

From Woofs to Words: Towards Intelligent Robotic Guide Dogs with Verbal Communication

Yohei Hayamizu*[†], David DeFazio*, Hrudayangam Mehta*, Zainab Altaweel, Jacqueline Choe,
Chao Lin, Jake Juettner, Furui Xiao, Jeremy Blackburn, Shiqi Zhang

The State University of New York at Binghamton
4400 Vestal Pkwy E, Binghamton, NY 13902 USA

Abstract

Assistive robotics is an important subarea of robotics that focuses on the well-being of people with disabilities. A robotic guide dog is an assistive quadruped robot that helps visually impaired people in obstacle avoidance and navigation. Enabling language capabilities for robotic guide dogs goes beyond naively adding an existing dialog system onto a mobile robot. The novel challenges include grounding language in the dynamically changing environment and improving spatial awareness for the human handler. To address those challenges, we develop a novel dialog system for robotic guide dogs that uses LLMs to verbalize both navigational plans and scenes. The goal is to enable verbal communication for collaborative decision-making within the handler-robot team. In experiments, we conducted a human study to evaluate different verbalization strategies and a simulation study to assess the efficiency and accuracy in navigation tasks.

Website — <https://sites.google.com/view/woofs-words>

1 Introduction

Guide dogs play a crucial role in the lives of visually impaired individuals, enhancing their independence, confidence, companionship, and mobility (Whitmarsh 2005). Unfortunately, breeding and training guide dogs takes multiple years and can be very costly (Tomkins, Thomson, and McGreevy 2011), and even then, guide dog training centers have less than a 50% graduation rate (Marks 2025). At the same time, many people with visual impairments need guide dogs but have difficulties caring for them, e.g., walking miles a day and dog allergies. As a result, only a very small portion of the visually impaired use guide dogs, e.g., about 2% in the U.S. (Marks 2025) and only about 400 guide dogs for more than 10 million visually impaired people in China (Wang and Wu 2024).

While biological guide dogs are helpful in leading the visually impaired individuals around obstacles, they have fundamental limitations in understanding and communicating with humans. Even after systematic training, guide dogs can

*These authors contributed equally.

[†]Corresponding author: yhayami1@binghamton.edu
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

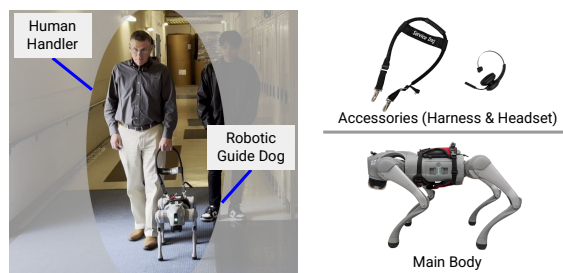


Figure 1: A legally blind person walking with our robotic guide dog system during a participant study.

only respond to very short, predefined verbal cues (Guiding-Eyes 2023). Robotic guide dogs have the potential to be valuable navigation aids for visually impaired individuals (Wang, Zhao, and Zhang 2021; Xiao et al. 2021; Chen et al. 2023b; Hwang et al. 2023; Kim et al. 2023; DeFazio, Hirota, and Zhang 2023; Hwang et al. 2024, 2025). To the best of our knowledge, there was no existing research focusing on natural language interactions for robotic guide dogs.

The integration of dialog systems into robotic guide dogs introduces unique challenges not present in virtual dialog agents. First, it is important for the dialog system to ground open-vocabulary language to the current environment, to generate relevant and feasible navigation goals, as well as navigational plans for realizing the goals. For instance, when a handler says “*I am thirsty*”, the dialog system should provide options that are available, e.g., by saying “*There is a vending machine that serves water bottles and a water fountain on this floor*”, and identify the handler’s preference after providing additional information, such as “*Navigating to the water fountain requires going through a long corridor and entering a kitchen room.*”. A second challenge is ensuring the spatial awareness of the visually impaired handler, which is crucial for the handler to make informed decisions about routes and destinations.¹

¹A common misunderstanding is that a visually impaired handler completely follows the biological guide dog in navigation. The reality is that they work as a team, where the handler makes decisions about the routes (e.g., using tactile and echolocation sensing), and the guide dog helps avoid obstacles.

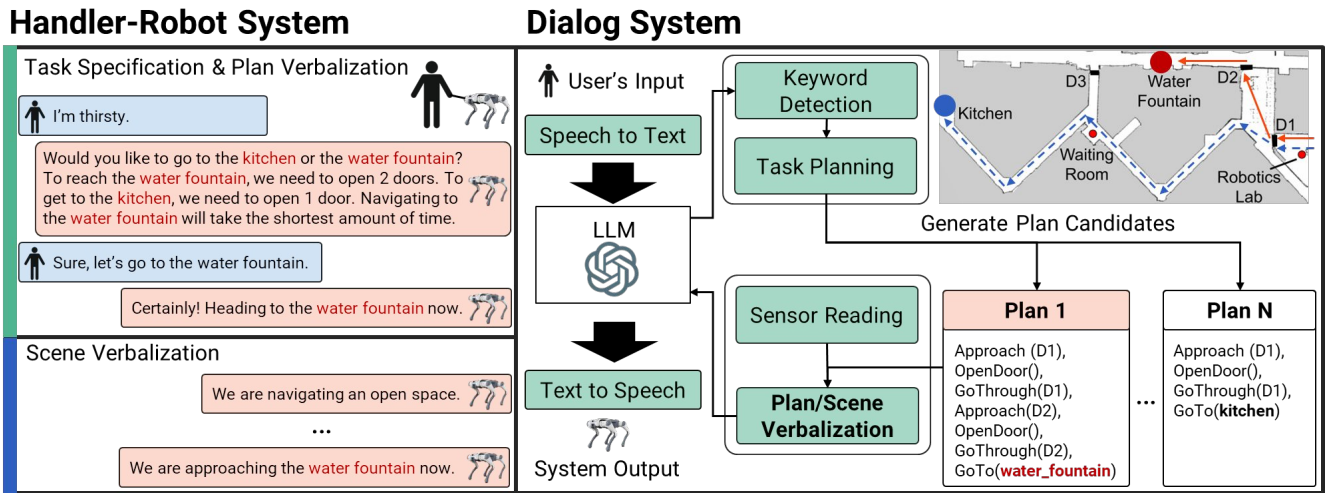


Figure 2: Our system uses human-robot dialog to define a formal service task. An LLM first determines relevant navigable locations, and a task planner generates multiple action sequences (plans) for each candidate. These plans are summarized for the human via **plan verbalization**, detailing metrics like navigation cost and door openings. After human selection, the robot executes the chosen plan, providing navigation guidance while providing **scene verbalization** to describe the surroundings.

