What If LLMs Can Smell: A Prototype

Xueyi Zhou[1], **Qi Lu**[2*], **Dong-Kyu Chae**[1*]

[1]Department of Computer Science, Hanyang University, Republic of Korea
[2]The Future Laboratory, Tsinghua University, China
{hokyeejau,dongkyu}@hanyang.ac.kr, luq@mail.tsinghua.edu.cn

Abstract

The olfaction is hardly mentioned in the studies of multi-modal Large Language Models (LLMs). This demo presents a prototypical framework to embody prevalent LLMs with smelling ability using a plug-and-play olfactory signal processing service. To this end, we collect a dataset on Korean beers by self-developed electronic noses (e-noses) and an open-source dataset. An olfaction-related question-answering corpus is also generated to fine-tune LLMs. A gas classification model is applied to identify the smelling liquor upon the e-nose data. We then adopt and fine-tune LLMs on the generated datasets. The results show that LLMs under this framework can interact with the environment by its 'nose' and provide olfaction-related answers augmented by our dataset. **To the best of our knowledge, this is the first work on embodying LLMs with artificial olfaction.** We additionally deployed the gas classification model and the trained LLM in a simple web-based system to show the feasibility of our prototype. Our demo video can be found at: https://bit.ly/4j8x6ZY.

1 Introduction

Large language models (LLMs) are empirically proven to effectively process and fuse multi-modal data, such as images, audio, videos, text and 3D point cloud [Lyu *et al.*, 2023; Wu *et al.*, 2023; Yang *et al.*, 2024; Wang *et al.*, 2024; Han *et al.*, 2023]. They use encoders for different modalities in parallel to encode and align the data and extract semantic understanding by LLM-based models. This kind of multi-modal LLMs is highly limited by the encoders they apply. An additional encoder might need more datasets for modal semantic alignment and far more time-consuming training or fine-tuning. LLMs, on the other hand, can serve as an omniscient Agent by means of generating and using diverse instructions for real-world service APIs and multi-step decision-making processes. For instance, ToolLLM [Qin *et al.*, 2023] facilitates LLMs, *e.g.*, LLaMA, with more than 16,000 real-world APIs to enable effective interaction to accomplish complex tasks. It could search reasoning paths and make deliberate decisions to generate generalized answers on an out-of-distribution (OOD) dataset. However, prevalent tool-alike LLMs [Qin *et al.*, 2023; Liu *et al.*, 2024; Gao *et al.*, 2023] do not consider the interaction with multi-modal applications. LLMs for robotics [Vemprala *et al.*, 2024; Brohan *et al.*, 2023] take into account this end, but they hardly talk about the *olfaction modality*.

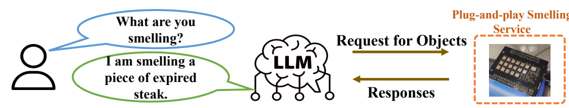

Figure 1: An overview of olfactory augmented LLMs

Olfactory augmented LLMs can help scent-based descriptions, scent recognition, personalized user experiences and robot interaction. However, few studies are proposed to embody olfaction-related capabilities in LLMs. In [Hassan *et al.*, 2024], the authors provide an effective algorithm which encodes the olfaction and vision sensor data into a multi-modal prompt and instruct the LLM to select a navigation behavior for robots. Sniff AI [Zhong *et al.*, 2024] conducts a comprehensive user study to investigate how well LLMs can interpret human descriptions of scents using an LLM-based guessing model and a scent delivery device. Combining an LLM with a text-image model, OdorAgent [Zhang *et al.*, 2024] can align video-odor semantic understanding and dependently display odor based on images. However, these studies neither *de facto* embody the smelling capability in LLMs nor consider plug-and-play interactive service-oriented e-nose tools.

Inspired by these works, we present an olfactory augmented prototype framework to enable LLMs smelling capabilities. We study liquor classification as an application scenario. Specifically, we customize a dataset *w.r.t.* three common Korean beers by self-developed four copies of e-noses collecting the spatio-temporal data in a distributed manner. We additionally adopt an open-source dataset `Liquor2021` [Hou *et al.*, 2022; Meng and Hou, 2021], which provides eight kinds of Chinese liquors on their self-developed e-nose. Six traditional machine learning models [Wijaya *et al.*, 2023] and three deep learning models [Wu *et al.*, 2024; Huang *et al.*, 2017; Karim *et al.*, 2019] are

[*]Corresponding authors.

