MDPI

Article

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

Sunzid Hassan [1,†], Lingxiao Wang [2*], and Khan Raqib Mahmud [1]

[1] 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.)
[2] 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 find an unknown odor source in a given environment. An effective navigation algorithm that guides the robot to approach the odor source is the key to successfully locating the odor source. Compared to traditional olfaction-only OSL method, our proposed method integrates vision and olfaction sensor modalities to localize odor sources even if olfaction sensing is disrupted by turbulent airflow or vision sensing is impaired by environmental complexities. The model 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 method, we implemented the proposed algorithm on a mobile robot in a complex, real-world search environment that presents challenges to both olfaction and vision-sensing modalities. 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 navigation algorithm. Experimental results demonstrate that multi-sensory navigation algorithms are statistically superior to single sensory navigation algorithms. The proposed algorithm outperformed the other algorithms in both laminar and turbulent airflow environments.

Keywords: odor source localization; multi-modal robotics; Large Language Models (LLMs); robot operating system (ROS)

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/

Received:
Revised:
Accepted:
Published:

1. Introduction

Sensory systems like olfaction, vision, audition, etc. allow animals to interact with their surroundings. Of these, olfaction is the most ancient sensory system to evolve in organisms [1]. It enables organisms with odorant receptors to detect food, potential mates, threats, and predators [2]. Similarly, a mobile robot with a chemical sensor can detect odors in the environment. Robotic OSL allows robots to mimic animals' olfaction-based behaviors. Specifically, it is the technology that allows robots to navigate toward an unknown target odor source in a given environment [3]. It has important applications, such as monitoring wildfires [4], locating air pollution [5], detecting chemical gas leaks [6], identifying unexploded mines and bombs [7], finding underground gas leaks [8], and conducting marine surveys like 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. Locating an unknown odor source

requires an effective OSL navigation algorithm to guide the robot based on sensor observations. Research on robotic OSL algorithms has garnered considerable interest in recent decades [10]. Traditional OSL algorithms include bio-inspired methods that imitate animal OSL behaviors, engineering-based methods that rely on mathematical models to estimate potential odor source locations, and machine learning-based methods that use a trained machine learning model to guide the robot toward the odor source. Typical bio-inspired methods include the Moth-inspired algorithm that imitates male moths' mate-seeking behaviors [11], where a robotic agent will follow a 'surge/casting' movements [12]. Typical engineering-based methods include the Particle Filter algorithm [13] that updates odor source location prediction based on observations. Finally, typical machine learning-based OSL methods include deep supervised [14] and reinforcement learning methods [15] to locate odor sources.

All of these approaches rely on olfaction (e.g., chemical and airflow) sensing to detect and navigate to the given odor source. However, approaches that rely solely on olfaction sensing struggle in environments where turbulent airflow disrupts olfaction sensing. Thus, in nature, while simple organisms without vision (e.g., nematodes) use vision-free OSL [16], most complex animals, including humans, mammals, raptors [17], invertebrates like fruit flies [18], mosquitoes [19], beetles [20], etc. utilize vision with olfaction sensing for OSL. The two modalities offer unique advantages in OSL – olfaction helps early awareness of odor source object, while vision helps with pinpointing the location of the object [21]. Similarly, a robot with both olfaction- and vision-sensing capabilities (e.g., with a camera and chemical sensor), and a navigation algorithm that can effectively integrate and utilize the sensory modalities, can find an unknown odor source more efficiently.

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. Such a combination of capabilities is exemplified by recent advanced multi-modal LLMs like GPT-4 [22], Gemini [23], etc. These models demonstrate state-of-the-art performance in reasoning over multiple sensory modalities like text, vision, and sound [24]. This multi-modal reasoning ability makes them promising for application in multi-modal robotics. 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.

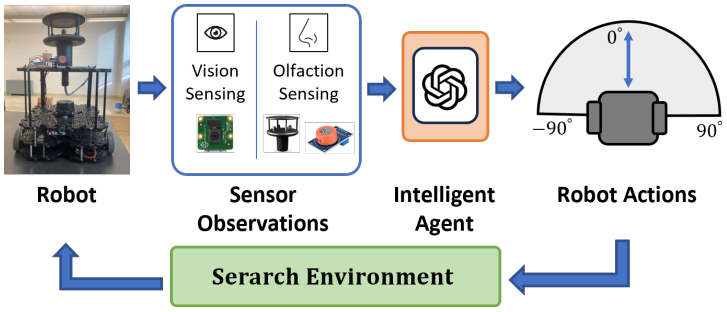


Figure 1. Flow diagram of the proposed method for the OSL experiment. We utilized the Turtlebot3 robot platform. We equipped it with a camera, Laser Distance Sensor, airflow sensor, chemical sensor, etc. The proposed algorithm utilizes multi-modal LLM for navigation decision-making.

Fig. 1 illustrates the proposed OSL navigation algorithm. The core of the proposed navigation algorithm 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 turbulent airflow, vision is challenged by obstacles and multi-modal reasoning is challenged by environment complexities.

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 with the supervised learning-based vision and olfaction fusion navigation algorithm [25].

In the remainder of this paper, Section 2 reviews the recent progress of OSL research; Section 3 reviews technical details of the proposed OSL algorithm; Section 4 presents details of the performed real-world experiments; and, finally, Section 5 includes overall conclusions of the work.

2. Related Works

2.1. Olfaction-only Methods

Various organisms, regardless of their size, rely on scent to locate objects. Whether it is a bacterium navigating an amino acid gradient or a wolf tracking prey, the ability to follow odors is vital for survival. Designing algorithms that replicate the navigation methods of biological organisms is a common 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 their movement in response to changes in chemical concentration. When they encounter higher levels of an attractive chemical, their likelihood of making temporary turns decreases, 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 [26]. This straightforward approach allows single-celled organisms to navigate a gradient of appealing chemicals through a guided random walk. Nematodes [16] and crustaceans [27] 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 [28]. Several early studies [29–32] confirmed the effectiveness of chemotaxis in laminar flow environments, characterized by low Reynolds numbers. However, in turbulent flow environments with high Reynolds numbers, alternative methods inspired by more complex biological navigation techniques and engineering techniques were proposed.

