**Responses to Reviewer 2:**

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

**Question 1:** The paper introduces the use of an LLM model, but the core functionality here relies on a Chain-of-Thought reasoning process. This reasoning process is straightforward and direct. Given this, why not implement the Chain-of-Thought reasoning process with a simpler, more efficient program instead of an LLM model?

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

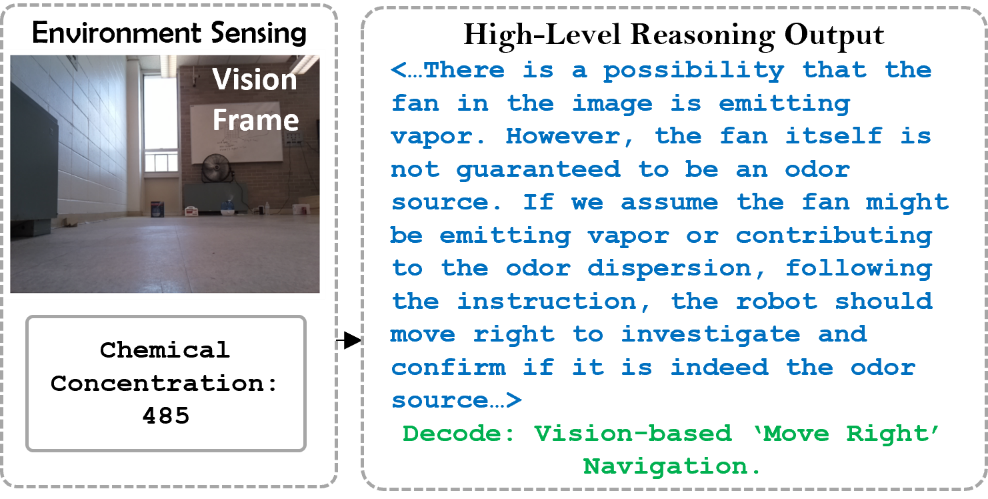
*We should mention that, chain-of-thought reasoning is a type of method in prompt engineering which studies how to ask questions (prompts) to LLM. The goal of chain-of-thought reasoning is to help LLM to decode a complex problem into several middle steps. Therefore, the most common usage of chain-of-thought reasoning is with LLMs.*

*To explain the question, we have added this explanation and reference in lines 284-288 of section 3.3 of our manuscript.*

**Question 2:** The LLM model not only demands significant computational resources, but also requires converting the robot’s sensory data into input for the model and then translating the model’s output back into executable robot commands. This double translation process seems unnecessary and could potentially be eliminated by using a program based on Chain-of-Thought reasoning (similar to traditional expert systems) instead of an LLM model. Could the authors clarify the distinctions between using an LLM model and a traditional expert system or a simpler program implementing Chain-of-Thought reasoning? What is the specific advantage of employing an LLM model?

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

*Compared to other expert systems, the motivation of using LLMs in this work is to utilize LLMs’ strong multi-modal semantic understanding and reasoning capabilities. For example, as shown in the Figure below, the LLM inferred that approaching the fan could lead the robot to the odor source, despite not detecting visible plumes.*

**

*The goal of this project was to leverage these reasoning capabilities to process visual and olfactory data, creating a more effective navigation algorithm. In the experiments, we demonstrated that the LLM-based navigation algorithm outperformed the traditional moth-inspired method and the simple rule-based vision and olfaction ‘Fusion’ navigation algorithm in terms of the average search time and success rate. To answer this question, we added a sentence in lines 65-67 of section 1 that mentions our motivation of using LLM over expert systems, and added comparison of our proposed algorithm to simpler rule-based algorithm in lines 556-566 of section 6.*

**Question 3:** When determining the number of experimental trials, ensuring statistical significance is essential. Based on best practices, if the anticipated effect size is small and significant differences are to be observed across various experimental conditions, a larger number of trials is generally required (e.g., 20-30 trials). With a larger effect size, approximately 10-15 trials per condition may be sufficient. In the current study, each method was tested only four times per scenario, which appears insufficient for statistical reliability. Could the authors address this concern and provide a rationale for the chosen number of trials?

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

*In our experiment design, we defined two searching environments, including laminar flow search and turbulent flow search. To ensure statistical significance, we conducted a total of 64 trials for each searching environment. For each environment, we used four distinct starting positions to demonstrate that our proposed method performs well from various initial positions and orientations.*

*To explain this, we added lines 409-413 of section 4.2.*

**Question 4:** In the conclusion section, although the paper summarizes its main contributions and innovations, specific findings—particularly any quantifiable results—are not clearly articulated. I recommend presenting the conclusions in a list format to clearly emphasize key findings, making it easier for readers to identify the primary outcomes of the study.

