# A Multimodal AI-Powered Travel Assistant Geo-App for BLV Users

Tyler Ortiz[1], Mamuna Chaudhry[1], Jianing Qi[2], Richard Perez[1], William H. Seiple[3,4], Hao Tang[5,2], Zhigang Zhu[1,2]

The City College of New York, CUNY[1], The Graduate Center, CUNY[2], Lighthouse Guild[3], New York University[4], Borough of Manhattan Community College, CUNY[5]

tortiz003@citymail.cuny.edu, mchaudh001@citymail.cuny.edu, jqi@gradcenter.cuny.edu, rperez023@citymail.cuny.edu, wseiple@lighthouseguild.org, htang@bmcc.cuny.edu, zzhu@ccny.cuny.edu

## Abstract

Buddy Walk is a multimodal AI-powered web application designed to enhance vision assistance technology for users who are blind or have low vision (BLV), by combining real-time image analysis with contextual information from the Google Maps API. The app aims to help users travel more independently, confidently, and safely by providing detailed, accessible descriptions of their surroundings. Our novel approach integrates computer vision, geolocation services, and a large language model to provide a variety of assistive services such as accurately identifying and detecting objects and landmarks, responding to queries ranging anywhere from descriptions of images or videos captured by the user or from navigation-related queries, and returning relevant environmental details through an intuitive user interface compliant with the Web Content Accessibility Guidelines. The app is developed through continuous user testing sessions and feedback from BLV users and organizations such as Lighthouse Guild and Visions. By integrating multiple data sources and interactive feedback, Buddy Walk focuses on making independent travel safer and more accessible for BLV users and demonstrates the potential for AI-assisted mobility tools to bridge accessibility gaps in urban navigation, paving the way for broader adoption and future integration with emerging assistive technologies. The source code is publicly available, and developers are welcome to contribute to this open-source project.

## Keywords

Vision assistance technology, environmental awareness, accessibility, computer vision, contextual awareness, real time support, assistive technology, adaptive technology

**Introduction**

In the United States, nearly 6 million individuals experience some degree of vision loss, with around 1 million classified as legally blind (Centers for Disease Control and Prevention). For individuals who are blind or have low vision (BLV), independently navigating outdoor environments remains one of the most daunting challenges. The inability to see signage and visual landmarks can make travel much more difficult and stressful for a BLV individual than for a sighted individual.

Mainstream navigation mobile Apps like Google Maps and Apple Maps assist with route planning but are not designed for BLV users. These apps do not take into consideration key elements such as contextual details and landmarks, lack real-time data such as the presence of construction, and have issues directing users to the exact entrance of an establishment, an issue known as the "last few meters" challenge (Saha). Additionally, they are not optimized for such a use case and can be cumbersome to navigate efficiently during travel (Sanchez et al.).

Specialized navigation applications for BLV users, such as Lazarillo (Lazarillo) and BlindSquare (BlindSquare), provide audio-based routing and nearby point-of-interest information. While these tools improve accessibility compared to mainstream navigation apps, they primarily focus on routing and location cues and only offer limited integration of real-time visual context.

Existing vision assistance tools address complementary aspects of this challenge. Applications such as Be My Eyes (Be My Eyes) provide human-mediated visual assistance, while Seeing AI (Seeing AI) and Be My AI – a Be My Eye app (Be My Eyes) offer automated image-based descriptions. However, these approaches lack integrated geographic context, limiting their ability to support independent travel planning and situational understanding.

*Problem Statement*

Through interactions with BLV individuals, we found that they often face significant limitations with existing assistive technologies, which typically either offer basic object recognition without up-to-date contextual data or provide location information without integrating visual assistance. Insights from discovery phase interviews, where we informally interviewed three BLV individuals, revealed that this gap requires users to rely on multiple apps to gather a full understanding of their surroundings, leading to inefficiencies and added barriers to independent navigation.

The application we have been developing, called ***Buddy Walk***, aims to bridge this critical gap by combining rich, contextual geolocation data and real-time image recognition giving BLV users a seamless, conversational style, all-in-one tool for them to better understand and navigate outdoor environments, without providing turn-by-turn navigation.

*Rationale*

Buddy Walk is designed as a travel assistant. Rather than providing continuous routing guidance, it supports travel planning, destination discovery, and situational awareness by helping BLV users identify nearby locations, explore route options at a high level, and recognize storefronts and landmarks.

