LLM-Glasses: GenAI-driven Glasses with Haptic Feedback for Navigation of Visually Impaired People

Issatay Tokmurziyev, Miguel Altamirano Cabrera, Muhammad Haris Khan,
Yara Mahmoud, Luis Moreno, and Dzmitry Tsetserukou

***Abstract*— We present LLM-Glasses, a wearable navigation system designed to assist visually impaired individuals by combining haptic feedback, YOLO-World object detection, and GPT-4o-driven reasoning. The system delivers real-time tactile guidance via temple-mounted actuators, enabling intuitive and independent navigation. Three user studies were conducted to evaluate its effectiveness: (1) a haptic pattern recognition study achieving an 81.3% average recognition rate across 13 distinct patterns, (2) a VICON-based navigation study in which participants successfully followed predefined paths in open spaces, and (3) an LLM-guided video evaluation demonstrating 91.8% accuracy in open scenarios, 84.6% with static obstacles, and 81.5% with dynamic obstacles. These results demonstrate the system's reliability in controlled environments, with ongoing work focusing on refining its responsiveness and adaptability to diverse real-world scenarios. LLM-Glasses showcases the potential of combining generative AI with haptic interfaces to empower visually impaired individuals with intuitive and effective mobility solutions.**

I. Introduction

Navigating dynamic environments presents significant challenges for visually impaired individuals, particularly in real-time obstacle detection and avoidance. While GPS-based systems provide valuable guidance, their reliance on visual or auditory feedback often increases cognitive load and reduces effectiveness in crowded or unpredictable settings. These limitations highlight the need for intuitive, non-intrusive navigation aids.

Haptic feedback offers a promising alternative by delivering spatial information through tactile sensations, eliminating the need for visual or auditory reliance [1]. Prior research has shown its efficacy in wearable and handheld systems [2], [3], [4], enabling enhanced spatial awareness and navigation confidence. Despite their potential, many existing systems face practical constraints, such as bulky designs or limited integration into daily life, reducing their usability and effectiveness.

Recent advancements in Artificial Intelligence (AI) have opened new opportunities for assistive technologies. Large Language Models (LLMs), like GPT-4o, excel in contextual reasoning and generating adaptive, user-specific guidance [5], [6]. Combined with object detection systems, such as YOLO-based frameworks, and multimodal AI approaches, these models enable personalized, real-time feedback [7], [8].

The authors are with the Intelligent Space Robotics Laboratory, Center for Digital Engineering, Skolkovo Institute of Science and Technology. {issatay.tokmurziyev, m.altamirano, haris.khan, yara.mahmoud, luis.moreno, d.tsetserukou}@skoltech.ru

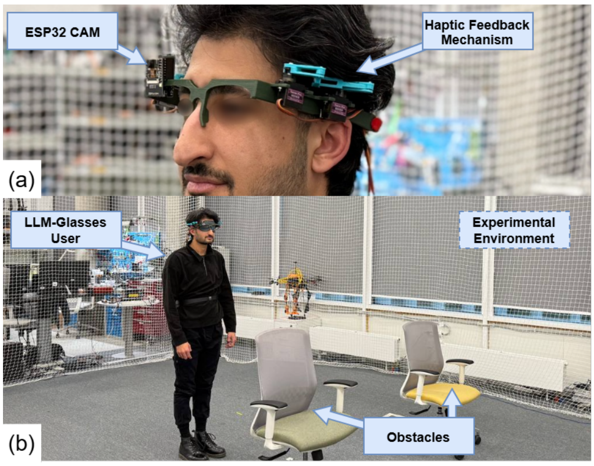

Fig. 1. (a) LLM-Glasses prototype, highlighting key components: ESP32 CAM and a haptic feedback mechanism with a five-bar linkage design. (b) An experimental setup with the user navigating through an obstacle course with real-time guidance provided by the system.

These innovations pave the way for intelligent, responsive systems tailored to the needs of visually impaired users.

