and combines answers from different consumers. LLM does not simply memorize the training data but really understands the review data.

5. Future Work

We plan to focus on the following areas to enhance our research:

- We will try higher quality data and better-fine-tuned algorithms [PLH+23, HWL+23] to further improve the accuracy of simulateting consumers. Right now the accuracy is just better than the unfine-tuned model, and there is still a long way to go before we can fully simulate consumers.
- Explore the potential of LLM to the analysis of unstructured text data (like UGC), especially topic analysis. The topic analysis (LDA) [JWY+19] is based on the assumption of bag-of-words model, thus ignoring the contextual semantics [BHL+20]. In contrast, LLM is good at recognizing contextual semantics. However, the main drawback of LLM is that it is not accurate enough to count the frequency of subject terms. In the future, techniques such as RAG [GXG+24] can be considered to improve the recognition accuracy.
- Explore more areas of LLM for simulateting consumers. Since it is possible to manipulate LLMs via prompt, we expect to be able to study some biased responses that are difficult to get with conventional questionnaire methods. For example, the domains of heuristic cues [CB12], demand characteristics [MDW12], social desirability [SDB10], and racial discrimination [PS08], which are usually difficult to get high-quality answers through questionnaires, might be attempted to get meaningful results from another aspect through the LLM.

6. Conclusion

Our study explores the potential of fine-tuning a Large Language Model (LLM) using unstructured data, such as consumer reviews, to capture specific consumer preferences. By training the LLM with review data from consumers of a particular product, we tested its ability to simulate the distribution of consumer preferences both in text and scores. The fine-tuned LLM is 5% better at predicting the score mean value and 35% better at predicting the std and variance value compared to the untrained LLM. This indicates that fine-tuning significantly enhances the model's ability to understand and capture consumer tendencies. Additionally, the fine-tuned LLM demonstrated the capability to perform complex combinatorial inferences, providing insights into consumer behavior that are difficult to obtain through traditional survey methods or conventional machine learning techniques. Our findings suggest that LLMs, when fine-tuned appropriately, offer a powerful and advanced approach for gaining deep consumer insights, presenting a significant advancement over traditional methods.

References

[AW10] H. Abdi and L. J. Williams. Principal component analysis. *Wiley interdisciplinary reviews: computational statistics*, 2(4): 433–459, 2010.

[BHL+20] J. Berger, A. Humphreys, S. Ludwig, W. W. Moe, O. Netzer, and D. A. Schweidel. Uniting the tribes: using text for marketing insight. *Journal of marketing*, 84(1): 1–25, 2020.

[BIN23] J. Brand, A. Israeli, and D. Ngwe. Using gpt for market research. *Available at SSRN 4395751*, 2023.

[CB12] D. A. Cohen and S. H. Babey. Contextual influences on eating behaviours: heuristic processing and dietary choices. *Obesity Reviews*, 13(9): 766–779, 2012.

[CWX+24] J. Chen, X. Wang, R. Xu, S. Yuan, Y. Zhang, W. Shi, J. Xie, S. Li, R. Yang, T. Zhu, et al. From persona to personalization: a survey on role-playing language agents. *arXiv preprint arXiv:2404.18231*, 2024.

[DQL+21] Z. Du, Y. Qian, X. Liu, M. Ding, J. Qiu, Z. Yang, and J. Tang. Glm: general language model pretraining with autoregressive blank infilling. *arXiv preprint arXiv:2103.10360*, 2021.

[EEA+23] J. L. Espejel, E. H. Ettifouri, M. S. Y. Alassan, E. M. Chouham, and W. Dahhane. Gpt-3.5, gpt-4, or bard? evaluating llms reasoning ability in zero-shot setting and performance boosting through prompts. *Natural Language Processing Journal*, 5: 100032, 2023.

[GS24] A. Goli and A. Singh. Frontiers: can large language models capture human preferences? *Marketing Science*, 2024.