This approach distinguishes Buddy Walk from existing tools. Navigation-focused systems such as Google Maps, Apple Maps, Lazarillo, and BlindSquare emphasize routing but provide limited visual context and often struggle with precise destination identification in urban settings. Vision-based assistants such as Seeing AI and Be My AI offer image descriptions without geographic grounding, while Be My Eyes relies on sighted volunteers, limiting scalability and independence.

Buddy Walk addresses these gaps through a multimodal approach, integrating data from multiple sources. The application leverages OpenAI's API with its vision language model (VLM), providing natural human-like responses, alongside the Google Maps API, which supplies rich geolocation data, including business names, categories, operating hours, and accessibility features such as ramps and accessible entrances.

Buddy Walk targets the specified use cases (Table 1). Buddy Walk is designed with simplicity as its core principle, enabling users to interact through voice commands supported by integrated speech-to-text and text-to-speech capabilities.

Table 1. Use Case List Targeted by Buddy Walk

| Category | Example User Queries |
|---|---|
| Location Identification and Context | • "Where am I right now? Can you describe the surrounding area?" <br> • "What type of neighborhood is this (residential, commercial, industrial)?" <br> • "Are there any landmarks or notable buildings nearby?" |
| Business and Service Information | • "What stores or businesses are near me?" <br> • "Can you tell me the name and type of this establishment?" <br> • "What time does this store close?" |
| Navigation and Directions | • "How can I get to the nearest public transit stop?" <br> • "How far am I from the nearest department store?" <br> • "What are the cross streets at my location?" |
| Transportation Options | • "What public transportation options are available here?" <br> • "Where's the nearest bus stop or subway station?" |
| Visual Description of Images and Videos | • Descriptions of captured images or short video clips to help users understand nearby objects, scenes, and environmental context |

## Discussion

Buddy Walk is intentionally designed to be simple, intuitive, and fully accessible for BLV users, prioritizing ease of use and effective navigation (Figure 1). We tested the app extensively with BLV users through continuous sessions, which helped us identify and refine the

optimal approach to developing the user interface. The app follows WCAG 2 (Web Content

Accessibility Guidelines) standards to meet high accessibility benchmarks (World Wide Web

Consortium). It offers full compatibility with screen readers such as Apple's VoiceOver on iOS,

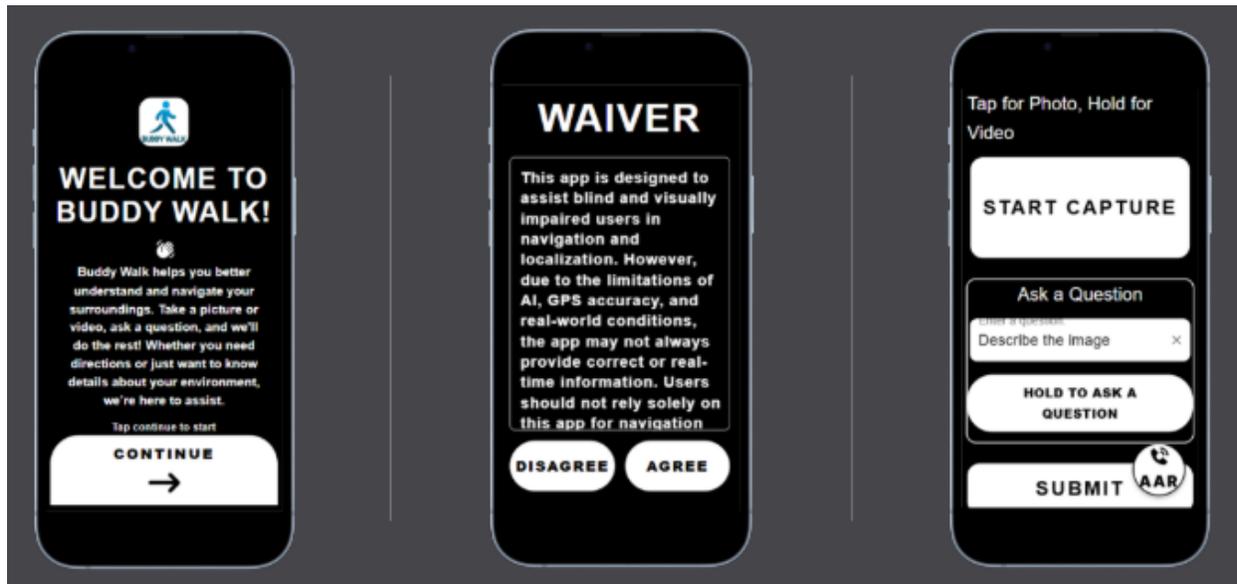allowing for seamless interaction.