In this paper, we develop a dialog system that leverages a large language model (LLM) (Achiam et al. 2023) for dialog management and extracting navigation tasks from open-vocabulary service requests. We integrate task planning into language synthesis to enrich LLM responses with candidate *navigation plans*, which is referred to as **plan verbalization**. After the handler-robot team agrees on a navigation task, the robot will take the sequence of actions generated by the planner, e.g., open door X , enter corridor Y , and approach room Z , to fulfill the service request. During navigation, our dialog interface verbalizes scene descriptions to improve the human handler’s spatial awareness and facilitate collaborative decision-making, referred to as **scene verbalization**. The plan verbalization (before navigation) and scene verbalization (during navigation) capabilities together identify the main novelty of our proposed dialog system for robotic guide dogs. We demonstrate the capabilities of our guide dog robot (see Figure 1) through a human study with visually impaired individuals. This study, which involves real-world indoor navigation tasks, allowed us to evaluate the system’s performance and analyze its potential societal contributions. To supplement these real-world findings, we also conducted simulation experiments to quantitatively measure the accuracy of our dialog system in identifying navigation tasks from open-vocabulary language.

2 Related Work

In this section, we discuss existing robotic guide dog systems, human-robot dialog systems, and non-robotic navigational aid systems for visually impaired people, highlighting how our system differs in each of these categories.

Robotic Guide Dog Systems have been developed to assist visually impaired individuals in navigating, thanks to reduced hardware costs and recent advances in quadruped

locomotion control. Systems of this type perform human-robot co-navigation tasks, where a human holds a leash to establish a physical connection with a quadruped robot (Cohav et al. 2025; Hwang et al. 2025). Some of these systems only support unidirectional robot-to-human communication via the forces the human feels from the leash (Wang, Zhao, and Zhang 2021; Xiao et al. 2021; Chen et al. 2023b). In those systems, the robot helps the human avoid obstacles at the motion level, and the human completely follows the robot during navigation, assuming both sides know the goal ahead of time. Other systems allow for bidirectional communication, such that the human can indicate desired navigation directions via buttons (Hwang et al. 2023), predefined voice commands (Kim et al. 2023; Hata et al. 2024; Kim et al. 2025), or tugs on the leash (DeFazio, Hirota, and Zhang 2023) during navigation. Compared with those robotic guide dog systems, ours further supports human-robot communication via bi-directional, multi-turn dialog.

Robot Dialog Systems have greatly improved the interaction capabilities of robots (Tellex et al. 2020; Bisk et al. 2020; Hayamizu, Yu, and Zhang 2023). These systems are optimized for different goals, including mediating the human-robot perception gap (Chai et al. 2014), learning from dialog experiences (Amiri et al. 2019), social dialog (Chen et al. 2018), learning navigation actions from dialog histories (Thomason et al. 2020), and understanding ambiguous human instructions (Thomason et al. 2015). Earlier works have focused on constructing formal navigation actions from natural language instructions (Chen and Mooney 2011) or commands (Matuszek et al. 2013). While effective for various dialog tasks, none of the above works were designed for the context of guided navigation for visually impaired individuals, where communicating navigation goals, planning to achieve them, and maintaining spatial awareness

are critically important. Such challenges motivated the development of dialog systems for robotic guide dogs in this paper, where we leverage task planners (Silver et al. 2022; Liu et al. 2023) and LLMs (Brohan et al. 2023; Stark et al. 2023) to facilitate human-robot dialog on navigation tasks in open vocabulary.

Non-robotic Navigation Aids were developed to help visually impaired people avoid obstacles, e.g., through belts (Yang 2021), dialog systems (Balata, Mikovec, and Slavik 2018; Katz et al. 2012; Berka et al. 2022; Helal, Moore, and Ramachandran 2001; Chen and Shiu 2020), tele-assistance (Balata, Mikovec, and Maly 2015), and smart canes that support haptic feedback from the users (Agrawal, West, and Hayes 2022). None of those systems were able to ground ambiguous human requests to navigation plans or used a robot for physical guidance, which is believed to be more effective in a recent questionnaire study (Hwang et al. 2024). This paper showcases a novel dialog system for robotic guide dogs with the capabilities of language synthesis (using a task planner and LLMs) for plan and scene verbalization in handler-robot co-navigation tasks.

3 Problem Settings and Methodology

In this section, we present our problem setting and robotic guide dog system as summarized in Figure 2.

3.1 Problem Settings and Assumptions

In this paper, we assume the robot has full knowledge about the domain, including a map of the environment associated with semantics information (e.g., labels and coordinates of kitchens and conference rooms). We also assume access to a stable quadruped locomotion controller that can accurately track linear and angular velocity commands. On top of the locomotion controller, the robot is equipped with a global path planner for 2D trajectory planning, and a local path planner that minimizes the deviation from the global path while avoiding obstacles. Those software packages are widely used and publicly available in the Robot Operating System (ROS) community (Koubãa 2017). We do not discuss scenarios where the human and robot together explore a new environment, the human helps the robot dog recover from falls, or the robot disobeys human commands, which occurs on biological guide dogs.