Odor-gated anemotaxis navigation is a more sophisticated bio-inspired OSL method that uses both odor and airflow senses for navigation. Moths [33–35], birds [36,37], and other organisms utilize this type of navigation. Specifically, mimicking the mate-seeking behavior of male moths led to the development of the prevalent moth-inspired method in OSL research [38]. Additionally, diverse bio-inspired search strategies such as zigzag, spiral, fuzzy-inference, and multi-phase exploratory approaches have been introduced in recent time[39]. The method is also applied for complex three-dimensional search areas [40,41].

Engineering-based methods differ from bio-mimicking algorithms by relying on mathematical models to estimate odor source locations. They involve discretizing the search area and learning the likelihood of each region containing the odor source. Algorithms used for constructing such maps include Bayesian inference, particle filters, stochastic mapping [42], infotaxis [43,44], source term estimation [45], information-based search [46], partially observable Markov decision processes [47], reactive-probabilistic search [48], etc. After

predicting the odor source location, robots are then guided towards the estimated source via path-planning algorithms such as artificial potential fields and A-star [49,50].

Machine-learning (ML)-based methods are increasingly employed in OSL experiments. Recent advancements include the use of Deep Neural Networks (DNNs) and reinforcement learning for plume-tracing strategies. For example, Kim et al. [14] trained a Recurrent Neural Network (RNN) to predict potential odor source locations using data from stationary sensor networks obtained through simulation. Hu et al. [15] developed a plume-tracing algorithm based on model-free reinforcement learning, employing the deterministic policy gradient to train an actor-critic network for Autonomous Underwater Vehicle (AUV) navigation. Wang et al. [51] trained an adaptive neuro-fuzzy inference system (ANFIS) to address the OSL problem in simulations. The methods were validated in virtual environments through simulated flow fields and plume distributions, highlighting the need for real-world implementations to validate their effectiveness.

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 turbulent 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 [52]. 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 [25]. The algorithm was effective in localizing unknown odor sources in real-world obstacle-ridden environments with 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 [53]. 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 [54] 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 [55]. LLMs exceed previous RNN-based language models due to emergent abilities, including chain-of-thought reasoning [56], instruction understanding [57], and in-context learning [58]. Notable examples of LLMs include BERT [59], GPT-3 [58], LLaMA [60], 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 [61]. Unlike supervised vision classifiers, multi-modal LLMs are simultaneously trained with vision and language data. For example, the multi-modal LLM CLIP [62] 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 [63]. Thus, multi-modal LLMs are increasingly used in robotics tasks like generating robot action plans by reasoning over multi-modal sensor data [64].

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 [65] 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^2Nav [66] 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 [67] utilizes the GPT-4 model for zero-shot navigation in simulated indoor scenarios. VELMA [68] 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 [69], 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 [70] 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 [71] 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 [72] 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, allowing for more efficient training. For example, training a general LLM-driven robot, such as Google's RT-1 [73], 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, fine-tuning them specifically for the robotic OSL task, significantly reducing the need for extensive training data.

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:

$$a^t = F(s^t), \tag{1}$$

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

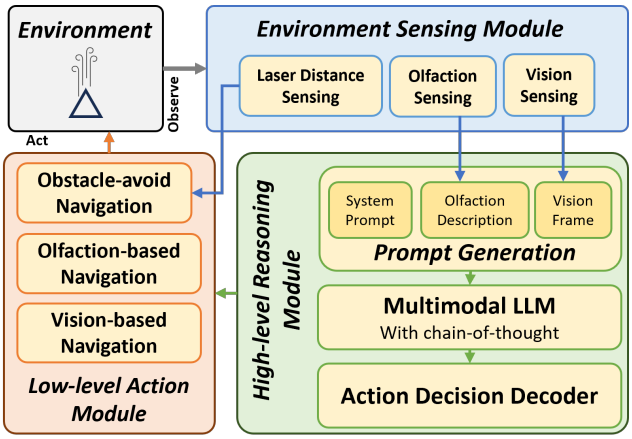


Figure 2. The framework of the proposed multi-modal LLM-based navigation algorithm. The three main modules are the 'Environment Sensing Module', 'High-level Reasoning Module', and 'Low-level Action Module'.

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

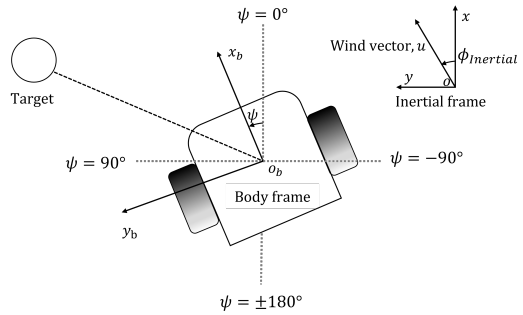


Figure 3. Robot notations. Robot position (x, y) and heading ψ are monitored by the built-in localization system. Wind speed u and wind direction are measured from the additional anemometer in the body frame. Wind direction in inertial frame $\phi_{Inertial}$ is derived from robot heading ψ and wind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

Table 1. Environment sensing parameters.

Symbols	Parameters
p	Visual Observation
u	Wind Speed
ϕ_b	Wind Direction
ρ	Chemical 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 detection. The visual frame captured by the camera is the visual observation p. An anemometer senses wind speed u m/s, and wind direction ϕ_b-degrees in the body frame. The chemical concentration ρ is expressed in ppm. At time t the observed state by the robot is $s^t = [p, u, \phi, \rho]^t$. The sensors used in real-world experimentation are discussed in subsection 4.3.

3.3. High-level Reasoning Module

'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.

Prompt-generation is the first step in this module. Formulating effective prompts is crucial for LLM's reasoning process. The proposed algorithm combines the 'System Prompt', current 'Olfaction Description', and current 'Vision Frame' to generate a multi-modal prompt. The 'System Prompt' provides the reasoning objective, action selection instructions, specific constraints, and expected output to the LLM. The current chemical concentration ρ is translated into a structured textual descriptor. The visual frame p is transformed using 'Binary-to-text' encoding before attaching it with the prompt.

The LLM was instructed to use the Chain-of-Thought reasoning process [56] to capture logical coherence in multi-modal reasoning process over complex multi-modal sensory inputs. 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.