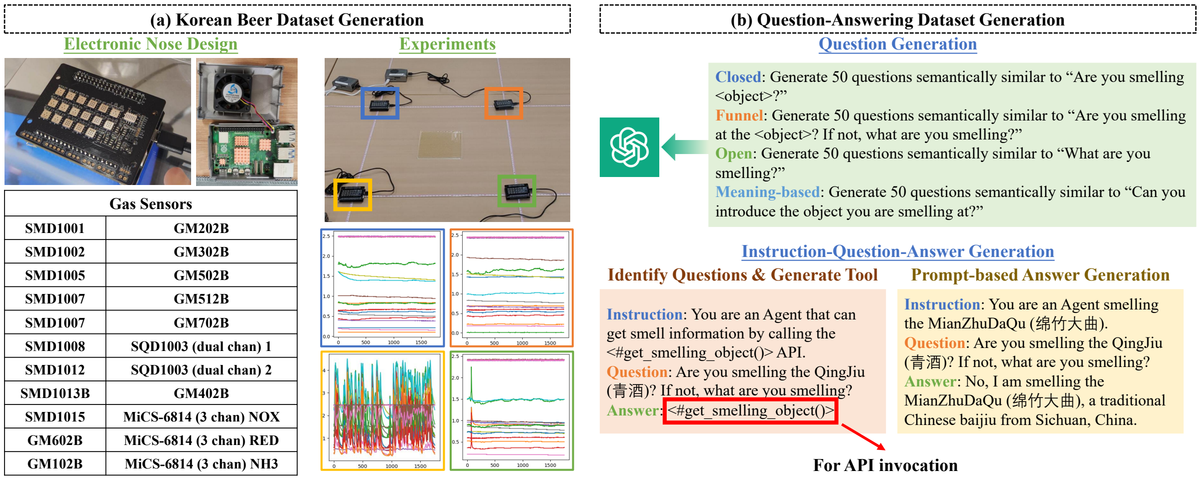

Figure 2: Dataset generation for both the Korean beer sensory data and olfaction-related question answering.

trained on the two datasets and selected based on their performance. Then the trained model is applied in a plug-and-play service API to infer the labels of smelling objects upon the sensor data. A small smelling-related question-answering corpus is generated by ChatGPT 4o mini and filtered by manual labor to fine-tune the selected LLMs (*i.e.*, DeepSeek: R1 Distill Qwen 1.5B and DeepSeek: R1 Distill Qwen 7B [DeepSeek-AI, 2025]). The experiments show that, fine-tuned within 5 epochs, the LLMs can identify the olfaction-related question; by two designed instructions, LLMs can generate the API invocation commands and answer the questions based on the requested smelling information. Since the olfaction modality is decoupled from the LLM semantic understanding, the instruction can yet boost LLMs to answer the OOD olfaction-related questions if the application scenario is changed. We ultimately deploy the fine-tuned LLM in a simple web-based system to show its feasibility and deployability (elucidated in the video only).

2 Experiment

This section introduces dataset generation for both the artificial olfaction on Korean beers and olfaction-related question answering, the gas classification models and their experimental results, and the means of fine-tuning LLMs.

2.1 Beer Olfactory Sensor Dataset

We briefly exhibit the design of our self-developed electronic nose and experiments in Fig. 2(a). Totally 19 gas sensors and 22 channels are embedded in each copy. Each of 4 Raspberry Pi 5s functions as an e-nose controller. We additionally developed a simple web-based management system to remotely activate the collection program and collect the sensory data. Since the diffusion patterns of the gases are also significant, we expose our e-noses to air instead of limiting the odors within a traditional pumping chamber. The second figure in Fig. 2(a) shows that the e-noses are located at the four corners of a 0.5m $\times$ 0.5m area. The sample container is a transparent acrylic container measuring 15cm $\times$ 15cm $\times$ 0.5cm. The container is always placed at the exact center of the area. We used five 1L bottles of Terra, Kelly, and Cass (these are the names of the three most popular beers in South Korea) beer each. Each bottle of beer was stored in a household refrigerator at a constant temperature of 5°C before the experiment. We extracted twelve samples of 50ml each from every bottle to conduct experiments. Each sample was left somewhere for 10 minutes and still at the target position for 1 minute before data collection began. Each experiment lasts for 10 minutes. The sampling rate is nearly 3Hz. All experiments were conducted in a private room with temperature (10∼16°C) and 20% humidity. We ultimately adopted $5 \times 3 \times 12 = 180$ data samples.