Our goal is to develop a robotic guide dog system that fulfills a handler’s request by recommending and executing navigation actions based on the service request in spoken language. This involves developing a dialog system that can map the service request to known locations in the environment, as well as generating an action sequence leading the handler-robot team to the goal location.

3.2 Our Dialog System for Robotic Guide Dogs

Our dialog system combines the flexible natural language understanding capabilities of an LLM with the long-horizon reasoning of task planners. This integration allows the system to engage in dialogue with users to specify navigation tasks and to deliver contextual, task-relevant information throughout execution. A key contribution of our work is the

Procedure 1: Robotic Guide Dog Dialogue and Navigation

```

    ▷ Dialogue for Task Specification
1: Greet the guide dog handler to initiate the conversation
2: while navigational plan is not finalized do
3:   Receive the handler’s service request
4:   Convert the service request to Task  $T$  in a formal language
5:   Enter  $T$  into a task planner and compute task plans  $\mathbf{P}$ 
6:   for each plan  $P \in \mathbf{P}$  do
7:     Verbalize plan  $P$  using an LLM    ▷ Plan Verbalization
8:     if User confirms a plan then Break the while loop
9:   end for
10:  Request the user to rephrase the service request
11: end while

    ▷ Plan Execution (concurrent): Thread 1 on navigation
12: for each navigation action  $a$  in plan  $P$  do
13:   Execute action  $a$ 
14: end for
15: Return: arrival at the final destination

    ▷ Plan Execution (concurrent): Thread 2 on dialog
16: while a scene change is detected or there is a long silence do
17:   Collect surrounding objects  $O$  using vision and world map
18:   Verbalize  $O$  with spatial info    ▷ Scene Verbalization
19: end while

```

ability to verbalize robotic guide dog’s outputs, e.g., scene understanding and decision-making process, into clear, accessible language. This verbal feedback equips visually impaired users with the critical information needed to make confident, informed decisions about their own navigation.

We first introduce how our system specifies a task and verbalizes plans based on the task, then describe the scene verbalization process during navigation. The control loops are shown in Procedure 1.

Plan Verbalization for Task Specification: Navigation begins when the user issues a request to task T in natural language through a headset microphone. The speech is converted to text using a speech-to-text model (Shmyrev et al. 2023) and fed to our LLM agent. The agent is set with a system prompt (Figure. 3) that defines its role as a guide dog, lists valid navigation locations, and specifies the expected output formats.

If the user’s command is ambiguous (e.g., “I’m thirsty”), the LLM clarifies the intent through dialogue. Instead of merely listing potential destinations (e.g., “kitchen” and “water fountain”), our system performs a unique process we call Plan Verbalization. For each potential destination, the system runs a symbolic planner (detailed in Section 3.3) in the background. The planner computes an optimal action sequence and extracts information about physical constraints, such as the total cost (i.e., travel time) and the number of doors that must be opened.

The LLM then incorporates this information into its response, generating an informative prompt like, “*We can go to the kitchen or the water fountain. The kitchen requires opening one door and will take about three minutes. The water fountain has no doors and will take about one minute. Where would you like to go?*” By verbalizing plan details, we

Robot Guide Dog Command Protocol

You are a robot guide dog assisting a visually impaired person in navigating an indoor environment. The human is holding a rigid connection attached to you. Your task is to navigate to the requested locations based on the human's commands. When the human makes a request, you must engage in communication to clarify their intentions. Ask clarification questions until you are confident about the request.

Command Syntax:

There is one command type: `goto <parameter>`

- robotics lab
- staircase
- elevator
- conference room
- bathroom
- waiting room
- water fountain
- kitchen

Valid Parameters:

Handling Requests:

- **Unavailable Locations:** Do not navigate to locations not on the list. If an unavailable location is requested, you must state this in your clarification question.
- **Ambiguous Requests:** If a request could refer to multiple valid locations, you must suggest each of them in your clarification question.

Required Output Format: After determining the correct action, provide your output in one of the two formats below.

To ask for more information:

CLARIFICATION QUESTION: <question>

To execute a command:

COMMAND <type> <parameter>

conversational statement

Figure 3: LLM prompt defining the role of our robot guide dog dialog system. Given possibly ambiguous human service tasks in natural language, the LLM must conduct a dialog and select the location that best satisfies the task.

enable the user to make an informed decision based on their preferences and circumstances. Since LLMs often struggle with spatial reasoning (Cohn and Hernandez-Orallo 2023), our dialog system employs a task planner to compute plans for achieving navigation goals by reasoning about both real-world constraints and the robot's action primitives. Once the user confirms their choice, the LLM generates a final formal task and a conversational statement to naturally conclude the dialogue (e.g., “*Okay, I will guide you to the kitchen.*”), ending the dialogue phase.

Scene Verbalization for Spatial Awareness: Once a destination is confirmed and navigation begins, the system continues to provide information to the user through Scene Verbalization *during* navigation. This is triggered by major scene changes or a long silence in the dialogue (see Procedure 1), such as when the robot's position crosses the boundary of a semantically labeled region on the map (e.g., moving from a “corridor” to a “kitchen”).

When the robot enters a new area, a message describing the current scene is played through a speaker, e.g., “*We are navigating in a long corridor... We are approaching an of-*

fice door... We just arrived at an office door.” This allows the visually impaired user to maintain awareness of their surroundings during navigation.

While our current implementation uses a simple strategy, we position it as a first step, drawing on the rich literature on verbalization in human-robot co-navigation (Rosenthal, Selvaraj, and Veloso 2016). We focus on evaluating its effectiveness in a human study in this paper, and plan to investigate more advanced methods, such as learning verbalization policies, in future work.

3.3 Implementation Details

Safeguard for Dialog Management: Due to the uncertainty of LLM responses and the lack of spatial awareness, we employ a language synthesis module with a safeguard that post-processes the authentic LLM responses before delivering them to the human user. Specifically, during the plan verbalization phase, if the LLM does not respond in the expected format (e.g., no “CLARIFICATION QUESTION” or “COMMAND” indicators in the response), or if it fails to suggest any valid locations as determined by keyword detection, then we replace the LLM response with the message:

“I can only assist with navigation requests of nearby locations that I know about. Would you like help navigating to somewhere nearby?”