Fig. 1. User Interface of Buddy Walk.

*User Interface*

The design of the app is minimalistic and structured, focusing only on essential elements

to reduce cognitive load, with larger, bolder text and clearly defined buttons to facilitate easy

navigation across diverse user demographics. Buddy Walk has been tested across multiple

devices, including iPhones and Android phones, to ensure consistent accessibility. Performance

is optimized with a fast text-to-speech response time of two to three seconds, providing smooth

and responsive user feedback. Together, these features make Buddy Walk a user-friendly and

highly accessible tool tailored to the specific needs of BLV users, standing apart from other apps

by minimizing clutter and focusing on core functionality.

*Multimodal Capture*

When it comes to multimodal capture modes, Buddy Walk offers the flexibility to capture images or videos and allows users to pose specific questions about the captured content in a streamlined manner. Videos (with audio) can be recorded for up to 30 seconds, which is then sent to and analyzed by the AI model to generate contextually relevant insights (Figure 2).
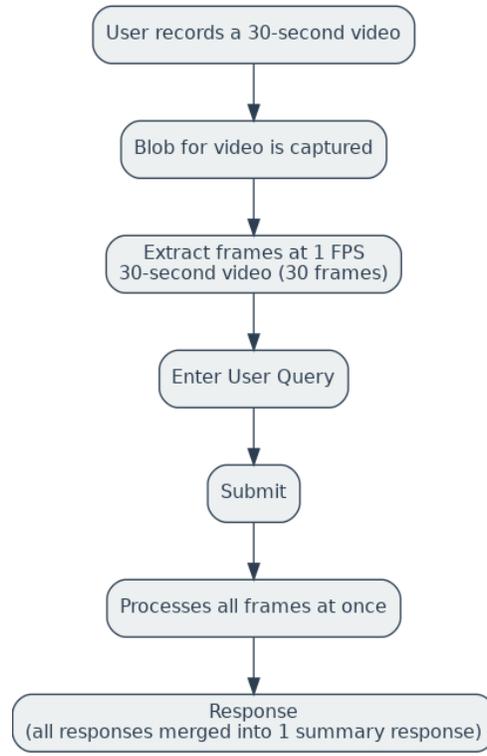


Fig. 2. Flowchart of Video Frame Extraction Feature.

Inline conversational context is provided so users can ask follow-up questions and clarify previous responses. Speech-to-text recognition ensures functionality on all types of smartphones without solely relying on built-in functionality like Apple's dictation feature, giving users the option to either speak or type their question into the textbox.
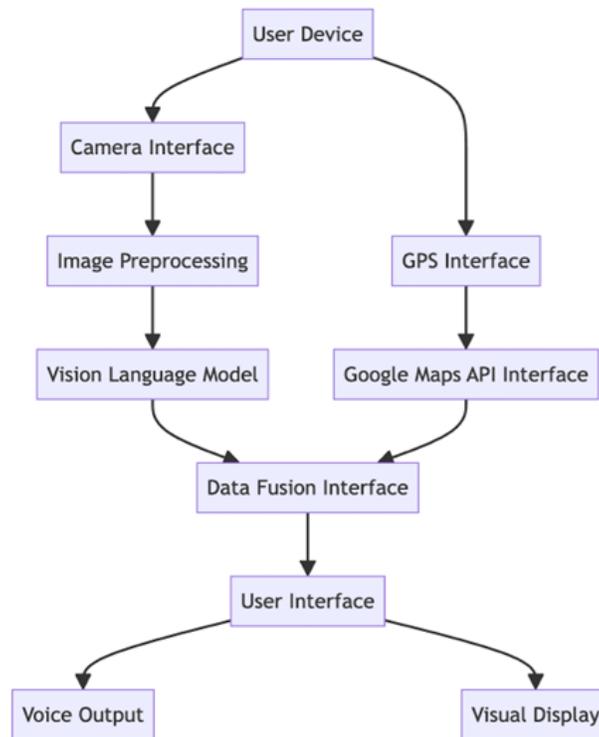
*Technical Tools*



Fig. 3. System Architecture Initial Design.