2.2 Gas Classification

We first downsample the `Liquor2021` dataset from 1000Hz to 10Hz. Then the `Liquor2021` and the beer olfactory sensor dataset are each divided into five parts for 5-fold validation. We adopt six traditional machine learning models (*i.e.*, MLP, k-NN, SVM, Decision Tree, Random Forest, and AdaBoost) [Wijaya *et al.*, 2023] and three deep learning models (*i.e.*, TCN [Wu *et al.*, 2024], DenseNet [Huang *et al.*, 2017] and LSTM-FCN [Karim *et al.*, 2019]) to seek an optimal gas classification model based on their average accuracies. According to our experiment, DenseNet achieved 100% average accuracies on 2 datasets. We thus apply it in its service API to infer the classification results on real-time sensory data.

2.3 Olfaction-Augmented Corpus

We categorize the olfaction-augmented questions into 4 types: closed, funnel, open and meaning-based questions. This categorization is dependent on the question difficulty. To boost the LLMs, we use ChatGPT 4o mini to generate semantically similar questions given manually designed questions. The first part in Fig. 2(b) demonstrates the prompt-based question generation. By manually filtering, we got a total of 783 questions (281 closed questions, 51 funnel questions, 299 open questions and 152 meaning-based questions). To form instruction-question-answer triplets, we design two delicate instructions to prompt the LLMs as shown in the sec-

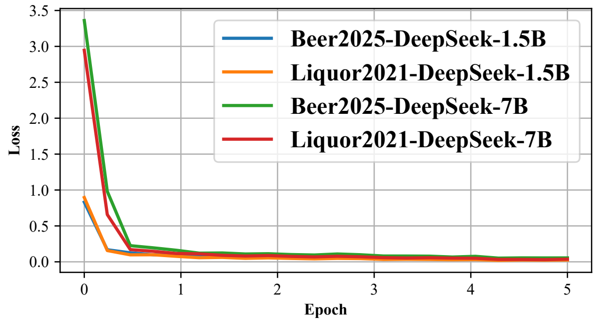

Figure 3: Validation loss while fine-tuning the DeepSeek models.

ond part in Fig. 2(b). For the message in the first round, the LLM identifies if the question is related to the smell; if so, it provides the command to activate the second round[1]. The message in the second round is designed for generating the real answers given the requested gas information. Before fine-tuning, the "⟨ object ⟩" tokens in the questions are substituted by a random label. In addition, we provide dozens of template answers in the second round. We manually split the triplets containing four kinds of questions into the train/valid sets. For the training set, we have four types of questions in the first round but funnel and open questions only in the second round. The same applies to the validation set. We further include all the questions without answers in the test set.

2.4 Olfactory Augmented LLMs

We adopt and fine-tune two DeepSeek models [DeepSeek-AI, 2025]: R1 Distill Qwen 1.5B and R1 Distill Qwen 7B within 5 epochs on the two datasets, respectively. The labels are randomly assigned to questions while training and evaluating. All the fine-tuning experiments share the same group of hyperparameters: 8 samples per training and validation batch, a learning rate of 0.0001, checkpoint saving per 100 steps, and gradient accumulation per 2 steps. The maximum sequence length of LLMs is set to be 384. The experiments were conducted on an HPC (DELL DSS8440) equipped with 8 NVIDIA A100 GPUs at the Biospecimen-Multiomics Digital Bioanalysis Core Facility of Hanyang University. The validation loss curve is shown in Fig. 3.

3 Use Cases

We provide a comprehensive use case study on our self-developed beer olfactory sensor dataset to demonstrate the proposed prototype is effective and feasible. Besides, we do not discuss funnel questions here since they are similar to closed questions. The use cases are exemplified in Fig. 4, where each question is randomly selected from each category.

Closed Questions. As shown in the blue blocks in Fig. 4, both models could answer "Yes/No" precisely to the closed questions; they can provide the correct beer labels. The answers generated are slightly different but semantically similar.

[1]We do not generate the instructions alike the others since we only consider real-time application scenarios where only one service API is equipped.