Plan Generation for Navigation: To complete navigation tasks specified by our dialog system, we generate navigation plans via an Answer Set Programming (ASP) planner (Lifschitz 2019). We leverage planners designed for navigation tasks (Hayamizu et al. 2021) to determine high-level action sequences. We consider the set of actions:

- 1 approach (door_id),
- 2 gothrough (door_id),
- 3 opendoor (door_id),
- 4 goto (location)

Arriving at the destination specified by the dialog system involves generating a sequence of actions while minimizing their total cost. Action costs between locations are considered and measured based on navigation distances, along with an additional constant cost added for door opening actions. Rules in the ASP encode the necessary domain knowledge for navigation. For example, the following rule states that P1 can be accessed by P2 via door D, if P1 and P2 are unique locations, D is a door, and P1 and P2 both have door D.

- 1 dooracc (P1, D, P2) :-
- 2 hasdoor (P1, D), hasdoor (P2, D),
- 3 P1 != P2, door (D),
- 4 location (P1), location (P2).

LLM-based task planning methods are increasingly mature (Kannan, Venkatesh, and Min 2024; Rana et al. 2023; Liu et al. 2023; Zhang et al. 2025), and robotic guide dog practitioners can feel free to implement the task planner using LLMs instead. Next, we present a human study for evaluating the effectiveness of our dialog system in collaborative settings between a human handler and a robotic guide dog (Section 4), followed by an extensive evaluation in simulation focusing on accuracy and efficiency (Section 5).

4 Real-World Effectiveness Evaluation: Human Study

In this section, we detail the human study, which was designed to test the following primary hypothesis: Combining scene and plan verbalization improves user satisfaction and perceived effectiveness for robot-assisted navigation.²

4.1 Participants

Table 1 shows the demographics of the participants. Seven legally blind individuals (5 Male, 2 Female; age range 40–68) were recruited for the study.³ One participant who completed the study self-identified as not meeting the vision requirement and was excluded from the final analysis. Two of the participants had prior experience with a guide dog.

Age	Gender	Impairment Duration (yrs)	Guide Dog Experience
68	Female	68	None
68	Male	60	None
59	Male	51	8 yrs, 2 dogs
40	Male	40	None
64	Male	35	18 yrs, 3 dogs
57	Male	20	None
52	Female	15	None

Table 1: Participant Demographics.

4.2 Experimental Setup

The experiment took place in a large, many-room office environment. The task involved navigating a route from an entrance to a conference room. We used a Unitree Go2 as a quadruped robot platform and integrated it with our dialog system. For safety reasons during the human study, the robot’s movements were controlled by an expert operator in a Wizard of Oz setup (Dahlbäck, Jönsson, and Ahrenberg 1993). This approach ensures safety and does not affect the evaluation of the effectiveness of the dialog strategies.

After signing a consent form, participants were briefed on the task. They would request to go to a conference room, and the robotic guide dog would guide them under three distinct verbalization conditions in a within-subjects design:

- **Minimum Verbalization:** A control condition with language-based task specification before navigation and then no verbal interaction during navigation.
- **Scene Verbalization:** The robot only verbalized information about the immediate surroundings (e.g., obstacles and environmental context).
- **Ours (Scene + Plan):** The robot provided both scene and plan verbalization.

²This section focuses on presenting the results of our robotic guide dog system under *controlled autonomy*. The system under *full autonomy* is demonstrated on the project webpage.

³A person is considered legally blind if they meet one of the following criteria in their better-seeing eye, even with the best conventional correction: 1) Visual acuity of 20/200 or less; 2) Visual field of 20 degrees or less.

Q1 Utility	The natural language interface is a useful component of the system.
Q2 Helpfulness	The robot is helpful for visually impaired navigation when compared to having no tools for visually impaired navigation.
Q3 Helpfulness	The robot is helpful for visually impaired navigation when compared to having other tools for visually impaired navigation.
Q4 Safety	The robot is capable of safely guiding me to a navigation goal.
Q5 Ease of comm.	I can easily communicate with the robot.
Q6 Preference	Given the same navigation task we performed today I prefer a guide dog to guide me instead of the robot.
Q7 Feedback	If you have any questions, comments, or concerns please leave them here.

Table 2: The list of questions in the questionnaire. The items assessed interface utility, helpfulness, safety, ease of communication, and preference relative to a real guide dog.

Condition	Q1	Q2	Q3	Q4	Q5	Q6
Mini. Verb.	4.33	4.67	3.50	4.00	4.33	3.17
Scene Verb.	4.67	4.83	4.00	4.00	4.00	3.17
Ours	4.83	4.83	4.33	3.83	4.50	3.67

Table 3: Average survey ratings for each verbalization condition. Ratings are on a 5-point Likert scale (1: Strongly Disagree, 5: Strongly Agree).

Given the age and the energy span of our participants, we decided to keep the trials and questionnaire within 30 minutes, and evaluate no more than three methods. After completing the task under each condition, participants filled out a 7-item questionnaire shown in Table 2. Q1 through Q6 were rated on a 5-point Likert scale, and Q7 was an open-ended feedback question. Participants self-administered the questionnaires after task completion, while we provided support by reading the instructions and questions.

4.3 Results

Figure 4 presents an instance of the experiment trials in the human study. Table 3 shows the average ratings for each condition. Our proposed method, **Ours (Scene + Plan)**, received the highest ratings on natural language interface utility (Q1). The result indicates the robot’s strong performance in helpfulness (Q2 and Q3). Furthermore, our method outperformed in ease of communication (Q5) and preferability over real guide dogs (Q6). The overall result suggests that providing both scene and plan information leads to higher effectiveness for navigation and a user-friendly approach.