[GXG+24] Y. Gao, Y. Xiong, X. Gao, K. Jia, J. Pan, Y. Bi, Y. Dai, J. Sun, M. Wang, and H. Wang. *Retrieval-Augmented Generation for Large Language Models: A Survey*. 2024. arXiv: 2312.10997 [cs.CL].

[HSW+21] E. J. Hu, Y. Shen, P. Wallis, Z. Allen-Zhu, Y. Li, S. Wang, L. Wang, and W. Chen. *LoRA: Low-Rank Adaptation of Large Language Models*. 2021. arXiv: 2106.09685 [cs.CL].

[HWL+23] Z. Hu, L. Wang, Y. Lan, W. Xu, E.-P. Lim, L. Bing, X. Xu, S. Poria, and R. K.-W. Lee. Llm-adapters: an adapter family for parameter-efficient fine-tuning of large language models. *arXiv preprint arXiv:2304.01933*, 2023.

[JWY+19] H. Jelodar, Y. Wang, C. Yuan, X. Feng, X. Jiang, Y. Li, and L. Zhao. Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. *Multimedia tools and applications*, 78: 15169–15211, 2019.

[LCKS24] P. Li, N. Castelo, Z. Katona, and M. Sarvary. Frontiers: determining the validity of large language models for automated perceptual analysis. *Marketing Science*, 2024.

[MDW12] J. McCambridge, M. De Bruin, and J. Witton. The effects of demand characteristics on research participant behaviours in non-laboratory settings: a systematic review. *PloS one*, 7(6): e39116, 2012.

[PLH+23] B. Peng, C. Li, P. He, M. Galley, and J. Gao. Instruction tuning with gpt-4. *arXiv preprint arXiv:2304.03277*, 2023.

[POC+23] J. S. Park, J. O'Brien, C. J. Cai, M. R. Morris, P. Liang, and M. S. Bernstein. Generative agents: interactive simulacra of human behavior. *Proceedings of the 36th Annual ACM Symposium on User Interface Software and Technology,* 1–22, 2023.

[PS08] D. Pager and H. Shepherd. The sociology of discrimination: racial discrimination in employment, housing, credit, and consumer markets. *Annu. Rev. Sociol*, 34: 181–209, 2008.

[SDB10] J.-B. E. Steenkamp, M. G. De Jong, and H. Baumgartner. Socially desirable response tendencies in survey research. *Journal of Marketing Research*, 47(2): 199–214, 2010.

[SMR23] M. Shanahan, K. McDonell, and L. Reynolds. Role play with large language models. *Nature*, 623(7987): 493–498, 2023.

[SY73] G. Salton and C. T. Yu. On the construction of effective vocabularies for information retrieval. *Acm Sigplan Notices*, 10(1): 48–60, 1973.

[TJ23] Z. Tan and M. Jiang. User modeling in the era of large language models: current research and future directions. *arXiv preprint arXiv:2312.11518*, 2023.

[WPQ+23] Z. M. Wang, Z. Peng, H. Que, J. Liu, W. Zhou, Y. Wu, H. Guo, R. Gan, Z. Ni, M. Zhang, et al. Rolellm: benchmarking, eliciting, and enhancing role-playing abilities of large language models. *arXiv preprint arXiv:2310.00746*, 2023.

[WTB+22] J. Wei, Y. Tay, R. Bommasani, C. Raffel, B. Zoph, S. Borgeaud, D. Yogatama, M. Bosma, D. Zhou, D. Metzler, et al. Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*, 2022.

[WWS+22] J. Wei, X. Wang, D. Schuurmans, M. Bosma, F. Xia, E. Chi, Q. V. Le, D. Zhou, et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems*, 35: 24824–24837, 2022.

[XCG+23] Z. Xi, W. Chen, X. Guo, W. He, Y. Ding, B. Hong, M. Zhang, J. Wang, S. Jin, E. Zhou, et al. The rise and potential of large language model based agents: a survey. *arXiv preprint arXiv:2309.07864*, 2023.