This paper introduces **LLM-Glasses**, a lightweight, wearable navigation system that integrates YOLO-based object detection, GPT-4o-driven reasoning, and a temple-mounted haptic feedback device. The system processes real-time environmental data to generate intuitive tactile signals. By combining the capabilities of vision and language models with haptic technologies, LLM-Glasses addresses existing limitations, offering a practical solution for enhancing independent mobility for users with impairments.

II. Related Works

Haptic feedback has demonstrated significant potential in aiding navigation for visually impaired individuals by providing non-visual spatial cues. Prior studies have validated the effectiveness of vibrotactile systems, such as multi-actuator handles achieving directional cue success rates of up to 100% [9], head-mounted systems providing obstacle detection via forehead vibrations [10], and walker handles offering bimanual vibrotactile feedback to reduce cognitive load [11]. Additionally, multimodal haptic devices like Gallo's augmented white cane [2] highlight the utility of combining tactile signals with mobility aids. However, these systems often rely on bulky designs, limiting their practicality for everyday use.

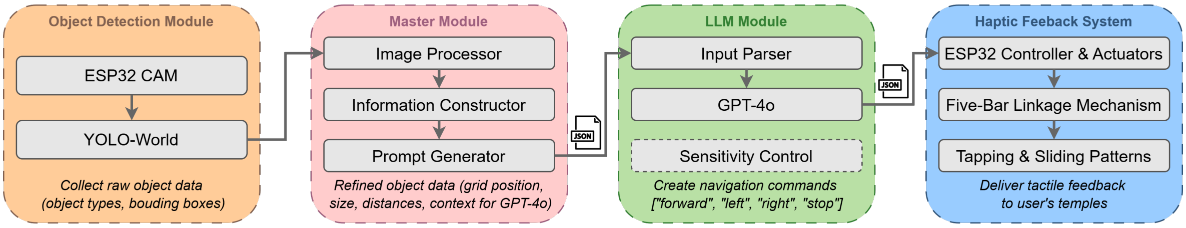

Fig. 2. System Architecture of the LLM-Glasses.

Wearable haptic devices have addressed some of these limitations by integrating compact mechanisms, such as skin stretch devices for navigation [3], multi-contact tactile feedback [12], and waist-mounted vibrotactile signals for wayfinding [4]. Handheld systems have also shown effectiveness in delivering directional guidance [13]. While these advances improve accessibility, they can still suffer from usability challenges and limited intuitiveness.

Recent works have explored the development of assistive robotic systems utilizing haptic feedback to enhance mobility and autonomy for visually impaired users. For instance, Agrawal et al. introduced a novel perceptive robotic cane that supports visually impaired individuals in social scenarios such as seat selection, using haptic cues to provide navigational assistance and environmental awareness [14]. In a similar direction, a robotic guidance system was proposed to support outdoor running for visually impaired users through continuous haptic feedback, ensuring safety and route adherence in dynamic environments [15].

Beyond individual prototypes, a comprehensive survey on navigation systems for individuals with visual impairment [16] reviews a wide range of assistive technologies, including wearable haptic devices, computer vision systems, and multimodal interfaces designed to improve environmental awareness and navigation autonomy. Complementing this, assistive robotics surveys [17] have mapped the landscape of robotic systems aimed at empowering visually impaired users, detailing a wide array of solutions that integrate haptic guidance with autonomy.

Moreover, a systematic review of over 20 years of research on haptic feedback for blind and low-vision individuals revealed a growing trend towards multimodal assistive tools that combine tactile, auditory, and visual modalities to support complex tasks such as navigation, education, and daily activities [18]. These insights reinforce the relevance of lightweight, hands-free solutions like LLM-Glasses, which leverage haptic feedback alongside advanced AI-based reasoning to create an intuitive, unobtrusive navigation aid.

Multimodal approaches integrating haptic feedback with AI technologies, such as Visual Language Models (VLMs) [19], [20] and Large Language Models [21], have further advanced navigation aids. Systems like VIAssist [5] and ChatGPT for Visually Impaired [8] combine object detection with language-based interaction, though they often lack the real-time responsiveness or compactness needed for practical application.

