ChitChatGuide: Conversational Interaction Using Large Language Models for Assisting People with Visual Impairments to Explore a Shopping Mall

YUKA KANIWA*, Waseda University, Japan
MASAKI KURIBAYASHI*, Waseda University, Japan
SEITA KAYUKAWA, IBM Research - Tokyo, Japan
DAISUKE SATO, Carnegie Mellon University, United States
HIRONOBU TAKAGI, IBM Research - Tokyo, Japan
CHIEKO ASAKAWA, Miraikan - The National Museum of Emerging Science and Innovation, Japan and IBM Research, United States
SHIGEO MORISHIMA, Waseda Research Institute for Science and Engineering, Japan

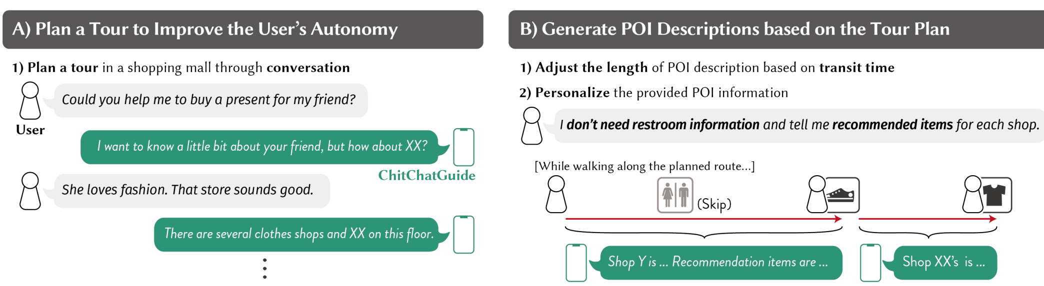

Fig. 1. Two functionalities of ChitChatGuide. (A) The system assists PVI in planning a tour to a destination of interest through conversational interaction. (B) Tailored descriptions of POIs are also generated depending on user interests or requests.

To enable people with visual impairments (PVI) to explore shopping malls, it is important to provide information for selecting destinations and obtaining information based on the individual's interests. We achieved this through conversational interaction by integrating a large language model (LLM) with a navigation system. ChitChatGuide allows users to plan a tour through contextual conversations, receive personalized descriptions of surroundings based on transit time, and make inquiries during navigation. We conducted a study in a shopping mall with 11 PVI, and the results reveal that the system allowed them to explore the facility with

*Both authors contributed equally to this research.

Authors' Contact Information: Yuka Kaniwa, Waseda University, Tokyo, Japan, ycrabring@ruri.waseda.jp; Masaki Kuribayashi, Waseda University, Tokyo, Japan, rugbykuribayashi@toki.waseda.jp; Seita Kayukawa, IBM Research - Tokyo, Tokyo, Japan, seita.kayukawa@ibm.com; Daisuke Sato, Robotics Institute, Carnegie Mellon University, Pittsburgh, Pennsylvania, United States, daisukes@cmu.edu; Hironobu Takagi, IBM Research - Tokyo, Tokyo, Japan, takagih@jp.ibm.com; Chieko Asakawa, Miraikan - The National Museum of Emerging Science and Innovation, Tokyo, Japan and IBM Research, Yorktown Heights, New York, United States, chiekoa@us.ibm.com; Shigeo Morishima, Waseda Research Institute for Science and Engineering, Tokyo, Japan, shigeo@waseda.jp.

ACM 2573-0142/2024/9-ART247
https://doi.org/10.1145/3676492

increased enjoyment. The LLM-based conversational interaction, by understanding vague and context-based questions, enabled the participants to explore unfamiliar environments effectively. The personalized and in-situ information generated by the LLM was both useful and enjoyable. Considering the limitations we identified, we discuss the criteria for integrating LLMs into navigation systems to enhance the exploration experiences of PVI.

CCS Concepts: • **Human-centered computing → Accessibility technologies**; • **Social and professional topics → People with disabilities**.

Additional Key Words and Phrases: visual impairment; orientation and mobility; large language model

ACM Reference Format:
Yuka Kaniwa, Masaki Kuribayashi, Seita Kayukawa, Daisuke Sato, Hironobu Takagi, Chieko Asakawa, and Shigeo Morishima. 2024. ChitChatGuide: Conversational Interaction Using Large Language Models for Assisting People with Visual Impairments to Explore a Shopping Mall. *Proc. ACM Hum.-Comput. Interact.* 8, MHCI, Article 247 (September 2024), 25 pages. https://doi.org/10.1145/3676492

1 Introduction

People often visit shopping malls without a specific purpose like finding a particular item at a specific store, and they enjoy casually exploring places that interest them, known as recreational window-shopping [17, 78]. People with visual impairments (PVI) often find it challenging to enjoy this exploration experience in shopping malls [41, 63, 73]. They must often rely on friends, family, or sighted passersby for information and directions [28, 56]. However, many PVI prefer independence, and assistance is not always available [7, 28, 56]. Enabling PVI to explore shopping malls more independently, with less reliance on others, directly enhances their social participation and improves their quality of life [41]. While many systems assist the mobility of PVI with turn-by-turn navigation [4, 60, 67, 83], few address exploration of unfamiliar indoor places like shopping malls. Effective exploration requires both mobility and an understanding of surroundings [38]. This research focuses on the latter, assuming the use of mobility assistance [47] (*e.g.*, navigation robots) and envisioning its integration in these types of systems.

To create opportunities for exploration, previous works [38, 43] proposed two major requirements for providing surrounding information. First, it is important to improve the autonomy of PVI, *i.e.*, the ability to select destinations independently at one's own pace, based on their knowledge and interests [43]. Previous works suggest that overviews and precise spatial information are crucial for wayfinding decisions and active exploration [6, 22, 38]. Existing navigation systems for public buildings (*e.g.*, shopping malls [69] and train stations [45]) use methods such as offering a list of store names from which to select destinations [45, 69] or allowing users to speak to the system that can detect specific words [69]. However, the main drawbacks of these systems are their passive nature and lack of flexibility. PVI have preferred conversational agents for actively understanding surroundings [31], suggesting that conversational interaction may better support exploration by actively providing overviews and recommending routes based on user interests. Second, it is essential to enable PVI to grasp surrounding points of interest (POIs) based on their preferences while navigating [38]. POI information could provide PVI with an enjoyable experience in public spaces such as a shopping mall [69] or an art museum [7]. Existing systems provide POI information when passing nearby [15, 34], but they deliver only predetermined information, which may be unnecessary or lack expected details in exploratory settings to fully meet user needs [1, 33]. Previous works highlighted the need for systems that provide flexible POI information based on user needs, emphasizing the potential of conversational interfaces that process and output contextual information [1, 31, 62]. These works demonstrate the need for technologies that permit exploration through versatile conversational interaction, thus supporting varied information needs. However, the actual design and impact of such interaction have never been examined. Based on these considerations, we present our initial research question: *How can we design conversational interaction to provide information flexibly to PVI who want to independently explore shopping malls?*

To answer this question, we hypothesized that integrating large language models (LLMs) into an assistance system for PVI could enrich the exploration experience. For the first requirement, LLMs may identify destinations of interest by flexibly interacting with PVI, demonstrating a deep understanding of conversation context [9, 82]. For the second requirement, LLMs may understand user preferences and create engaging and personalized sentences in various formats [9, 30]. These features can be enabled by prompting LLMs with information that traditional navigation systems have access to (*e.g.*, map information, localization results, and information on each POI) along with requests from users (*e.g.*, inquiries about stores in the facility or preferences for how each POI should be explained). The integration with LLMs can provide complex, versatile, and contextual conversational interaction between users and systems [75], thus offering personalized information to each PVI. These characteristics motivated us to design LLM-based conversational interaction and examine its use in a real shopping mall.

We developed *ChitChatGuide*, a system that assists users in exploring surroundings while navigating. We implemented two conversational functionalities to assist the exploration of PVI: planning tours and generating POI descriptions (Fig. 1). The system is designed to support users in selecting tours, *i.e.*, routes that help PVI explore a facility, based on their interests (Fig. 1–A). It provides interactive conversational interaction to enable users to understand an abstract of the facility by asking for an overview of each floor or recommendations for stores or inquiring about specific store details, assisting them in gaining autonomy. While navigating, the system describes the facility by generating descriptions whose lengths are adjusted based on the estimated transit time between POIs, thus allowing users to continuously receive information without an inefficient use of time (Fig. 1–B). Furthermore, users can request ChitChatGuide to provide specific information or exclude unnecessary information, allowing it to tailor the information. As the system describes nearby POIs, users may ask questions about them to gain in-depth knowledge not covered by the POI descriptions.

We conducted a user study at a public shopping mall with 11 visually impaired participants during regular business hours. They used the proposed system and a baseline system (Sec. 5.1), which is an existing navigation system that provides destination selection through a list of stores and concise descriptions of them while navigating. We employed an unstructured user study, which asked participants to freely select a route and then navigate it while interacting with the system to explore the facility. Through this study, we obtained the following findings:

- Most participants were able to engage in shopping mall exploration with increased enjoyment by planning their tour and listening to flexible POI descriptions using ChitChatGuide. Participants' comments highlighted the system's potential to motivate visually impaired shopping mall visitors to visit unknown stores and possibly purchase an item.
- For tour planning, the LLM's natural, contextual conversations helped users explore unfamiliar facilities effectively, especially those without specific destinations, because the system understood vague and context-based questions.
- The LLM's ability to generate personalized POI descriptions based on user preferences and location provided enjoyable, in-situ navigation information. Moreover, its Q&A handling conveniently responded to additional user requests.
- To enhance conversational interaction for exploration, we revealed three criteria for integrating an LLM into a navigation system for PVI: the balance of attractiveness and length of descriptions, trust in LLMs by PVI, and the requirement for the depth of information provided to LLMs.

2 Related Work

2.1 Wayfinding and Exploration in Indoor Navigation

To support PVI in unfamiliar indoor environments, navigation systems focus on both wayfinding (planning and following a route) and safe navigation (avoiding obstacles) [44, 47, 54], but wayfinding remains challenging [83]. Effective systems include localization and user interfaces for information and inputs [69, 83]. Localization methods (*e.g.*, Bluetooth Low Energy (BLE) beacons [37, 45, 69], radio frequency identification (RFID) tags [5, 29, 61], and visual features [48, 79]) on static route maps allow turn-by-turn navigation to destinations [4, 67, 69, 83].

While turn-by-turn navigation effectively guides users, recent research emphasizes the importance of exploration in unfamiliar environments [12, 38], which is crucial for spatial learning [20, 22], cognitive development and brain health [25, 46, 53], independence [24, 32], and decision-making agency [19, 21]. Facilitating exploration involves providing surrounding information [38, 74]. Current systems, encompassing portable devices (*e.g.*, smartphones [18, 45, 66, 69] and wearables [37, 50]) as well as navigation robots (*e.g.*, suitcase-shaped [34, 42, 43], wheeled [51, 81], and quadruped robots [76, 77, 81]) offer basic POI information through accessible user interfaces like VoiceOver, which permits basic autonomy [43, 69].

However, fixed information often contains undesired information [15, 69] and thus hinders active learning and environmental comprehension [26, 38]. Supporting active exploration, especially in large indoor areas like shopping malls [12, 31], remains challenging. Jain *et al.* [38] proposed general design implications for potential navigation systems to enable exploration, including fostering active engagement, supporting individual variations, and facilitating in-situ exploration. Vincenzi *et al.* [74] further illustrated the need for collaborative technologies. PVI prefer controllable "push" and "pull" interactions [12], receiving feedback on-demand [38], filtering information [33], and having conversational interaction to improve their exploration experience [40, 74].