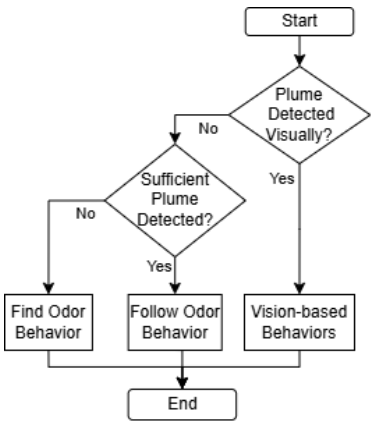


Figure 4. The flow diagram of the 'High-level Reasoning Module'. It illustrates how the proposed LLM-based agent integrates visual and olfactory sensory observations to make high-level navigation behavior decisions.

Fig. 4 illustrates the reasoning strategy, which was modeled after human odor search behaviors. 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. 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

$$a = [v_c, \omega_c], \tag{2}$$

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

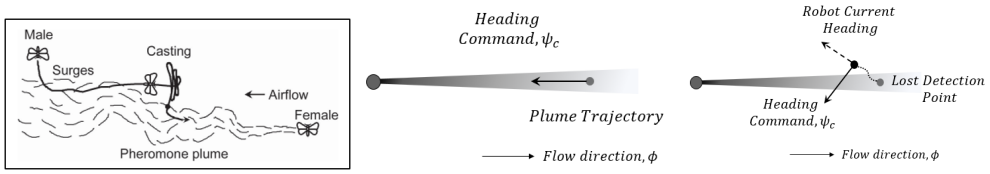
'Obstacle-avoid': This behavior is activated when a nearby obstacle is detected by the onboard LDS. The 'Obstacle-avoid' behavior directs the robot to navigate around the obstacle without deviating significantly from the direction the robot was following. Details of this navigation behavior are outlined in our previous paper [25].

'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.

$$\omega_c = \begin{cases} 0 & \text{if behavior = 'Front';} \\ constant & \text{if behavior = 'Left';} \\ -constant & \text{if behavior = 'Right'.} \end{cases} \tag{3}$$

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':



(a) Moth Behaviors (b) Follow Odor (c) Find Odor

Figure 5. (**a**) Moth mate seeking behaviors. (**b**) Moth-inspired 'Surge' and (**c**) 'Casting' navigation behaviors.

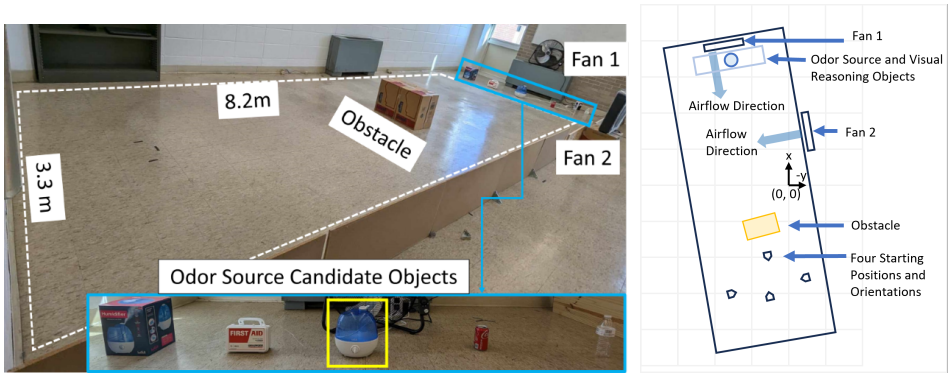
Finally, we utilize the moth-inspired 'Surge' movement for implementing the high-level 'Follow Odor' behavior, and the 'Casting' movement for implementing the high-level 'Find Odor' behavior [74]. In the 'Surge' behavior, the robot moves upwind toward the odor source. In 'Casting', the robot moves crosswind to increase the likelihood of encountering odor plumes.

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

Equation 4 shows the target heading ψ_c calculation for the two behaviors. Angular velocity ω_c is then adjusted to achieve the target heading ψ_c.

4. Experiment

4.1. Experiment setup



(a) Search Area (b) Search Area Schematic

Figure 6. **(a)** The experimental setup. The robot is initially placed in a downwind area with the objective of finding the odor source. A humidifier loaded with ethanol is employed to generate odor plumes. Two electric fans are placed perpendicularly to create artificial wind fields. An obstacle is placed in the search area. There are 5 objects to test the reasoning capability of the LLM-model. **(b)** Schematic diagram of the search area. The four robot starting positions are used for testing the performance of the proposed OSL algorithm.

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. 6 shows the search area used for the OSL navigation experiment. The search area has an obstacle in the 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. Thus, the navigation algorithm must select appropriate 'Olfaction-based' and 'Obstacle-avoid' navigation behaviors to approach the odor source without colliding with the obstacle at first. Once the robot passes the obstacle, the multi-modal LLM must 1) detect objects present in the visual frame, 2) reason from the provided goal which object is the potential odor source, and 3) select actions that

will guide the robot toward the odor source. The robot is considered successful if the robot reaches within 0.8 m of the odor source location within 120 s.

4.2. Comparison Algorithms

To determine the effectiveness of olfaction and vision integration in OSL, we compared the OSL performance of single and multi-sensory modality-based algorithms. For single sensory modality-based OSL, we tested 'Olfaction-only' and 'Vision-only' navigation algorithms. For multi-sensory modality-based OSL, we tested the supervised learning-based Fusion navigation algorithm [25] in addition to the proposed LLM-based navigation algorithm.

