

Article

# INTEGRATING VISION AND OLFACTION VIA MULTI-MODAL LLM FOR ROBOTIC ODOR SOURCE LOCALIZATION

Sunzid Hassan <sup>1,†</sup>, Lingxiao Wang <sup>2,\*</sup>, and Khan Raqib Mahmud <sup>1</sup>

<sup>1</sup> Department of Computer Science, Louisiana Tech University, 201 Mayfield Ave, Ruston, LA 71272, USA; sha040@latech.edu (S.H.); krm070@email.latech.edu (K.R.M.)

<sup>2</sup> Department of Electrical Engineering, Louisiana Tech University, 201 Mayfield Ave, Ruston, LA 71272, USA; lwang@latech.edu (L.W.)

\* Correspondence: lwang@latech.edu; Tel.: +1-318-257-2758

**Abstract:** Odor Source Localization (OSL) technology allows autonomous agents like mobile robots to localize target odor source in an unknown environment. This is achieved by an OSL navigation algorithm that processes agents sensor readings to calculate action commands to guide the robot to locate the odor source. Compared to traditional ‘Olfaction-only’ OSL algorithms, our proposed OSL algorithm integrates vision and olfaction sensor modalities to localize odor sources even if olfaction sensing is disrupted by non-unidirectional airflow or vision sensing is impaired by environmental complexities. The algorithm leverages the zero-shot multi-modal reasoning capabilities of large language models (LLMs), negating the requirement of manual knowledge encoding or custom-trained supervised learning models. A key feature of the proposed algorithm is the ‘High-level Reasoning’ module, which encodes the olfaction and vision sensor data into a multi-modal prompt and instructs the LLM to employ a hierarchical reasoning process to select an appropriate high-level navigation behavior. Subsequently, the ‘Low-level Action’ module translates the selected high-level navigation behavior into low-level action commands that can be executed by the mobile robot. To validate our algorithm, we implemented it on a mobile robot in a real-world environment with non-unidirectional airflow environments and obstacles, to mimic a complex, practical search environment. We compared the performance of our proposed algorithm to single sensory modality-based ‘Olfaction-only’ and ‘Vision-only’ navigation algorithms, and a supervised learning-based ‘Vision and Olfaction Fusion (Fusion)’ navigation algorithm. Experimental results show that the proposed LLM-based algorithm outperformed the other algorithms in terms of success rates and average search times in both unidirectional and non-unidirectional airflow environments. The code for this work can be found at: [https://github.com/SunzidHassan/24\\_LLM-OSL](https://github.com/SunzidHassan/24_LLM-OSL).

**Citation:** Hassan, S.; Wang, L.; Mahmud, K. Integrating Vision and Olfaction via multi-modal LLM for Robotic Odor Source Localization. *Sensors* **2024**, *1*, 0. <https://doi.org/10.3390/sensors2024001>

Received:

Revised:

Accepted:

Published:

**Copyright:** © 2024 by the authors. Submitted to *Sensors* for possible open access publication under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

## 1. Introduction

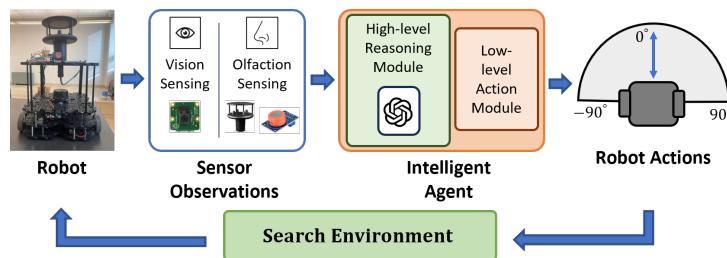
Humans sense the external environment using sensory systems such as vision, olfaction, audition, etc. The data is then used for decision-making within the environment. Similarly a mobile robot, can perceive the environment using artificial sensory devices like a camera, chemical sensor, microphone, etc. Of the sensory systems, olfaction is the first to evolve in organisms [1], allowing them to detect predators, food, potential mates, etc. [2]. However, the application of olfaction in robotics is not well studied. Robotic OSL is the technology that allows robots to localize an unknown odor source in the surrounding environment [3]. The technology is used in monitoring wildfires [4], chemical gas leaks [5], air pollution [6], underground gas leaks [7], identifying unexploded mines and bombs [8], locating hydrothermal vents [9], among others.

Advancements in robotics and autonomous systems have enabled the deployment of mobile robots to locate odor or chemical sources. Identifying the source of an unknown odor necessitates a proficient OSL navigation algorithm that directs the robot based on sensor readings [10]. Conventional OSL algorithms comprise animal behavior mimicking bio-inspired techniques, mathematical model-based engineering-based approaches, and machine learning-based strategies. Notable bio-inspired techniques include the Moth-inspired algorithm, which emulates the mate-seeking behaviors of male moths [11], allowing a robotic agent to perform 'surge/casting' movements [12] to localize the odor source. Engineering-based approaches include the Particle Filter algorithm [13], infotaxis [14,15], stochastic mapping [16], etc. These methods update predictions regarding the odor source's location based on olfactory observations. Lastly, machine learning-based OSL methods feature reinforcement learning [17] and deep supervised learning [18] techniques.

Most of the traditional OSL algorithms rely on olfactory (i.e., chemical and airflow) sensing to detect and navigate to the target odor source. However, methods that depend exclusively on olfactory sensing tend to underperform in environments where non-unidirectional airflow disrupts olfactory detection. The integration of vision and olfaction offers a great advantage to 'Olfaction-only' systems. For example, fruit flies use a combination of vision and olfaction to locate aromatic and colorful food [19]. They integrate olfactory information with visual landmark information to 'see' where the fruit aroma is coming from. This sensory integration allows them to precisely pinpoint the location of the fruit in their environment. Similarly, a robot equipped with both olfactory and vision-sensing abilities (such as a camera and a chemical sensor) and a navigation algorithm capable of effectively integrating these sensory modalities can more efficiently locate an unknown odor source in complex environments.

Humans often recognize visual objects in the surrounding environment and assume their relationship to the goal of making navigation decisions. A navigation system that tries to imitate such behavior needs to have several complex abilities – the ability to understand navigation objectives, the ability to detect objects in sensory inputs like vision, the ability to deduce contextual relation of those objects to the navigation goal, etc. Multimodal LLMs demonstrate state-of-the-art performance in reasoning over multiple sensory modalities like text, vision, and sound [20]. 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. However, applying these models in robotics introduces additional challenges, such as converting robot sensor readings into a format that can be processed by the LLMs, and subsequently translating the LLM's textual outputs into actionable robot commands.

The proposed OSL system is shown in Fig. 1. The core of the system is an intelligent agent, which encodes vision and olfaction observations with a hierarchical navigation behavior selection instruction set for an LLM. The LLM then applies reason over the multi-modal input and selects a high-level navigation behavior. Finally, a low-level action module translates the navigation behavior for the mobile robot. To validate the proposed algorithm, we conducted tests in a real-world environment where olfaction is challenged by non-unidirectional airflow, vision is challenged by obstacles and multi-modal reasoning is challenged by environment complexities.



**Figure 1.** Flow diagram of the OSL system. The robot platform is equipped with a camera for vision and a chemical detector and an anemometer for olfaction sensing. The proposed algorithm utilizes multi-modal LLM for navigation decision-making.

The main contributions of this work can be summarized as follows:

1. Integrating vision and olfaction sensing to localize odor source in complex real-world environments.
2. Developing an OSL navigation algorithm that utilizes zero-shot multi-modal reasoning capability of multi-modal LLMs for OSL. This includes designing modules to process inputs to and outputs from the LLM model.
3. Implementing the proposed intelligent agent in real-world experiments and comparing its search performance to the supervised learning-based Fusion navigation algorithm [21].

In the following sections of this paper, Section 2 includes literature review of recent OSL algorithms; Section 3 reviews details of the implemented OSL algorithm; Section 4 details experiment setup and results; Section 5 includes limitations and future research directions; and, finally, Section 6 presents the conclusion of the work.

## 2. Related Works

### 2.1. Olfactory-based Methods

Various organisms utilize olfaction to localize odor sources. This includes a bacterium navigating gradient in amino acid or a lion tracking prey. Algorithms that mimic olfaction-based navigation behaviors of organisms is an effective approach in robotic OSL research.

Chemotaxis represents the simplest OSL strategy in biological organisms, where navigation relies solely on olfaction. For instance, bacteria demonstrate chemotaxis by altering movement against odor concentration changes. They make less temporary turns in the presence of attractive chemical, resulting in straighter movement. Conversely, in the absence of a gradient or when moving away from higher concentrations, their default turning probability remains the same [22]. This straightforward approach allows single-celled organisms to navigate a gradient of appealing chemicals through a guided random walk. Nematodes [23] and crustaceans [24] also utilize chemotaxis-based OSL. Early OSL efforts focused on implementing such simple gradient-following chemotaxis algorithms. Typically, these methods used a pair of chemical sensors on plume-tracing robots, guiding them towards areas with higher concentration readings [25]. While early studies [26–29] validated chemotaxis in unidirectional flow environments. However, alternative OSL methods were proposed for complex non-unidirectional flow environments.

Anemotactic is a more sophisticated bio-inspired OSL method that uses both chemical and airflow senses for navigation. A wide variety of organisms, including moths [30–32], birds [33,34], etc. utilize this approach. Specifically, prevalent moth-inspired method was developed by mimicking the mate-seeking behavior of male moths [35]. This was reinforced with additional bio-inspired search strategies in recent times, including zigzag, fuzzy-inference, and multi-phase exploratory [36] search behaviors. Jin et al [37] trained a neural network to predict gas flow patterns in environment with obstacles, and combined it with the probabilistic Source Term Estimation (STE) algorithm to localize gas sources in simulated and built environments. Ojeda et al [38] utilized predictive dispersion filament model to predict gas-hit maps from airflow data, and compared it to measured gas-hit

map to update source probability distribution. Bio-inspired methods are also applied for complex three-dimensional search areas [39,40]. Notice that both chemical concentration and wind direction data is used in our proposed olfactory-based behaviors.

Engineering-based methods differ from bio-mimicking algorithms by the utilization of mathematical models to estimate locations of odor sources. They involve discretizing the search area and learning the likelihood of each region containing the odor source. Algorithms used for generating such maps include particle filters, infotaxis [14,15], stochastic mapping [16], information-based search [41], source term estimation [42], partially observable Markov decision processes [43], reactive-probabilistic search [44], etc. After predicting the odor source location, robots are then guided to the source through path-planning algorithms like the artificial potential fields and the A-star [45,46].