2.2 Conversational Interaction for PVI

While inclusive designs like VoiceOver on accessible user interfaces [4, 18, 45, 69] have enabled PVI to receive information, conversational systems are gaining attention for their engaging [23, 31] and hands-free interaction, allowing users to interact on the go [2, 11, 13, 55, 64, 69]. Simple conversational interaction has been used in navigation assistance to specify destinations [13, 55, 64, 69], access traffic information [55], and perceive surrounding objects [64] through voice input and predetermined commands. For instance, NavCog3 [69] used Watson Assistant API [8] to recognize specific words corresponding to stores (*e.g.*, *"coffee"*).

However, simple voice input and output may not suffice for flexible conversational interaction in exploration scenarios. The results of previous investigations highlight the need for human-like questions and answering (Q&A) models in contextual scene conversations [14, 31], which is still lacking in current voice-based assistants such as Siri or Alexa [23]. In web browsing, Pucci *et al.* [62] argued that conversational interfaces should support both exploration and direct queries. Abdolrahmani *et al.* [1] emphasized that conversations should offer transactional features to support practical goals, such as planning navigation among POIs [23]. These interactions should also provide flexible, personalized experiences while addressing privacy concerns and customizing information formats to individual needs [1, 11, 31, 62].

Nevertheless, these design implications stem from stakeholder discussions based on past experiences. It remains unclear how such systems perform in real-world scenarios. We present the design and implementation of a conversational interface in a shopping mall, investigating its advantages and future directions in providing assistance to explore information-rich, large indoor spaces.

2.3 Large Language Models

LLMs [82] generate text based on training with vast amounts of text data. They respond naturally to user inputs, called prompts, and leverage their knowledge from the training data. Expressing the desired output within the prompt can enhance performance on new tasks, known as few-shot learning [16, 35]. Research shows that LLMs excel in downstream tasks like text summarization [80] and text classification [71]. Their use in interactive systems has been explored for scientific writing [9] and controlling graphical user interfaces via conversational input [75]. However, LLMs' ability to access and manipulate knowledge is limited. Thus, retrieval-augmented generation (RAG) methods, which collaborate with external memory (*e.g.*, Wikipedia), have been developed to access more recent information [49].

LLMs show their advantages in versatility and generalization [75], which can be used to achieve the design considerations of the conversational interaction described in Sec. 2.2. First, LLMs can easily achieve human-level natural conversations in various tasks [1, 23]. Second, unlike traditional agents, which struggle with context and inferring from previous interactions [52], LLMs can provide contextual conversations through prompt engineering [16, 68]. Third, traditional conversational agents need users to switch the ways of expressing questions [52], but LLMs can comprehend various languages and expressions, including slang [72]. Moreover, LLM-generated responses can be tailored to individual user preferences [27].

However, the feasibility of using LLM-powered conversational interaction for navigation, specifically wayfinding and exploration, has not been investigated. We explore how such a system can enhance exploration by PVI compared to existing navigation systems.

3 System Design of ChitChatGuide

We designed ChitChatGuide to facilitate two tasks in exploratory navigation: tour planning and perceiving POI descriptions. The purpose of the system is to assist users in shopping mall walkways in exploring the facility by enabling them to grasp the existence and characteristics of stores of their interests. We define situations where PVI lack a specific cue (such as a store name) to narrow down their destination in a shopping mall as shopping mall exploration. We do not concentrate on scenarios where they have clear destinations in mind before commencing exploration because they can utilize existing navigation systems to reach those specific destinations (Sec. 2.1). We anticipate a setting where users are aided by assistants such as autonomous navigation robots for PVI, which guide them to their destination. We designed ChitChatGuide based on investigative works that proposed implications for designing conversational interaction to enhance exploration as follows.

3.1 Tour Planning

Allowing PVI to independently select destinations at their own pace, according to their understanding and preferences (*i.e.*, autonomy), is important for exploratory navigation for PVI [43]. Prior investigative works in exploration highlighted needs for enabling PVI's exploration experience through versatile conversational interaction and support varied needs when acquiring information [1, 31, 38, 40]. In this task, we utilize LLM-based conversational interaction to allow users to set a tour to navigate, that meets their interest through iterative conversations. To do so, we provide the system with a static route map, a database used for navigation, and the localization result of users. The system first provides users with a general overview of the facility, serving as a catalyst for them to ask questions and identify stores that pique their interest [12, 70]. Utilizing the POI information in the map and the location of users, the system dynamically summarizes the surroundings and the building (*e.g.*, *"Our shopping mall offers a wide variety of stores. The facility has various stores selling daily necessities, food, bread, and jewelry on the first floor. On the second floor, there are..."*). This is

achieved through LLMs' robust capability to comprehend and summarize given prompt data [80]. Users can interact with the system through iterative conversations using free-form speech input, which is more flexible than the system based on fixed conversational scenarios [11, 13, 55, 64, 69]. To facilitate users in easily deciding on a tour, the system guides the conversation by responding to vague questions or requests from users and suggesting several stores to narrow down through multiple rallies. This aligns with the expectations of PVI regarding conversational interaction [52]. Then, users can select tours to navigate from one of the two types of routes:
(1) Single Destination Route Users could select a store in the facility that meets their interest during the conversation with the system. Once the store of interest is set through the conversation, the system calculates the shortest route.
(2) Predefined Route Users who have no knowledge of the environment and want to explore the facility could request the system to navigate based on the predefined route. Users could select a route that navigates all around the floor, enabling the system to describe all stores on that floor. Since providing information about surrounding POI while walking helps PVI locate stores of interest [1], we anticipated that users would be able to find a store of interest by following a predefined route and listening to POI descriptions of stores along the way. The route for each floor is determined manually prior to the navigation.

3.2 Perceiving POI Descriptions

In exploration settings, grasping the POIs in the surrounding environment is indeed the most important task. The limitation of previous systems was that they only provided fixed descriptions of POIs (Sec. 2.1). In a content-rich environment such as a shopping mall, PVI often desire continuous access to detailed information without the need to ask specific questions. Fixed descriptions, such as the name, direction, and type of surrounding facilities, do not suffice in meeting this need [12]. Thus, we designed the system to adjust the amount of information that an LLM generates based on the estimated time users travel on the route. Also, as preferred information may vary by individual [1, 12, 33, 69, 70], we designed the system to personalize the provided information. Users could request to read out specific types of POIs or to read out specific information for each POI. For example, if users say, *"I do not need a description of the existence of elevator,"* the system removes elevators from the next POI description. Another example is *"Always include an explanation of recommended products."* As a result, the LLM generates descriptions of each POI whose length and contents are both adjusted. This would allow the system to customize the information provided to users, depending on their preferences or situation.

It is also important to make information that was not described in the POI descriptions accessible for users [12, 70]. Thus, we designed the system to allow users to ask questions about the facilities of interest at any time. When users want to ask questions to gain more knowledge of the facility, users can stop and start questioning the system (*e.g.*, *"What are business hours of the ABC cafe?"*). The system answers the question by referring to the database of the facility.

4 Implementation

Fig. 2 illustrates the components of our system, ChitChatGuide. We implemented the system on a smartphone (iPhone 11 Pro, Fig. 2–C). This app uses the framework called Human-scale Localization Platform (HULOP) [36]. We note that even though it is developed as an iPhone app, it is possible to integrate ChitChatGuide into other navigation devices. HULOP uses BLE beacons (Fig. 2–A) for localization [57] and a map server (Fig. 2–D) for planning routes and providing store information. The LLM used for the implementation is GPT-4 model [58] which is provided by OpenAI[1] (Fig. 2–E).

[1]Fig. 2– 4 use OpenAI's logo, following the guidelines at https://openai.com/brand/. The logo belongs to OpenAI.

Fig. 2. The components of ChitChatGuide. A) BLE beacons are implemented at the shopping mall for localization. B) and C) Users manipulate our system installed on an iPhone 11 Pro. D) Map Server contains the information of the mall and plans a route. E) For each functionality, the system communicates with the GPT-4 model on an LLM server.

We customized LLMs through prompting. Examples of prompts and store information provided to the LLM are shown in our supplementary materials (Sec. A). While all instructions via prompting to LLM were done in English, as the research was conducted in a non-English country, we instructed the LLM to respond in the native language.

We used a double tap with two fingers for initiating/ending speech input to the system (Fig. 2–B). This gesture is used by VoiceOver, a screen reading software on iOS, to initiate/end speech input, preventing users from triggering the conversation by accidentally tapping the screen.

4.1 Map Server and Localization

We prepared a map server to provide the system with map information (Fig. 2–D). The server contains the locations and topological connections of POIs. Each store has an *identifier index (ID) number* and a *long description* presented as a list, which includes the descriptive store's concept, several recommended items, business hours, and other basic information (*e.g.*, phone number and the number of seats). Additionally, there's a *short description* formatted as a list, briefly highlighting a store's concise concept, a single recommended item, and its category. The information on each POI was mainly sourced from the official website of the experimental facility. Experimenters also added information they could obtain in the field, such as temporary recommended items. To ensure correctness, in this study, we prepared the information manually without utilizing RAG. We used BLE beacons placed in the building for localization in the map (Fig. 2–A) [57]. The system calculates users' speed at 0.1-second intervals using localization results and the distance traveled over the past 10 seconds. If the speed exceeds (or falls below) 0.25 m/s for 10 consecutive intervals, the system determines that users are walking (or standing still).

4.2 Planning a Tour

We show the overview of planning a tour in Fig. 3. When the system is initiated, the system describes the facility (Fig. 3–A). We realize this by providing the LLM with a list of names, a short description, and the floor of each POI from the map server as a prompt, and giving it an instruction to summarize the content by α words (a prompt example: Sec. A.1). In this study, $\alpha = 300$.

The system allows users to select a destination or a predefined route, through conversational interaction (Fig. 3–B). To do so, we provide the LLM with the same information used for summarization, ID numbers of each POI, the floor where users are located, and names of three stores located near users. We also provide predefined route information in the format of a list of names, ID numbers, and a short description of each route. We prepared four predefined routes, each navigating one entire floor. A prompt example is shown in Sec. A.2.

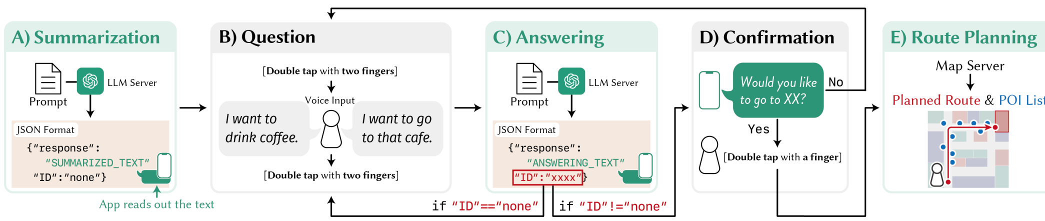

Fig. 3. The overview of the process of planning a tour. A) The system first reads out a summarization description of the facility. B) Users tell what they want to do to select their tours. C) The LLM generates the response in JSON format. D) If the "ID" key is not "none," the system confirms whether to set it as a tour. E) The system plans a route by using the map server.