The main functionality of the application comes from the fusion of the OpenAI API with the Google Maps API. We designed a system that can parse a user's input and pick the most relevant Google Maps API call, execute that call, then feed the information to the OpenAI API to provide it with the relevant information to formulate a detailed and relevant response that is sent back to the user. The app is a type of Retrieval-Augmented Generation (RAG) (Lewis) system designed to be scalable, allowing us to include more API calls to different sources beyond just the Google Maps API. All code is open source and readily available on GitHub (Ortiz).

*User Driven Feedback Loop*

Our app development has consistently been focused on addressing the needs of BLV users first. As such, we have maintained contact with end-users throughout the entire process. Since the inception of the project, one of the authors who has low vision has closely worked with

the research team to design and test the app. Another blind user has also given feedback on multiple occasions. A bi-weekly collaboration with VISIONS has enabled us to reach additional BLV users, and to obtain input directly from an orientation and mobility specialist. VISIONS is a service center providing programs to assist BLV people to lead independent and active lives (VISIONS).

Our dedicated low-vision collaborator has helped to guide the development of this application. His feedback, including suggestions such as effective color choices, text and vocabulary recommendations, spacing, button sizes, interface layout, additional features, and much more, has been instrumental to the development of this app. However, this singular collaborator was not the only voice we listened to. The needs of disparate users must be addressed, so as this collaborator still had some vision that he relied on, his experience would not always be congruous with other users who may have less vision or be completely blind. As such, we aimed to receive numerous disparate opinions as we developed the app. Also important to note is this tester's familiarity with using Google Maps despite his low vision as well as frequent use of ChatGPT, granting a unique perspective on Buddy Walk's integration of both systems.

During our development, we had the opportunity to visit VISIONS on five occasions in 2025 to meet with a small focus group of interested BLV individuals. The dates we visited are as follows: March 19, April 2, April 30, May 28, and June 27. Each visit, we presented to a group of around 10 BLV users, first giving a general overview of the app then performing hands-on testing by sending the link to the web application via text message then observing how the user interacted with the app. For the hands-on testing, usually one BLV individual would use the app on their own smartphone while other BLV individuals would gather around the tester to listen to responses and understand the functionality of the application, giving suggestions for prompts to

input and ways to interact with the app. The total number of users who directly tested the application over the five visits is 10, and more than 30 BLV participants indirectly tested the app and offered their suggestions. Over the course of the five visits, we recorded feedback from users. Some of the feedback we received, and the actions taken, are listed in Table 2.

Table 2. Feedback from BLV Testers.

| Feedback from testers and observed behavior | Issue to address |
|---|---|
| "The response time is too long." | Adjust the API calls to OpenAI to speed up response time (Done) |
| "I would like to know the cross streets of a specific address." | Add a new feature that allows the AI to see nearby cross streets (Done) |
| "The response is incorrect when asking for the nearest Home Depot store." | Error when determining user starting location (Done) |
| "When I ask for the nearby store, it gives numbers instead of an address." | Returning latitude and longitude rather than address (Done) |
| Confusion due to poorly worded accessibility labels | Alter the accessibility labels to accurately describe the object in question (Done) |
| "Please add a way to see when the bus arrives at the bus stop and which bus will come next." | Add real-time MTA data into app (To Do) |
| "The provided answer was insufficient until I rephrased the question to be more detailed" | Adjust AI prompt to produce response that aligns closer to what users expect (To Do) |

*Specific Functionality Testing*

To determine the accuracy of certain features, we met with BLV individuals at VISIONS again on November 12th to receive feedback on certain app functionalities. Our goal was to test the accuracy of three specific features currently implemented into the app, as well as the rate of AI hallucinations. The tested features are as follows:

Q1. Directions to a nearby location (travel planning)

Q2. Generating a list of locations that fit into a specific category

Q3. Image description

During testing, eight BLV volunteers were asked to pose questions for which they knew the answers, and their feedback was recorded. Their demographics are listed in Table 3.

Table 3. Participant Demographics.

| ID | Age | Level of Vision |
|---|---|---|
| P1 | 74 | Totally Blind |
| P2 | 63 | Totally Blind |
| P3 | 69 | Low Vision |
| P4 | 64 | Totally Blind |
| P5 | 49 | Low Vision |
| P6 | N/A | Totally Blind |
| P7 | 61 | Totally Blind |
| P8 | 73 | Totally Blind |

Each volunteer chose a destination to generate directions (Q1), a category of locations to generate a list of nearby places (Q2), and an object for which they wanted a visual description (Q3). For Q2, the participants were asked to have a destination that fits within their chosen category in mind, and we recorded where the destination landed in the generated list. For example, a Q2 result of "2 out of 5" for P1 indicates that the correct destination appeared as the second item in the generated list of five items.