Our work builds on these foundations, addressing the limitations of prior systems by introducing a lightweight, temple-mounted haptic feedback device powered by YOLO-based [22] object detection and LLM-driven contextual reasoning. This approach offers real-time, intuitive guidance tailored to the needs of visually impaired users.

III. System Architecture

The LLM-Glasses system is a lightweight and wearable device designed to assist visually impaired individuals with real-time navigation. It integrates three main modules – Object Detection, Information Processing (Master Module), Language Processing, and Haptic Feedback – working together seamlessly to provide intuitive, non-intrusive guidance. The overall system architecture is illustrated in Fig. 2, while the prototype of the wearable device is shown in Fig. 3a.

A. Object Detection Module

The Object Detection Module serves as the system's sensory input, capturing real-time environmental data. An ESP32 CAM is used to stream video frames. This camera was chosen for its compact size and wireless connectivity, which ensures minimal interference with the user's movement. The module processes these frames using the YOLO-World object detection model, an efficient algorithm known for its ability to identify objects across an open vocabulary.

Detected objects are classified and localized within a 2x3 spatial grid, and their bounding boxes are calculated. Each detected object is assigned a priority based on its distance and position within the grid to ensure that immediate obstacles receive higher attention. This spatial mapping enables the system to associate detected obstacles with specific regions, such as "bottom-center" or "top-left", which helps in prioritizing immediate hazards closer to the user. The raw detection outputs (object type, bounding boxes' location) are then passed to the Master Module for further refinement.

B. Master Module

The Master Module acts as an intermediary between object detection and language processing, refining raw detection outputs into structured, meaningful data. It first preprocesses the YOLO outputs using an Image Processor, ensuring that object sizes, positions, and distances are calculated consistently across frames. The Multi-Frame Information Constructor collects data from consecutive frames to mitigate momentary occlusions or detection errors.

The refined data is then used by the Prompt Generator, which converts object detection results into structured textual prompts for the Language Model. These prompts incorporate grid locations, object types, and proximity information, contextualized for GPT-4o. To ensure accurate obstacle avoidance, obstacles detected in the "bottom-center" grid that are closer than 1 meter are flagged as immediate hazards. This ensures that the system prioritizes actionable guidance over non-urgent detections.

C. Language Module

The navigation system leverages computer vision to detect objects and obstacles in real-time, providing essential spatial data. However, to enhance the user experience and ensure more intuitive guidance, the system integrates a Language Module responsible for analyzing spatial data and generating user-friendly navigation commands.

At its core, the module uses GPT-4o, a state-of-the-art large language model, to interpret object information and provide concise navigation instructions. When multiple objects are detected, GPT-4o prioritizes hazards based on their proximity and generates simplified directional commands like "left", "right", "forward", or "stop". These instructions are designed to minimize cognitive load while ensuring safety during navigation.

Moreover, the system incorporates adaptive sensitivity settings, allowing users to adjust the level of detail in the feedback. Sensitivity settings range from "low" (reporting only critical hazards) to "high" (providing detailed obstacle descriptions). This adaptability enables the system to function effectively in various environments, such as crowded streets or quieter indoor spaces.

D. Haptic Feedback System

The Haptic Feedback System translates navigation commands into intuitive tactile signals delivered through temple-mounted actuators. As shown in Fig. 3a, the actuators are integrated into a lightweight glasses frame using a five-bar linkage mechanism [23], driven by servo motors controlled by an ESP32 microcontroller. This mechanism provides precise and comfortable feedback, ensuring intuitive tactile guidance for navigation. The glasses, including the five-bar linkage mechanism, ESP32 CAM camera, and frame, weigh a total of 122.5 grams, making them lightweight and suitable for prolonged use without causing user discomfort.

The system supports two types of feedback patterns: tapping and sliding. Tapping patterns are used for localized directional cues [`right tap, left tap, front tap, back tap, center tap`], while sliding patterns provide continuous guidance [`front slide, back slide, right slide, left slide`]. These patterns are depicted in Fig. 3b. To enhance perception, a calibration process is performed to adjust haptic intensity based on individual user sensitivity and temple geometry.