We instruct the LLM to give a response in JSON format, which contains two keys, "response" and "ID" (Fig. 3–C, Sec. A.1, and Sec. A.2). The value in the "response" key contains the LLM's response to user input, such as a store recommendation or confirmation of users' desired action, to be read by the system using text-to-speech (TTS). The "ID" key contains the ID number of the desired POI or predefined route. Initially, its value is set to "none" until users select a tour. When users express interest in a specific POI or predefined route, such as saying *"I want to go to that store,"* or *"I want to walk a tour on the first floor,"* the LLM provides the ID number of the chosen store or predefined route (Fig. 5). Once users specify the tour and the LLM returns a value in the "ID" key, the system asks for confirmation from users by saying, *"Would you like to go to XX?"* (Fig. 3–D). Consequently, users can double tap the screen with a single finger to confirm the tour. Finally, the system plans a route to the destination or to the start point of the predefined route in the map server (Fig. 3–E).

4.3 Generating POI Descriptions

4.3.1 Classification of User Intent. After confirming the tour, the system always asks users if they have any preference for how the description is generated (Fig. 4). Users could make a speech input if they desire to make a request (Fig. 4–A) or skip this phase if they do not have any requests. The spoken content can be either a request to read/filter out specific POIs (*e.g.*, *"Only explain restaurants,"* or *"I do not need a description of toilets."*) or a request for how each POI is described (*e.g.*, *"Tell me a recommended item for each store."*). We determine the type of request users make by using the LLM as a classifier (Fig. 4–B and Sec. A.3). We prompt the LLM to classify spoken content by users based on POI information including the name, the floor location, and the short description. The LLM outputs the result of the classification in JSON format, which contains a "type" key. The value of this key is "filtering" in case of a request to read/filter out specific POIs, or "content" in case of a request for how each POI is described. If the key is "filtering" ("content"), the spoken content is appended to "Filtering (Content) List", which is used in Sec. 4.3.2 (Sec. 4.3.3). To make the LLM robustly classify, we use the few-shot prompting method. If the spoken preference is requested for how each POI should be described, the spoken content is used for the generation of a POI description, which is described in the following Sec. 4.3.3. The abovementioned algorithm can also be employed when users want to interact with the system while navigating. Users could stop at any time to initiate speech input and ask questions about POIs or make additional requests based on previous descriptions. In such case, we instruct the LLM to classify three intents: a question about a specific POI , a request to read/filter out specific POIs, and a request for how each POI is described (a prompt example: Sec. A.6). Based on the classification result, the system answers the question (Sec. 4.3.4), refines the list of POIs (Sec. 4.3.2), or generates updated POI descriptions (Sec. 4.3.3).

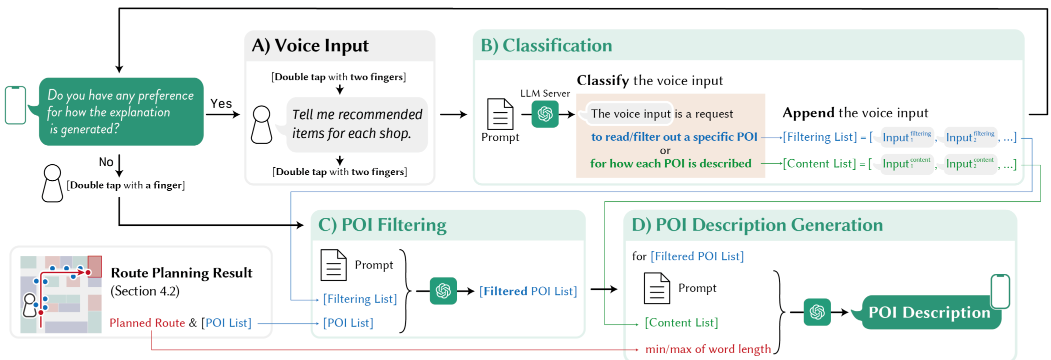

Fig. 4. The overview of the process of generating POI descriptions. A) Users can make a speech input if they have any preference for POI descriptions. B) The system classifies the spoken text according to the type of request. C) The system determines whether each POI should be explained. D) The system generates personalized POI descriptions, with their lengths adjusted based on transit time.

4.3.2 POI Filtering Functionality. If the spoken preference by users is a request to read/filter out a specific POI, the system determines which POI should be described (Fig. 4–C). We utilize the LLM to identify POIs along the route that align with user requests. Specifically, we provide the LLM with the name and a short description without its category of each POI, with user requests ("Filtering List" described in Sec. 4.3.1), and instruct the LLM to determine whether each POI should be described (a prompt example: Sec. A.4). If the LLM classifies the POI as not to be described, the system excludes the POI from being described along the route.

4.3.3 POI Description Generation. To generate descriptions, the system considers two factors: user preferences and the duration of the spoken description (Fig. 4–D). To consider the first factor, we provide the LLM with the name, direction (*i.e.*, left or right from users), and long description of each store along with the preference of users ("Content List" described in Sec. 4.3.1). For the second factor, we specify the number of words the LLM generates. Let us denote the speed of the TTS engine as S_{TTS} [words/second], the distance between each store on the planned route as d [meters], and users' walking speed as V [meters/second]. To calculate the length L of the spoken description, we use the equation, $L = \frac{S_{TTS}(\beta d)}{V}$, β is a constant representing the percentage of the distance the system speaks while walking between each POI. S_{TTS} is a number of words per second converted from the speed of TTS. The speed can be adjusted by users according to their preference. We set constant $\beta = 0.7$, to provide some margin between each description of POI. When providing the LLM with a number of words, we instruct the LLM to generate a description of POI within $0.8L$ words to $1.2L$ words. We chose to specify the word length as an interval as this approach yielded stable results, rather than using a single, fixed word count. As a result, the LLM generates a description for each store that takes into account transit time and personal preferences (a prompt example: Sec. A.5). Of the POIs on the route, the description of the first POI is read out when users begin walking (Sec. 4.1), and the descriptions of subsequent POIs are read out each time users pass a previous POI.

4.3.4 Questions and Answering During Navigation. While navigating, users could stop and ask questions about POIs being read out. To do so, we provide the LLM with each POI's name and short description in the order they are read out, along with the name and long description currently being read out. Users can inquire about the system, and the LLM returns the answer referring to

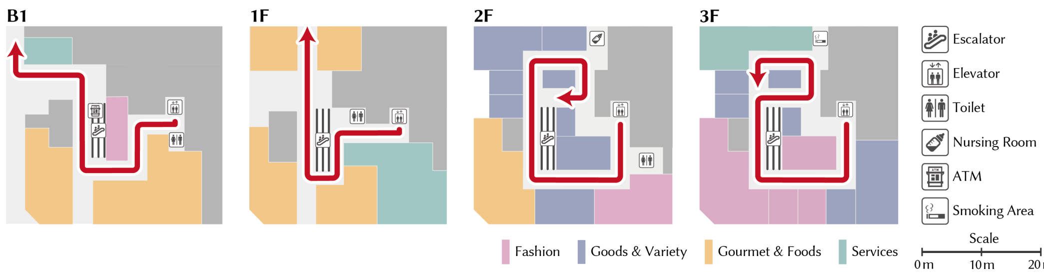

Fig. 5. Floor map of shopping mall and predefined routes

the given prompt (a prompt example: Sec. A.7). The system can answer questions such as *"What is the recommended item in the store you just described?"* since we feed the LLM with information about which POI was being read.

4.4 Real-Time Voice Feedback Using LLMs

We developed an algorithm that offers real-time voice feedback from the response of the LLM, which typically doesn't instantly produce complete responses. To achieve this, we utilize OpenAI's API that makes LLMs return tokens, a unit of character LLMs use, as soon as they are generated. Our system accumulates chunks of response as a partial response. Once the system receives punctuation marks, like commas or periods, it reads aloud the segment of text from the last point it read out (initially set at the start of the sentence) up to the punctuation mark. While the TTS engine is vocalizing this segment, the system continues to receive chunks from the LLM, allowing the next segment to be read aloud immediately after the current one is completed, enabling real-time voice output of the LLM responses. Note that vocalizing each chunk with the TTS engine will collapse the pronunciation of each individual word. We also implemented an algorithm to realize the abovementioned algorithm even though the response was in JSON format.

5 User Study

To investigate the impact of ChitChatGuide and expand the understanding of using LLMs for PVI, we performed an in-the-wild study in a public shopping mall with 11 visually impaired participants who are legally blind (three females and eight males, 46.9 years old on average in Tab. 1). As shown in Fig. 5, the user study was conducted at a shopping mall building (COREDO Muromachi3[2]), which has four floors (B1–3F) with 29 stores. This user study was approved by the university's IRB, and informed consent was obtained from every participant.

5.1 Baseline System

To understand the differences between the proposed system and existing systems, we used Inclusive Navi [3] as a baseline, which is a navigation system for PVI publicly available in the experimental location. Users can set a destination in Inclusive Navi by selecting a store name from a list using VoiceOver, which is the primary method of destination selection for users. Due to the data handling constraints at the facility, we were unable to use the functionality of selecting a destination with voice input. When a destination is selected, the system plans a route to the destination. While navigating, the system provides directions and store names (*e.g.*, *"On your right, there is ABC cafe."*) as they pass by, with voice feedback. Inclusive Navi is designed particularly for navigating users to specified destinations, so navigation instructions (*e.g.*, *"Go straight."*) are also provided. The

[2]https://mitsui-shopping-park.com/urban/coredo-muromachi/e/index.html

instructions were excluded from Inclusive Navi to ensure that the type of feedback aligns with that provided by the proposed system while navigating.

5.2 Task and Procedure

We first explained the purpose of the study and conducted a 30-minute interview asking about their demographic information, including the number of visits to the shopping mall and whether they had used Inclusive Navi (Tab. 1), and their experience in exploring unfamiliar facilities (Sec. 6.4.1). We then conducted a 30-minute training session on the baseline system and the proposed system in a space prepared for training. During the training, they adjusted the speed of the screen reader of the baseline system, and the speed of TTS of the proposed system resulting in S_{TTS}. Then, all participants were asked to use the baseline system and then the proposed system to conduct the main task. As the main task, they were asked to freely navigate and explore the facility using systems for a specified time (10 minutes and 30 minutes for the baseline system and the proposed system, respectively). Until the specified time elapsed, participants repeated selecting a tour using conversational interaction for the proposed system or VoiceOver for the baseline system and walking to the end of the tour while listening to POIs in the surrounding environment. Due to constraints on facility usage, each participant was allocated 40 minutes to conduct the main task. To obtain as many opinions and findings about the proposed system as possible, we designed the task so that participants use the proposed system longer than the baseline system. As the functionalities of the baseline system are simple, we considered that ten minutes of usage would be sufficient for participants to understand the system's functionality and compare it with the proposed system. The baseline condition was conducted first to ensure both conditions were completed in a limited time. Due to a scheduling issue, P11 was given 7 minutes for the baseline system and 15 minutes for the proposed system. Throughout the study, an experimenter guided participants along a route planned by the system (the walking speed was assumed as $V = 0.7$ meters/second). After completing the main tasks, participants took a post-questionnaire. We asked the participants to answer a set of questions using a seven-point Likert scale (1: strongly disagree, 4: neutral, 7: strongly agree), which is reported in Fig. 7. We also asked open-ended questions about reasons for their rating of the questions, the advantages and issues of our system for each functionality, their strategies for exploring the shopping mall, and suggestions for improvement (Sec. 6.4.2–Sec. 6.4.5). The whole study was recorded with a video camera and audio recorder. All participants were compensated with $25 per hour.

