LLM Inference to obtain large-scale olfactory data

Yuki Harada, Manuel Aleixandre, Manabu Okumura, Takamichi Nakamoto

Institute of Innovative Research, Institute of Science Tokyo, Tokyo, Japan
harada.y.ai@m.titech.ac.jp

Large-scale odor map that accurately reflects both the diversity and specificity of human olfaction is essential for advancing digital scent technology. Although it is indispensable for the developing that technology, the collection of large-scale data of olfaction needs huge amount of time and labor: Although the sensory dataset made by Dravnieks[1] has been frequently used in this field, it was made about 40 years ago and the amount of the dataset is not enough for modern study of digital scent technologies. The application of large language models (LLMs) to infer large-scale odor maps attracts growing interest, as the access to such data could significantly accelerate the development of digital scent technologies.

We studied two types of odor maps inferred by LLMs. First one was based on LLM-derived similarity scores using short prompts[2] inspired from Kurfali's study[3]. Second one was based on embeddings derived by LLM inference of verbal description.

We performed statistical tests to compare the two types of LLM inferences with Dravnieks' dataset. We observed that attempting to obtain textual descriptions of odors without prompt optimization resulted in variability in expression patterns, leading to inconsistent results with LLM-derived similarity scores. We also found that prompt optimization could reduce inconsistencies in verbal descriptions.

The digitalization of odors and virtual reality (VR) implementation in olfaction have lagged due to the difficulty in making large-scale odor map. This technology here can contribute to olfactory content since the intended scent in the olfactory content can be easily generated based on odor map.

Acknowledgment

This work was supported by the Japan Science and Technology Agency (JST) Future Society Creation Program (Grant Number JPMJMI25H1).

References

1. Andrew Dravnieks. 1985. Atlas of odor character profiles. ASTM, Philadelphia, PA.
2. Yuki Harada, Manuel Aleixandre, Manabu Okumura, and Takamichi Nakamoto. Odor Maps from the LLM-derived similarity scores. arXiv preprint arXiv:2604.20310 (2026).
3. Murathan Kurfalı, Pawel Herman, Stephen Pierzchajlo, Jonas Olofsson, and Thomas Hörberg. Representations of smells: The next frontier for language models? Cognition 264 (2025), 106243.