Machine learning (ML) based methods have achieved state-of-the-art performance in tasks of multiple data modalities, including vision, audio and text. These methods have increasingly been applied for OSL tasks as well. In these types of method, artificial deep neural network processes robot sensor data and generates robot headings [10]. These models are trained using supervised or reinforcement learning methods. In supervised learning-based methods, ANN is trained as a classifier to predict the location of odor source. For example, Kim et al. [18] trained a Recurrent Neural Network (RNN) with a long-short-term memory (LSTM) module to predict odor source in simulation data. Bilgera et al. [47] used a convolutional LSTM model to detect odor source in data measured from an anemometer and chemical sensors. Thrift et al. [48] trained a convolutional neural network (CNN) and a support vector machine (SVN) to classify directions of multiple odor sources. In reinforcement learning, ANN is trained to generate action decisions to approach odor source location. For example, Hu et al. [17] used deterministic policy gradient actor-critic network for Autonomous Underwater Vehicle (AUV) navigation. Wang et al. [49] developed an adaptive neuro-fuzzy inference system (ANFIS) for OSL in a simulated environment. Both of these methods were validated in simulated environments, highlighting the need for real-world implementations.

## 2.2. Vision and Olfaction Integration in OSL

The bio-inspired, engineering-based, and learning-based methods discussed above are 'Olfaction-only'. Olfaction-based approaches suffer if olfaction sensing is disturbed by non-unidirectional airflow, which is a common occurrence in real-world environments. Additionally, olfaction data is typically represented as the concentration level or detection rate of a chemical (e.g., ethanol). These representations inherently contain limited information about the location of the odor source. Thus, it is unclear if more complex algorithms can extract the ever-increasing amount of information from the olfaction data. Thus, it can be argued that the addition of vision sensing is the next paradigm in OSL research. Among the existing literature that utilized vision sensing in OSL, Monroy et al. discussed using vision sensing with olfaction sensing for Gas Source Localization [50]. They defined the odor footprint of some predefined objects using web ontology language (WOL). They used the You Only Look Once v3 (YOLOv3) model for those objects and looked up the odor footprint of those objects from the predefined knowledge base. The requirement of knowledge definition makes the model less scalable for complex environments.

In our previous work, we fused vision and olfaction for OSL using a custom-trained YOLOv6 model that directly detects visible plumes in the vision frame [21]. The algorithm was effective in localizing odor sources in real-world environments with obstacles and complex airflow. However, the vision model required visible odor plumes, and the algorithm followed olfaction-based navigation if odor plumes were invisible or obstructed. But even without visible odor plumes, vision data can still contain latent odor source location information that can help narrow search boundaries. For example, we may narrow our odor source search area to a restaurant without directly seeing the odor-emitting food. This information extraction requires the visual reasoning ability that multi-modal LLMs possess. This work aims to mitigate the limitations of previous vision and olfaction-based OSL

models, i.e., to replace manual knowledge-based and supervised learning-based models with multi-modal reasoning-based models.

### 2.3. *LLM in Robotics*

Large Language Models are a major milestone in the research of Natural Language Processing (NLP). LLMs are specialized models for natural language generation [51]. These models are trained in a self-supervised learning approach - which negates the requirement for labeled training data. This allows the models to be trained on vast textual data on the internet. Additionally, it has been shown that there are similarities between the vision understanding by mammalian brains and by self-supervised learning approach [52] that is utilized by LLMs. The models are based on transformer architecture with a self-attention mechanism that allows them to learn complex interrelations in textual data [53]. LLMs exceed previous RNN-based language models due to emergent abilities, including chain-of-thought reasoning [54], instruction understanding [55], and in-context learning [56]. Notable examples of LLMs include BERT [57], GPT-3 [56], LLaMA [58], etc.

To further enhance the applications of LLMs in embodied intelligence tasks, researchers are training these models with multi-modal data - like text, image, audio, etc. These models are termed as Vision Language Models (VLM) or multi-modal LLM [59]. Unlike supervised vision classifiers, multi-modal LLMs are simultaneously trained with vision and language data. For example, the multi-modal LLM CLIP [60] is trained to minimize the distance of related images and texts in a high-dimensional representation space. Training over massive multi-modal datasets allows these models to learn complex interrelationships among textual concepts and visual objects. This allows LLM-based robots to make zero-shot or few-shot reasoning over visual objects and states in a complex environment [61]. Thus, multi-modal LLMs are increasingly used in robotics tasks like generating robot action plans by reasoning over multi-modal sensor data [62].

In recent years, a rich collection of work has been published in the field of LLM-based robot navigation. These works can be broadly categorized into planning and semantic understanding models. Planning-based methods directly generate action decisions to guide the agent. Examples of such models include Clip-Nav [63] which utilizes an LLM for extracting location key phrases from the provided navigation objective, and uses CLIP VLM to ground the key phrases in the visual frame for navigation. A<sup>2</sup>Nav [64] has five pre-defined actions, and separate navigators are trained for each of those actions. It utilizes the GPT-3 model for predicting action, and the BERT model for aligning the predictions with the pre-defined actions. NavGPT [65] utilizes the GPT-4 model for zero-shot navigation in simulated indoor scenarios. VELMA [66] identifies landmarks from human-authored navigation instructions, and uses CLIP to ground them in panoramic view of the robot. The model then generates a textual representation of the environment for textual command-based navigation. Semantic understanding models process sensor inputs, and the insights are then used to generate agent actions. Examples of such models include LM-Nav [67], which uses GPT-3 to translate verbal instructions into a series of textual landmarks. CLIP grounds the landmarks to a topological map, and a self-supervised robotic control model executes the physical actions. L3MVN [68] uses a language module to handle natural language instructions - generating a semantic map embedded with general physical world knowledge. Another module employs the semantic map to guide robotic exploration. ESC [69] conducts zero-shot object navigation by leveraging commonsense knowledge from pre-trained language models. It uses LLM to ground objects and to deduce the semantic relationship of those objects in an indoor environment. Exploration techniques like 'Frontier-based exploration' are used to navigate based on the semantic map. Concept fusion [70] utilizes a multi-modal LLM to generate a multi-modal semantic map of the environment. The model can perform navigation using textual, visual, or audio cues.

## 2.4. Research Niche

The proposed LLM-based intelligent agent distinguishes itself from current LLM-driven robotic applications in two key ways. (i) First, our system differs in its input requirements. Rather than relying solely on visual observations, our model is designed to process both visual and olfactory sensory data. These multi-modal inputs provide the robot with a more comprehensive understanding of its environment, enabling richer interactions. (ii) Second, our model is purpose-built for a specific task: robotic OSL. Unlike generalized LLM-driven robots, which require vast amounts of training data and substantial computational resources, our system focuses on a specialized task. For example, training a general LLM-driven robot, such as Google's RT-1 [71], for various object manipulation tasks involved data collection from 13 robots over 17 months, a costly process. In contrast, our system will leverage pre-trained multi-modal LLMs.

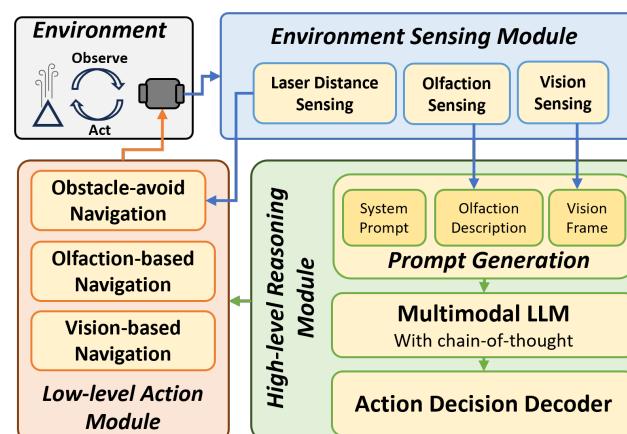
## 3. Methodology

### 3.1. Problem Statement

The objective of robotic OSL is to develop a navigation algorithm that can subscribe to environment observations (i.e., state) from a mobile robot and process the state to generate action instructions for the robot to localize an unknown odor source in the robot's surrounding environment. This process can be represented as:

$$a^t = F(s^t), \quad (1)$$

where  $s^t$  is the robot observations at time  $t$ , and  $a^t$  is the action output by the OSL function  $F$ .

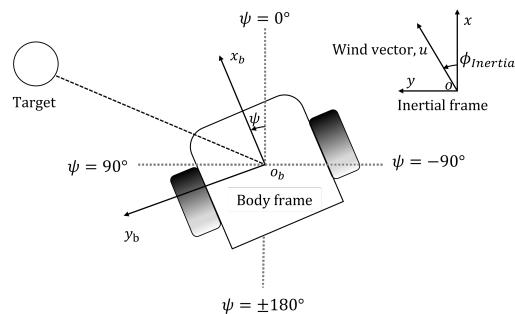


**Figure 2.** The framework of the proposed multi-modal LLM-based navigation algorithm. The three main modules are the 'Environment Sensing Module', 'High-level Reasoning Module', and 'Low-level Action Module'.

Fig. 2 illustrates the proposed Robotic OSL framework. The algorithm has three primary modules – the 'Environment Sensing' module (subsection 3.2) that processes robot sensory inputs, the 'High-level Reasoning' module (subsection 3.3) that reasons over the input and decide a high-level navigation behavior, and the 'Low-level Action' module (subsection 3.4) translates those high-level behaviors into low-level actions that are executable by the robot.

### 3.2. Environment Sensing Module

249



**Figure 3.** Robot notations. Robot position ( $x, y$ ) and heading  $\psi$  are monitored by the built-in localization system. Wind speed  $u$  and wind direction are measured from the additional anemometer in the body frame. Wind direction in inertial frame  $\phi_{Inertial}$  is derived from robot heading  $\psi$  and wind direction in body frame.

**Table 1.** Environment sensing parameters.

Symbols	Parameters
$p$	Visual Observation
$u$	Wind Speed
$\phi_b$	Wind Direction
$\rho$	odor concentration

Fig. 3 illustrates the environment sensing notations for this project. The agent is placed in an environment with a  $x - o - y$  inertial frame. The agent senses the environment in terms of its body frame  $x_b - o_b - y_b$ . Table 1 Includes the parameter definitions and sensors. The mobile robot used in this work has a camera for visual detection, an anemometer and a chemical sensor for olfactory detection, and a laser distance sensor (LDS) for obstacle distance detection. The visual frame captured by the camera is the visual observation  $p$ . An anemometer senses wind speed  $u$  m/s, and wind direction  $\phi_b$ -degrees in the body frame. The odor concentration  $\rho$  is expressed in ppm. At time  $t$  the observed state by the robot is  $s^t = [p, u, \phi_b, \rho]^t$ . The sensors used in real-world experimentation are discussed in subsection 4.3.

250  
251  
252  
253  
254  
255  
256  
257  
258  
259

### 3.3. High-level Reasoning Module

260

'High-level Reasoning Module' is the core of our proposed algorithm. The proposed algorithm uses a multi-modal LLM to perform zero-shot reasoning over multi-modal sensory inputs and decide high-level navigation behavior. Fig. 2 shows the three main sub-modules: (1) prompt generation; (2) multi-modal reasoning; and (3) action decoding.