6 Result

6.1 System Performance of ChitChatGuide

6.1.1 Planning the Tour. We report the accuracy of the system's response to participant questions during the tour planning phase. Participants asked 143 questions, with the system providing 106 correct responses. Fourteen incorrect responses were due to speech recognition errors, which participants identified and asked again. Among other questions, 30 were general exploration questions (*e.g.*, *"What is on this floor?"* or *"What is around me?"*) with 93.3% correct responses, 51 were category-specific (*e.g.*, *"I want to go to the second floor's cafe,"* or *"Any places to eat?"*) with 72.5% correct responses, and 48 questions were specific inquiries (*e.g.*, *"Tell me about the bank,"* or *"The Italian one you mentioned."*) with 85.4% correct responses.

In 54 planned tours, 74.1% involved asking more than one question. Among them, participants asked 2.90 questions per planning session on average (SD = 1.43). Participants typically began with vague or general questions (*e.g.*, P09: *"Any recommendations on the third floor?"*). The system responded with a broad overview and invited more specific queries. Then, participants established

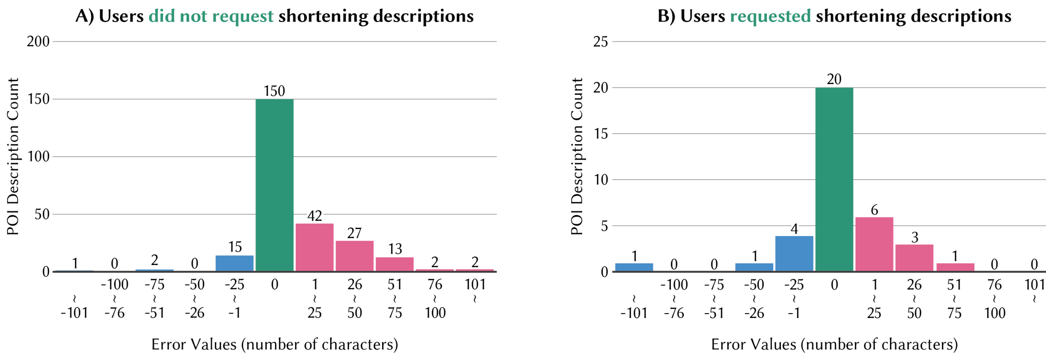

Fig. 6. The distributions of the length differences between generated descriptions and system-specified values (error values), measured in 25-character intervals. Characters are in their original non-English language.

a category (*e.g.*, P09: *"Any women's clothing stores?"*) and received a list of available stores. Next, participants asked for specific guidance (*e.g.*, P09: *"Take me to the third one."*). See the conversation example in Sec. B. During these conversations, user input was often contextual and vague. The LLM could identify flexible references (*e.g.*, *"Italian one"* or *"others"*) to maintain natural dialogue.

Of the 23 incorrect responses, 17 were mistakes providing false information. Nine were misidentifying the category of stores, five were offering incorrect floor information, and three were hallucinations introducing something that did not exist (*e.g.*, items or predefined route tours). Participants identified three incorrect category responses. Six errors occurred when users referenced specific orders from previous conversations (*e.g.*, *"the second one"*), with the system selecting the wrong ID number. Participants noticed five of these errors.

6.1.2 Generating Descriptions During the Tour. **(1) Filtering POIs.** During the study, three participants (P04, P07, and P08) used the POI filtering functionality a total of six times, and the system was able to handle the request without failure. The requirements included excluding banks (once), banks and toilets (once), and banks, toilets, and information centers (once), as well as including stores selling alcohol (once) and restaurants only (twice).

(2) Personalizing Contents. We report whether each generated POI description satisfied participants' requirements regarding content personalization. Across all participants, 23 requests were asked for personalization, with six requests involving two specific requests. Without manually dividing mixed requests in the "Content List" (Sec. 4.3.3), each specific request was distinguished and understood by the LLM. In total, 29 specific requests were made, affecting 142 POI descriptions. Requests were classified into six categories: (1) excluding contact address (25 descriptions, 100% correct), (2) excluding business hours (93 descriptions, 91.4% correct), (3) including recommended items (7 descriptions, 71.4% correct), (4) shortening description length (36 descriptions, 16.7% correct), (5) including budget (4 descriptions, 0% correct), and (6) including price of recommended items (5 descriptions, 0% correct). Some generated POI descriptions overlapped across categories. Overall, 66.9% of descriptions fully met the requirements. Examples of generated descriptions are in Sec. C. Besides two POI descriptions that were due to network errors, 45 failures fell into three categories: (1) Inclusion errors: The system included information users wanted to exclude, primarily business hours. (2) Exclusion errors: The system omitted the requested information. This included ignoring specific requests (2 recommended items, 1 budget) or lacking information in the database (3 budget, 5 price). (3) Length control failures: The system ignored requests to shorten descriptions, detailed in the "Controlling Length" part. All participants appeared to be aware of each failure.

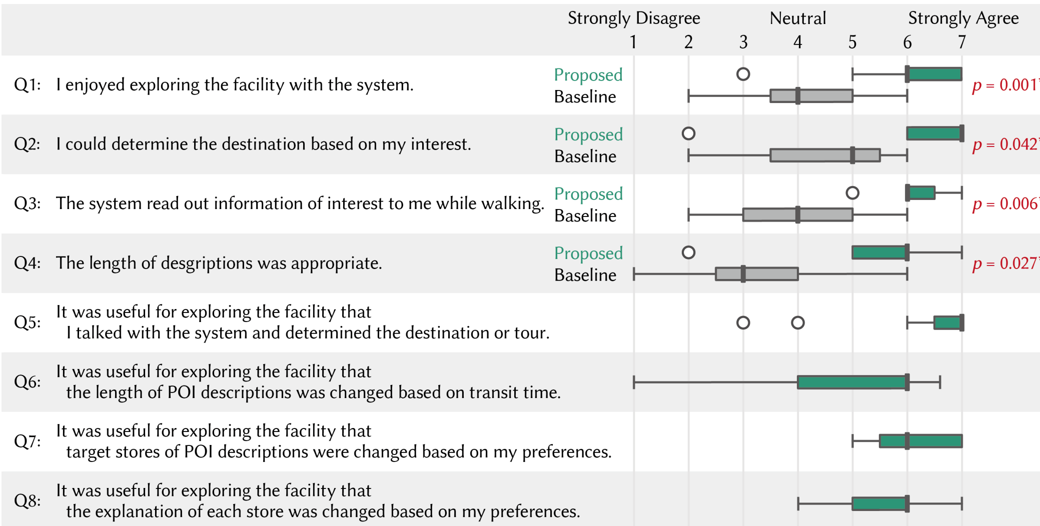

Fig. 7. Summary of Likert scale responses. p is the value of the Wilcoxon signed-rank test done on each question (* and ** indicate the significance found at levels of 0.05 and 0.01, respectively).

(3) Controlling Length. We evaluate whether the system was able to generate descriptions of POIs within a specified length (Sec. 4.3.3) for all 290 POI descriptions. As three participants (P03, P06, and P08) made a request regarding the length of descriptions, affecting 36 POI descriptions, we report the performance separately for those with and without length requests. Fig. 6 displays the deviation of the generated description length from the specified length with an interval of 25 characters. Fig. 6–A displays the distribution when no length request was made. While 59.1% of the generated descriptions were within the specified length, a tendency to generate longer descriptions was observed. When users requested shorter descriptions, the system dismissed many of these requests (described in the "Personalizing Contents" part). Fig. 6–B shows the distribution when such a request was made. Most of the time, the system generated descriptions within an internally specified length, ignoring user requests. This failure was possibly due to a conflict between internal transit-time-based length and user-specified length requirements in the same prompt provided as "Content List."

6.1.3 Further Questions and Answering During the Tour. We report the accuracy of the system's feedback to questions asked by participants while navigating. Participants asked 20 questions, with 55% correct answers. Four responses were not generated due to system or network errors. Out of the remaining 16 questions, 11 were about specific store details (recommended items and characteristics: 100% correct, budget: 50% correct, inexpensive items: 0% correct), three were about overviews of previously passed stores (100% correct), and two were about nearby POI information (0% correct). Among the errors, four (POI information near users and inexpensive items) were because the LLM lacked the necessary knowledge to provide an answer, and one (the budget of stores) was due to a mistake in the LLM.

6.2 Questionnaire

Fig. 7 shows the questionnaire results. Statistical analysis using the Wilcoxon signed-rank test revealed that ChitChatGuide was rated significantly higher than the baseline system for Q1–4.

Table 1. Demographic information and participants' behaviors when using the baseline system. The total number of tours set by participants ($\#_{\text{Tours}}$), the total time spent in setting tours (T_{ST}) and reading out POI while walking (T_{POI}), and the ratio of reading out POI description duration to walking duration (Ratio).

Demographic Information						Baseline System			
ID	Age	Age of onset	Gender	# of visits	Experience of Inclusive Navi	$\#_{\text{Tours}}$	Interactions while setting tours T_{ST} (sec)	POI description while walking T_{POI} (sec)	Ratio (percent)
P01	23	0	Male	0	No	3	84	9.1	5.2
P02	37	14	Female	Above 10	Yes	3	24	15.5	4.4
P03	31	10	Male	3 or 4	Yes	3	59	14.1	6.2
P04	36	1	Male	1	Yes	3	114	10.3	4.1
P05	28	14	Male	1	Yes	3	82	8.5	4.8
P06	59	54	Female	1	Yes	3	65	34.1	10.6
P07	58	56	Male	Few	No	4	74	12.4	6.0
P08	57	7	Male	0	No	4	133	20.4	9.7
P09	63	50	Female	Few	Yes	3	105	20.5	8.8
P10	71	63	Male	Few	No	3	74	37.1	11.9
P11	53	8	Male	3	No	3	100	14.7	8.8

Table 2. Participants' behaviors when using ChitChatGuide. The total number of tours set by participants ($\#_{\text{Tours}}$), in the format of a single destination route (S)/predefined route (P). The total time spent in and number of Q&A ($T_{\text{Q\&A}}$ and $\#_{\text{Q\&A}}$), filtering request (T_{FR} and $\#_{\text{FR}}$), content request (T_{CR} and $\#_{\text{CR}}$), and all interactions (T_{Total} and $\#_{\text{Total}}$), in a format of while setting tours/while navigating. The total time of reading out POI while walking (T_{POI}), and the ratio of reading out POI description duration to walking duration (Ratio).