1.5B	**Question**: Do you smell the Terra beer? **Smelling Object**: Terra beer **Answer**: Yes, I can sense through my nose that I'm smelling Terra beer.
	Question: What are you smelling? **Smelling Object**: Cass beer **Answer**: The aroma I'm picking up suggests it's Cass beer. Cass beer is a popular South Korean lager, known for its light, crisp, and refreshing taste. It has a smooth, mildly bitter flavor, making it easy to drink and a favorite choice for casual dining or social gatherings. Often enjoyed with Korean BBQ or fried foods, Cass is one of the leading beer brands in Korea..
	Question: Can you introduce what you are smelling? **Smelling Object**: Terra beer **Answer**: Yes, I can sense through my nose that I'm smelling Terra beer.
7B	**Question**: Do you smell the Kelly beer? **Smelling Object**: Cass beer **Answer**: No, The scent I'm catching indicates the presence of Cass beer.
	Question: What are you smelling? **Smelling Object**: Kelly beer **Answer**: I can tell from the scent that it's Kelly beer I'm smelling. Kelly beer is a South Korean brand known for its smooth, light lager. It has a crisp, refreshing taste with a mild bitterness, making it an easy-drinking beer. Popular for casual gatherings and meals, Kelly beer is enjoyed for its balanced flavor and wide appeal, offering a satisfying option for beer enthusiasts.
	Question: Can you introduce the thing that you are smelling? **Smelling Object**: Terra beer **Answer**: I can tell from the smell that I'm detecting Terra beer. Terra beer is a popular South Korean lager known for its clean, crisp taste and refreshing qualities. It is brewed with a focus on purity, offering a light, smooth flavor with a subtle bitterness. Terra is often enjoyed with Korean cuisine, particularly BBQ, and is known for its easy-drinking style, making it a favorite choice for casual social occasions.

Figure 4: Use cases of the LLMs on three kinds of questions.

Open Questions. As shown in the green blocks in Fig. 4, it can answer what exactly it is smelling, *e.g.*, "I can tell from the scent that it's Kelly beer". Besides, it additionally provides a short introduction to the beer.

Meaning-based Questions. The answers from the DeepSeek R1 Distill Qwen 7B model provide the exact introduction of the beer without answering "Yes/No". On the contrary, the smaller LLM can answer politely, starting with "yes". It also names the correct label of the beers but loses the exact introduction. We ascribe these limitations to insufficient training and low diversity of the answers. Nevertheless, it indicates that olfactory embodiment in LLMs is feasible, and we are inspired to diversify both the gas dataset and the corpus.

4 Conclusion and Future Work

This demo presents an olfactory augmented prototype framework for LLMs. The olfaction modality therein is embodied in the LLMs primarily by plug-and-play smelling services and designed instructions. The plug-and-play services are powered by the gas classification models, each trained on one open-source dataset and one customized dataset using self-developed e-noses. Two LLMs are fine-tuned on the self-constructed corpus, which includes four kinds of questions and delicately designed instructions. The use case study and the web demonstration evidence the feasibility and deployability of the proposed prototype.

In the future, we will expand the corpus *w.r.t.* more odors and collect their sensory data. We will explore more powerful LLMs with retrieval-augmented generation (RAG) to answer questions towards any gas. Further, we consider that this and related applications can inspire many domains, such as virtual reality, robotics or other human-centric products.

Acknowledgments

This work was partly supported by the Foundation of National Key Laboratory of Human Factors Engineering (Grant NO. HFNKL2023J08), the National Natural Science Foundation of China (Grant No. 62441219 and No. 62172252), and the Institute of Information & communications Technology Planning & Evaluation (IITP) grant funded by the Korea government (MSIT)(RS-2020-II201373, Artificial Intelligence Graduate School Program (Hanyang University)).

References

[Brohan *et al.*, 2023] Anthony Brohan, Yevgen Chebotar, Chelsea Finn, Karol Hausman, Alexander Herzog, Daniel Ho, Julian Ibarz, Alex Irpan, Eric Jang, Ryan Julian, et al. Do as i can, not as i say: Grounding language in robotic affordances. In *Conference on robot learning*, pages 287–318. PMLR, 2023.

[DeepSeek-AI, 2025] DeepSeek-AI. Deepseek-r1: Incentivizing reasoning capability in llms via reinforcement learning, 2025.

[Gao *et al.*, 2023] Difei Gao, Lei Ji, Luowei Zhou, Kevin Qinghong Lin, Joya Chen, Zihan Fan, and Mike Zheng Shou. Assistgpt: A general multi-modal assistant that can plan, execute, inspect, and learn. *arXiv preprint arXiv:2306.08640*, 2023.

[Han *et al.*, 2023] Jiaming Han, Renrui Zhang, Wenqi Shao, Peng Gao, Peng Xu, Han Xiao, Kaipeng Zhang, Chris Liu, Song Wen, Ziyu Guo, et al. Imagebind-llm: Multi-modality instruction tuning. *arXiv preprint arXiv:2309.03905*, 2023.