[ZWV24] W. Zhao, W. Wang, and S. Viswanathan. Spillover effects of generative ai on human-generated content creation: evidence from a crowd-sourcing design platform. *Available at SSRN 4693181*, 2024.

7. Supplementary Material

Table 7.1: Origin Score Value of Hotpot, Cake, and Barbecue Categories.

Name	**Mean**	**Std**	**Min**	**25%**	**50%**	**75%**	**Max**	**Variance**
				Barbecue Categories				
taste_llm_barbecue	4	0	4	4	4	4	4	0
environment_llm_barbecue	3.4944	0.0727	3	3.5	3.5	3.5	4	0.0053
service_llm_barbecue	4.4881	0.0762	4	4.5	4.5	4.5	4.5	0.0058
taste_human_GT_barbecue	3.9553	1.0551	0.5	3.5	4	5	5	1.1132
environment_human_GT_barbecue	3.8883	1.1349	0.5	3.5	4	5	5	1.2880
service_human_GT_barbecue	4.0307	1.1307	0.5	3.5	4.5	5	5	1.2785
taste_llm_trained	3.9968	0.3189	3	4	4	4	5	0.1017
environment_llm_trained	3.5960	0.5214	3	3	4	4	7	0.2719
service_llm_trained	3.9330	0.6213	1	4	4	4	5	0.3860
				Cake Categories				
taste_llm_cake	4.14	0.2246	4	4	4	4.5	4.5	0.0505
environment_llm_cake	3.6275	0.2306	3	3.5	3.5	4	4	0.0532
service_llm_cake	4.4931	0.1014	4	4.5	4.5	4.5	5	0.0103
taste_human_GT_cake	4.4283	0.6523	0.5	4	4.5	5	5	0.4254
environment_human_GT_cake	4.045	0.7145	0.5	3.5	4	4.5	5	0.5105
service_human_GT_cake	4.3167	0.6655	0.5	4	4.5	5	5	0.4429
taste_llm_trained	4.2313	0.2495	4	4	4	4.5	4.5	0.0622
environment_llm_trained	3.9191	0.3809	3	3.5	4	4	5	0.1451
service_llm_trained	4.0646	0.4119	1	4	4	4.5	5	0.1697
				Hotpot Categories				
taste_llm_hotpot	4	0	4	4	4	4	4	0
environment_llm_hotpot	3.4863	0.1052	3	3.5	3.5	3.5	4	0.0111
service_llm_hotpot	4.5019	0.0530	4	4.5	4.5	4.5	5	0.0028
taste_human_GT_hotpot	4.4866	0.7474	0.5	4.5	4.5	5	5	0.5585
environment_human_GT_hotpot	4.5217	0.7387	0.5	4.5	5	5	5	0.5457
service_human_GT_hotpot	4.6204	0.6955	0.5	4.5	5	5	5	0.4838
taste_llm_trained	4.8349	0.2532	3.5	4.5	5	5	5	0.0641
environment_llm_trained	4.8302	0.2663	3.5	4.5	5	5	5	0.0709
service_llm_trained	4.8463	0.2531	3	4.5	5	5	5	0.0640