ID	$\#_{\text{Tours}}$	Conversations while setting tours / navigating								POI description while walking	
		Q&A		Filtering Request		Content Request		Total			
	S / P	$T_{\text{Q\&A}}$ (sec)	$\#_{\text{Q\&A}}$	T_{FR} (sec)	$\#_{\text{FR}}$	T_{CR} (sec)	$\#_{\text{CR}}$	T_{Total} (sec)	$\#_{\text{Total}}$	T_{POI} (sec)	Ratio (percent)
P01	3 / 3	730 / 0	19 / 0	0 / 0	0 / 0	45 / 0	4 / 0	775 / 0	23 / 0	300	49
P02	3 / 1	351 / 38	9 / 2	0 / 0	0 / 0	8 / 0	1 / 0	359 / 38	10 / 2	258	52
P03	4 / 3	508 / 14	15 / 1	0 / 0	0 / 0	34 / 0	2 / 0	542 / 14	17 / 1	332	47
P04	2 / 2	223 / 31	9 / 2	18 / 0	2 / 0	11 / 0	1 / 0	252 / 31	12 / 2	219	40
P05	4 / 1	542 / 86	17 / 3	0 / 0	0 / 0	0 / 0	0 / 0	542 / 86	17 / 3	230	55
P06	3 / 3	457 / 31	10 / 2	0 / 0	0 / 0	52 / 32	4 / 3	509 / 63	14 / 5	475	52
P07	2 / 2	668 / 149	19 / 6	16 / 0	1 / 0	24 / 0	1 / 0	708 / 149	21 / 6	220	33
P08	3 / 2	390 / 0	9 / 0	33 / 0	3 / 0	23 / 6	2 / 1	446 / 6	14 / 1	310	56
P09	3 / 1	596 / 113	14 / 3	0 / 0	0 / 0	67 / 0	3 / 0	663 / 113	17 / 3	248	63
P10	7 / 0	726 / 0	16 / 0	0 / 0	0 / 0	0 / 0	0 / 0	726 / 0	16 / 0	335	60
P11	1 / 1	159 / 28	6 / 1	0 / 0	0 / 0	9 / 0	1 / 0	168 / 28	7 / 1	129	40

6.3 Behavior While Using ChitChatGuide and the Baseline System

Tab. 1 and Tab. 2 show the number of tours set by participants, the time taken to set tours and read out POI, and the ratio of the system's voice feedback while walking when using the baseline system and ChitChatGuide. Tab. 2 also provides details of tours and conversations with the proposed system. On average, participants took 99.1 seconds (SD = 51.9) to plan one tour using the proposed system, compared to 26.1 seconds (SD = 18.3) with the baseline system. The proposed system offered voice feedback for an average of 49.7% (SD = 8.77) while walking in the navigation phase, while the baseline system did so for 7.3% (SD = 2.60) of the time on average.

6.4 Qualitative Feedback

6.4.1 Daily Experience of Exploration. Based on the interview conducted before the main task, to explore a facility, all participants stated that they rely on sighted companions, such as the facility's information desk clerks and guide helpers for PVI. They mentioned that they visit facilities with a specific purpose, such as purchasing a specific item, and to do so, eight participants mentioned that they conducted research on the internet about facilities beforehand. P04 described exploring a facility such as a shopping mall as follows: **C1**: (Comment number 1) *"I cannot see my surroundings and get information, so I don't enjoy it (exploration). While I can ask around about the location of stores, it is tiring and inefficient."* (P04) Finally, they all agreed that they would be glad to explore independently if possible.

6.4.2 Exploration Experience with the Baseline System. All participants felt that the baseline system lacked information in the overall experience as it mainly provided store names. On the other hand, some participants favored the simplicity of the functionality of the baseline system. For example, five participants (P02, P03, P07, P10, and P11) mentioned that they were able to select the destination from the list of store names based on their interests to some extent: **C2**: *"(The list) was good for quickly understanding the store in the facility because I didn't know much about them."* (P07)

6.4.3 Exploration Experience with ChitChatGuide. Overall, participants generally expressed their enjoyment when using ChitChatGuide, as it provided them with continuous and informative descriptions, and a conversation feature to ask for stores that interest them: **C3**: *"There was a fresh aspect for me, and it was the part that enabled me to search for certain information really easily. It provided various information on POIs when I came close, which made me comfortable. Through the conversation, the system told me a lot about what was around the destination, which helped those who had difficulty accessing information."* (P05) Also, seven participants (P03-P08 and P11) mentioned that the system enables a window-shopping-like experience, and increases their motivation to visit a new store: **C4**: *"It was exactly like window-shopping. The detailed descriptions are good for me as I can get information other than a destination. This descriptive information made me want to go to stores which I did not intend to visit before using the system."* (P06) In addition, six participants (P03-P07 and P11) mentioned that this system would allow them to explore the facility when they have time or visit the facility with no specific purpose: **C5**: *"I can get a general sense of this place and thus would use it when I have time. Usually, I visit with a purpose in mind, so I think it is useful if I have extra time after going to a specific destination. It's a kind of window-shopping technique to find out what kind of stores are around."* (P03) Moreover, three of them (P04, P05, and P11) felt the system would make it easier to visit unfamiliar places: **C6**: *"When I go around a certain floor by selecting a predefined route, I can enjoy and see what kind of stores are on the floor, like window-shopping. It is difficult for PVI to do something like window-shopping and we always decide a specific destination before going out. Thus, I think it is a lot of fun for PVI to be able to drop by a place and explore there with the system."* (P05)

6.4.4 Conversational Interaction for Inquiries with ChitChatGuide. All participants mentioned that the conversational interaction with ChitChatGuide enhanced their understanding of stores within the facility because it offered easy access to various information and allowed for questions for in-depth information: **C7**: *"I narrowed down the restaurant options based on my preferences and their characteristics. The system's detailed information improved my understanding of the facility compared to the baseline system."* (P11) and **C8**: *"For us VoiceOver users, searching on smartphones takes a lot of time and effort. This voice interaction is so easy and comfortable, allowing us to casually ask questions and get information from the AI."* (P05) Also, P04 and P07 expressed the convenience of the Q&A functionality during tours: **C9**: *"Usually, I have to enter a store and ask clerks for details. It (the Q&A functionality) would help me get information beforehand and decide whether to enter. Details like recommended items, cuisine, and prices would be very helpful,"* (P04) and **C10**: *"For bargain and sale items, getting a shop assistant to show you those is difficult. And sometimes it's hard to ask for cheaper options. The system naturally, and fairly, provides that information, which is probably something* *we often miss out on."* (P07) For tour planning, ten participants stated the proposed system better assisted them in selecting destinations in unfamiliar environments, even responding to ambiguous requests when their purpose was unclear: **C11**: *"I think the nice feature of (conversational interaction with) GPT is to handle vague instructions. For example, (the system) could understand 'The first one' after providing multiple recommendations. Google Assistant won't start to work unless you say the exact name and what to do."* (P01) and **C12**: *"I think the advantage of conversational interaction (with the system) is that I can easily start walking even without a specific purpose. This is well-suited for scenarios like exploration and learning about the surroundings."* (P03) On the other hand, three participants (P02, P06, and P09) sometimes felt the tour planning through conversational interaction was cumbersome, as it takes more time than setting a destination from a list: **C13**: *"I usually want to go through all the names of the store before determining my destination. For example, it is easier for me to go through the list of stores and pick my own destination,"* (P02) and **C14**: *"If there were an option to select a simplified mode when I have a clear idea of what I want to see, and a full conversational mode when I'm not sure, it would be less frustrating."* (P09) Meanwhile, two of them (P02 and P06) and P11 felt that the weakness of the interaction as follows: **C15**: *"I felt like it was doubtful that no store matched my question. I thought the responses were not reliable."* (P11)

6.4.5 Descriptions Provided by ChitChatGuide. Participants generally expressed enjoyment when using ChitChatGuide, particularly the descriptions that made stores more intriguing: **C16**: *"The system made it fun and easy to understand the store's concept and goods, whereas the baseline system only provided the store's name."* (P01) P03 and P07, who didn't prefer the proposed method, suggested as follows: **C17**: *"The system produces lengthy descriptions for stores, regardless of their size or product range. I prefer concise summaries with the option to request more details. I also recommend shorter narrations while users are walking, with automatic expansion when they stop out of interest,"* (P07) and **C18**: *"I would prefer all descriptions to be the same length and short. I want to ask questions if the short description interests me."* (P03)

Three participants (P04, P07, and P08) used the functionality of filtering specific POIs, as they wanted to know only what interested them: **C19**: *"(I made the request because) I was looking for a restaurant for drinking, so I wasn't interested in other stores."* (P07) On the other hand, this functionality was not used by the others because the facility was not large, and the purpose of exploration was to find a store of interest: **C20**: *"I didn't use the functionality since the facility wasn't very large, and I wanted to understand all the stores."* (P01) Nine participants (P01-P04, P06-P09, and P11) requested to personalize the POI descriptions. P03 described the advantages of the functionality as follows: **C21**: *"I think it would be better to have the functionality to customize the contents because many different users would want to use the system in various ways, not just myself."* (P03) Also, P06 gave the following comment on the advantage of LLMs in generating variant descriptions: **C22**: *"(It is good) because I don't want to listen to the same description when I pass by the same POI twice in a route."* (P06)

7 Discussion

7.1 Addressing RQ: How can we design conversational interaction to provide information flexibly to PVI who want to independently explore shopping malls?

We designed ChitChatGuide, an LLM-based conversational interface integrated into a navigation app, enabling tour planning and POI learning during navigation. Next, we discuss how it provides flexible information for PVI's exploration.

7.1.1 Experience When Using ChitChatGuide. ChitChatGuide allowed participants to enjoy exploration, resulting in an improved exploration experience (median for Fig. 7–Q1 was 6). The questions regarding the enhancement of autonomy received high ratings (Fig. 7–Q2 and Q5). The conversational interaction enabled users to find stores based on their interests even if they were unfamiliar with the facility (C12). Although the conversational tour setting took more time than the baseline system (Sec. 6.3), participants appreciated taking several rounds to narrow down the information to set their destination (C7). Meanwhile, three participants sometimes felt tour planning through conversational interaction cumbersome, expressing a preference for using a simple method such as a list of store names if they had a specific destination in mind (C13 and C14). Considering that one of the three participants was accustomed to the facility (Tab. 1), the adoption of conversational interaction seems to be mainly effective when users want to explore unfamiliar facilities.

The descriptive information that the system provided during navigation was another reason users appreciated the system (Fig. 7–Q3). Ten participants enjoyed receiving information they couldn't get from store names alone (C16). They also found details about pricing and sales useful, as they were often hesitant to ask store clerks (C10). In addition, they were also able to acquire additional information by asking questions, which appeared to be crucial in determining whether to enter a store (C9). In essence, our system provides a window-shopping-like experience for PVI, allowing them to leisurely explore and grasp the facility's details, regardless of their familiarity (C6).

7.1.2 Behavioral Change. The system could encourage PVI to visit facilities independently without a specific purpose. Currently, they travel to specific stores with a predetermined purpose and conduct web research beforehand (Sec. 6.4.1). However, comments indicated they would enjoy traveling without a specific purpose (C6), since the system allows window-shopping previously unavailable to them. A participant expressed interest in using spare time to explore even when visiting for a specific purpose (C5). Additionally, when using the system with a specific destination, participants commented that descriptive information about surrounding stores could motivate them to visit new stores (C4). In short, the results suggest the system can motivate PVI to change their behaviors, visit facilities without a specific purpose, and explore unknown stores. It also potentially encourages them to engage with and purchase at new stores.

7.1.3 LLM's Impact. For the conversational tour planning functionality, the natural and contextual conversation style powered by the LLM (Sec. 6.1.1) was preferred by ten participants (C7) and proved particularly effective for users without a specific destination who wanted to explore unfamiliar facilities (C12). The LLM generally provided conversational interaction by understanding vague requests that referred to previous conversations (C11), a common occurrence for those unfamiliar with a facility, since remembering the newly learned keywords could be cognitively demanding.

The LLM's ability to generate POI descriptions based on various personalization requests (Sec. 6.1.2–(1) and (2)) was another significant finding. By leveraging map information, user location, and user requests to include or exclude information in the prompt, it could output in-situ information during navigation. Participants appreciated the customized POI descriptions tailored to their preferences (C19 and C21). Moreover, the expressive and variant descriptions generated by the LLM could make navigation in a shopping mall enjoyable (C16 and C22). In addition to personalizing the description, the LLM can also easily incorporate environmental information and user requests in the prompt to output additional information (Sec. 6.1.3), which participants found to be convenient (C9 and C10).