[Hassan *et al.*, 2024] Sunzid Hassan, Lingxiao Wang, and Khan Raqib Mahmud. Integrating vision and olfaction via multi-modal llm for robotic odor source localization. *Sensors*, 24(24):7875, 2024.

[Hou *et al.*, 2022] Hui-Rang Hou, Qing-Hao Meng, Pei-Feng Qi, and Li-Cheng Jin. A double triangular feature-based sensor sequence coding approach for identifying chinese liquors using an e-nose system. *IEEE Sensors Journal*, 22(5):3878–3887, 2022.

[Huang *et al.*, 2017] Gao Huang, Zhuang Liu, Laurens Van Der Maaten, and Kilian Q Weinberger. Densely connected convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4700–4708, 2017.

[Karim *et al.*, 2019] Fazle Karim, Somshubra Majumdar, Houshang Darabi, and Samuel Harford. Multivariate lstm-fcns for time series classification. *Neural networks*, 116:237–245, 2019.

[Liu *et al.*, 2024] Xukun Liu, Zhiyuan Peng, Xiaoyuan Yi, Xing Xie, Lirong Xiang, Yuchen Liu, and Dongkuan Xu. Toolnet: Connecting large language models with massive tools via tool graph. *arXiv preprint arXiv:2403.00839*, 2024.

[Lyu *et al.*, 2023] Chenyang Lyu, Minghao Wu, Longyue Wang, Xinting Huang, Bingshuai Liu, Zefeng Du, Shuming Shi, and Zhaopeng Tu. Macaw-llm: Multi-modal language modeling with image, audio, video, and text integration. *arXiv preprint arXiv:2306.09093*, 2023.

[Meng and Hou, 2021] Qing-Hao Meng and Hui-Rang Hou. Electronic nose dataset for recognition of eight liquor types, 2021.

[Qin *et al.*, 2023] Yujia Qin, Shihao Liang, Yining Ye, Kunlun Zhu, Lan Yan, Yaxi Lu, Yankai Lin, Xin Cong, Xiangru Tang, Bill Qian, et al. Toolllm: Facilitating large language models to master 16000+ real-world apis. *arXiv preprint arXiv:2307.16789*, 2023.

[Vemprala *et al.*, 2024] Sai H Vemprala, Rogerio Bonatti, Arthur Bucker, and Ashish Kapoor. Chatgpt for robotics: Design principles and model abilities. *IEEE Access*, 2024.

[Wang *et al.*, 2024] Yabing Wang, Le Wang, Qiang Zhou, Zhibin Wang, Hao Li, Gang Hua, and Wei Tang. Multimodal llm enhanced cross-lingual cross-modal retrieval. In *Proceedings of the 32nd ACM International Conference on Multimedia*, pages 8296–8305, 2024.

[Wijaya *et al.*, 2023] Dedy Rahman Wijaya, Rini Handayani, Tora Fahrudin, Guntur Prabawa Kusuma, and Farah Afianti. Electronic nose and optimized machine learning algorithms for non-infused aroma-based quality identification of gambung green tea. *IEEE Sensors Journal*, 2023.

[Wu *et al.*, 2023] Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. Next-gpt: Any-to-any multimodal llm. *arXiv preprint arXiv:2309.05519*, 2023.

[Wu *et al.*, 2024] Fan Wu, Ruilong Ma, Yiran Li, Fei Li, Shukai Duan, and Xiaoyan Peng. A novel electronic nose classification prediction method based on tetcn. *Sensors and Actuators B: Chemical*, 405:135272, 2024.

[Yang *et al.*, 2024] Yijun Yang, Tianyi Zhou, Kanxue Li, Dapeng Tao, Lusong Li, Li Shen, Xiaodong He, Jing Jiang, and Yuhui Shi. Embodied multi-modal agent trained by an llm from a parallel textworld. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 26275–26285, 2024.

[Zhang *et al.*, 2024] Yu Zhang, Peizhong Gao, Fangzhou Kang, Jiaxiang Li, Jiacheng Liu, Qi Lu, and Yingqing Xu. Odoragent: Generate odor sequences for movies based on large language model. In *2024 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*, pages 105–114. IEEE, 2024.

[Zhong *et al.*, 2024] Shu Zhong, Zetao Zhou, Christopher Dawes, Giada Brianz, and Marianna Obrist. Sniff ai: Is my'spicy'your'spicy'? exploring llm's perceptual alignment with human smell experiences. *arXiv preprint arXiv:2411.06950*, 2024.