Table 4. Participant Responses and System Performance Across Three Query Types (Q1–Q3).

| ID | Q1 Location | Q1 Results | Q2 Category | Q2 Results | Q3 Subject | Q3 Results |
|---|---|---|---|---|---|---|
| P1 | PC Richards | Further chain store than expected. | Barbecue | Correct. 2 out of 5. | P1 | "Pretty good, I don't like that it said old man. Not too much information" |

| ID | Q1 Location | Q1 Results | Q2 Category | Q2 Results | Q3 Subject | Q3 Results |
|---|---|---|---|---|---|---|
| P2 | Dallas BBQ | Correct location and directions. | Supermarket | Correct. 2 out of 10. | Crocheted bag | "Very good!" |
| P3 | Men's Warehouse | Further chain store than expected. | Electronic stores | Correct. 2 out of 14. | P3's sweater | "OK." |
| P4 | Papaya | Further chain store than expected. | Shopping Centers | Incorrect. | Pocket book | "Nice, very good." |
| P5 | Target | Correct location, incorrect directions | Libraries | Correct. 6 out of 6. | Auditorium stage | "Nice, not too wordy, the more the better." |
| P6 | CVS | Correct location, incorrect directions | Banks | Correct. 3 out of 19. | P6's shirt | "That is cool. Not too wordy." |
| P7 | Planet Fitness | Could not locate the nearest chain store. | Fitness Centers | Correct. 3 out of 20. | P7 | "I don't like that it called me 'older'. Good description." |
| P8 | Trader Joes | Correct location and directions | Restaurants | Correct. P8 had no destination in mind. | P8 | "Very good. Age description is a bit judgmental." |

From the experiment, we discovered that two of the three features were working well,

particularly the image description and the nearby places features. The directions feature was only

working to our satisfaction about half the time. The image description feature worked well

according to the volunteers, providing enough detail without being overly verbose. Volunteers

said the more information given by the AI, the better they could understand their surroundings.

However, the testing of this feature was limited to indoor building locales, so further testing is

needed in outdoor street settings to test how well it can describe potentially hazardous conditions.

The nearby locations testing also went well. When the volunteers suggested a category of locations, they typically had a specific location in mind, which the app listed within the first six options for 7 out of 8 volunteers. Some issues arose depending on how the user categorized the location they were thinking of, which sometimes did not align with the way Google Maps categorized that location. For example, participant 4 was looking for Burlington Coat Factory by using the descriptor "Clothing store." However, when searching for nearby clothing stores, the desired destination did not appear on the list. This is because Google Maps qualifies the store as a "department store" rather than a "clothing store." Additional work is required to alter the way we retrieve nearby stores using keywords inputted by the user; generalization may be required to give the user what they are looking for.

Finally, the generation of directions to a specific location needs additional work to enhance accuracy. Only half of the tests correctly identified the address of the location the user wanted to travel to, with only two responses providing the route the user would take themselves. The correctly generated routes did reach the destination but were deemed less efficient by the participant than the route they would take themselves. Some of the other issues that arose were incorrect directions, inability to find the desired location, providing directions to a chain store that was not the closest location, and difficulty discerning directionality due to confusing cardinal direction descriptors. These inaccuracies occurred due to algorithmic limitations and the complexity of the task.

*Hallucination Risks*

The implementation of Buddy Walk safeguards against AI hallucinations in multiple different ways. First, our prompt to the AI specifies to only use data provided from Google Maps, ensuring that the AI does not try to answer questions primarily using the data it was trained on. It requires the RAG system to retrieve the relevant data. If no data is returned from Google Maps, the AI responds to the user with "I'm sorry, there is not enough data to answer this question."

When configuring the AI, there exists a setting to fine-tune the AI's "temperature". An AI's temperature refers to the amount of creativity it is allowed to use in the response, with one introducing more randomness for diverse and creative outputs, something the Buddy Walk does not need. A temperature close or equal to 0 delivers more deterministic outputs, which allows the AI responses to be more consistent. This greatly alleviates the risk of AI-induced hallucinations in the response. Incorrect responses are more likely to be caused by incorrect data in the Google Maps database, low signal quality resulting in inaccurate geolocation, or errors in the retrieval algorithm within the Buddy Walk app.