In the *Olfaction-Only Navigation Algorithm*, the robot uses the olfaction-based 'surge' and 'casting' behaviors with the 'Obstacle-avoid' navigation behavior discussed in subsection 3.4. If there are obstacles in the robot's path, the algorithm follows 'Obstacle-avoid' behavior to navigate around the obstacles. Otherwise, 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 detecting sufficient plume concentration. If the detected plume concentration is above the threshold, the algorithm follows 'surge' behavior to move upwind towards the 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

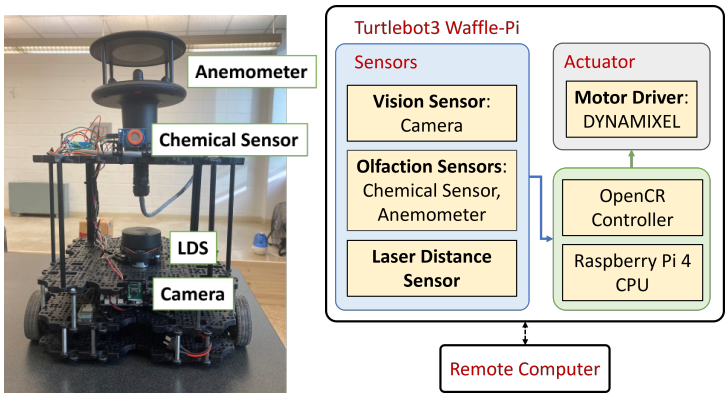
The *Vision and Olfaction Fusion Navigation Algorithm* utilizes a hierarchical control mechanism to select 'surge', 'casting', 'Obstacle-avoid', or 'vision-based' navigation behaviors. In contrast to the 'vision-based' navigation behavior of the proposed LLM-based navigation algorithm, the 'vision-based' navigation of the fusion algorithm is triggered if a supervised learning model detects an 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 laminar and turbulent airflow environments. 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.

4.3. Robot Platform

Table 2. Type, name, and specification of the built-in camera, laser distance sensor, and added anemometer and chemical sensor.

Source	Sensor Type	Module Name	Specification
Built-in	Camera	Raspberry Pi Camera v2	Video Capture: 1080p30, 720p60 and VGA90.
	Laser Distance Sensor	LDS-02	Detection Range: 360-degree. Distance Range: 160∼8000 mm.
Added	Anemometer	WindSonic, Gill Inc.	Speed: 0–75 m/s. Wind direction: 0–360 degrees.
	Chemical Sensor	MQ3 alcohol detector	Concentration: 25–500 ppm.



(a) Mobile Robot (b) System Configuration

Figure 7. (a) Turtlebot3 platform based mobile robot is used in this work. In addition to the built-in camera and Laser Distance Sensor, the robot is equipped with a chemical sensor and an anemometer for measuring chemical concentration, wind speeds, and airflow directions. (b) System configuration consisting of the robot platform and a remote PC. The solid line indicates a physical connection, while the dotted line represents a wireless link.

Fig. 7(a) shows the robotic platform used in the real-world experiments. Fig. 7(b) illustrates the system configuration, where the Robot Operating System (ROS) connects the Turtlebot3-based robot platform with a remote computer using a local area network. Table 2 shows the built-in and added sensors for OSL experiments. Raspberry Pi Camera V2 was used for capturing video, LDS-02 Laser Distance Sensor was used for obstacle detection, WindSonic Anemometer was used for wind speed and wind direction measurements in the body frame, and MQ3 alcohol detector was used for detecting chemical plume concentration. The robot platform development was more detailed in our previous paper [75].

4.4. Sample Run

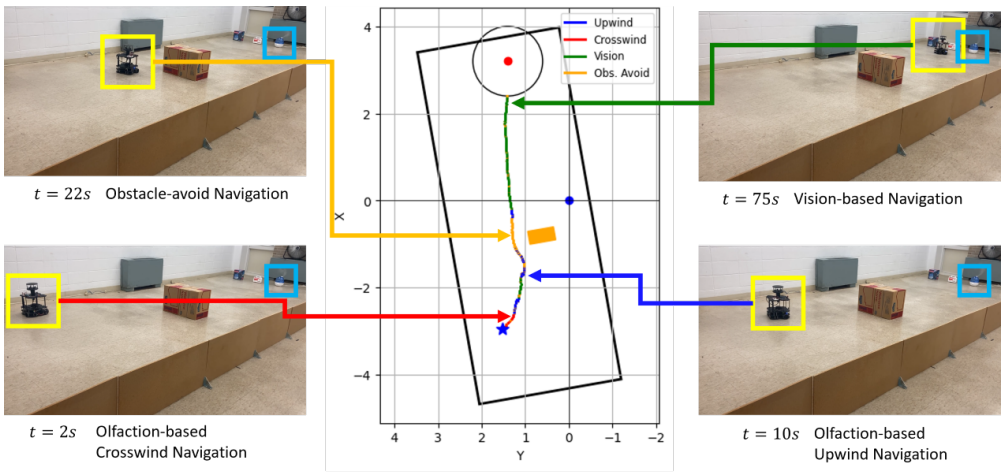


Figure 8. Trajectory graph of a successful experiment run with the proposed multi-modal LLM-based OSL algorithm in laminar airflow environment. The navigation behaviors are color-separated. The obstacle is indicated by orange box, and the odor source is represented by a red point with the surrounding circular source declaration region.