Our method scored slightly lower on perceived safety (Q4: 3.83) than the other conditions. Qualitative feedback (Q7) implies that the reason might be due to initial unfamiliarity with walking with the robot. The human study confirms the real-world value of our proposed natural language interface for robotic guide dogs.

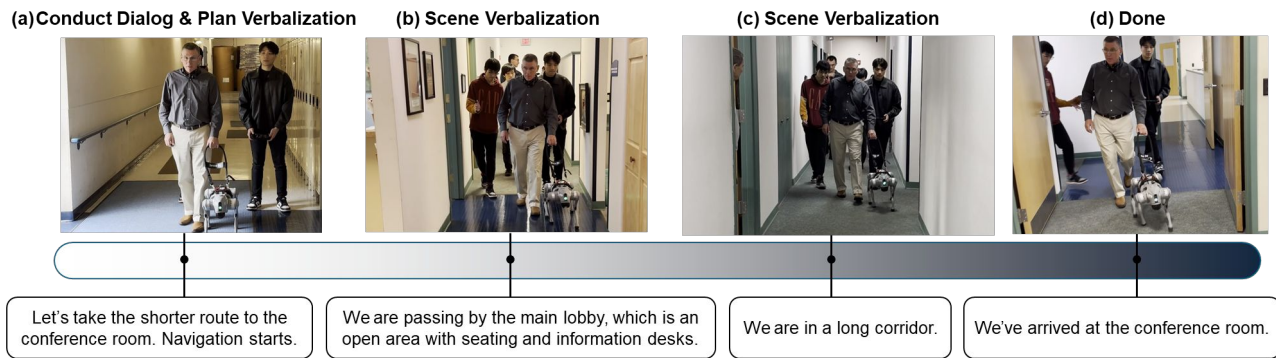


Figure 4: An illustrative example of our robotic guide dog assisting a legally blind participant to navigate to a conference room. The sequence shows: (a) The robot verbalizes the generated navigation plans for a user request, and starts navigation after the user chooses one of the plans. (b, c) During navigation, the system provides real-time scene verbalization, such as passing a lobby and entering a corridor. (d) The robot announces that they have arrived at the destination.

Category	Locations
Places to eat	food court, cafeteria, restaurant, dining hall, shawarma joint, cafe, vending machine, food truck, fast food restaurant
Restroom	restroom, bathroom, the loo
Change Floor	elevator/stairs, elevator/escalator, elevator
Cut Hair	barber shop, salon

Table 4: Samples in our collected service task dataset: locations in each category considered functionally equivalent.

5 System Performance Evaluation: Simulation Analysis

While the human study demonstrated the real-world usability of our approach, we conducted a large-scale simulation analysis to test the underlying system’s performance. This analysis tests three hypotheses: our system can accurately and efficiently resolve ambiguous service requests (**Hypothesis I**), is robust to noisy speech recognition (**Hypothesis II**), and can leverage a navigation planner to reduce travel costs (**Hypothesis III**). Inspired by recent work using LLMs for data synthesis (Chen et al. 2023a; Tan et al. 2024), we simulated a visually impaired user with GPT-4 (Achiam et al. 2023). To ground our simulation in reality, the user’s goals were based on a library of navigation-related service requests collected from human participants, as described next.

5.1 Simulation Setup

Service Task Library To create a realistic task library, we surveyed 16 university students (ages 19-30). Participants were asked to imagine being a visually impaired person who needs help navigating and provide five $\langle \text{location}, \text{purpose} \rangle$ pairs for potential navigation requests, such as (“water fountain”, “I want something to drink”). After filtering three overly vague entries, we obtained 77 unique samples. We normalized location names and grouped functionally equivalent locations (e.g., “cafeteria,” “dining

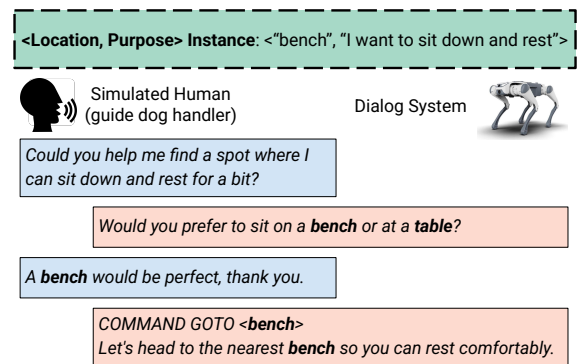


Figure 5: An example simulated conversation. The handler implicitly states a purpose sampled from $\langle \text{location}, \text{purpose} \rangle$ pair. The dialog system suggests relevant locations. After confirming the handler is looking for a bench, the system generates a formal task and concludes the task specification conversation.

hall”), as shown in Table 4. The unique locations can be inserted in the system prompt as the set of valid parameters in Figure 3. We report accuracy and efficiency for evaluation.

Simulated Human We used GPT-4 to simulate a visually impaired user, instructing with the following prompt:

“You are an AI chatbot pretending to be a visually impaired person. You need to navigate to *location*. Do not say the name of where you want to go, unless asked. You are only interested in going to *location* and must not show interest in any other locations.”

This prompt yields ambiguity by discouraging the direct mention of the location. Each conversation was initiated by providing the simulated human with the purpose from a $\langle \text{location}, \text{purpose} \rangle$ pair (e.g., “Begin the conversation by indicating that you want something to drink”). An example dialog is shown in Figure 5.

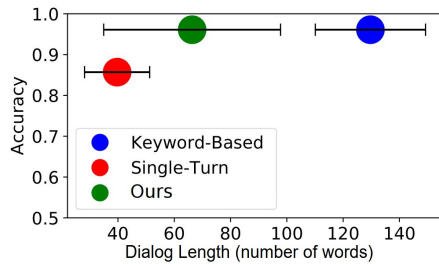


Figure 6: Accuracy vs. Average Dialog Length without noise. Our approach provides the best trade-off between high accuracy and efficient conversation.