Although the LLM responded successfully to most questions, it exhibited occasional mistakes in responses, difficulties in altering the output length, and failure to fulfill some personalization needs (Sec. 6.1). We elaborate on these in Sec. 7.2. In summary, the LLM's conversational and generative capabilities facilitated natural and effective exploration and navigation in unfamiliar facilities, leading to an enjoyable experience.

7.2 Criteria to Consider When Integrating LLMs into Navigation Systems for PVI

While results revealed that integrating an LLM into a navigation system can potentially enhance the exploration, three essential criteria were implied through the study, which should be considered during the design phase.

7.2.1 Balance of Attractiveness and Length of Descriptions. The first criterion to consider when implementing navigation systems for PVI using LLMs is the balance between attractiveness and length of descriptions. We designed conversational interaction for planning a tour and perceiving POI descriptions based on investigative works that proposed implications for designing conversational interaction to enhance exploration (Sec. 3). Participants preferred long and varying descriptions of our system to shorter ones of the baseline system (Fig. 7–Q4). On the other hand, the ratings for the question about controlling the description length in relation to its appropriateness differed among participants (Fig. 7–Q6). Some participants found value in longer descriptions, particularly when they were adjusted according to transit time, saying it provided in-depth information without needing to take any action (C3). Nonetheless, to better reflect individual interests, some participants preferred concise descriptions (C18) or descriptions varying in length based on their interests (C17). Also, the LLM failed to meet user requests for excluding certain information or shortening the length of descriptions (Sec. 6.1.2–(2) and (3)). This may have resulted in the descriptions becoming less attractive. One reason for the failures arose because the LLM tried to meet internal length requirements first. Therefore, to better meet users' expectations, future designs integrating LLMs should consider preferentially incorporating their interests instead of simply increasing the amount of information provided. For an exploration scenario, an alternative design could be an interactive design where the system would initially offer a short description for each store, and then provide more detailed information with a long description based on users' actions, such as interrupting to ask questions or stopping to indicate interest. This aligns with guidelines for providing information to explore text-based interfaces [62, 70]. Note that even for the short description, LLMs should still be used, as it could consider users' requirements and generate diverse descriptions (C21 and C22).

7.2.2 Trustworthiness of Response of Conversation Systems. The second criterion to consider is determining how we could ensure the system provides accurate and trustful responses. Tour planning through LLM-based conversations was generally appreciated (Fig. 7–Q5), as participants were able to ask questions in a variety of formats and granularities (*i.e.*, general exploration, category-specific, and specific questions in Sec. 6.1.1). On the other hand, three participants felt that the system's response could not be fully trusted (C15). This occurred when the system offered simple negative responses (*e.g.*, *"I'm sorry, but there is no coffee shop,"*) which contradicted the participant's common sense. Also, false information was occasionally observed in the LLM in conversational interaction, but participants were often unaware of it (Sec. 6.1.1). The difficulty of judging whether information is correct from responses might lead users to distrust the system's answers. Therefore, when integrating LLMs into assistance systems, it's essential to establish methods for ensuring their accuracy and trustworthiness. First, it is crucial to ensure that the data provided to the LLM is comprehensive and reliable for PVI exploring shopping malls (further discussed in Sec. 7.2.3). If stores that perfectly match user demand do not exist, trustworthiness can be improved by providing relevant recommendations on similar or alternative stores based on details included in the data (*e.g.*, *"There is no coffee shop at the facility, but the bakery offers drinks as a set menu so that coffee may be available."*). Second, it is necessary to reduce mistakes in LLM responses, as they can cause ethical problems, such as misdirecting users' conversations and decision-making, increasing bias introduced by using an LLM as a recommendation system, and harming stores by spreading misinformation and disinformation. Since PVI rely solely on audio information and cannot access visual cues, false information in conversational interaction can become a more important issue. Examples of methods to mitigate response mistakes such as hallucinations include using RAG to ground the model with an external vector database [59] and fine-tuning the model with additional data [39, 49]. In particular, RAG is a popular method and can be considered a natural future extension of the current system design, giving information as prompts. By only providing the LLM with information that is relevant to user input, this approach is expected to reduce most of the mistakes found in our study, that is regarding floor and category of stores (Sec. 6.1.1). Since it is impossible to achieve perfect accuracy in LLM responses, we need to make users aware of the potential risk of encountering inaccurate content. Additionally, implementing frameworks for fact-checking or detecting hallucinations and malicious content can be a possible safeguard [65]. Thus, LLM-based information provision systems must strive to provide accurate and convincing information to build user trust.

7.2.3 Depth of Data. The third criterion to consider is ensuring to have sufficient data depth to satisfy user demand. This criterion arose from the tendency observed in participants' questions while navigating. The Q&A functionality was only used 20 times while navigating, but participants felt the need for this functionality (C18). In this study, the sources of information were mainly the data from the facility's official website, and some were also collected in the field by the experimenter. Still, there were cases when the information requested by the participants was not included in the database as the amount of information was insufficient for users (Sec. 6.1.2–(2) and Sec. 6.1.3). This observation shows that the information users need is often difficult to obtain with systems designed for navigation. When designing a navigation system like the one used in this study, it may be necessary to establish a framework that stores could cooperate with developers of assistive technologies and reflect the descriptive information to the map database. As our system may motivate PVI to visit new stores by providing engaging information, such a cooperative framework could also benefit stores with increased purchases. The framework would further motivate purchases by making previously inaccessible or hesitant-to-ask information apparent, reducing barriers to buying sale items (C9 and C10). For example, by integrating the map server with an existing stock management system whose stock information updates in real-time, the system will be able to answer what is being offered in each store and their availability. Additionally, even if a sufficient amount and depth of data were available, a challenge remains in making LLMs understand such data effectively. While we customized LLMs using prompt engineering, the information provided to the LLM through prompts is limited by the model's token constraints. To address these issues, methods such as RAG and fine-tuning can improve the performance of LLMs for specific purposes (*e.g.*, PVI's shopping mall exploration) while reducing the amount of prompts [10, 49]. Furthermore, fine-tuning can enhance computational efficiency for sustainable use, addressing a potential challenge of LLMs [10].

7.3 Limitation and Future Work

The study was conducted on four floors of a building with 29 stores. As the experimental location was not large, this may have been why the functionality to filter out specific POIs was only used six times (Tab. 2), as also indicated by P01 (C20). Thus, we were unable to discuss the usability of this functionality. Also, six participants (P01, P03-P05, P08, and P09) tended to listen to the names of all stores on the list and then specify the destination when they used the baseline system (C2 and C13). Due to the limited size of the experimental location, the aforementioned strategy was feasible. However, in larger facilities, listening to all stores in the list with the baseline system would take more time. Therefore, the experiences with both the proposed system and the baseline system in larger facilities may differ from those observed in smaller ones. As there is a possibility that interface design tailored to the scale of the facility becomes important, for future work, we aim to conduct studies in facilities of various sizes and structures, as well as in places other than shopping malls, such as museums, airports, and theme parks.

This study, conducted at a public shopping mall during regular business hours, faced time constraints on facility usage due to agreements with the facility (Sec. 5.2). Within time constraints, the longer time spent using the proposed system may have introduced a bias, making it more memorable which could influence participant feedback. However, the baseline system took less time to set a tour than the proposed system (Sec. 6.3), allowing participants to experience several tours within the specified time. Furthermore, since the baseline system could only provide single destination routes with fixed POI descriptions, participants could fully understand its characteristics after a few tours. Thus, we consider the bias was minimized and the results and findings were reliable.

For voice input, participants could use it properly and found its advantage in easily accessing information (C8). In this study, we instructed participants to interact with the system using as few proper nouns as possible, but faced several errors due to the poor recognition accuracy of proper nouns (Sec. 6.1.1). Considering this issue and the fact that voice input is sometimes difficult to use (*e.g.*, in noisy places), the system should provide both voice input and accessible text input user interfaces in actual operation. Another issue was that the LLM struggled to generate appropriate sentences within a specified length, especially when the length was particularly long (Sec. 6.1.2–(3)). We attempted to enhance our system by increasing the width of the length specification, but it remained challenging to control without error, and better prompts for the length control should be explored.

8 Conclusion

In this study, we investigated how conversational interaction can be used to enable exploration that requires flexible information in shopping malls for PVI. To this end, we developed ChitChatGuide by integrating LLMs into navigation systems and conducted a user study with 11 visually impaired participants. Through the study, we obtained four main findings: the potential of ChitChatGuide to increase enjoyment and encourage behavioral change, the usefulness of LLM's natural and contextual conversations for tour planning, the effectiveness of POI descriptions personalized by LLMs, and three criteria to consider when integrating LLMs into navigation systems for PVI. The findings highlight directions to fully enable PVI's exploration in information-rich indoor places through LLM-based conversational interaction.

Acknowledgments

We would like to thank all participants who took part in our user study. We would also thank Xiyue Wang, Gary Vierheller, Kentaro Fukuda, Mitsui Fudosan Co., Ltd., and Miraikan - The National Museum of Emerging Science and Innovation for their support. This work was supported by JSPS KAKENHI (JP23KJ2048).

References

[1] Ali Abdolrahmani, Maya Howes Gupta, Mei-Lian Vader, Ravi Kuber, and Stacy Branham. 2021. Towards More Transactional Voice Assistants: Investigating the Potential for a Multimodal Voice-Activated Indoor Navigation Assistant for Blind and Sighted Travelers. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Article 495, 16 pages. https://doi.org/10.1145/3411764.3445638