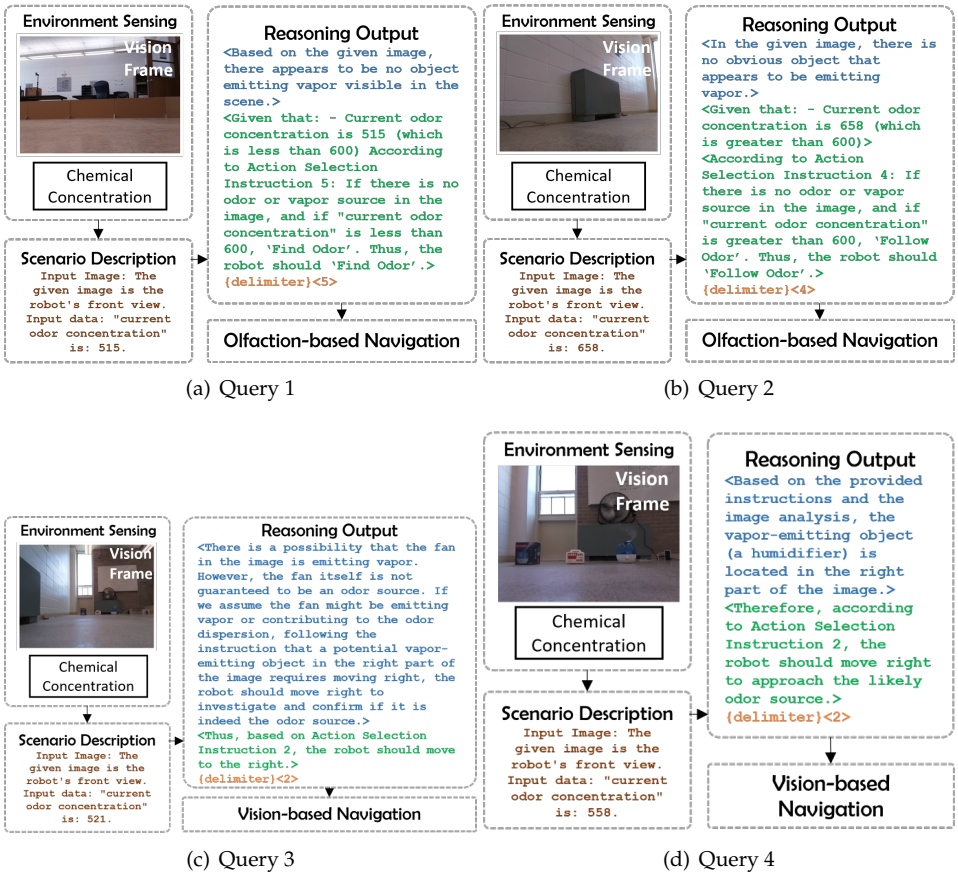
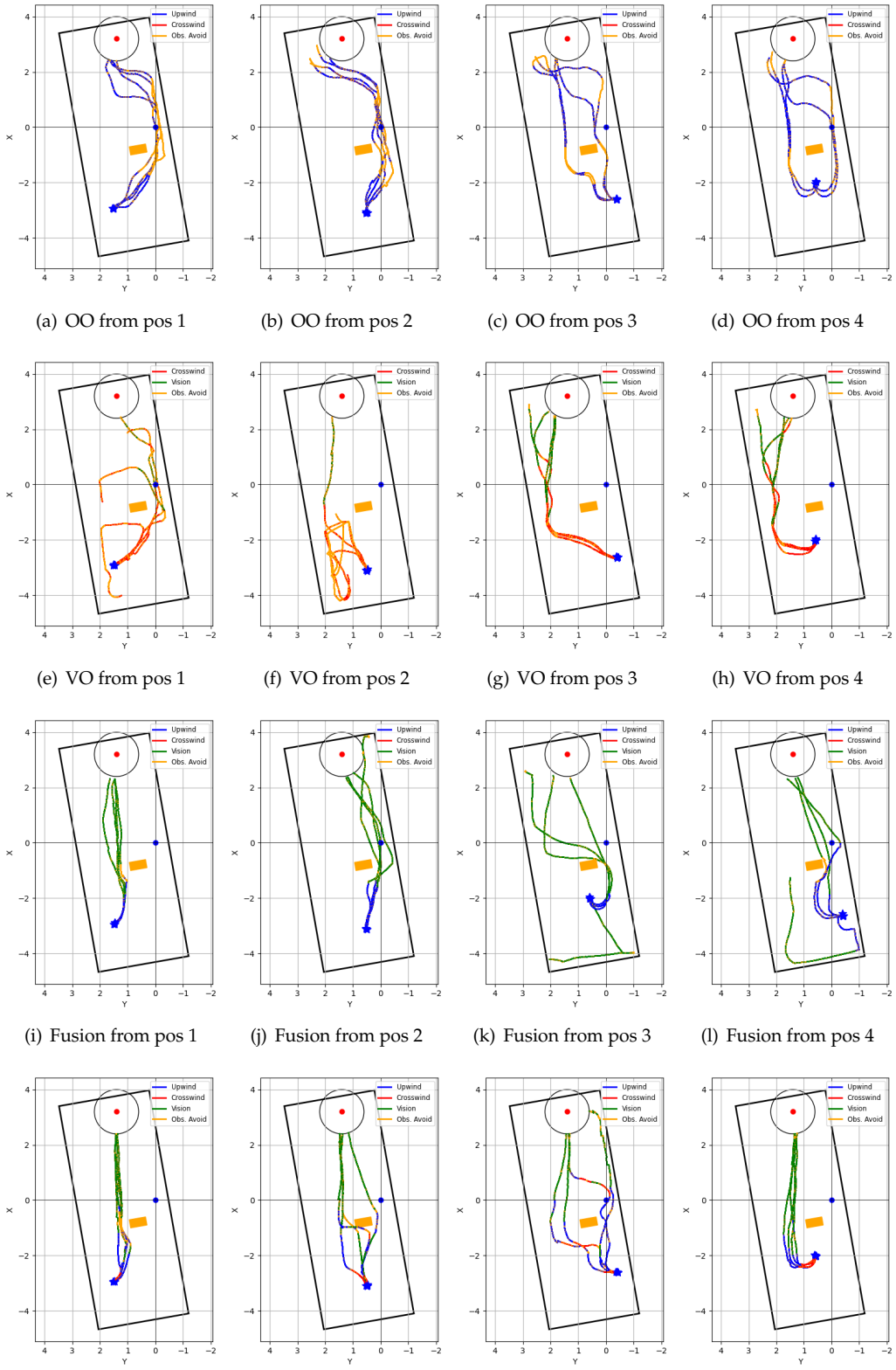


Figure 9. Examples of 'Environment Sensing' and 'Reasoning Output' by the GPT-4o model.

Fig. 8 shows the robot trajectory and snapshots of a successful experiment run with the proposed algorithm in a laminar airflow environment. In this run, the robot initialized at t = 1 s and followed 'Olfaction-based' crosswind navigation. At t= 3s it found sufficient chemical concentration and started following 'Olfaction-Based' upwind navigation. At t = 16 s, the robot faced the obstacle and followed 'Obstacle-avoid' navigation. Afterward, it followed 'Vision-based' navigation to reach the odor source at t = 79 s. Fig. 9 illustrates prompt input and reasoning output by the 'GPT-4o' model from three points of the sample run. In the first query, the model finds no possible odor source in the visual frame. Then it checks the chemical concentration and finds it to be less than the pre-defined threshold. Thus, the model outputs 'Find Odor' navigation behavior. In the second query, there was still no possible odor source in the visual frame. However, the model output was 'Follow Odor' navigation behavior as chemical concentration was above the threshold. In the third query, the model fails to distinguish any odor source in the visual frame. However, it detects an electric fan and deduces that the fan may contribute to odor propagation. Thus, the model selects vision-based navigation behavior to approach the fan. In this case, despite not visually detecting the odor source, the model selected the action that approached the odor source. In the third query, the model detects odor plumes, correctly identifies the humidifier as the odor source, and selects vision-based navigation behavior to approach the humidifier. While we used the Chain-of-Thought technique to validate LLM reasoning in the sample run, we turned off the Chain-of-Thought reasoning output in the repeated tests to reduce the inference time of the 'GPT-4o' LLM. This brought the inference time down to under 3 seconds for most multi-modal queries.