**Conclusion**

Buddy Walk demonstrates a practical and accessible solution to enhance safer mobility and environmental awareness for Blind and Low Vision users by combining real-time visual analysis with contextual geolocation data. Through user-centered design and rigorous testing, the app aims to support independent navigation, reduce cognitive load, and address critical gaps in existing assistive technologies. While challenges remain in perfecting the accuracy of responses, future work will focus on integrating additional environmental cues and expanding user testing to further refine and scale the solution. Our next steps include exploring additional datasets to

implement into the app for better navigation and geospatial awareness as well as conducting experiments involving field testing with BLV users to evaluate their use while travelling. Ultimately, Buddy Walk represents a meaningful step toward empowering BLV individuals to travel more confidently and safely in urban environments by integrating cutting-edge technology with existing solutions.

*Acknowledgments*

**Works Cited**

American Foundation for the Blind. "Demographics of Americans with Vision Difficulty."

*American Foundation for the Blind*, https://www.afb.org/research-and-

initiatives/statistics/demographics-americans-vision-difficulty. Accessed 12 Aug. 2025.

Anderson, Raeda K., et al. "Artificial Intelligence Use and Perceptions by Disabled Adults."

*Shepherd Center*, Georgia State University.

Apple Inc. "Turn on and Practice VoiceOver on iPhone." *iPhone User Guide*, Apple Support,

n.d., https://support.apple.com/guide/iphone/turn-on-and-practice-voiceover-

iph3e2e415f/ios.

Be My Eyes. "Accessibility Technology for Blind & Low Vision People." *Be My Eyes*, n.d.,

https://www.bemyeyes.com/.

BlindSquare. "What Is BlindSquare?" *BlindSquare*, n.d.,

https://www.blindsquare.com/ufaqs/what-is-blindsquare/.

Centers for Disease Control and Prevention. "VEHSS Modeled Estimates: Prevalence of Vision

Loss and Blindness." *Vision and Eye Health Surveillance System*, National Center for

Chronic Disease Prevention and Health Promotion, 15 May 2024,

https://www.cdc.gov/vision-health-data/prevalence-estimates/vision-loss-

prevalence.html.

iAccessibility Solutions. "Seeing AI Blind Accessibility App." *iAccessibility*, n.d.,

https://www.iaccessibility.com/apps/blind/index.cgi/product?ID=321.

Lazarillo. *Lazarillo*. n.d., https://lazarillo.app/. Accessed 30 Aug. 2025.

Lewis, Patrick, et al. "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks." *Proceedings of the 34th International Conference on Neural Information Processing Systems*, Curran Associates Inc., 2020, pp. 9459–9474.

Microsoft. "Seeing AI." *Microsoft Garage*, n.d., https://www.microsoft.com/en-us/garage/wall-of-fame/seeing-ai/. Accessed 12 Aug. 2025.

New York State Office of General Services. "New York State Language Access Law." *Office of General Services*, New York State, n.d., https://ogs.ny.gov/new-york-state-language-access-law.

Ortiz, Tyler. "Buddy_Walk." *GitHub*, 31 July 2025, https://github.com/tort8678/Buddy_Walk. Accessed 12 Aug. 2025.

Saha, Manaswi, et al. "Closing the Gap: Designing for the Last-Few-Meters Wayfinding Problem for People with Visual Impairments." *Proceedings of the 21st International ACM SIGACCESS Conference on Computers and Accessibility*, ACM, 2019. *Microsoft Research*, https://www.microsoft.com/en-us/research/wp-content/uploads/2019/07/LandmarkAIASSETS.pdf.

Sanchez, Jaime, et al. "Understanding In-Situ Use of Commonly Available Navigation Technologies by People with Visual Impairments." *ACM Transactions on Accessible Computing*, vol. 11, no. 3, Oct. 2018, pp. 1–27. *ACM Digital Library*, https://doi.org/10.1145/3230683.

VISIONS. "VISIONS – Services for the Blind and Visually Impaired." *VISIONS Services*, n.d., https://visionsvcb.org/. Accessed 12 Aug. 2025.

World Wide Web Consortium (W3C). "Web Content Accessibility Guidelines (WCAG) 2

Overview." *Web Accessibility Initiative (WAI)*, n.d., https://www.w3.org/WAI/standards-

guidelines/wcag/. Accessed 30 Aug. 2025.