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

*According to your feedback, we have presented our contributions in a list format in the conclusion section. The experimental results indicate that the proposed LLM-based navigation algorithm outperformed the three comparison algorithms.*

* *In laminar search scenarios, the proposed method reduces average search time by 18.1 seconds compared to the ‘Olfaction-only’ navigation algorithm, 14.9 seconds compared to the vision-based algorithm, and 3.87 seconds compared to the simple rule-based ‘Fusion’ of vision and olfaction. The proposed method also achieves a 100% success rate, outperforming ‘Olfaction-only’ (60%), ‘Vision-only’ (50%), and the rule-based ‘Fusion’ algorithm (75%).*
* *In turbulent search scenarios, the proposed method shortens average search time by 5.37 seconds relative to the vision-based algorithm and 12.49 seconds compared to the rule-based ‘Fusion’ approach. It achieves a 75% success rate, whereas the ‘Olfaction-only’ algorithm has 0% success, the ‘Vision-only’ algorithm 12.5%, and the rule-based ‘Fusion’ 50%.*

*According to your feedback, the quantifiable results are added in lines 556-566 of section 6.*

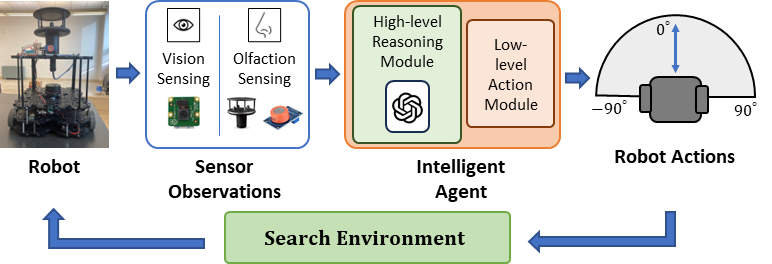
**Question 5:** In Figures 6, 8, and other illustrations, as well as in the main text, describing indoor airflow as "laminar" is inaccurate. Even without ventilation, indoor airflow is typically turbulent. If you intend to describe the flow’s pattern, it is usually classified as either a unidirectional flow environment or a non-unidirectional turbulent flow environment; however, both are turbulent environments. Please review and revise the relevant descriptions accordingly.

***Author’s response***: *Thank you for your feedback! In this project, we define 'laminar airflow' as the condition when only one fan is used, creating a unidirectional flow, and 'turbulent airflow' as the condition when two perpendicularly placed fans are used, creating a non-unidirectional flow. We have added this explanation in the lines 362-364 of the section 4.1 to clarify our terminology.*

**Minor Comments**

**Comment 1**: In the introduction, the visual elements of Figure 1 and its textual description do not align well. It may be beneficial to redraw Figure 1 to visually highlight key components, such as the LLM, high-level navigation behavior, and low-level action module, for better clarity.

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

**

*According to your feedback we have updated the Figure 1.*

**Comment 2**. In Figure 1, the term “serarch” should be corrected to “search.” Please review the entire manuscript for similar typographical errors, as they reflect the overall thoroughness of the work.

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

*We have corrected this typographical error, and reviewed the entire manuscript for similar errors. And we did not find other grammar or spelling issues.*

**Comment 3**. Since arXiv papers have not undergone peer review, it may be beneficial to reduce reliance on these sources, such as references 20-22, particularly for critical citations.

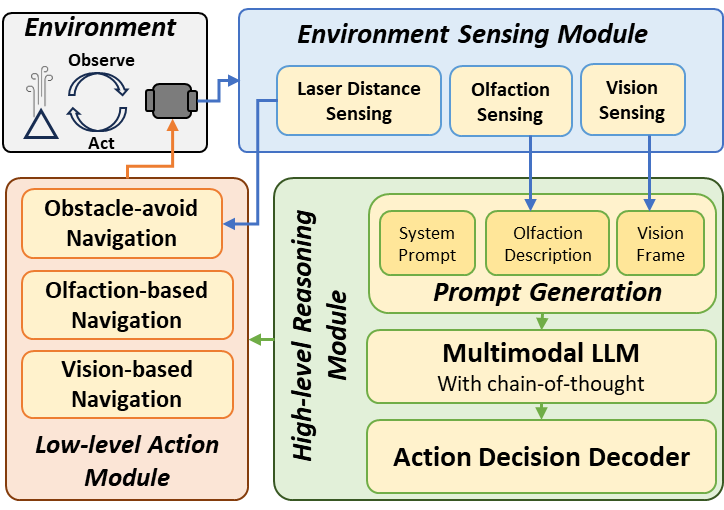
**Author’s response**: *Thank you for your feedback!*

*According to your feedback, we removed arXiv paper citations for references 20-22. We used some highly-cited arXiv paper in other sections.*

**Comment 4**. In Figure 2, consider adding an icon representing the robot. I understand that the Low-level Action Module directs the robot’s actions.

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

*According to your feedback we have updated Figure 2.*

**

**Comment 5**. Throughout the manuscript, I suggest placing descriptive text for Figures before the Figures themselves.

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

*According to your feedback we have updated the organization of the manuscript.*

**Comment 6**. In describing Figure 4, it is stated: “If a valid odor source object is later identified visually, the system will switch back to vision-based navigation again.” However, this flow is not evident in the flowchart depicted in Figure 4. Please clarify this in the diagram.

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

*The Figure shows the hierarchical decision making process of one timestep. This decision making process will repeat until the robot finds the odor source, moves out of the search area, or gets out of time.*

*Based on your feedback, we have updated the Figure 4 and added some explanations in lines 292-298 of section 3.3.*

**Comment 7**. In Section 4.2, the authors mention: “For each navigation algorithm, the mobile robot was initialized from the same four starting positions-orientations. Four test runs were recorded from each starting position, totaling 96 test runs.” Please verify whether this total should actually be 128 test runs.

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

*We have corrected the total number of tests in lines 412-413 of section 4.2.*