261  
262  
263  
264  
265  
266  
267

**Prompt-generation** is the first step in this module. Formulating effective prompts is crucial for LLM's reasoning process. Fig. 4 shows the prompt design, that includes the 'System Prompt' and the 'Olfaction Description'. Specifically, the system prompt includes:

- **Task:** Describes the objective and process the LLM should follow.
- **Actions:** Lists the actions available for the LLM to choose from.
- **Hints:** Guides the LLM to select appropriate vision-based or olfaction-based actions based on multi-modal reasoning.
- **Output Instruction:** Directs the LLM to generate only the action without additional reasoning.

268  
269  
270  
271  
272  
273

The olfaction description includes current odor concentration  $\rho$  and odor concentration threshold as numeric values. The final prompt integrates all these instructions.

274  
275

Fig. 5 shows process of querying the LLM. In this work, we employed GPT-4 as the multi-modal LLM. By default, requests are sent to GPT-4 as a JSON payload, where the

276  
277

image input  $p$  is encoded to text string using the default 'base64' function. Upon receiving the payload, GPT-4 then decodes this string back into an image format for processing. All these processes are encapsulated inside the GPT-4. That means that the GPT-4 is able to analyze the graphic information.

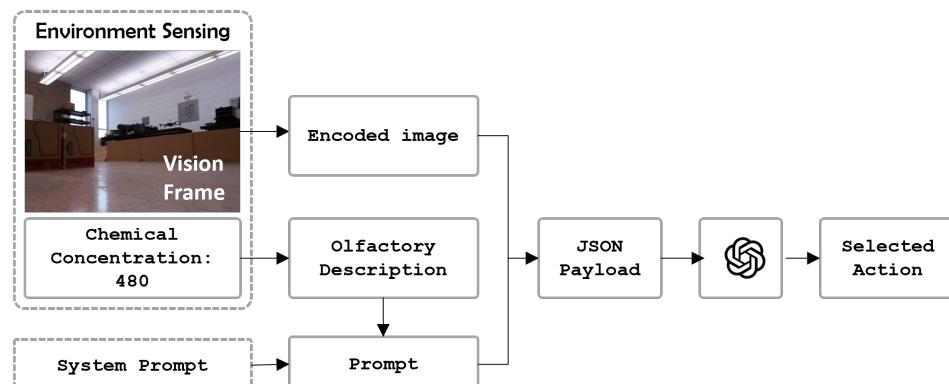
<b>task</b> = """"	<b>System Prompt</b>
Select the best action for a mobile robot to move towards an odor/plume/vapor emitting object. You will select the action by analyzing given robot egocentric visual frame and odor concentration data. You will be given: - A list of actions to select from - Hints to select the actions - Instructions of how to respond to user. All of these inputs are delimited by {delimiter}. """	
<b>actions</b> = """"	
1: Move forward. (Action_id = 1). 2: Move right. (Action_id = 2). 3: Move left. (Action_id = 3). 4: Follow high odor. (Action_id = 4). 5: Find odor emitting object. (Action_id = 5). """	
<b>hints</b> = """"	
Hint 1: If there is an odor/plume/vapor emitting object in the image, select actions from (1-3) to approach the object. - If the object is in the center of the image, select Move forward (Action_id = 1); - if it's in the right half of the image, select Move right (Action_id = 2); - if it's in the left half of the image, select Move left (Action_id = 3). Hint 2: If there is no odor/plume/vapor emitting object in the image, then check the 'odor_concentration' and 'odor_concentration_threshold' values. - If the 'odor_concentration' value is greater than the 'odor_concentration_threshold' value, then select action 'Follow high odor' (Action_id = 4). - Otherwise, select action 'Find odor emitting object' (Action_id = 5). """	
<b>output_instructions</b> = """"	
Response to user:{delimiter} <Respond with the corresponding numerical value of the 'Action_id' (1, 2, 3, 4, or 5) without any additional text or punctuation.> """	
<b>odor_concentration</b> = <Numerical Value> <b>odor_concentration_threshold</b> = <Numerical Value>	<b>Olfaction Description</b>
<b>delimiter</b> : #####	<b>Prompt</b>
<b>prompt</b> = f""" {delimiter} Task: {task} {delimiter} Available Actions: {actions} {delimiter} Hints: {hints} {delimiter} Output Instructions: {output_instructions} {delimiter} "odor_concentration" Value: {chemical} {delimiter} "odor_concentration_threshold" Value: {chmThr} """	

**Figure 4.** Implementation of the prompt. The System Prompt includes the task, actions, hints and output instructions. The final Prompt includes the system Prompt and the Olfactory Description.

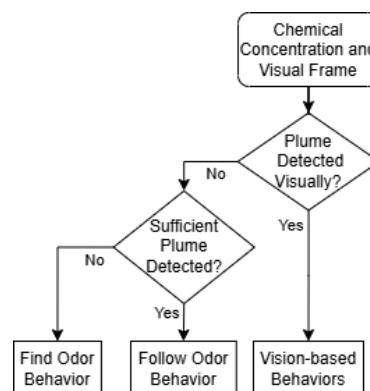
The LLM was instructed to use the Chain-of-Thought reasoning process [54] to capture logical coherence in multi-modal reasoning process over complex multi-modal sensory inputs. We should mention that, chain-of-thought reasoning is a type of method in prompt engineering which studies how to ask questions (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. Based on the provided prompt, the multi-modal LLM model selects appropriate high-level 'Vision-based' or high-level 'Olfaction-based' navigation behaviors. The 'System

Prompt' contains instructions for the LLM to follow a hierarchical order while selecting the high-level navigation behaviors.

Fig. 6 illustrates the hierarchical reasoning strategy for one time-step, which was modeled after human odor search behaviors, where vision and olfaction are used sequentially rather than simultaneously. Upon detecting an odor, vision is used first to locate the odor source. If visual reasoning fails to identify the source, the olfaction-based approach is employed to guide the robot toward the odor source. This decision making process will repeat until the robot finds the odor source, moves out of the search area, or gets out of time. Humans typically utilize vision to narrow down the odor source location. Based on 'common sense', humans can infer which objects within their visual field are likely to be odor sources. For instance, if we smell gas in a kitchen, we can deduce that the stove is a likely odor source. In this case, visual reasoning is utilized to pinpoint the odor source. Similarly, LLMs possess this kind of multi-modal 'common sense' reasoning, allowing them to deduce potential odor sources in their visual field.



**Figure 5.** Querying the LLM with image and prompt. The input of the model is the visual frame and the prompt. The output of the model is the high-level action selection.



**Figure 6.** The flow diagram of the 'High-level Reasoning Module'. It illustrates how the proposed LLM-based agent integrates visual and olfactory sensory observations to make high-level navigation behavior decisions.

The implemented reasoning module performs two primary visual reasoning tasks – 1) finding odor source location information in the visual frame, i.e., odor source location or possible odor source direction, and 2) selecting appropriate 'Vision-based' navigation behavior, i.e., forward, left or rightward movement, to directly approach the odor source location. Otherwise, it will analyze the olfaction description and select either the 'Follow Odor' or 'Find Odor' navigation behavior. If a valid odor source object is later identified visually, the system will switch back to vision-based navigation again. Lastly, the 'Action Decoder' extracts the output navigation behavior from the LLM and passes it to the 'Low-level Action Module'.

### 3.4. Low-level Action Module

The proposed algorithm has three high-level navigation behaviors - 'Obstacle-avoid', 'Vision-based', and 'Olfaction-based' navigation behaviors. Of these, the 'Obstacle-avoid' behavior is triggered directly if the LDS reading indicates that the robot is approaching an obstacle. The 'Vision-based' and the 'Olfaction-based' navigation behaviors are selected by the 'High-level Reasoning Module'. The 'Low-level Action' module then translates those high-level behaviors into low-level action vector

$$a = [v_c, \omega_c], \quad (2)$$

where  $v_c$  is the linear velocity (m/s) and  $\omega_c$  is the angular velocity (rad/s). The action vector is transmitted to and directly executed by the mobile robot.

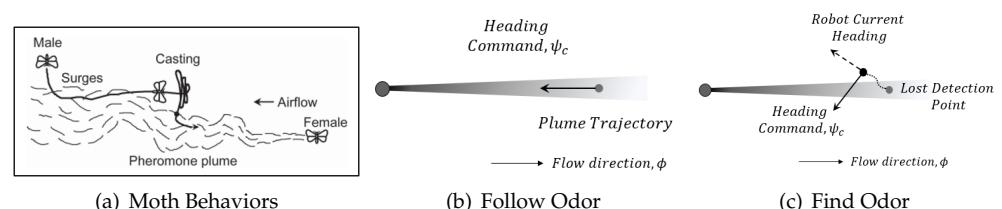
**'Obstacle-avoid':** This behavior is activated when a nearby obstacle is detected by the onboard LDS. The 'Obstacle-avoid' behavior directs the robot to navigate around the obstacle without deviating significantly from the direction the robot was following. Details of this navigation behavior are outlined in our previous paper [21].

**'Vision-based':** is a class of behaviors that are selected and returned from the 'High-level Reasoning' module. The core strategy of vision-based navigation is to keep the detected target in the middle of the image. If the 'High-level Reasoning' module selects 'Vision-based Navigation' behavior, it returns one of three values for 'behavior' - 'Front', 'Left' or 'Right', indicating if the robot should approach straight ahead, move towards right or left to approach the odor source. Equation 3 is used by the 'Low-level Action' module for calculating linear and angular velocities, where the velocities are fixed as constant values.

$$\omega_c = \begin{cases} 0 & \text{if action} = \text{'Front'}; \\ \text{constant} & \text{if action} = \text{'Left'}; \\ -\text{constant} & \text{if action} = \text{'Right'}. \end{cases} \quad (3)$$

This means if 'behavior' is 'Front', the robot will go straight ahead with a constant linear velocity without any angular velocity. If 'behavior' is returned as 'Right' or 'Left', the robot will execute both constant linear and angular velocity to rotate to the right or left to face the odor source.

**'Olfaction-based':** Finally, we utilize the moth-inspired 'Surge' movement for implementing the high-level 'Follow Odor' behavior, and the 'Casting' movement for implementing the high-level 'Find Odor' behavior [72]. Fig. 7 shows the moth-inspired behaviors. In the 'Surge' behavior, the robot moves upwind toward the odor source. In 'Casting', the robot moves crosswind to increase the likelihood of encountering odor plumes.



**Figure 7.** (a) Moth mate seeking behaviors. This figure was retrieved from [73] (b) Moth-inspired 'Surge' and (c) 'Casting' navigation behaviors.

Equation 4 shows the target heading  $\psi_c$  calculation for the two behaviors. Angular velocity  $\omega_c$  is then adjusted to achieve the target heading  $\psi_c$ .