Training Dataset

[
{“instruction”: “If you are a random (positive or negative) consumer, please describe one possible taste, environment, and service at a cake restaurant. Then give a final specific score for each from low to high (1.0 to 5.0) score.”, “input”: “”, “output”: ‘‘口味描述: 壹海城 onemall 的韩式烤肉味道一般。Taste Score: 3.0/5.0. 环境描述: 壹海城 onemall 的环境人还蛮多的, 需要排队等号。Environment Score: 3.0/5.0. 服务描述: 虽然服务人员态度还可以，但是需要改进服务效率。Service Score: 4.0/5.0.”},
{“instruction”: “If you are a random (positive or negative) consumer, please describe one possible taste, environment, and service at a cake restaurant. Then give a final specific score for each from low to high (1.0 to 5.0) score.”, “input”: “”, “output”: ‘‘口味描述: 这家小料台的卫生环境较差，而且补货速度很慢，周末更是忙碌，应该提前做好准备。Taste Score: 3.0/5.0. 环境描述: 这家小料台的空间非常狭小，桌子特别小，让人感觉很拥挤。Environment Score: 3.5/5.0. 服务描述: 服务态度差，没有让人感到满意的地方。Service Score: 0.5/5.0.”},
{“instruction”: “If you are a random(positive or negative) consumer, please describe one possible taste, environment, and service at a cake restaurant. Then give a final specific score for each from low to high (1.0 to 5.0)score.”, “input”: “”, “output”: ‘‘口味描述: 大众点评上的双人套餐又划算又好吃，还点了冷面和鳗鱼石锅拌饭，店员也很勤快和贴心，教我们怎么吃冷面和包烤肉，赞！Taste Score: 4.5/5.0. 环境描述: 店内整洁干净，装修风格简约时尚，氛围舒适宜人。Environment Score: 4.5/5.0. 服务描述: 店员服务热情周到，耐心解答我们的问题，还主动帮助我们拍照留念，非常贴心。Service Score: 4.5/5.0.”},
...
]

Fig. 7 Example format of training dataset. We merge the original review text data with the scores of “taste, environment, and service” as the data pair of “instruction and output” for training.

Question

If you are a random (positive or negative) consumer, please describe one possible service at a barbecue restaurant. Why did they get a lower score in service? Give me 5 specific reasons.

Before Training

1. Staff were rude or unhelpful.
2. The food was of poor quality.
3. The menu was limited and not varied.
4. The consistency of quality was poor (i.e., some orders were better than others).
5. The attention to detail was lacking, such as incorrect orders or late deliveries.

After Finetuning

1. Overcrowding: Too crowded, leading to long wait times and frustrated customers.
2. Inefficient kitchen staff: Insufficient staff, causing slow food preparation and delivery.
3. Poor communication: Miscommunication among staff, leading to incorrect or incomplete orders.
4. Lack of personalization: No personalized service, making customers feel undervalued.
5. Overpriced menu items: High prices without corresponding quality, leading to poor value perception.

Fig. 7 Comparison of LLM before and after training. After training, LLMs can not only customize responses but also merge comments and answers from multiple low-scoring consumers. LLMs can act as a product that allows virtual consumers to understand comments rather than just remembering them.

Question

If you are a random(positive or negative) consumer at a cake restaurant. What are Customers' attitude towards queuing?

Before Training

Customers' attitudes towards queuing can vary based on factors like queue length, reason for waiting, and overall restaurant experience. Generally, customers have a negative attitude towards queuing, feeling frustration, anxiety, and inconvenience. However, some may view it as an opportunity to socialize or enjoy the restaurant atmosphere.

Research shows that customers who wait more than 15 minutes tend to have a negative attitude towards queuing. Conversely, customers who are kept informed about wait times and estimated food arrival tend to have a more positive attitude.

In conclusion, while queuing is an inevitable part of the restaurant experience, customers' attitudes can be influenced by various factors. Effective queue management is crucial for ensuring customer satisfaction.

After Training

Some customers may view queuing as a necessary part of the dining experience, and they may be willing to wait in line if they feel that the wait will be worth it in terms of the quality of the food or service. Other customers may not be willing to wait in line for extended periods of time and may be more likely to look for alternative dining options.

There are also factors that can influence customer attitudes towards queuing, such as the size of the queue, the speed of service, and the level of comfort provided while waiting. In general, customers are more likely to have a positive attitude towards queuing if they feel that their needs and expectations are being met and if they are provided with a comfortable and enjoyable dining experience while waiting.

Fig. 7 Comparison of LLM before and after training. After training, LLMs are able to respond appropriately to specific questions posed in comments. This flexibility to ask questions about "a particular issue" (e.g., queuing) was not possible with previous ML representational methods (LDA or LIWC).