4.5. Repeated Test Result



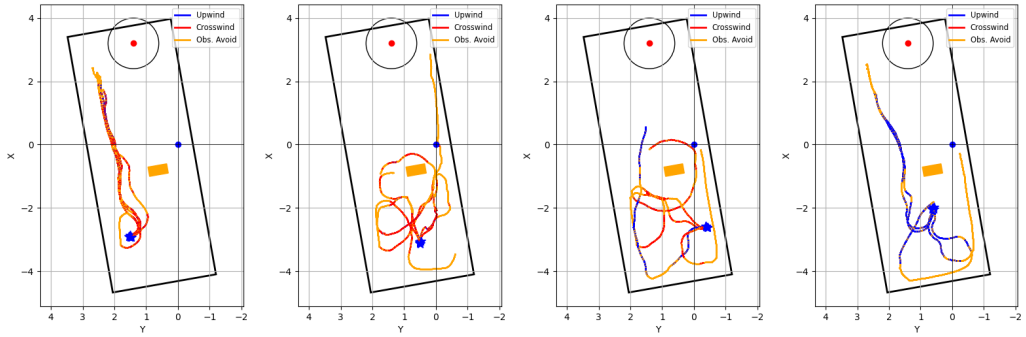
(a) OO from pos 1 (b) OO from pos 2 (c) OO from pos 3 (d) OO from pos 4

(e) VO from pos 1 (f) VO from pos 2 (g) VO from pos 3 (h) VO from pos 4

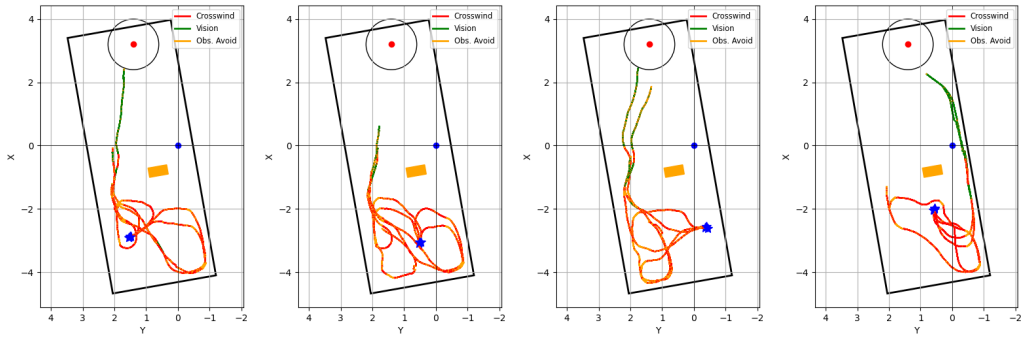
(i) Fusion from pos 1 (j) Fusion from pos 2 (k) Fusion from pos 3 (l) Fusion from pos 4

(m) LLM from pos 1 (n) LLM from pos 2 (o) LLM from pos 3 (p) LLM from pos 4

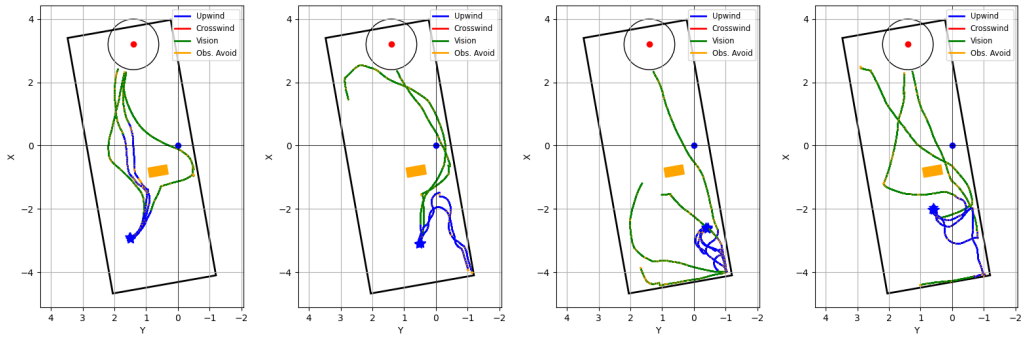
Figure 10. Robot trajectories of repeated tests in laminar airflow environment: **(a−d)** 'Olfaction-only Navigation Algorithm' (OO); **(e−h)** 'Vision-only Navigation Algorithm' (VO); **(i−l)** 'Vision and Olfaction Fusion Navigation Algorithm' (Fusion); and **(m−p)** 'LLM-based Navigation Algorithm' (LLM).



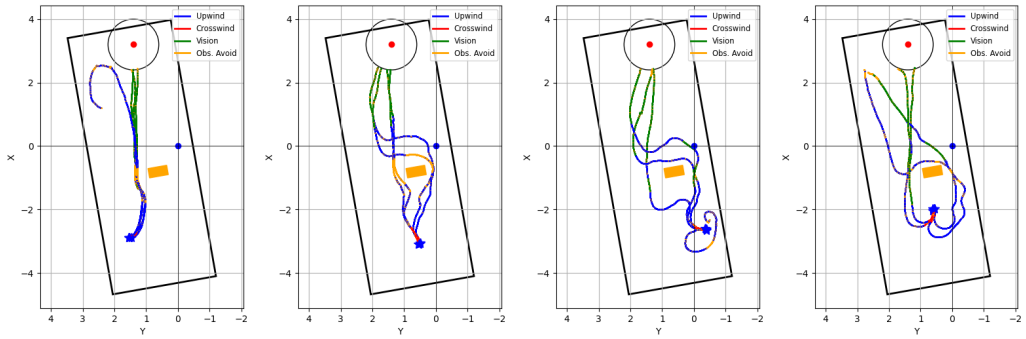
(a) OO from pos 1 (b) OO from pos 2 (c) OO from pos 3 (d) OO from pos 4



(e) VO from pos 1 (f) VO from pos 2 (g) VO from pos 3 (h) VO from pos 4



(i) Fusion from pos 1 (j) Fusion from pos 2 (k) Fusion from pos 3 (l) Fusion from pos 4



(m) LLM from pos 1 (n) LLM from pos 2 (o) LLM from pos 3 (p) LLM from pos 4

Figure 11. Robot trajectories of repeated tests in turbulent airflow environment: **(a−d)** 'Olfaction-only Navigation Algorithm' (OO); **(e−h)** 'Vision-only Navigation Algorithm' (VO); **(i−l)** 'Vision and Olfaction Fusion Navigation Algorithm' (Fusion); and **(m−p)** 'LLM-based Navigation Algorithm' (LLM).