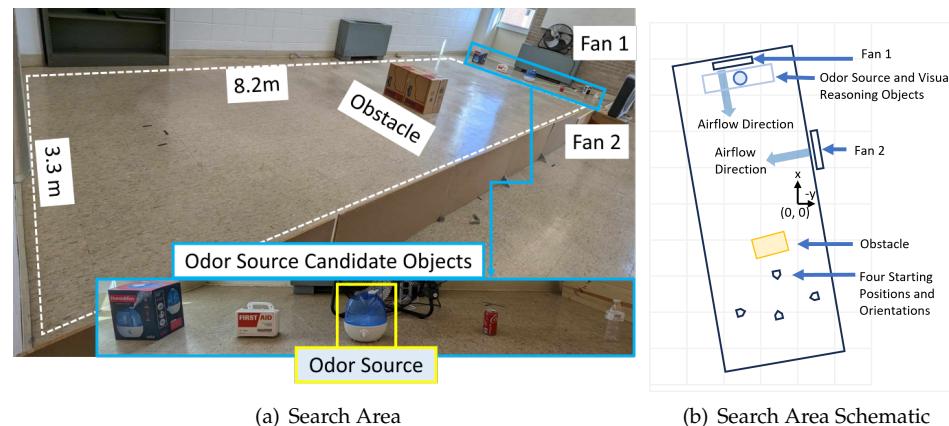
$$\psi_c = \begin{cases} \phi_{\text{Inertial}} + 180 & \text{if action} = \text{'Follow Odor'}; \\ \phi_{\text{Inertial}} + 90 & \text{if action} = \text{'Find Odor'}. \end{cases} \quad (4)$$

It should be noted that we used a reactive method due to its simplicity and lower computational cost compared to engineering-based methods (such as those involving mapping or memory). Adding mapping and memory would require more computational resources and could slow down the decision-making process. However, it should be noted that engineering-based methods can be implemented in the proposed 'Low-level Action' module without requiring any changes to the existing 'High-level Reasoning' module.

## 4. Experiment

### 4.1. Experiment setup

The focus of the experiment is to test if the proposed navigation algorithm can reason over vision and olfaction sensory inputs to determine the actions to localize an unknown odor source. Fig. 8 shows the search area used for the OSL navigation experiment. The search area has an obstacle in the middle to simulate complex indoor environments. There are multiple candidate odor source objects placed in the upwind area. The LLM must use reason to determine the correct odor source object from the candidate objects. The search area has an obstacle in middle. The purpose of the obstacle is to mimic constructed indoor environments, such as household environments, office environments, etc. The obstacle also blocks initial odor source vision. In order to succeed in localizing odor source in this search area, the navigation algorithm must integrate vision and olfaction sensing and reason over them effectively. In this project, we define 'unidirectional airflow' as the condition when only one fan is used, and 'non-unidirectional airflow' as the condition when two perpendicularly placed fans are used. The odor concentration threshold is set to the background concentration, which is determined when the chemical sensor is not within the alcohol plume environment. The task is concluded successfully if the robot reaches within 0.8 m of the odor source within 120 s. It is important to note that this work primarily focuses on the design of the navigation algorithm. The key research question we aim to address is how to process multi-modal sensory inputs to compute robot actions that guide the robot toward the odor source. Source declaration in this work is defined by a distance threshold, meaning that reaching the threshold is considered as detecting the source.



**Figure 8.** (a) Figure of the search area. The size of the search area is  $8.2m \times 3.3m$ . The odor source is a humidifier that generates ethanol plumes. An obstacle prevents initial plume vision and obstructs navigation. Two perpendicular electric fans are used to create unidirectional or non-unidirectional airflow. There are objects to test the visual reasoning capability of the LLM-model. (b) Schematic diagram of the search area. We selected four different robot initial positions at the downwind area in the repeated tests.

### 4.2. Comparison Algorithms

To validate the proposed OSL navigation algorithm, we have compared it with single-modality and multi-modality OSL navigation algorithms. Single-modality OSL navigation algorithms include the 'Olfaction-only' and the 'Vision-only' navigation algorithms. We

also compared the performance of the multi-modal ‘Fusion’ navigation algorithm [21]. Unlike the visual reasoning-based navigation of the proposed LLM-based algorithm, the Fusion navigation algorithm utilizes a custom-trained ‘You Only Look Once version 7’ (YOLOv7)-model to navigate to visible odor plumes. The primary goal of our experiments is to demonstrate that the proposed multi-modal OSL navigation algorithm outperforms single-modal algorithms, which include both olfaction-based and vision-based approaches. Most recent advancements in OSL are single-modal algorithms, primarily olfaction-based. For this study, we selected a representative olfaction-based algorithm, the moth-inspired method, due to the availability of its control code, which allowed us to implement it in our robotic agent.

The Olfaction-Only Navigation Algorithm comprises olfaction-based ‘surge’ and ‘casting’ behaviors with the ‘Obstacle-avoid’ navigation behavior discussed in subsection 3.4. The algorithm follows ‘Obstacle-avoid’ behavior to navigate around the obstacles. In absence of obstacles, the algorithm tests the current odor concentration level against a threshold. If the detected plume concentration is below the threshold, the algorithm follows ‘casting’ behavior to maximize the chance of finding greater plume concentration. Otherwise, the algorithm follows ‘surge’ behavior to approach the upwind odor source.

In the *Vision-Only Navigation Algorithm*, the robot used the ‘casting’, ‘vision-based’, and ‘Obstacle-avoid’ behaviors discussed in subsection 3.4. The algorithm follows ‘Obstacle-avoid’ behavior to navigate around obstacles. Otherwise, the algorithm checks if there is any potential odor source cue in the visual frame. The algorithm follows ‘Vision-based’ navigation if it finds visual cues towards the odor source. Otherwise, the algorithm moves perpendicular to the wind direction, resembling a ‘zigzag’ exploration movement to increase the chance of detecting plume vision

The *Fusion Navigation Algorithm* utilizes a hierarchical control mechanism to select ‘surge’, ‘casting’, ‘Obstacle-avoid’, or ‘vision-based’ navigation behaviors. In contrast to the ‘vision-based’ navigation behavior of the proposed LLM-based navigation algorithm, that uses zero-shot visual reasoning to identify potential odor source object in the visual frame, the ‘vision-based’ navigation of the ‘Fusion’ algorithm is triggered if a custom trained YOLOv7 model detects a visible odor plume in the visual frame. In that case, the ‘vision-based’ navigation behavior tries to approach the visible plume directly.

The four navigation algorithms were tested in both unidirectional and non-unidirectional airflow environments. For each environment, we used four distinct starting positions to demonstrate that our proposed method performs well from various initial positions and orientations, and four test runs were recorded from each starting positions. Our experiments focus on analyzing how unidirectional and non-unidirectional flow environments impact the performance of each navigation algorithm, rather than the influence of the robot’s starting point. While only four trials were conducted per method for each starting point, each method was tested 16 times per environment (unidirectional or non-unidirectional), exceeding the commonly accepted trial range of 10-15. In our statistical analysis, we evaluate the effects of airflow environments and navigation methods, not the starting points. We conducted a total of 128 test runs, covering two airflow scenarios, four navigation algorithms, and 16 trials per scenario.

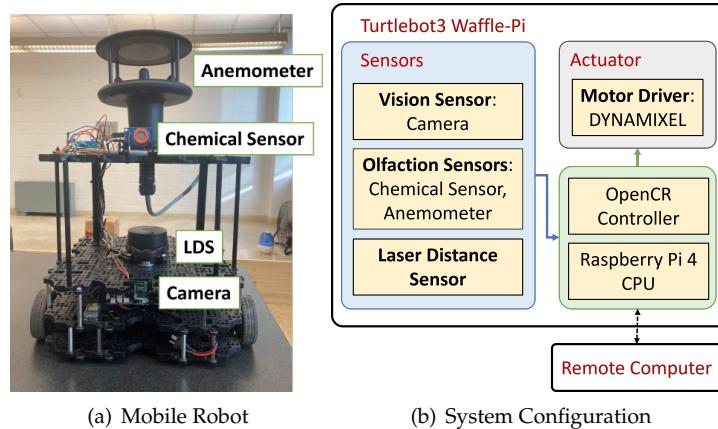
#### 4.3. Robot Platform

Fig. 9(a) shows the robotic platform used in the real-world experiments. We used Turtlebot3 as the base for our robotic agent. In addition to the onboard vision sensors, we added an anemometer and a chemical sensor for olfactory detection. These sensors include:

- Camera: Raspberry Pi Camera v2, that can record 1080p video at 30 Frames Per Second (FPS). This was used to capture the robot’s egocentric vision frame.
- LDS: LDS-02, that can detect 160 – 800mm distance at 360-degrees. This was used to detect distances from obstacles.
- Anemometer: WindSonic, Gill Inc., that can sense 0 – 75m/s wind at 360-degrees. This was used to record wind direction and speed.

- Chemical Sensor: MQ3 alcohol detector, that can sense 25-500 ppm alcohol concentration. This was used to record odor concentration.

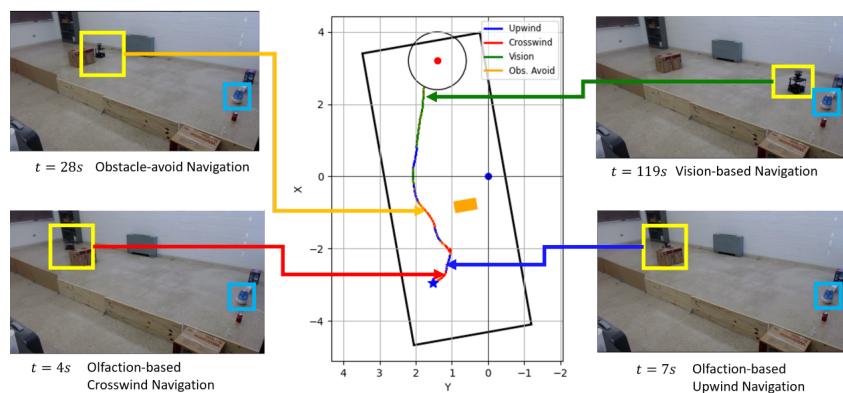
Fig. 9(b) illustrates the system configuration, where the Robot Operating System (ROS) connects the robot platform to a remote computer over a local area network. ROS publishes the sensor readings from the robot, which are subscribed to by the navigation algorithm running on the remote computer. The algorithm uses these readings to calculate and publish heading commands that the robot then executes. The robot subsequently collects a new set of sensor readings, and the cycle continues until it locates the odor source. The robot platform development was more detailed in our previous paper [74].



**Figure 9.** (a) The robot platform includes a camera for vision sensing and a chemical sensor and an anemometer for olfaction sensing. (b) The computation system consists of the robot platform and a remote PC. The dotted line represents wireless link and the solid line represents physical connection.

