**Responses to Reviewer 1:**

*Thank you for reviewing our manuscript! Here are our responses to your questions:*

**Question 1:** The article only compares with random walk, vision-only and olfactory-only methods, and does not fully compare with existing advanced OSL algorithms (such as those based on reinforcement learning or deep learning), and cannot prove its performance advantage.

***Author’s response****:* *Thank you for your feedback!*

*The primary objective of our work is to demonstrate semantic Odor Source Localization (OSL) navigation by integrating both olfactory and visual sensory modalities. Many existing OSL algorithms (such as those based on RL or DL) primarily focus on olfactory-only methods. Our experiments show that integration of vision with olfaction outperforms single sensory modality-based vision-only and olfactory-only methods.*

***Add above Conclusion.***

*We added this description in section 4.*

**Question 2:** The reasoning process of LLM relies on pre-trained language models and may not fully understand complex visual and olfactory information, especially when faced with ambiguous perceptual data.

***Author’s response****:* *Thank you for your response!*

*The main innovation of this project lies in integrating vision and olfaction through a multimodal LLM. The multimodal GPT-4 model is capable of reasoning about the robot’s next actions based on current visual and olfactory observations, using common-sense knowledge. Our experiments indicate that highly sophisticated reasoning is not required to guide the robot effectively toward the odor source. Consequently, the multimodal LLM was able to infer potential odor sources in the environment, select appropriate exploratory actions, and ultimately localize the true odor source with reliable performance.*

*We added this description in section 5.*

**Question 3:** There are some grammatical problems in the article.

***Author’s response****:* *Thank you for your feedback!*

*We have carefully reviewed the manuscript and corrected the identified grammatical errors.*