Power for the system is supplied by two separate battery packs: one for the ESP32 controller and one for the servo motors, each using four 1.5V batteries connected in series to provide 6V. The ESP32 controller can operate for approximately 10 hours, while the servo motors can function for around 2.5 hours under moderate load. These runtime durations are sufficient for typical short-term navigation tasks, and further optimization of power consumption is planned for future iterations of the system.

IV. Experimental Evaluation

A. Haptic Pattern Recognition Study

The proposed wearable device provides tactile guidance to facilitate navigation in unknown environments. By leveraging tactile sensations on the temples, the system guides users in adjusting head positions to change walking directions. The inverse five-bar linkage mechanism enables the creation of tactile patterns across a 7 cm range, with independent tap and sliding patterns varying in force, position, direction, and speed. The objective of this study is to identify the patterns with the highest and most intuitive recognition rate for use during user navigation.

1) Pattern Design: We designed nine different tactile patterns: five tapping at different temple regions [`front, center, back, right, left`] and four sliding patterns in different directions [`back-to-front, front-to-back, left-to-right, right-to-left`], as shown in Fig. 3b. Sliding patterns were rendered at two speeds: slow (1.5 seconds per trajectory) and fast (1 second per trajectory). These patterns were designed for spatial and directional intuitiveness, aiding users in navigation.

2) Experimental Setup: Thirteen participants (12 males, 1 female, aged 22-35 years, mean 25.7 $\pm$3.8) completed the study. None of them reported any deficiencies in sensorimotor function. The participants were informed about the experiments and agreed to the consent form. Participants underwent a training session to familiarize themselves with the task. Each pattern was rendered three times during training, and a visual reference of the patterns was provided throughout the session.

During the evaluation, participants wore the device and used a graphical user interface (GUI) on a PC to identify the patterns they perceived. Each of the 13 patterns (five taps and four sliding patterns at two speeds) was presented five times in a randomized order, resulting in 65 trials per participant. Participants provided feedback on the perceived sensations at the end of the study.

3) Results: The recognition accuracy of the haptic patterns averaged 81%, with the back sliding slow pattern achieving the highest recognition rate (95%) and the right tap pattern the lowest (65%). A confusion matrix summarizing the results is shown in Table I.

Statistical analysis was conducted using a single-factor repeated-measures ANOVA ($\alpha < 0.05$), revealing a significant difference in recognition rates across patterns: $F(12, 156) = 1.9902, p = 0.0284$. Pairwise t-tests with Bonferroni correction identified significant differences between specific patterns, such as back slide slow vs. right tap ($p = 0.0228$). While a significant difference was observed between

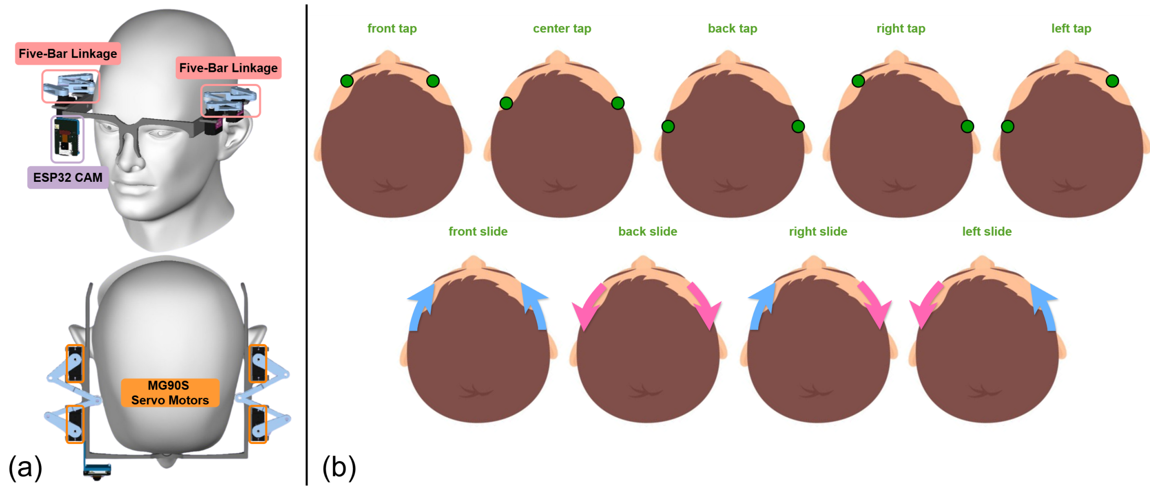

Fig. 3. (a) 3D model of the LLM-Glasses haptic navigation system. (b) The nine haptic feedback patterns used in the user study, illustrating tapping and sliding motions across different regions of the user's temples. Each pattern provides distinct sensory input designed to aid in directional navigation.