#### 4.4. Sample Run

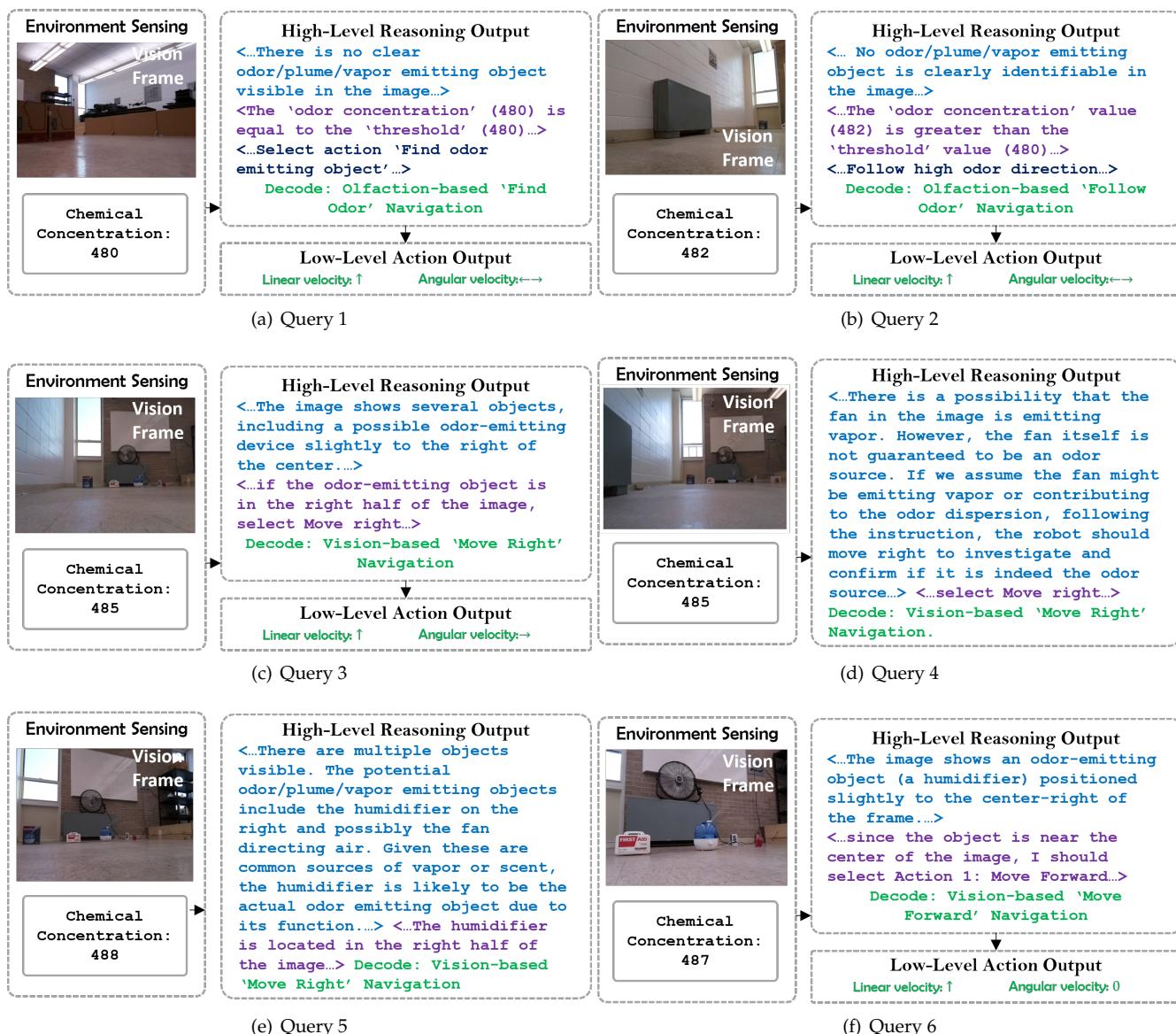
Fig. 10 shows the robot trajectory and snapshots of a successful sample experiment run with the proposed algorithm in a unidirectional airflow environment. In this run, the robot was following 'Olfaction-based' crosswind navigation at  $t = 5$  s. At  $t = 7$  s it was sensing sufficient odor concentration and following 'Olfaction-based' upwind navigation. At  $t = 28$  s, the robot was following 'Obstacle-avoid' navigation. Afterward, it followed 'Olfaction-Based' upwind and 'Vision-based' navigation to reach the odor source at  $t = 119$  s.



**Figure 10.** Trajectory graph of a successful sample run with the proposed multi-modal LLM-based OSL algorithm in unidirectional airflow environment. The navigation behaviors are color-separated. The obstacle is indicated by orange box, and the odor source is represented by a red point with the surrounding circular source declaration region.

We then extracted the robot's egocentric visual frames and chemical reading and used them to query the 'High-level Reasoning' module for navigation decisions with detailed reasoning output. Fig. 11 illustrates prompt input and reasoning output by the 'GPT-4o' model from six time-steps of the sample run. It should be noted that in this project, the LLM is instructed to navigate towards a single odor source in zero-shot manner, no model training is involved in this process. The model does not provide a prioritized list of potential odor sources or any confidence scores. This is because in this work, the LLM is not commanded to choose which object is odor source, but to choose which action the robot should select to approach the odor source.

446  
447  
448  
449  
450  
451  
452  
453  
454



**Figure 11.** Examples of 'Environment Sensing' and 'Reasoning Output' by the GPT-4o model.

In Query 1, the model finds no possible odor source in the visual frame. Then it checks the odor concentration and finds it to be less than the pre-defined threshold. Thus, the model outputs 'Find Odor' navigation behavior. In Query 2, there was still no possible odor source in the visual frame. However, the model output was 'Follow Odor' navigation behavior as odor concentration was above the threshold. In Query 3, the model could not detect any clear odor source. But it detected the fan and deduced that approaching the fan could lead the robot closer to the odor source. In Query 4, the model finds a potential

455  
456  
457  
458  
459  
460  
461

odor-emitting device in the right part of the visual frame and selects vision-based 'Move Right' navigation behavior. In Query 5 the model found potential odor source objects, and decided to approach the humidifier based on its semantic understanding. In Query 6, the model correctly detects the humidifier as the odor source and selects vision-based 'Move Forward' navigation behavior. Based on the navigation behaviors, the 'Low-level Action Module' calculated proper linear and angular velocities for the robot. It should be noted that during experiment runs, we instructed the LLM not to generate textual output of the chain-of-thought reasoning to reduce inference time. This brought the inference time down to under 3 seconds for most multi-modal queries.

#### 4.5. Repeated Test Result

Fig. 12 shows trajectories of the four algorithms in unidirectional airflow environment, and Fig. 13 shows trajectories of the four algorithms in non-unidirectional airflow environment. Each algorithm was tested from four fixed starting positions and four trials were recorded from each starting position. Table 2 shows the performance comparison of the four navigation algorithms in a unidirectional airflow environment and Table 3 shows the performance comparison in a non-unidirectional airflow environment.

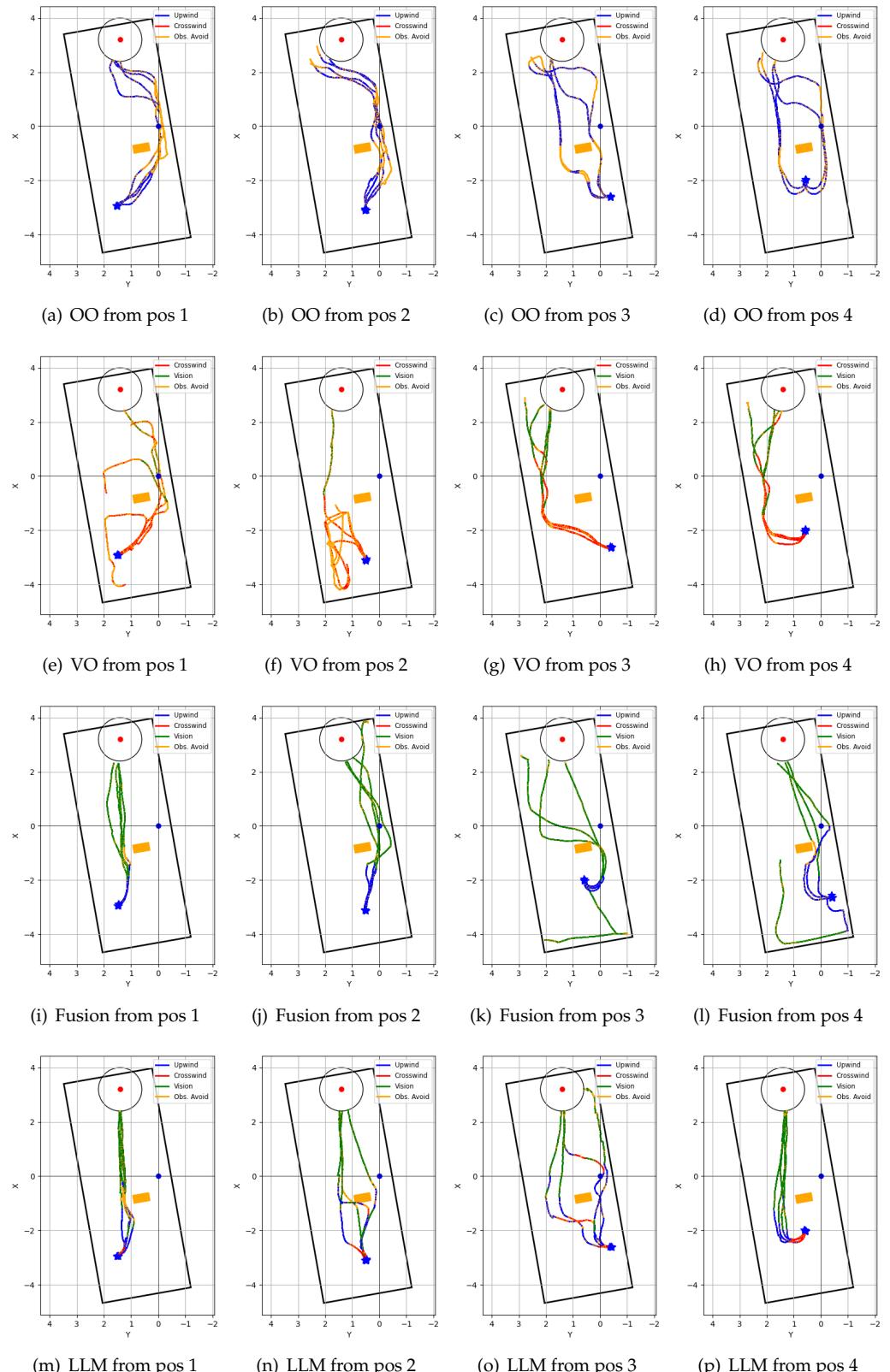
**Table 2.** Comparison of search time (mean and std. dev.), travelled distance (mean and std. dev.), and success rates of the four tested algorithms in unidirectional airflow environment.