5.2 Evaluation of Accuracy and Efficiency

To test **Hypothesis I**, we simulated conversations for all 77 task pairs and compared three approaches:

1. **Keyword-Based:** A multi-turn template-based dialog system using keyword detection. If no keywords are found, it lists all known locations.
2. **Single-Turn:** Our LLM-based system, but restricted to a single turn (no clarification questions).
3. **Ours:** Our full multi-turn LLM-based dialog system, described in Section 3.

For multi-turn approaches, dialogs were limited to six turns; exceeding this limit counted as a failure. A trial was successful if the dialog concluded with the correct (or functionally equivalent) formal task. We measured accuracy and efficiency by the dialog length and its success rate.

As shown in Figure 6, the result highlights the trade-offs between each approach in terms of accuracy, average number of dialog turns. It achieves the highest accuracy (94.8%) while being more concise than the Keyword-Based baseline, which is overly verbose as it often must enumerate all locations. These results support **Hypothesis I**.

5.3 Evaluation of Robustness to Noisy Speech

To test robustness to speech recognition errors that may arise from speech-to-text models (**Hypothesis II**), we simulated noisy input by perturbing each character in the user’s responses with 0.3 probability, randomly applying a deletion, insertion, or substitution. This is particularly relevant to a guide dog setting, in which the visually impaired person can navigate in crowded and noisy environments.

Figure 7 shows the results that suggest our dialog system is robust to perturbations, only losing 5.2% accuracy, and marginally increasing dialog length. In contrast, the Keyword-Based’s performance collapsed, as it failed to match the perturbed keywords. This supports **Hypothesis II**.

5.4 Evaluation of Planner-Informed Dialog

Finally, we evaluated if providing plan information (e.g., travel distance) in the dialog helps the user choose more efficient routes (**Hypothesis III**). We used three tasks (Table 5) where multiple functionally equivalent destinations existed at different distances.

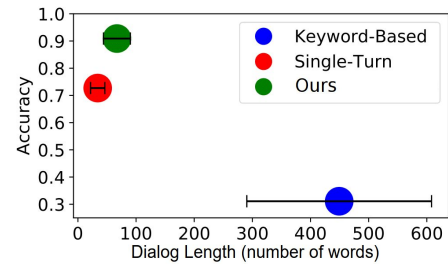


Figure 7: Accuracy vs. Dialog Length with perturbed (noisy) inputs. Our approach remains highly accurate, while the keyword-based method fails.

Location	Purpose
elevator	“I want to go to the 2nd floor”
water fountain	“I want something to drink”
waiting room	“I want to sit down”

Table 5: The three location-purpose pairs used for planner-informed evaluation.

Method	Dialog Len.	Nav. Cost	Total Time
w/ Plan Info	115.67	55.17	230.15
w/o Plan Info	84.83	73.00	277.27

Table 6: Effect of Plan Information on Performance Metrics

We compared our system with and without plan information. Table 6 shows that when the dialog included navigation costs, the simulated user consistently chose the closer destination. Although this increased dialog length, the reduction in navigation cost led to a lower overall task time (calculated assuming 2.5 words/sec talking and 0.3 m/s walking speeds). These results support **Hypothesis III**.

6 Conclusion

In this paper, we developed a novel robotic guide dog system that can verbally convey both navigation plans and environmental scenes using a task planner and an LLM. The effectiveness was validated through two evaluations. The human study with visually impaired participants confirmed that combined plan and scene verbalization is an effective and preferred communication strategy. The simulation analysis supported this, demonstrating high accuracy, robustness, and efficiency by leveraging planner knowledge, which accounts for the high user satisfaction.

Despite these promising results, the human study highlighted key challenges, particularly a lower safety rating compared to biological counterparts, underscoring the need for further enhancements. Future work will involve expanded user studies, increased system autonomy, and deployment in more complex environments. The ultimate goal is to integrate robotic guide dogs into daily life, enhancing independent navigation for both visually impaired and sighted individuals.

Ethical Statement

The research protocol involving human participants was reviewed and approved by the Institutional Review Board (IRB). Seven legally blind individuals were recruited for the study. All participants provided written informed consent after being fully briefed on the experimental task and were compensated for their time. Participants were explicitly informed that their participation was voluntary and that they retained the right to withdraw from the study at any time without penalty. To ensure participant safety during the real-world navigation component, a Wizard-of-Oz approach was used. In this setup, an expert operator remotely controlled all physical movements of the robot. This protocol was implemented to mitigate risks and isolate the evaluation to the system's dialog strategies. Researchers provided support by reading instructions and questionnaire items to participants. Data was stored securely in accordance with institutional data protection policies. This work contributes to the field of assistive robotics, which focuses on the well-being of people with disabilities. The primary objective of this research is to enhance the independence and mobility of visually impaired individuals by addressing the limitations posed by the scarcity of guide dogs and the challenges of their care.

Acknowledgments

The authors thank Thomas Panek, Bill Ma, and Ken Fernald for their guidance in this research, and the participants for their valuable feedback. A portion of this work has taken place at the Autonomous Intelligent Robotics (AIR) Group, SUNY Binghamton. AIR research is supported in part by the NSF (IIS-2428998, NRI-1925044), Ford Motor Company, DEEP Robotics, OPPO, Guiding Eyes for the Blind, and SUNY RF. Research reported in this publication was supported in part by SUNY System Administration using the SUNY AI Platform.