TABLE I
CONFUSION MATRIX FOR ACTUAL AND PERCEIVED PATTERN RECOGNITION.

%			Answers (Predicted Class)												
			Tap					Slide							
								Fast				Slow			
			front	center	back	left	right	front	back	right	left	front	back	right	left
Patterns	Tap		front	0.85	0.11	0.02	0.03	-	-	-	-	-	-	-	-	-
			center	0.06	0.80	0.05	0.08	0.02	-	-	-	-	-	-	-	-
			back	-	0.22	0.71	0.02	0.05	-	0.02	-	-	-	-	-	-
			left	0.08	0.12	0.05	0.72	0.03	-	-	-	-	-	-	-	-
			right	0.03	0.22	0.09	-	0.65	-	-	-		-	-	-	-
	Slide	Fast	front	-	-	-	-	-	0.80	-	0.09	0.02	0.09	-	-	-
			back	-	-	-	-	-	-	0.83	0.03	0.03	-	0.09	0.02	-
			right	-	-	0.02	-	-	-	0.02	0.89	-	-	-	0.08	-
			left	-	0.02	-	-	-	-	-	0.02	0.89	-	-	0.02	0.06
		Slow	front	-	-	0.02	0.02	-	0.11	-	-	-	0.82	0.05	-	-
			back	-	-	-	-	-	-	0.03	-	-	-	0.95	0.02	-
			right	-	-	-	-	-	-	-	0.08	0.03	-	0.02	0.83	0.05
			left	-	-	-	-	-	0.02	0.02	0.02	0.08	-	0.02	0.03	0.83

tapping and sliding patterns overall ($F(1, 167) = 10.9792$, $p = 0.0011$), no statistical differences were found within each category (tapping or sliding patterns).

Participants frequently reported experiencing the "hanger effect" [24], where haptic patterns, particularly sliding motions, naturally prompted them to move their heads in the indicated direction. This intuitive response underscores the potential of the system for navigational assistance, enabling users to orient themselves without visual or auditory input.

B. Haptic Feedback Navigation Performance Evaluation

This evaluation aimed to assess the effectiveness of haptic rendering as a navigational aid by analyzing user responses to four distinct haptic patterns: slide-front, slide-left, slide-right, and tap-front during navigation. These patterns were selected based on their recognition rates in the study from the previous section.

1) Experimental Setup: Six participants (4 males and 2 females, aged 21–35 years, mean 25.83 $\pm$5.45) were recruited for the study. None of them reported any deficiencies in sensorimotor function. The participants were informed about the experiments and agreed to the consent form. Before the evaluation, participants completed a training session. During the training session, the device was calibrated for accurate feedback delivery, and the patterns were presented 3 times each to familiarize themselves with the task.

This experiment was conducted in a 6 $\times$ 6 m room equipped with a VICON motion capture system for precise tracking of participant position and orientation. Participants navigated two predefined paths, consisting of 6 and 5 waypoints, respectively. The system calculated the position and orientation required to reach each waypoint. If participants deviated beyond a $\pm$0.3 m tolerance or $\pm$15 deg orientation threshold, corrective haptic patterns were rendered: slide-left to rotate left, and slide-right to rotate to the right. If there was no deviation, the pattern slide-front was rendered to indicate the users to walk forward. Upon reaching each waypoint, a tap-front pattern was delivered to confirm successful navigation.