Navigation Algorithm	Search Time (s)		Travelled Distance (m)		Success Rate ↑
	Mean	Std. dev.	Mean	Std. dev.	
↓	↓	↓	↓	↓	
<b>Olfaction-only</b>	98.46	11.87	6.86	0.35	10/16
<b>Vision-only</b>	95.23	3.91	6.68	0.27	8/16
<b>Fusion</b>	84.2	12.42	6.12	0.52	12/16
<b>Proposed LLM-based</b>	<b>80.33</b>	4.99	6.14	0.34	<b>16/16</b>

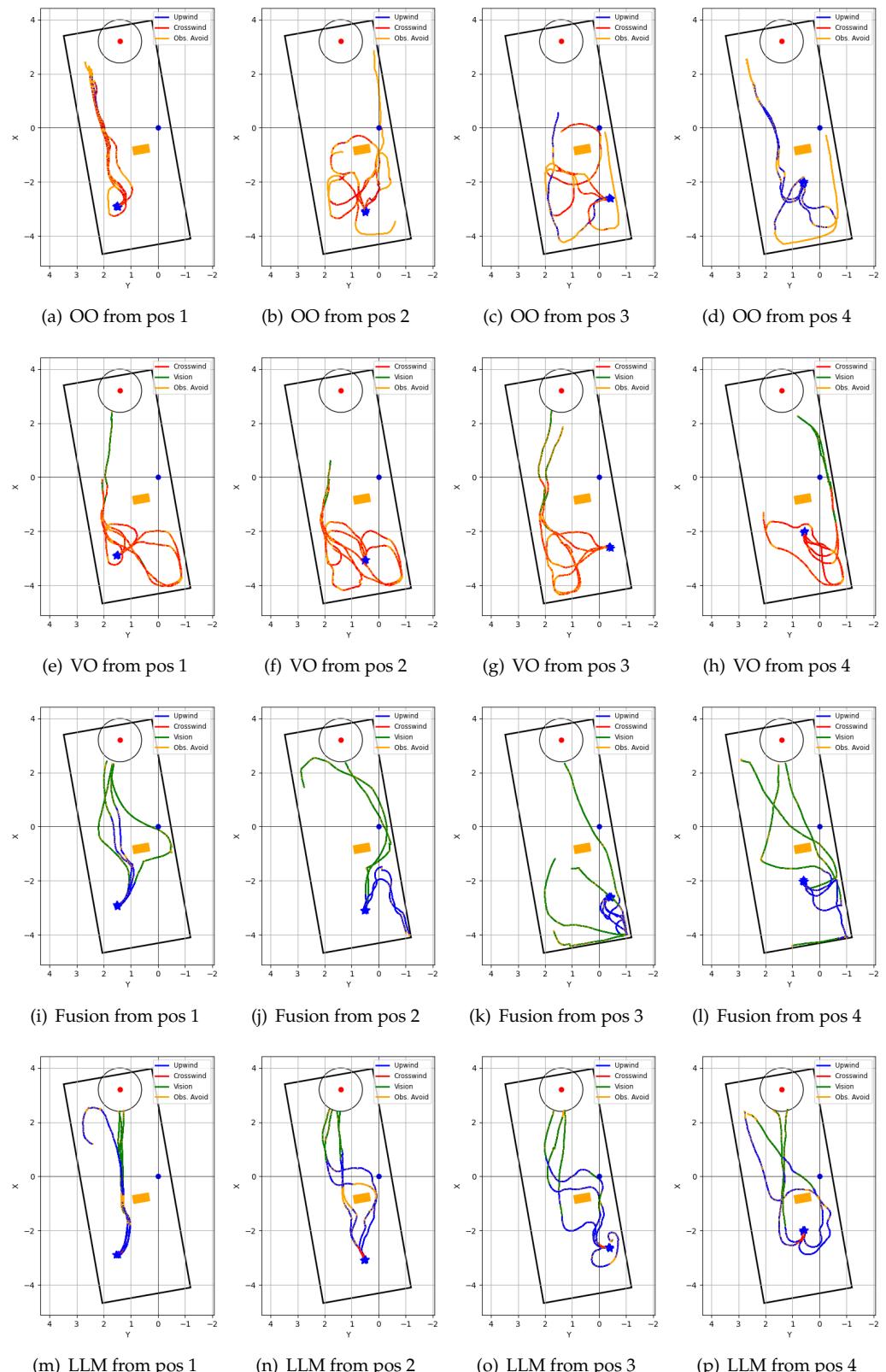
**Table 3.** Comparison of search time (mean and std. dev.), travelled distance (mean and std. dev.), and success rates of the four tested algorithms in non-unidirectional airflow environment.

Navigation Algorithm	Search Time (s)		Travelled Distance (m)		Success Rate ↑
	Mean	Std. dev.	Mean	Std. dev.	
↓	↓	↓	↓	↓	
<b>Olfaction-only</b>	-	-	-	-	0/16
<b>Vision-only</b>	90.67	-	6.69	-	2/16
<b>Fusion</b>	97.79	4.69	7.08	0.53	8/16
<b>Proposed LLM-based</b>	<b>85.3</b>	5.03	<b>6.37</b>	0.31	<b>12/16</b>

In a unidirectional airflow environment, both the Olfaction-only and the Vision-only navigation algorithms performed poorly compared to the Fusion and proposed LLM-based navigation algorithms in terms of both mean search time and mean traveled distance. The proposed navigation algorithm performed better than all other algorithms in terms of success rate and mean search time. In a non-unidirectional airflow environment, the Olfaction-only navigation algorithm failed to localize the odor source in all trial runs. The proposed navigation algorithm again outperformed other algorithms in terms of mean search time, mean distance traveled, and success rate.



**Figure 12.** Robot trajectories of repeated tests in unidirectional airflow environment: **(a–d)** ‘Olfaction-only’ (OO); **(e–h)** ‘Vision-only’ (VO); **(i–l)** ‘Vision and Olfaction Fusion’ (Fusion); and **(m–p)** ‘LLM-based’ (LLM) navigation algorithms.

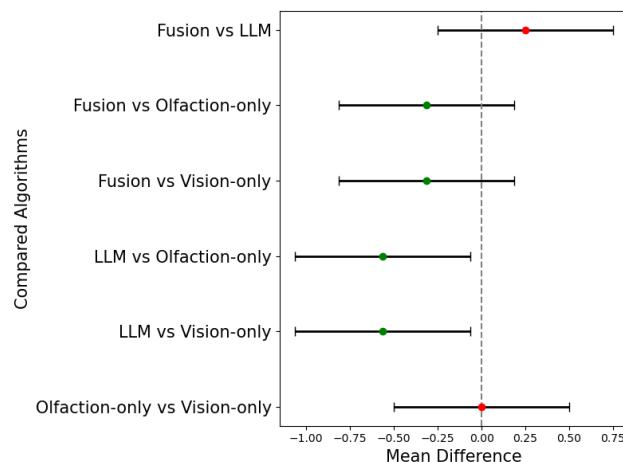


**Figure 13.** Robot trajectories of repeated tests in non-unidirectional airflow environment: (a–d) ‘Olfaction-only’ (OO); (e–h) ‘Vision-only’ (VO); (i–l) ‘Fusion’ (Fusion); and (m–p) ‘LLM-based’ (LLM) navigation algorithms.

While Olfaction-only navigation algorithm had 62.5% success rate in unidirectional, the success rate went down to 0% in non-unidirectional airflow environment. The algorithm relies upon sufficient odor concentration detection and upon the assumption that the odor source is in the upwind direction. Complex airflow from multiple directions affects both of these aspects - it can dilute odor concentration, and non-unidirectional airflow from multiple directions can prevent OSL by upwind navigation.

Vision-based algorithm can only navigate towards the odor source if it's within its visual frame. The algorithm utilized crosswind movement, that resembles 'zigzag' like exploration movement perpendicular to the wind direction. This allowed the model to acquire initial plume vision in a unidirectional airflow environment. However, the algorithm often got sidetracked and lost plume vision while avoiding obstacles in the environment. This resulted in a 50% success rate. However, in a non-unidirectional airflow environment, the casting movement resulted in chaotic exploration of the environment. Thus, the success rate of the algorithm dropped down to 12.5%.

Both the Fusion navigation algorithm and the proposed LLM-based navigation algorithm utilize both vision and olfaction for localizing the odor source. Without proper visual cues, both of these algorithms follow olfaction-based crosswind movement to find, and olfaction-based upwind movement to approach the odor source. Thus, their performance dropped in non-unidirectional airflow environments compared to unidirectional airflow environments. The Fusion navigation algorithm utilizes a deep learning-based vision model and follows a visible odor plume. In contrast, the proposed LLM-based navigation algorithm can reason over the vision frame to deduce possible odor source direction. Thus, it can follow efficient vision-based navigation even without clearly discerning visible odor plumes or odor sources. In a unidirectional airflow environment, the proposed algorithm outperformed the Fusion algorithm in terms of both average success rate (100% vs. 75%) and average search time (80.3 s vs. 84.2 s). In a non-unidirectional airflow environment, the proposed multi-modal LLM-based navigation algorithm far exceeded the performance of the Fusion navigation algorithm in terms of average success rate (75% vs. 50%), average travel time (85.3 s vs. 97.7 s), and average traveled distance (6.4 m vs. 7.1 m).



**Figure 14.** Mean differences of success rates of the four navigation algorithms. The positive differences are statistically significant at Family-wise error rate (FWER) of 5%.

Fig. 14 shows Tukey's honestly significant difference test (Tukey's HSD) test results among the success rates of the four algorithms. In the six one-to-one comparisons, the null hypothesis,  $H_0$ , states that the difference in the mean success rates of the two algorithms is not statistically significant with FWER of 5%. The results show that the null hypothesis is not rejected for comparison with similar sensory modality algorithms, i.e., Olfaction-only Vs. Vision-only and Fusion Vs. LLM-based navigation algorithms. However, the differences are statistically significant for the comparison among mixed modality algorithms. This

indicates that the success rates of multi-sensory modality-based navigation algorithms are statistically superior to the single-sensory modality-based navigation algorithms.

## 5. Limitations and Future Works

One important limit for the LLM-based agent is the inference time. In our application, the inference time is three seconds. This can be reduced by using a smaller LLM. Secondly, our evaluation field is on a small scale. This field is sufficient to validate our proposed work, but a larger search area is required to test real-world imitating applications. In the future, the 2-D OSL method discussed in this paper can further be extended to 3-D OSL scenarios using drones in larger search areas. Experiments can also be extended to actual environments, such as office or household environments. Reasoning-based vision processing can also be used to localize odor sources in 3-D spaces after initial plume recognition. The single-agent OSL discussed here can also be extended to multi-agent odor source localization or monitoring tasks over a larger area in the future. Example applications of multi-agent 3-D OSL include monitoring wildfire outbreaks in large forests or locating chemical leak sources in an indoor setting. Semantic search is another future direction, where next-generation olfaction sensors could detect the type of odor (e.g., chemical leak, food odor). In this case, the high-level reasoning module could search for specific odor sources in the visual field based on the detected odor type. Additionally, a more sophisticated anemotactic method can be implemented in the 'Low-level Action Module' for more effective olfactory-based navigation. A more sophisticated source declaration algorithm that can rank potential odor sources can be incorporated with this model. Finally, reliance on LLM-based reasoning can be minimized by generating a semantic representation of the environment. In this approach, the LLM would only be queried when the robot senses unexplored parts of the environment. The semantic representation could then be used by alternative control mechanisms, such as reinforcement learning, for OSL navigation.