[2] Ali Abdolrahmani, Ravi Kuber, and Amy Hurst. 2016. An empirical investigation of the situationally-induced impairments experienced by blind mobile device users. In *Proceedings of the 13th International Web for All Conference*. ACM, New York, NY, USA, 1–8. https://doi.org/10.1145/2899475.2899482
[3] ACCESSIBLEJAPAN. 2024. Nihonbashi, Inclusion and Technology. Retrieved in February, 2024 from https://www.accessible-japan.com/nihonbashi-inclusion-and-technology/.
[4] Dragan Ahmetovic, Cole Gleason, Chengxiong Ruan, Kris Kitani, Hironobu Takagi, and Chieko Asakawa. 2016. NavCog: a navigational cognitive assistant for the blind. In *Proceedings of the 18th International Conference on Human-Computer Interaction with Mobile Devices and Services*. ACM, New York, NY, USA, 90–99. https://doi.org/10.1145/2935334.2935361
[5] Saleh Alghamdi, Ron van Schyndel, and Ahmed Alahmadi. 2013. Indoor navigational aid using active RFID and QR-code for sighted and blind people. In *2013 IEEE Eighth International Conference on Intelligent Sensors, Sensor Networks and Information Processing*. IEEE, Piscataway, NJ, USA, 18–22. https://doi.org/10.1109/ISSNIP.2013.6529756
[6] Abdulrhman Alkhanifer and Stephanie Ludi. 2015. Disorientation Factors that Affect the Situation Awareness of the Visually Impaired Individuals in Unfamiliar Indoor Environments. In *Universal Access in Human-Computer Interaction. Access to the Human Environment and Culture*. Springer International Publishing, Cham, 89–100. https://doi.org/10.1007/978-3-319-20687-5_9
[7] Saki Asakawa, João Guerreiro, Dragan Ahmetovic, Kris M Kitani, and Chieko Asakawa. 2018. The present and future of museum accessibility for people with visual impairments. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 382–384. https://doi.org/10.1145/3234695.3240997
[8] Watson Assistant. 2024. IBM Watson. Retrieved in February, 2024 from https://www.ibm.com/watson.
[9] Amos Azaria, Rina Azoulay, and Shulamit Reches. 2023. ChatGPT is a Remarkable Tool – For Experts. arXiv:2306.03102 [cs.HC]
[10] Guangji Bai, Zheng Chai, Chen Ling, Shiyu Wang, Jiaying Lu, Nan Zhang, Tingwei Shi, Ziyang Yu, Mengdan Zhu, Yifei Zhang, Carl Yang, Yue Cheng, and Liang Zhao. 2024. Beyond Efficiency: A Systematic Survey of Resource-Efficient Large Language Models. arXiv:2401.00625 [cs.LG]
[11] Jan Balata, Zdenek Mikovec, and Pavel Slavik. 2018. *Conversational Agents for Physical World Navigation*. Springer International Publishing, Cham, 61–83. https://doi.org/10.1007/978-3-319-95579-7_4
[12] Nikola Banovic, Rachel L. Franz, Khai N. Truong, Jennifer Mankoff, and Anind K. Dey. 2013. Uncovering information needs for independent spatial learning for users who are visually impaired. In *Proceedings of the 15th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, Article 24, 8 pages. https://doi.org/10.1145/2513383.2513445
[13] Jakub Berka, Jan Balata, Catholijn M Jonker, Zdenek Mikovec, M Birna van Riemsdijk, and Myrthe L Tielman. 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 (2022), 1909–1925. https://doi.org/10.1080/10447318.2022.2059925
[14] Jeffrey P. Bigham, Chandrika Jayant, Hanjie Ji, Greg Little, Andrew Miller, Robert C. Miller, Robin Miller, Aubrey Tatarowicz, Brandyn White, Samual White, and Tom Yeh. 2010. VizWiz: nearly real-time answers to visual questions. In *Proceedings of the 23nd Annual ACM Symposium on User Interface Software and Technology*. ACM, New York, NY, USA, 333–342. https://doi.org/10.1145/1866029.1866080
[15] Jeffrey R Blum, Mathieu Bouchard, and Jeremy R Cooperstock. 2011. What's around me? Spatialized audio augmented reality for blind users with a smartphone. In *International Conference on Mobile and Ubiquitous Systems: Computing, Networking, and Services*. Springer Berlin Heidelberg, Berlin, Heidelberg, 49–62. https://doi.org/10.1007/978-3-642-30973-1_5
[16] Tom B. Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel M. Ziegler, Jeffrey Wu, Clemens Winter, Christopher Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language models are few-shot learners. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, 1877–1901.
[17] Kristina Bäckström. 2006. Understanding recreational shopping: A new approach. *The International Review of Retail, Distribution and Consumer Research* 16, 2 (2006), 143–158. https://doi.org/10.1080/09593960600572167
[18] Hsuan-Eng Chen, Yi-Ying Lin, Chien-Hsing Chen, and I-Fang Wang. 2015. BlindNavi: A navigation app for the visually impaired smartphone user. In *Proceedings of the 33rd annual ACM conference extended abstracts on human factors in computing systems*. ACM, New York, NY, USA, 19–24. https://doi.org/10.1145/2702613.2726953
[19] Elizabeth R Chrastil, Katherine R Sherrill, Michael E Hasselmo, and Chantal E Stern. 2015. There and back again: hippocampus and retrosplenial cortex track homing distance during human path integration. *Journal of Neuroscience* 35, 46 (2015), 15442–15452. https://doi.org/10.1523/JNEUROSCI.1209-15.2015

[20] Elizabeth R Chrastil and William H Warren. 2012. Active and passive contributions to spatial learning. *Psychonomic bulletin & review* 19 (2012), 1–23. https://doi.org/10.3758/s13423-011-0182-x

[21] Elizabeth R Chrastil and William H Warren. 2013. Active and passive spatial learning in human navigation: acquisition of survey knowledge. *Journal of experimental psychology: learning, memory, and cognition* 39, 5 (2013), 1520–1537. https://doi.org/10.1037/a0032382

[22] Elizabeth R Chrastil and William H Warren. 2015. Active and passive spatial learning in human navigation: acquisition of graph knowledge. *Journal of experimental psychology: learning, memory, and cognition* 41, 4 (2015), 1162. https://doi.org/10.1037/xlm0000082

[23] Leigh Clark, Nadia Pantidi, Orla Cooney, Philip Doyle, Diego Garaialde, Justin Edwards, Brendan Spillane, Emer Gilmartin, Christine Murad, Cosmin Munteanu, Vincent Wade, and Benjamin R. Cowan. 2019. What makes a good conversation? Challenges in designing truly conversational agents. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Article 475, 12 pages. https://doi.org/10.1145/3290605.3300705

[24] D. D. Clark-Carter, A. D. Heyes, and C. I. Howarth. 1986. The efficiency and walking speed of visually impaired people. *Ergonomics* 29, 6 (1986), 779–789. https://doi.org/10.1080/00140138608968314

[25] Gregory D. Clemenson, Caden M. Henningfield, and Craig E. L. Stark. 2019. Improving hippocampal memory through the experience of a rich Minecraft environment. *Frontiers in Behavioral Neuroscience* 13, Article 57 (2019), 13 pages. https://doi.org/10.3389/fnbeh.2019.00057

[26] Sanorita Dey, Karrie Karahalios, and Wai-Tat Fu. 2018. Getting there and beyond: Incidental learning of spatial knowledge with turn-by-turn directions and location updates in navigation interfaces. In *Proceedings of the 2018 ACM Symposium on Spatial User Interaction*. ACM, New York, NY, USA, 100–110. https://doi.org/10.1145/3267782.3267783

[27] Joel Eapen and Adhithyan V S. 2023. Personalization and Customization of LLM Responses. *International Journal of Research Publication and Reviews* 4, 12 (2023), 2617–2627. https://doi.org/10.55248/gengpi.4.1223.123512

[28] Christin Engel, Karin Müller, Angela Constantinescu, Claudia Loitsch, Vanessa Petrausch, Gerhard Weber, and Rainer Stiefelhagen. 2020. Travelling More Independently: A Requirements Analysis for Accessible Journeys to Unknown Buildings for People with Visual Impairments. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, Article 27, 11 pages. https://doi.org/10.1145/3373625.3417022

[29] Navid Fallah, Ilias Apostolopoulos, Kostas Bekris, and Eelke Folmer. 2012. The user as a sensor: navigating users with visual impairments in indoor spaces using tactile landmarks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 425–432. https://doi.org/10.1145/2207676.2207735

[30] Mohammadreza Farrokhnia, Seyyed Kazem Banihashem, Omid Noroozi, and Arjen Wals. 2024. A SWOT analysis of ChatGPT: Implications for educational practice and research. *Innovations in Education and Teaching International* 61, 3 (2024), 460–474. https://doi.org/10.1080/14703297.2023.2195846

[31] Bhanuka Gamage, Thanh-Toan Do, Nicholas Seow Chiang Price, Arthur Lowery, and Kim Marriott. 2023. What do Blind and Low-Vision People Really Want from Assistive Smart Devices? Comparison of the Literature with a Focus Study. In *Proceedings of the 25th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, Article 30, 21 pages. https://doi.org/10.1145/3597638.3608955

[32] Nicholas A Giudice. 2018. Navigating without vision: Principles of blind spatial cognition. In *Handbook of behavioral and cognitive geography*. Edward Elgar Publishing, United Kingdom, 260–288. https://doi.org/10.4337/9781784717544.00024

[33] Cole Gleason, Alexander J. Fiannaca, Melanie Kneisel, Edward Cutrell, and Meredith Ringel Morris. 2018. FootNotes: Geo-referenced Audio Annotations for Nonvisual Exploration. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 2, 3, Article 109 (2018), 24 pages. https://doi.org/10.1145/3264919

[34] João Guerreiro, Daisuke Sato, Saki Asakawa, Huixu Dong, Kris M Kitani, and Chieko Asakawa. 2019. CaBot: Designing and Evaluating an Autonomous Navigation Robot for Blind People. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 68–82. https://doi.org/10.1145/3308561.3353771

[35] Tanmay Gupta and Aniruddha Kembhavi. 2023. Visual Programming: Compositional Visual Reasoning Without Training. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. IEEE, Piscataway, NJ, USA, 14953–14962. https://doi.org/10.1109/CVPR52729.2023.01436

[36] HULOP. 2018. Human-scale Localization Platform (HULOP). Retrieved in May, 2024 from https://github.com/hulop.

[37] Dhruv Jain. 2014. Path-guided indoor navigation for the visually impaired using minimal building retrofitting. In *Proceedings of the 16th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 225–232. https://doi.org/10.1145/2661334.2661359

[38] Gaurav Jain, Yuanyang Teng, Dong Heon Cho, Yunhao Xing, Maryam Aziz, and Brian A Smith. 2023. "I Want to Figure Things Out": Supporting Exploration in Navigation for People with Visual Impairments. *Proceedings of the ACM on Human-Computer Interaction* 7, CSCW1, Article 63 (2023), 28 pages. https://doi.org/10.1145/3579496

[39] Ziwei Ji, Nayeon Lee, Rita Frieske, Tiezheng Yu, Dan Su, Yan Xu, Etsuko Ishii, Ye Jin Bang, Andrea Madotto, and Pascale Fung. 2023. Survey of Hallucination in Natural Language Generation. *Comput. Surveys* 55, 12, Article 248