References

- Achiam, J.; Adler, S.; Agarwal, S.; Ahmad, L.; Akkaya, I.; Aleman, F. L.; Almeida, D.; Altschmidt, J.; Altman, S.; Anadkat, S.; et al. 2023. Gpt-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Agrawal, S.; West, M. E.; and Hayes, B. 2022. A novel perceptive robotic cane with haptic navigation for enabling vision-independent participation in the social dynamics of seat choice. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- Amiri, S.; Bajracharya, S.; Goktolgal, C.; Thomason, J.; and Zhang, S. 2019. Augmenting knowledge through statistical, goal-oriented human-robot dialog. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*.
- Balata, J.; Mikovec, Z.; and Maly, I. 2015. Navigation problems in blind-to-blind pedestrians tele-assistance navigation. In *Human-Computer Interaction—INTERACT 2015: 15th IFIP TC 13 International Conference, Bamberg, Germany, September 14–18, 2015, Proceedings, Part I 15*, 89–109. Springer.
- Balata, J.; Mikovec, Z.; and Slavik, P. 2018. Conversational agents for physical world navigation. *Studies in Conversational UX Design*, 61–83.
- Berka, J.; Balata, J.; Jonker, C. M.; Mikovec, Z.; van Riemsdijk, M. B.; and Tielman, M. L. 2022. Misalignment in semantic user model elicitation via conversational agents: a case study in navigation support for visually impaired people. *International Journal of Human–Computer Interaction*, 38(18-20): 1909–1925.
- Bisk, Y.; Holtzman, A.; Thomason, J.; Andreas, J.; Bengio, Y.; Chai, J.; Lapata, M.; Lazaridou, A.; May, J.; Nisnevich, A.; et al. 2020. Experience Grounds Language. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 8718–8735.
- Brohan, A.; Chebotar, Y.; Finn, C.; Hausman, K.; Herzog, A.; Ho, D.; Ibarz, J.; Irpan, A.; Jang, E.; Julian, R.; et al. 2023. Do as i can, not as i say: Grounding language in robotic affordances. In *Conference on Robot Learning*, 287–318. PMLR.
- Chai, J. Y.; She, L.; Fang, R.; Ottarson, S.; Little, C.; Liu, C.; and Hanson, K. 2014. Collaborative effort towards common ground in situated human-robot dialogue. In *Proceedings of the 2014 ACM/IEEE international conference on Human-robot interaction*, 33–40.
- Chen, C.-H.; and Shiu, M.-F. 2020. Smart guiding glasses with descriptive video service and spoken dialogue system for visually impaired. In *IEEE International Conference on Consumer Electronics-Taiwan (ICCE-Taiwan)*.
- Chen, C.-Y.; Yu, D.; Wen, W.; Yang, Y. M.; Zhang, J.; Zhou, M.; Jesse, K.; Chau, A.; Bhowmick, A.; Iyer, S.; et al. 2018. Gunrock: Building a human-like social bot by leveraging large scale real user data. *Alexa prize proceedings*.
- Chen, D.; and Mooney, R. 2011. Learning to interpret natural language navigation instructions from observations. In *Proceedings of the AAAI Conference on Artificial Intelligence (AAAI)*.
- Chen, M.; Papangelis, A.; Tao, C.; Kim, S.; Rosenbaum, A.; Liu, Y.; Yu, Z.; and Hakkani-Tur, D. 2023a. PLACES: Prompting language models for social conversation synthesis. *arXiv preprint arXiv:2302.03269*.
- Chen, Y.; Xu, Z.; Jian, Z.; Tang, G.; Yang, L.; Xiao, A.; Wang, X.; and Liang, B. 2023b. Quadruped guidance robot for the visually impaired: A comfort-based approach. In *IEEE International Conference on Robotics and Automation (ICRA)*.
- Cohav, A. L.; Gong, A. X.; Kim, J. T.; Zeagler, C.; Ha, S.; and Walker, B. N. 2025. Do Looks Matter? Exploring Functional and Aesthetic Design Preferences for a Robotic Guide Dog. *arXiv preprint arXiv:2503.16450*.
- Cohn, A. G.; and Hernandez-Orallo, J. 2023. Dialectical language model evaluation: An initial appraisal of the common-sense spatial reasoning abilities of LLMs. *arXiv preprint arXiv:2304.11164*.
- Dahlbäck, N.; Jönsson, A.; and Ahrenberg, L. 1993. Wizard of Oz studies: why and how. In *Proceedings of the 1st international conference on Intelligent user interfaces*, 193–200.