## 6. Conclusion

This paper presents a novel methodology to integrate vision and olfaction sensing in robotic OSL. The dual-modality integration allows the localization of odor sources even if olfaction or vision sensing is disrupted by environmental complexities. The innovation of this paper is the utilization of multi-modal LLM for zero-shot OSL navigation reasoning. We introduced a 'High-level Reasoning' module that generates a multi-modal prompt from robot sensor readings. This prompt is used to query the multi-modal LLM. The reasoning output of the LLM is then decoded and passed to the 'Low-level Action' module. The module then calculates commands that can be executed by the robot. To validate the performance of the proposed algorithm, we implemented the algorithm in a real-world environment. The environment's non-unidirectional airflow challenges olfaction sensing, and obstacles and odor source candidate objects that challenge visual reasoning. We compared the performance of the proposed algorithm to single sensory modality-based 'Olfaction-only' and 'Vision-only' algorithms, and multi-sensory modality-based 'Fusion' navigation algorithms.

- Our proposed method is a multi-modal navigation algorithm that integrates olfactory and visual sensors. Unlike single-modal algorithms, such as the moth-inspired navigation, our approach leverages visual inputs to enhance performance. For instance, when the robot visually identifies the odor source, it can approach it directly, significantly improving the success rate in locating odor sources in both unidirectional and non-unidirectional flow environments. Moth-inspired navigation relies primarily on wind measurements, achieving high success rates in unidirectional flows (10/16 in Table 2) but struggling in non-unidirectional flows (0/16 in Table 3). In contrast, with the help of visual detection, our method demonstrates robust performance across both scenarios. Additionally, in unidirectional search environments, 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. In non-

unidirectional search environments, the proposed method shortens average search time by 5.37 seconds relative to the vision-based algorithm. 574

- Furthermore, compared to rule-based vision and olfaction fusion algorithms (Fusion 575 in Tables 2 and 3), our approach incorporates the reasoning and semantic understanding 576 capabilities of LLMs. This allows for more intelligent decision-making beyond 577 predefined rules. For example, when presented with an electrical fan (Query 3 in 578 Fig. 11), the LLM can deduce that the odor source is likely near the fan—an inference 579 unattainable by rule-based Fusion algorithms, which rely solely on recognizing 580 visible odor plumes. As a result, compared to the Fusion navigation algorithm, the 581 proposed method reduces average search time by 3.87 seconds in unidirectional search 582 environments, and by 12.49 seconds in non-unidirectional search environments. 583

The results also show that the success rates of the multi-sensory algorithms are significantly 584 better than the success rates of single-sensory modality-based algorithms. Overall, the 585 results validate the proposed LLM-based vision and olfaction integration for OSL. In the 586 future, this algorithm can be further expanded for multi-agent three-dimensional OSL 587 environments for both indoor and outdoor OSL tasks. Furthermore, the method can be 588 augmented with reinforcement learning methods to reduce computational cost. 589

**Author Contributions:** Conceptualization, S.H. and L.W.; methodology, S.H. and L.W.; software, S.H. 590 and K.R.M.; validation, L.W.; formal analysis, S.H.; investigation, S.H.; resources, L.W.; data curation, 591 S.H.; writing—original draft preparation, S.H.; writing—review and editing, L.W.; visualization, S.H.; 592 supervision, L.W.; project administration, L.W.; funding acquisition, L.W. All authors have read and 593 agreed to the published version of the manuscript. 594

**Funding:** This work is supported by the Louisiana Board of Regents with grant ID: LEQSF(2024-27)- 595 RD-A-22. 596

**Institutional Review Board Statement:** Not applicable. 598

**Informed Consent Statement:** Not applicable. 599

**Data Availability Statement:** The raw data supporting the conclusions of this article can be found at: 600 [https://github.com/SunzidHassan/24\\_LLM-OSL](https://github.com/SunzidHassan/24_LLM-OSL). 601

**Conflicts of Interest:** The authors declare no conflicts of interest. 602

## Abbreviations

The following abbreviations are used in this manuscript: 604

ANN	Artificial Neural Network
CNN	Convolutional Neural Netwrk
OSL	Odor Source Localization
LLM	Large Language Model
LSTM	Long Short-term Memory
Fusion	Vision and Olfaction Fusion
FWER	Family-Wise Error Rate
Tukey's HSD	Tukey's Honestly Significant Difference Test
VLM	Vision Language Models
WOL	Web Ontology Language

## References

1. Purves, D.; Augustine, G.; Fitzpatrick, D.; Katz, L.; LaMantia, A.; McNamara, J.; Williams, S. The Organization of the Olfactory System. *Neuroscience* **2001**, pp. 337–354. 608
2. Sarafoleanu, C.; Mella, C.; Georgescu, M.; Perederco, C. The importance of the olfactory sense in the human behavior and 609 evolution. *Journal of Medicine and life* **2009**, 2, 196. 610
3. Kowadlo, G.; Russell, R.A. Robot odor localization: a taxonomy and survey. *The International Journal of Robotics Research* **2008**, 611 27, 869–894. 612
4. Wang, L.; Pang, S.; Noyela, M.; Adkins, K.; Sun, L.; El-Sayed, M. Vision and Olfactory-Based Wildfire Monitoring with Uncrewed 613 Aircraft Systems. In Proceedings of the 2023 20th International Conference on Ubiquitous Robots (UR). IEEE, 2023, pp. 716–723. 614

5. Burgués, J.; Hernández, V.; Lilienthal, A.J.; Marco, S. Smelling nano aerial vehicle for gas source localization and mapping. *Sensors* **2019**, *19*, 478. 616  
6. Fu, Z.; Chen, Y.; Ding, Y.; He, D. Pollution source localization based on multi-UAV cooperative communication. *IEEE Access* **2019**, *7*, 29304–29312. 617  
7. Chen, Z.; Wang, J. Underground odor source localization based on a variation of lower organism search behavior. *IEEE Sensors Journal* **2017**, *17*, 5963–5970. 618  
8. Russell, R.A. Robotic location of underground chemical sources. *Robotica* **2004**, *22*, 109–115. 619  
9. Wang, L.; Pang, S.; Xu, G. 3-dimensional hydrothermal vent localization based on chemical plume tracing. In Proceedings of the Global Oceans 2020: Singapore–US Gulf Coast. IEEE, 2020, pp. 1–7. 620  
10. Jing, T.; Meng, Q.H.; Ishida, H. Recent progress and trend of robot odor source localization. *IEEJ Transactions on Electrical and Electronic Engineering* **2021**, *16*, 938–953. 621  
11. Cardé, R.T.; Mafra-Neto, A. Mechanisms of flight of male moths to pheromone. In *Insect pheromone research*; Springer, 1997; pp. 275–290. 622  
12. López, L.L.; Vouloumtsi, V.; Chimeno, A.E.; Marcos, E.; i Badia, S.B.; Mathews, Z.; Verschure, P.F.; Ziyatdinov, A.; i Lluna, A.P. Moth-like chemo-source localization and classification on an indoor autonomous robot. In *On Biomimetics*; IntechOpen, 2011. <https://doi.org/10.5772/19695>. 623  
13. Zhu, H.; Wang, Y.; Du, C.; Zhang, Q.; Wang, W. A novel odor source localization system based on particle filtering and information entropy. *Robotics and autonomous systems* **2020**, *132*, 103619. 624  
14. Vergassola, M.; Villermaux, E.; Shraiman, B.I. ‘Infotaxis’ as a strategy for searching without gradients. *Nature* **2007**, *445*, 406. 625  
15. Luong, D.N.; Kurabayashi, D. Odor Source Localization in Obstacle Regions Using Switching Planning Algorithms with a Switching Framework. *Sensors* **2023**, *23*, 1140. 626  
16. Jakuba, M.V. Stochastic mapping for chemical plume source localization with application to autonomous hydrothermal vent discovery. PhD thesis, Massachusetts Institute of Technology, 2007. <https://doi.org/10.1575/1912/1583>. 627  
17. Hu, H.; Song, S.; Chen, C.P. Plume Tracing via Model-Free Reinforcement Learning Method. *IEEE transactions on neural networks and learning systems* **2019**. 628  
18. Kim, H.; Park, M.; Kim, C.W.; Shin, D. Source localization for hazardous material release in an outdoor chemical plant via a combination of LSTM-RNN and CFD simulation. *Computers & Chemical Engineering* **2019**, *125*, 476–489. 629  
19. Frye, M.A.; Duistermars, B.J. Visually mediated odor tracking during flight in *Drosophila*. *JoVE (Journal of Visualized Experiments)* **2009**, p. e1110. 630  
20. Huang, D.; Yan, C.; Li, Q.; Peng, X. From Large Language Models to Large Multimodal Models: A Literature Review. *Applied Sciences* **2024**, *14*, 5068. 631  
21. Hassan, S.; Wang, L.; Mahmud, K.R. Robotic Odor Source Localization via Vision and Olfaction Fusion Navigation Algorithm. *Sensors* **2024**, *24*, 2309. 632  
22. Berg, H.C. Motile behavior of bacteria. *Physics today* **2000**, *53*, 24–29. 633  
23. Lockery, S.R. The computational worm: spatial orientation and its neuronal basis in *C. elegans*. *Current opinion in neurobiology* **2011**, *21*, 782–790. 634  
24. Radvansky, B.A.; Dombeck, D.A. An olfactory virtual reality system for mice. *Nature communications* **2018**, *9*, 839. 635  
25. Sandini, G.; Lucarini, G.; Varoli, M. Gradient driven self-organizing systems. In Proceedings of the Proceedings of 1993 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS’93). IEEE, 1993, Vol. 1, pp. 429–432. 636  
26. Grasso, F.W.; Consi, T.R.; Mountain, D.C.; Atema, J. Biomimetic robot lobster performs chemo-orientation in turbulence using a pair of spatially separated sensors: Progress and challenges. *Robotics and Autonomous Systems* **2000**, *30*, 115–131. 637  
27. Russell, R.A.; Bab-Hadiashar, A.; Shepherd, R.L.; Wallace, G.G. A comparison of reactive robot chemotaxis algorithms. *Robotics and Autonomous Systems* **2003**, *45*, 83–97. 638  
28. Lilienthal, A.; Duckett, T. Experimental analysis of gas-sensitive Braitenberg vehicles. *Advanced Robotics* **2004**, *18*, 817–834. 639  
29. Ishida, H.; Nakayama, G.; Nakamoto, T.; Moriizumi, T. Controlling a gas/odor plume-tracking robot based on transient responses of gas sensors. *IEEE Sensors Journal* **2005**, *5*, 537–545. 640  
30. Murlis, J.; Elkinton, J.S.; Carde, R.T. Odor plumes and how insects use them. *Annual review of entomology* **1992**, *37*, 505–532. 641  
31. Vickers, N.J. Mechanisms of animal navigation in odor plumes. *The Biological Bulletin* **2000**, *198*, 203–212. 642  
32. Cardé, R.T.; Willis, M.A. Navigational strategies used by insects to find distant, wind-borne sources of odor. *Journal of chemical ecology* **2008**, *34*, 854–866. 643  
33. Nevitt, G.A. Olfactory foraging by Antarctic procellariiform seabirds: life at high Reynolds numbers. *The Biological Bulletin* **2000**, *198*, 245–253. 644  
34. Wallraff, H.G. Avian olfactory navigation: its empirical foundation and conceptual state. *Animal Behaviour* **2004**, *67*, 189–204. 645  
35. Shigaki, S.; Sakurai, T.; Ando, N.; Kurabayashi, D.; Kanzaki, R. Time-varying moth-inspired algorithm for chemical plume tracing in turbulent environment. *IEEE Robotics and Automation Letters* **2017**, *3*, 76–83. 646  
36. Shigaki, S.; Shiota, Y.; Kurabayashi, D.; Kanzaki, R. Modeling of the Adaptive Chemical Plume Tracing Algorithm of an Insect Using Fuzzy Inference. *IEEE Transactions on Fuzzy Systems* **2019**, *28*, 72–84. <https://doi.org/10.1109/tfuzz.2019.2915187>. 647