(2023), 38 pages. https://doi.org/10.1145/3571730
[40] Vaishnav Kameswaran, Alexander J. Fiannaca, Melanie Kneisel, Amy Karlson, Edward Cutrell, and Meredith Ringel Morris. 2020. Understanding In-Situ Use of Commonly Available Navigation Technologies by People with Visual Impairments. In *Proceedings of the 22nd International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, Article 28, 12 pages. https://doi.org/10.1145/3373625.3416995
[41] Rie Kamikubo, Hernisa Kacorri, and Chieko Asakawa. 2024. "We are at the mercy of others' opinion": Supporting Blind People in Recreational Window Shopping with AI-infused Technology. arXiv:2405.06611 [cs.HC]
[42] Seita Kayukawa, Daisuke Sato, Masayuki Murata, Tatsuya Ishihara, Akihiro Kosugi, Hironobu Takagi, Shigeo Morishima, and Chieko Asakawa. 2022. How Users, Facility Managers, and Bystanders Perceive and Accept a Navigation Robot for Visually Impaired People in Public Buildings. In *Proceedings of the 31st IEEE International Conference on Robot & Human Interactive Communication*. IEEE, Piscataway, NJ, USA, 546–553. https://doi.org/10.1109/RO-MAN53752.2022.9900717
[43] Seita Kayukawa, Daisuke Sato, Masayuki Murata, Tatsuya Ishihara, Hironobu Takagi, Shigeo Morishima, and Chieko Asakawa. 2023. Enhancing Blind Visitor's Autonomy in a Science Museum Using an Autonomous Navigation Robot. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Article 541, 14 pages. https://doi.org/10.1145/3544548.3581220
[44] Sulaiman Khan, Shah Nazir, and Habib Ullah Khan. 2021. Analysis of Navigation Assistants for Blind and Visually Impaired People: A Systematic Review. *IEEE Access* 9 (2021), 26712–26734. https://doi.org/10.1109/ACCESS.2021.3052415
[45] Jee-Eun Kim, Masahiro Bessho, Shinsuke Kobayashi, Noboru Koshizuka, and Ken Sakamura. 2016. Navigating visually impaired travelers in a large train station using smartphone and bluetooth low energy. In *Proceedings of the 31st Annual ACM Symposium on Applied Computing*. ACM, New York, NY, USA, 604–611. https://doi.org/10.1145/2851613.2851716
[46] Kyoko Konishi and Véronique D Bohbot. 2013. Spatial navigational strategies correlate with gray matter in the hippocampus of healthy older adults tested in a virtual maze. *Frontiers in aging neuroscience* 5 (2013), 28885. https://doi.org/10.3389/fnagi.2013.00001
[47] Bineeth Kuriakose, Raju Shrestha, and Frode Eika Sandnes. 2020. Tools and Technologies for Blind and Visually Impaired Navigation Support: A Review. *IETE Technical Review* 39, 1 (2020), 3–18. https://doi.org/10.1080/02564602.2020.1819893
[48] Masaki Kuribayashi, Seita Kayukawa, Hironobu Takagi, Chieko Asakawa, and Shigeo Morishima. 2021. LineChaser: A Smartphone-Based Navigation System for Blind People to Stand in Lines. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Article 33, 13 pages. https://doi.org/10.1145/3411764.3445451
[49] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen-tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2020. Retrieval-augmented generation for knowledge-intensive NLP tasks. In *Proceedings of the 34th International Conference on Neural Information Processing Systems*. Curran Associates Inc., Red Hook, NY, USA, Article 793, 16 pages.
[50] Bing Li, J Pablo Munoz, Xuejian Rong, Jizhong Xiao, Yingli Tian, and Aries Arditi. 2016. ISANA: wearable context-aware indoor assistive navigation with obstacle avoidance for the blind. In *European Conference on Computer Vision*. Springer International Publishing, Cham, 448–462. https://doi.org/10.1007/978-3-319-48881-3_31
[51] Chen-Lung Lu, Zi-Yan Liu, Jui-Te Huang, Ching-I Huang, Bo-Hui Wang, Yi Chen, Nien-Hsin Wu, Hsueh-Cheng Wang, Laura Giarré, and Pei-Yi Kuo. 2021. Assistive Navigation Using Deep Reinforcement Learning Guiding Robot With UWB/Voice Beacons and Semantic Feedbacks for Blind and Visually Impaired People. *Frontiers in Robotics and AI* 8, Article 654132 (2021), 15 pages. https://doi.org/10.3389/frobt.2021.654132
[52] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288
[53] Eleanor A Maguire, David G Gadian, Ingrid S Johnsrude, Catriona D Good, John Ashburner, Richard SJ Frackowiak, and Christopher D Frith. 2000. Navigation-related structural change in the hippocampi of taxi drivers. *Proceedings of the National Academy of Sciences* 97, 8 (2000), 4398–4403. https://doi.org/10.1073/pnas.070039597
[54] Kanak Manjari, Madhushi Verma, and Gaurav Singal. 2020. A survey on Assistive Technology for visually impaired. *Internet of Things* 11, Article 100188 (2020). https://doi.org/10.1016/j.iot.2020.100188
[55] Madalin Matei, Lenuta Alboaie, and Adrian Iftene. 2022. Safety Navigation using a Conversational User Interface For Visually Impaired People. *Procedia Computer Science* 207 (2022), 1164–1173. https://doi.org/10.1016/j.procs.2022.09.172
[56] Karin Müller, Christin Engel, Claudia Loitsch, Rainer Stiefelhagen, and Gerhard Weber. 2022. Traveling More Independently: A Study on the Diverse Needs and Challenges of People with Visual or Mobility Impairments in Unfamiliar Indoor Environments. *ACM Transactions on Accessible Computing* 15, 2, Article 13 (2022), 44 pages. https://doi.org/10.1145/3514255
[57] Masayuki Murata, Dragan Ahmetovic, Daisuke Sato, Hironobu Takagi, Kris M. Kitani, and Chieko Asakawa. 2018. Smartphone-based Indoor Localization for Blind Navigation across Building Complexes. In *2018 IEEE International*

Conference on Pervasive Computing and Communications (PerCom). IEEE, Piscataway, NJ, USA, 1–10. https://doi.org/10.1109/PERCOM.2018.8444593

[58] OpenAI. 2024. Gpt-4. Retrieved in February, 2024 from https://openai.com/research/gpt-4.

[59] James Jie Pan, Jianguo Wang, and Guoliang Li. 2023. Survey of Vector Database Management Systems. arXiv:2310.14021 [cs.DB]

[60] Suraj R. Pardeshi, Vikul J. Pawar, Kailas D. Kharat, and Sachin Chavan. 2021. Assistive Technologies for Visually Impaired Persons Using Image Processing Techniques – A Survey. In *International Conference on Recent Trends in Image Processing and Pattern Recognition*. Springer Singapore, Singapore, 95–110. https://doi.org/10.1007/978-981-16-0507-9_9

[61] Manoj Penmetcha, Arabinda Samantaray, and Byung-Cheol Min. 2017. Smartresponse: Emergency and non-emergency response for smartphone based indoor localization applications. In *International Conference on Human-Computer Interaction*. Springer International Publishing, Cham, 398–404. https://doi.org/10.1007/978-3-319-58753-0_57

[62] Emanuele Pucci, Isabella Possaghi, Claudia Maria Cutrupi, Marcos Baez, Cinzia Cappiello, and Maristella Matera. 2023. Defining Patterns for a Conversational Web. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Article 118, 17 pages. https://doi.org/10.1145/3544548.3581145

[63] Aubrey Ramatla and Anne Mastamet-Mason. 2013. The decision-making processes of visually impaired consumers in an apparel retail environment. In *2013 DEFSA conference Design Cultures: Encultured Design*. Design Education Forum of South Africa, 220–228.

[64] Lisa Ran, Sumi Helal, and Steve Moore. 2004. Drishti: an integrated indoor/outdoor blind navigation system and service. In *Second IEEE Annual Conference on Pervasive Computing and Communications*. IEEE, Piscataway, NJ, USA, 23–30. https://doi.org/10.1109/PERCOM.2004.1276842

[65] Traian Rebedea, Razvan Dinu, Makesh Sreedhar, Christopher Parisien, and Jonathan Cohen. 2023. NeMo Guardrails: A Toolkit for Controllable and Safe LLM Applications with Programmable Rails. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*. Association for Computational Linguistics, Singapore, 431–445. https://doi.org/10.18653/v1/2023.emnlp-demo.40

[66] Microsoft Research. 2023. Microsoft Soundscape. Retrieved in February, 2024 from https://www.microsoft.com/en-us/research/product/soundscape/.

[67] Timothy H Riehle, P Lichter, and Nicholas A Giudice. 2008. An indoor navigation system to support the visually impaired. In *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. IEEE, Piscataway, NJ, USA, 4435–4438. https://doi.org/10.1109/IEMBS.2008.4650195

[68] Bishal Santra, Sakya Basak, Abhinandan De, Manish Gupta, and Pawan Goyal. 2023. Frugal Prompting for Dialog Models. arXiv:2305.14919 [cs.CL]

[69] Daisuke Sato, Uran Oh, João Guerreiro, Dragan Ahmetovic, Kakuya Naito, Hironobu Takagi, Kris M Kitani, and Chieko Asakawa. 2019. NavCog3 in the wild: Large-scale blind indoor navigation assistant with semantic features. *ACM Transactions on Accessible Computing* 12, 3, Article 14 (2019), 30 pages. https://doi.org/10.1145/3340319

[70] Ben Shneiderman. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages*. IEEE Computer Society, USA, 336–343. https://doi.org/10.1109/VL.1996.545307

[71] Xiaofei Sun, Xiaoya Li, Jiwei Li, Fei Wu, Shangwei Guo, Tianwei Zhang, and Guoyin Wang. 2023. Text Classification via Large Language Models. In *Findings of the Association for Computational Linguistics: EMNLP 2023*. Association for Computational Linguistics, Sentosa, Singapore, 8990–9005. https://doi.org/10.18653/v1/2023.findings-emnlp.603

[72] Zhewei Sun, Qian Hu, Rahul Gupta, Richard Zemel, and Yang Xu. 2024. Toward Informal Language Processing: Knowledge of Slang in Large Language Models. arXiv:2404.02323 [cs.CL]

[73] Sandra Tullio-Pow, Hong Yu, and Megan Strickfaden. 2021. Do You See What I See? The shopping experiences of people with visual impairment. *Interdisciplinary Journal of Signage and Wayfinding* 5, 1 (2021), 42–61. https://doi.org/10.15763/issn.2470-9670.2021.v5.i1.a69

[74] Beatrice Vincenzi, Alex S. Taylor, and Simone Stumpf. 2021. Interdependence in Action: People with Visual Impairments and their Guides Co-constituting Common Spaces. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1, Article 69 (2021), 33 pages. https://doi.org/10.1145/3449143

[75] Bryan Wang, Gang Li, and Yang Li. 2023. Enabling Conversational Interaction with Mobile UI Using Large Language Models. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Article 432, 17 pages. https://doi.org/10.1145/3544548.3580895

[76] Luyao Wang, Qihe Chen, Yan Zhang, Ziang Li, Tingmin Yan, Fan Wang, Guyue Zhou, and Jiangtao Gong. 2023. Can Quadruped Navigation Robots be Used as Guide Dogs? arXiv:2210.08727 [cs.HC]

[77] Liyang Wang, Jinxin Zhao, and Liangjun Zhang. 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*. ACM, New York, NY, USA, 815–818. https://doi.org/10.1145/3412841.3442098

[78] Lan Xia. 2010. An examination of consumer browsing behaviors. *Qualitative Market Research: An International Journal* 13, 2 (2010), 154–173. https://doi.org/10.1108/13522751011032593

[79] Chris Yoon, Ryan Louie, Jeremy Ryan, MinhKhang Vu, Hyegi Bang, William Derksen, and Paul Ruvolo. 2019. Leveraging Augmented Reality to Create Apps for People with Visual Disabilities: A Case Study in Indoor Navigation. In *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, New York, NY, USA, 210–221. https://doi.org/10.1145/3308561.3353788

[80] Tianyi Zhang, Faisal Ladhak, Esin Durmus, Percy Liang, Kathleen McKeown, and Tatsunori B. Hashimoto. 2024. Benchmarking Large Language Models for News Summarization. *Transactions of the Association for Computational Linguistics* 12 (2024), 39–57. https://doi.org/10.1162/tacl_a_00632

[81] Yan Zhang, Ziang Li, Haole Guo, Luyao Wang, Qihe Chen, Wenjie Jiang, Mingming Fan, Guyue Zhou, and Jiangtao Gong. 2023. "I am the follower, also the boss": Exploring Different Levels of Autonomy and Machine Forms of Guiding Robots for the Visually Impaired. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, Article 542, 22 pages. https://doi.org/10.1145/3544548.3580884

[82] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, Yifan Du, Chen Yang, Yushuo Chen, Zhipeng Chen, Jinhao Jiang, Ruiyang Ren, Yifan Li, Xinyu Tang, Zikang Liu, Peiyu Liu, Jian-Yun Nie, and Ji-Rong Wen. 2023. A Survey of Large Language Models. arXiv:2303.18223 [cs.CL]

[83] Yuhang Zhao, Elizabeth Kupferstein, Hathaitorn Rojnirun, Leah Findlater, and Shiri Azenkot. 2020. The Effectiveness of Visual and Audio Wayfinding Guidance on Smartglasses for People with Low Vision. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. ACM, New York, NY, USA, 1–14. https://doi.org/10.1145/3313831.3376516

Received February 2024; revised May 2024; accepted June 2024