- DeFazio, D.; Hirota, E.; and Zhang, S. 2023. Seeing-Eye Quadruped Navigation with Force Responsive Locomotion Control. In *Conference on Robot Learning*, 2184–2194. PMLR.
- Guiding-Eyes, f.t.-B. 2023. Guiding Eyes for the Blind Class Lectures (Chapter 4: Vocabulary of a Guide Dog). <https://www.guidingeyes.org/class-lectures/>. Accessed June 19, 2025.
- Hata, R.; Trikasemsak, N.; Giudice, A.; and Doore, S. A. 2024. See Spot Guide: Accessible Interfaces for an Assistive Quadruped Robot. *arXiv:2402.11125*.
- Hayamizu, Y.; Amiri, S.; Chandan, K.; Takadama, K.; and Zhang, S. 2021. Guiding robot exploration in reinforcement learning via automated planning. In *Proceedings of the International Conference on Automated Planning and Scheduling*, volume 31, 625–633.
- Hayamizu, Y.; Yu, Z.; and Zhang, S. 2023. Learning Joint Policies for Human-Robot Dialog and Co-Navigation. In *2023 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 7893–7898. IEEE.
- Helal, A.; Moore, S. E.; and Ramachandran, B. 2001. Drishti: An integrated navigation system for visually impaired and disabled. In *Proceedings fifth international symposium on wearable computers*, 149–156. IEEE.
- Hwang, H.; Adhikari, K.; Shodhaka, S.; and Kim, D. 2025. Synthetic Data Augmentation for Robotic Mobility Aids to Support Blind and Low. In *Robot Intelligence Technology and Applications 9: Results from the 12th International Conference on Robot Intelligence Technology and Applications*, volume 1419, 92. Springer Nature.
- Hwang, H.; Jung, H.-T.; Giudice, N. A.; Biswas, J.; Lee, S. I.; and Kim, D. 2024. Towards Robotic Companions: Understanding Handler-Guide Dog Interactions for Informed Guide Dog Robot Design. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1–20.
- Hwang, H.; Xia, T.; Keita, I.; Suzuki, K.; Biswas, J.; Lee, S. I.; and Kim, D. 2023. System configuration and navigation of a guide dog robot: Toward animal guide dog-level guiding work. In *IEEE International Conference on Robotics and Automation (ICRA)*.
- Kannan, S. S.; Venkatesh, V. L.; and Min, B.-C. 2024. Smart-llm: Smart multi-agent robot task planning using large language models. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 12140–12147. IEEE.
- Katz, B. F.; Kammoun, S.; Parseihian, G.; Gutierrez, O.; Brilhault, A.; Auvray, M.; Truillet, P.; Denis, M.; Thorpe, S.; and Jouffrais, C. 2012. NAVIG: Augmented reality guidance system for the visually impaired: Combining object localization, GNSS, and spatial audio. *Virtual Reality*, 16: 253–269.
- Kim, J. T.; Byrd, M.; Crandell, J. L.; Walker, B. N.; Turk, G.; and Ha, S. 2025. Understanding expectations for a robotic guide dog for visually impaired people. In *2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 262–271. IEEE.
- Kim, J. T.; Yu, W.; Kothari, Y.; Tan, J.; Turk, G.; and Ha, S. 2023. Transforming a Quadruped into a Guide Robot for the Visually Impaired: Formalizing Wayfinding, Interaction Modeling, and Safety Mechanism. In *Conference on Robot Learning*. PMLR.
- Koubâa, A. 2017. *Robot Operating System (ROS)*. Springer.
- Lifschitz, V. 2019. *Answer set programming*. Springer Heidelberg.
- Liu, B.; Jiang, Y.; Zhang, X.; Liu, Q.; Zhang, S.; Biswas, J.; and Stone, P. 2023. Llm+ p: Empowering large language models with optimal planning proficiency. *arXiv preprint arXiv:2304.11477*.
- Marks, P. 2025. Guide Dogs Are Expensive and Scarce. Could Robots Do Their Job? *Communications of the ACM*, 68(4): 12–14.
- Matuszek, C.; Herbst, E.; Zettlemoyer, L.; and Fox, D. 2013. Learning to parse natural language commands to a robot control system. In *Experimental robotics: the 13th international symposium on experimental robotics*, 403–415. Springer.
- Rana, K.; Haviland, J.; Garg, S.; Abou-Chakra, J.; Reid, I.; and Suenderhauf, N. 2023. Sayplan: Grounding large language models using 3d scene graphs for scalable robot task planning. *arXiv preprint arXiv:2307.06135*.
- Rosenthal, S.; Selvaraj, S. P.; and Veloso, M. M. 2016. Verbalization: Narration of Autonomous Robot Experience. In *IJCAI*.
- Shmyrev, N. V.; et al. 2023. Vosk Speech Recognition Toolkit. <https://github.com/alphacep/vosk-api>. Accessed Nov 10, 2024.
- Silver, T.; Hariprasad, V.; Shuttleworth, R. S.; Kumar, N.; Lozano-Pérez, T.; and Kaelbling, L. P. 2022. PDDL planning with pretrained large language models. In *NeurIPS 2022 foundation models for decision making workshop*.
- Stark, C.; Chun, B.; Charleston, C.; Ravi, V.; Pabon, L.; Sunkari, S.; Mohan, T.; Stone, P.; and Hart, J. 2023. Dobby: A Conversational Service Robot Driven by GPT-4. *arXiv preprint arXiv:2310.06303*.
- Tan, Z.; Li, D.; Wang, S.; Beigi, A.; Jiang, B.; Bhattacharjee, A.; Karami, M.; Li, J.; Cheng, L.; and Liu, H. 2024. Large language models for data annotation and synthesis: A survey. *arXiv preprint arXiv:2402.13446*.
- Tellex, S.; Gopalan, N.; Kress-Gazit, H.; and Matuszek, C. 2020. Robots that use language. *Annual Review of Control, Robotics, and Autonomous Systems*, 3: 25–55.
- Thomason, J.; Murray, M.; Cakmak, M.; and Zettlemoyer, L. 2020. Vision-and-dialog navigation. In *Conference on Robot Learning*, 394–406. PMLR.
- Thomason, J.; Zhang, S.; Mooney, R. J.; and Stone, P. 2015. Learning to interpret natural language commands through human-robot dialog. In *Twenty-Fourth International Joint Conference on Artificial Intelligence (IJCAI)*.
- Tomkins, L. M.; Thomson, P. C.; and McGreevy, P. D. 2011. Behavioral and physiological predictors of guide dog success. *Journal of Veterinary Behavior*, 6(3): 178–187.
- Wang, L.; Zhao, J.; and Zhang, L. 2021. Navdog: robotic navigation guide dog via model predictive control and

human-robot modeling. In *Proceedings of the 36th Annual ACM Symposium on Applied Computing*, 815–818.

Wang, X.; and Wu, W. 2024. Robot shows it can really work like man’s best friend. Published July 19, 2024.

Whitmarsh, L. 2005. The benefits of guide dog ownership. *Visual impairment research*, 7(1): 27–42.

Xiao, A.; Tong, W.; Yang, L.; Zeng, J.; Li, Z.; and Sreenath, K. 2021. Robotic guide dog: Leading a human with leash-guided hybrid physical interaction. In *IEEE International Conference on Robotics and Automation (ICRA)*.

Yang, X. 2021. Project Guideline: Enabling Those with Low Vision to Run Independently. <https://research.google/blog/project-guideline-enabling-those-with-low-vision-to-run-independently/>. Accessed June 19, 2025.

Zhang, X.; Ding, Y.; Hayamizu, Y.; Altaweel, Z.; Zhu, Y.; Zhu, Y.; Stone, P.; Paxton, C.; and Zhang, S. 2025. LLM-GROP: Visually grounded robot task and motion planning with large language models. *The International Journal of Robotics Research*, 02783649251378196.