37. Jin, W.; Rahbar, F.; Ercolani, C.; Martinoli, A. Towards efficient gas leak detection in built environments: Data-driven plume modeling for gas sensing robots. In Proceedings of the 2023 IEEE International Conference on Robotics and Automation (ICRA). IEEE, 2023, pp. 7749–7755. 673  
674  
675
38. Ojeda, P.; Monroy, J.; Gonzalez-Jimenez, J. Robotic gas source localization with probabilistic mapping and online dispersion simulation. *IEEE Transactions on Robotics* **2024**. 676  
677
39. Rahbar, F.; Marjovi, A.; Kibleur, P.; Martinoli, A. A 3-D bio-inspired odor source localization and its validation in realistic environmental conditions. In Proceedings of the 2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2017, pp. 3983–3989. 678  
679  
680
40. Shigaki, S.; Yoshimura, Y.; Kurabayashi, D.; Hosoda, K. Palm-sized quadcopter for three-dimensional chemical plume tracking. *IEEE Transactions on Instrumentation and Measurement* **2022**, *71*, 1–12. 681  
682
41. Hutchinson, M.; Liu, C.; Chen, W.H. Information-based search for an atmospheric release using a mobile robot: Algorithm and experiments. *IEEE Transactions on Control Systems Technology* **2018**, *27*, 2388–2402. <https://doi.org/10.1109/TCST.2018.2860548>. 683  
684
42. Rahbar, F.; Marjovi, A.; Martinoli, A. An algorithm for odor source localization based on source term estimation. In Proceedings of the 2019 International Conference on Robotics and Automation (ICRA). IEEE, 2019, pp. 973–979. 685  
686
43. Jiu, H.; Chen, Y.; Deng, W.; Pang, S. Underwater chemical plume tracing based on partially observable Markov decision process. *International Journal of Advanced Robotic Systems* **2019**, *16*, 1729881419831874. 687  
688
44. Luong, D.N.; Tran, H.Q.D.; Kurabayashi, D. Reactive-probabilistic hybrid search method for odour source localization in an obstructed environment. *SICE Journal of Control, Measurement, and System Integration* **2024**, *17*, 2374569. 689  
690
45. Pang, S.; Zhu, F. Reactive planning for olfactory-based mobile robots. In Proceedings of the 2009 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2009, pp. 4375–4380. 691  
692
46. Wang, L.; Pang, S. Chemical Plume Tracing using an AUV based on POMDP Source Mapping and A-star Path Planning. In Proceedings of the OCEANS 2019 MTS/IEEE SEATTLE. IEEE, 2019, pp. 1–7. 693  
694
47. Bilgera, C.; Yamamoto, A.; Sawano, M.; Matsukura, H.; Ishida, H. Application of convolutional long short-term memory neural networks to signals collected from a sensor network for autonomous gas source localization in outdoor environments. *Sensors* **2018**, *18*, 4484. 695  
696
48. Thrift, W.J.; Cabuslay, A.; Laird, A.B.; Ranjbar, S.; Hochbaum, A.I.; Ragan, R. Surface-enhanced Raman scattering-based odor compass: Locating multiple chemical sources and pathogens. *ACS sensors* **2019**, *4*, 2311–2319. 697  
698
49. Wang, L.; Pang, S. An Implementation of the Adaptive Neuro-Fuzzy Inference System (ANFIS) for Odor Source Localization. In Proceedings of the 2020 IEEE/RSJ International Conference on Intelligent Robots and Systems. IEEE, 2021. 700  
701
50. Monroy, J.; Ruiz-Sarmiento, J.R.; Moreno, F.A.; Melendez-Fernandez, F.; Galindo, C.; Gonzalez-Jimenez, J. A semantic-based gas source localization with a mobile robot combining vision and chemical sensing. *Sensors* **2018**, *18*, 4174. 702  
703
51. Chowdhary, K. Natural language processing for word sense disambiguation and information extraction. *arXiv preprint arXiv:2004.02256* **2020**. 704  
705
52. Nayebi, A.; Rajalingham, R.; Jazayeri, M.; Yang, G.R. Neural foundations of mental simulation: Future prediction of latent representations on dynamic scenes. *Advances in Neural Information Processing Systems* **2024**, *36*. 706  
707
53. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems* **2017**, *30*. 708  
709
54. Wei, J.; Wang, X.; Schuurmans, D.; Bosma, M.; Xia, F.; Chi, E.; Le, Q.V.; Zhou, D.; et al. Chain-of-thought prompting elicits reasoning in large language models. *Advances in neural information processing systems* **2022**, *35*, 24824–24837. 710  
711
55. Ouyang, L.; Wu, J.; Jiang, X.; Almeida, D.; Wainwright, C.; Mishkin, P.; Zhang, C.; Agarwal, S.; Slama, K.; Ray, A.; et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems* **2022**, *35*, 27730–27744. 712  
713
56. Brown, T.; Mann, B.; Ryder, N.; Subbiah, M.; Kaplan, J.D.; Dhariwal, P.; Neelakantan, A.; Shyam, P.; Sastry, G.; Askell, A.; et al. Language models are few-shot learners. *Advances in neural information processing systems* **2020**, *33*, 1877–1901. 714  
715
57. Devlin, J.; Chang, M.W.; Lee, K.; Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* **2018**. 716  
717
58. Touvron, H.; Lavril, T.; Izacard, G.; Martinet, X.; Lachaux, M.A.; Lacroix, T.; Rozière, B.; Goyal, N.; Hambro, E.; Azhar, F.; et al. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* **2023**. 718  
719
59. Li, C.; Gan, Z.; Yang, Z.; Yang, J.; Li, L.; Wang, L.; Gao, J.; et al. Multimodal foundation models: From specialists to general-purpose assistants. *Foundations and Trends® in Computer Graphics and Vision* **2024**, *16*, 1–214. 720  
721
60. Radford, A.; Kim, J.W.; Hallacy, C.; Ramesh, A.; Goh, G.; Agarwal, S.; Sastry, G.; Askell, A.; Mishkin, P.; Clark, J.; et al. Learning transferable visual models from natural language supervision. In Proceedings of the International conference on machine learning. PMLR, 2021, pp. 8748–8763. 722  
723
61. Shi, Y.; Shang, M.; Qi, Z. Intelligent layout generation based on deep generative models: A comprehensive survey. *Information Fusion* **2023**, p. 101940. 724  
725
62. Wang, J.; Wu, Z.; Li, Y.; Jiang, H.; Shu, P.; Shi, E.; Hu, H.; Ma, C.; Liu, Y.; Wang, X.; et al. Large language models for robotics: Opportunities, challenges, and perspectives. *arXiv preprint arXiv:2401.04334* **2024**. 726  
727
63. Dorbala, V.S.; Sigurdsson, G.; Piramuthu, R.; Thomason, J.; Sukhatme, G.S. Clip-nav: Using clip for zero-shot vision-and-language navigation. *arXiv preprint arXiv:2211.16649* **2022**. 728  
729  
730  
731

64. Chen, P.; Sun, X.; Zhi, H.; Zeng, R.; Li, T.H.; Liu, G.; Tan, M.; Gan, C. **AA2 Nav**: Action-Aware Zero-Shot Robot Navigation by 732  
Exploiting Vision-and-Language Ability of Foundation Models. *arXiv preprint arXiv:2308.07997* **2023**. 733
65. Zhou, G.; Hong, Y.; Wu, Q. Navgpt: Explicit reasoning in vision-and-language navigation with large language models. In 734  
Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2024, Vol. 38, pp. 7641–7649. 735
66. Schumann, R.; Zhu, W.; Feng, W.; Fu, T.J.; Riezler, S.; Wang, W.Y. Velma: Verbalization embodiment of llm agents for vision and 736  
language navigation in street view. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2024, 737  
Vol. 38, pp. 18924–18933. 738
67. Shah, D.; Osinski, B.; Ichter, B.; Levine, S. Robotic Navigation with Large Pre-Trained Models of Language. *Vision, and Action* 739  
**2022**. 740
68. Yu, B.; Kasaei, H.; Cao, M. L3mvn: Leveraging large language models for visual target navigation. In Proceedings of the 2023 741  
IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2023, pp. 3554–3560. 742
69. Zhou, K.; Zheng, K.; Pryor, C.; Shen, Y.; Jin, H.; Getoor, L.; Wang, X.E. Esc: Exploration with soft commonsense constraints for 743  
zero-shot object navigation. In Proceedings of the International Conference on Machine Learning. PMLR, 2023, pp. 42829–42842. 744
70. Jatavallabhula, K.M.; Kuwajerwala, A.; Gu, Q.; Omaha, M.; Chen, T.; Maalouf, A.; Li, S.; Iyer, G.; Saryazdi, S.; Keetha, N.; et al. 745  
Conceptfusion: Open-set multimodal 3d mapping. *arXiv preprint arXiv:2302.07241* **2023**. 746
71. Brohan, A.; Brown, N.; Carbajal, J.; Chebotar, Y.; Dabis, J.; Finn, C.; Gopalakrishnan, K.; Hausman, K.; Herzog, A.; Hsu, J.; et al. 747  
Rt-1: Robotics transformer for real-world control at scale. *arXiv preprint arXiv:2212.06817* **2022**. 748
72. Farrell, J.A.; Pang, S.; Li, W. Chemical plume tracing via an autonomous underwater vehicle. *IEEE Journal of Oceanic Engineering* 749  
**2005**, *30*, 428–442. 750
73. Ishida, H.; Moriizumi, T. Machine olfaction for mobile robots. *Handbook of Machine Olfaction: Electronic Nose Technology* **2002**, pp. 751  
399–417. 752
74. Hassan, S.; Wang, L.; Mahmud, K.R. Multi-Modal Robotic Platform Development for Odor Source Localization. In Proceedings 753  
of the 2023 Seventh IEEE International Conference on Robotic Computing (IRC). IEEE, 2023, pp. 59–62. 754

**Disclaimer/Publisher's Note:** The statements, opinions and data contained in all publications are solely those of the individual 755  
author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to 756  
people or property resulting from any ideas, methods, instructions or products referred to in the content. 757