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 laminar airflow environment.

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

Table 4. Comparison of search time (mean and std. dev.), travelled distance (mean and std. dev.), and success rates of the four tested algorithms in turbulent airflow environment.

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

Fig. 10 shows trajectories of the four algorithms in laminar airflow environment, and Fig. 11 shows trajectories of the four algorithms in turbulent airflow environment. Each algorithm was tested from four fixed starting positions, and four trials were recorded from each starting position. Table 3 shows the performance comparison of the four navigation algorithms in a laminar airflow environment, and Table 4 shows the performance comparison in a turbulent airflow environment.

In a laminar 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 turbulent 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.

While Olfaction-only navigation algorithm had 62.5% success rate in laminar, the success rate went down to 0% in turbulent airflow environment. The algorithm relies upon sufficient chemical 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 chemical concentration, and turbulent 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 laminar 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 turbulent 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 turbulent airflow environments compared to laminar 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. Thus, in a laminar 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 turbulent 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).

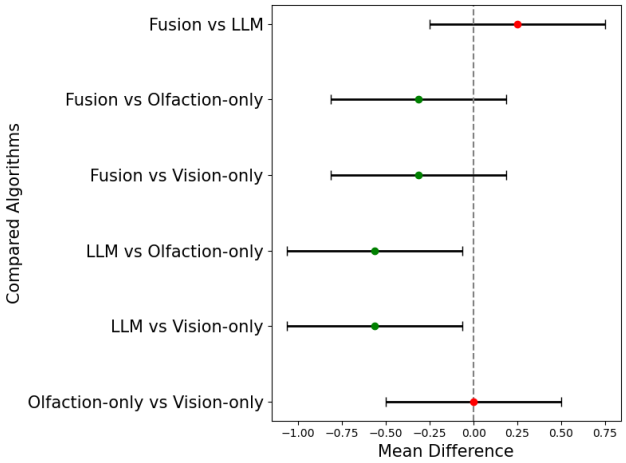


Figure 12. Mean differences of success rates of the four navigation algorithms. The positive differences are statistically significant at Family-wise error rate (FWER) of 5%.

Fig. 12 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. Conclusion

The results of section 4 indicate that a multi-modal LLM-based algorithm can successfully integrate vision and olfaction for zero-shot OSL navigation. The experimental setup presented mimics indoor environments with obstacles and odor sources. Therefore, the results can be generalized to other real-world indoor OSL scenarios, such as detecting indoor gas leaks in office or household settings with obstacles and potential gas sources. It is also feasible to extend the proposed method to outdoor applications, such as detecting wildfire locations using both vision (flame detection) and olfaction (smoke or other fire-related gases).

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

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

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw data supporting the conclusions of this article can be found at: `https://sunzidhassan.github.io/24_Vision-Olfaction-LLM/`.

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

Abbreviations

The following abbreviations are used in this manuscript:

OSL	Odor Source Localization
LLM	Large Language Model
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.
2. Sarafoleanu, C.; Mella, C.; Georgescu, M.; Perederco, C. The importance of the olfactory sense in the human behavior and evolution. *Journal of Medicine and life* **2009**, *2*, 196.
3. Kowadlo, G.; Russell, R.A. Robot odor localization: a taxonomy and survey. *The International Journal of Robotics Research* **2008**, *27*, 869–894.
4. Wang, L.; Pang, S.; Noyela, M.; Adkins, K.; Sun, L.; El-Sayed, M. Vision and Olfactory-Based Wildfire Monitoring with Uncrewed Aircraft Systems. In Proceedings of the 2023 20th International Conference on Ubiquitous Robots (UR). IEEE, 2023, pp. 716–723.
5. Fu, Z.; Chen, Y.; Ding, Y.; He, D. Pollution source localization based on multi-UAV cooperative communication. *IEEE Access* **2019**, *7*, 29304–29312.
6. Burgués, J.; Hernández, V.; Lilienthal, A.J.; Marco, S. Smelling nano aerial vehicle for gas source localization and mapping. *Sensors* **2019**, *19*, 478.
7. Russell, R.A. Robotic location of underground chemical sources. *Robotica* **2004**, *22*, 109–115.
8. Chen, Z.; Wang, J. Underground odor source localization based on a variation of lower organism search behavior. *IEEE Sensors Journal* **2017**, *17*, 5963–5970.
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.
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.
11. Cardé, R.T.; Mafra-Neto, A. Mechanisms of flight of male moths to pheromone. In *Insect pheromone research*; Springer, 1997; pp. 275–290.
12. López, L.L.; Vouloutsi, 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.
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.
14. 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.
15. Hu, H.; Song, S.; Chen, C.P. Plume Tracing via Model-Free Reinforcement Learning Method. *IEEE transactions on neural networks and learning systems* **2019**.
16. Lockery, S.R. The computational worm: spatial orientation and its neuronal basis in C. elegans. *Current opinion in neurobiology* **2011**, *21*, 782–790.
17. Potier, S.; Duriez, O.; Célérier, A.; Liegeois, J.L.; Bonadonna, F. Sight or smell: which senses do scavenging raptors use to find food? *Animal Cognition* **2019**, *22*, 49–59.
18. Frye, M.A.; Duistermars, B.J. Visually mediated odor tracking during flight in Drosophila. *JoVE (Journal of Visualized Experiments)* **2009**, p. e1110.

19. Van Breugel, F.; Riffell, J.; Fairhall, A.; Dickinson, M.H. Mosquitoes use vision to associate odor plumes with thermal targets. *Current Biology* **2015**, *25*, 2123–2129.
20. L. Yv, F.; Hai, X.; Wang, Z.; Yan, A.; Liu, B.; Bi, Y. Integration of visual and olfactory cues in host plant identification by the Asian longhorned beetle, Anoplophora glabripennis (Motschulsky)(Coleoptera: Cerambycidae). *PLoS One* **2015**, *10*, e0142752.
21. Kuang, S.; Zhang, T. Smelling directions: olfaction modulates ambiguous visual motion perception. *Scientific reports* **2014**, *4*, 1–5.
22. Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F.L.; Almeida, D.; Altenschmidt, J.; Altman, S.; Anadkat, S.; et al. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774* **2023**.
23. Team, G.; Anil, R.; Borgeaud, S.; Wu, Y.; Alayrac, J.B.; Yu, J.; Soricut, R.; Schalkwyk, J.; Dai, A.M.; Hauth, A.; et al. Gemini: a family of highly capable multimodal models. *arXiv preprint arXiv:2312.11805* **2023**.
24. Zhang, D.; Yu, Y.; Li, C.; Dong, J.; Su, D.; Chu, C.; Yu, D. Mm-llms: Recent advances in multimodal large language models. *arXiv preprint arXiv:2401.13601* **2024**.
25. Hassan, S.; Wang, L.; Mahmud, K.R. Robotic Odor Source Localization via Vision and Olfaction Fusion Navigation Algorithm. *Sensors* **2024**, *24*, 2309.
26. Berg, H.C. Motile behavior of bacteria. *Physics today* **2000**, *53*, 24–29.
27. Radvansky, B.A.; Dombeck, D.A. An olfactory virtual reality system for mice. *Nature communications* **2018**, *9*, 839.
28. 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.
29. 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.
30. 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.
31. Lilienthal, A.; Duckett, T. Experimental analysis of gas-sensitive Braitenberg vehicles. *Advanced Robotics* **2004**, *18*, 817–834.
32. 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.
33. Murlis, J.; Elkinton, J.S.; Carde, R.T. Odor plumes and how insects use them. *Annual review of entomology* **1992**, *37*, 505–532.
34. Vickers, N.J. Mechanisms of animal navigation in odor plumes. *The Biological Bulletin* **2000**, *198*, 203–212.
35. 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.
36. Nevitt, G.A. Olfactory foraging by Antarctic procellariiform seabirds: life at high Reynolds numbers. *The Biological Bulletin* **2000**, *198*, 245–253.
37. Wallraff, H.G. Avian olfactory navigation: its empirical foundation and conceptual state. *Animal Behaviour* **2004**, *67*, 189–204.
38. 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.
39. 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.
40. 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.
41. 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.
42. 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.
43. Vergassola, M.; Villermaux, E.; Shraiman, B.I. 'Infotaxis' as a strategy for searching without gradients. *Nature* **2007**, *445*, 406.
44. Luong, D.N.; Kurabayashi, D. Odor Source Localization in Obstacle Regions Using Switching Planning Algorithms with a Switching Framework. *Sensors* **2023**, *23*, 1140.
45. 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.
46. 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.
47. 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.
48. 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.
49. 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.
50. 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.
51. 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.

52. 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.
53. Chowdhary, K. Natural language processing for word sense disambiguation and information extraction. *arXiv preprint arXiv:2004.02256* **2020**.
54. 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*.
55. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, Ł.; Polosukhin, I. Attention is all you need. *Advances in neural information processing systems* **2017**, *30*.
56. 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.
57. 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.
58. 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.
59. 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**.
60. 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**.
61. 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.
62. 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.
63. Shi, Y.; Shang, M.; Qi, Z. Intelligent layout generation based on deep generative models: A comprehensive survey. *Information Fusion* **2023**, p. 101940.
64. 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**.
65. 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**.
66. Chen, P.; Sun, X.; Zhi, H.; Zeng, R.; Li, T.H.; Liu, G.; Tan, M.; Gan, C. A2 Nav: Action-Aware Zero-Shot Robot Navigation by Exploiting Vision-and-Language Ability of Foundation Models. *arXiv preprint arXiv:2308.07997* **2023**.
67. Zhou, G.; Hong, Y.; Wu, Q. Navgpt: Explicit reasoning in vision-and-language navigation with large language models. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2024, Vol. 38, pp. 7641–7649.
68. Schumann, R.; Zhu, W.; Feng, W.; Fu, T.J.; Riezler, S.; Wang, W.Y. Velma: Verbalization embodiment of llm agents for vision and language navigation in street view. In Proceedings of the Proceedings of the AAAI Conference on Artificial Intelligence, 2024, Vol. 38, pp. 18924–18933.
69. Shah, D.; Osinski, B.; Ichter, B.; Levine, S. Robotic Navigation with Large Pre-Trained Models of Language. *Vision, and Action* **2022**.
70. Yu, B.; Kasaei, H.; Cao, M. L3mvn: Leveraging large language models for visual target navigation. In Proceedings of the 2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). IEEE, 2023, pp. 3554–3560.
71. Zhou, K.; Zheng, K.; Pryor, C.; Shen, Y.; Jin, H.; Getoor, L.; Wang, X.E. Esc: Exploration with soft commonsense constraints for zero-shot object navigation. In Proceedings of the International Conference on Machine Learning. PMLR, 2023, pp. 42829–42842.
72. Jatavallabhula, K.M.; Kuwajerwala, A.; Gu, Q.; Omama, M.; Chen, T.; Maalouf, A.; Li, S.; Iyer, G.; Saryazdi, S.; Keetha, N.; et al. Conceptfusion: Open-set multimodal 3d mapping. *arXiv preprint arXiv:2302.07241* **2023**.
73. Brohan, A.; Brown, N.; Carbajal, J.; Chebotar, Y.; Dabis, J.; Finn, C.; Gopalakrishnan, K.; Hausman, K.; Herzog, A.; Hsu, J.; et al. Rt-1: Robotics transformer for real-world control at scale. *arXiv preprint arXiv:2212.06817* **2022**.
74. Farrell, J.A.; Pang, S.; Li, W. Chemical plume tracing via an autonomous underwater vehicle. *IEEE Journal of Oceanic Engineering* **2005**, *30*, 428–442.
75. Hassan, S.; Wang, L.; Mahmud, K.R. Multi-Modal Robotic Platform Development for Odor Source Localization. In Proceedings of the 2023 Seventh IEEE International Conference on Robotic Computing (IRC). IEEE, 2023, pp. 59–62.