2) Results: Fig. 4 illustrates the trajectories of all participants for both paths, with the orange area representing the $\pm$0.3 m tolerance zone. All participants successfully completed both paths, reaching each waypoint. The average

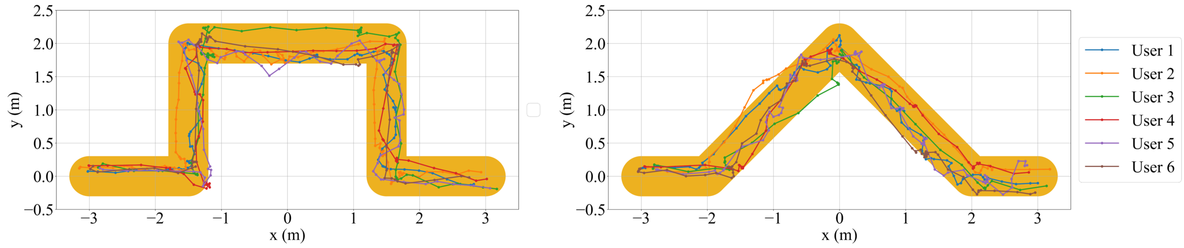

Fig. 4. Participants' trajectories for the two guiding paths. The orange area represents the ±0.3 m tolerance zone. Path 1 is shown on the left, and path 2 on the right.

Fig. 5. Example scenarios for the navigation system: (A) No obstacles, where the user moves through an open environment. (B) Static obstacles, where the user navigates around fixed objects. (C) Dynamic obstacles, involving interaction with moving objects.

completion time for path 1 was 2 minutes and 38.3 seconds, while path 2 averaged 2 minutes and 9 seconds. Participants spent an average of 4.27% of the time outside the tolerance zone for path 1 and 6.66% for path 2. The average number of times participants exited and re-entered the path was 1.33 for path 1 and 1.0 for path 2.

These findings demonstrate that participants successfully interpreted and followed the haptic feedback to navigate the paths. However, a key limitation was the pattern execution time of 1.25 seconds, which constrained responsiveness during rapid positional changes. Future work will explore alternative haptic techniques to enhance response times, particularly in scenarios involving variable user speeds or imminent collisions.

C. LLM and Video-Based Navigation Evaluation

This experiment assessed the integrated system's effectiveness in generating navigation cues through interaction between the video component and the language processing module. The study evaluated obstacle detection accuracy and the system's ability to provide timely haptic feedback under various environmental conditions. Three scenarios were tested: no obstacles, fixed barriers, and moving obstacles.

Twenty trials were conducted for each scenario, during which the system's decisions were compared against human judgment. Fig. 5 illustrates example scenarios, demonstrating the system's responses across different conditions.

1) Results: The system's performance across scenarios is summarized below:

- **No Obstacles**: In 20 trials without obstacles, the system made correct decisions in 91.8% of cases, closely matching human decision-making.
- **Static Obstacles**: For 20 trials involving fixed barriers, the system accurately guided users in 84.61% of cases, demonstrating reliable obstacle avoidance.
- **Dynamic Obstacles**: In 20 trials with moving obstacles, such as randomly moving chairs, the system achieved a correct response rate of 81.5%, despite increased complexity.

These results highlight the system's robustness in various scenarios, demonstrating high accuracy in obstacle detection and navigational cue generation, particularly in environments without or with static obstacles.

V. Conclusions and Future Work

This work presents a novel wearable navigation system designed to assist individuals with visual impairments by integrating haptic feedback, generative AI, and real-time object detection. The system features temple-mounted actuators driven by a five-bar linkage mechanism to deliver intuitive tactile cues. An ESP32 CAM camera captures video data, which is processed by a YOLO-World object detection module and a GPT-4o LLM to generate navigation commands. By combining these technologies, the system provides effective, real-time guidance without relying on auditory feedback, promoting accessibility and independence. Additionally, the hanger effect hypothesis is leveraged to enhance the intuitiveness of haptic feedback, guiding users naturally through tactile cues.

Three experimental evaluations validated the system's effectiveness. The first study assessed the recognition of 13 distinct haptic patterns, including tapping and sliding sensations at varying speeds. Participants achieved an average recognition rate of 81.3%, with sliding patterns showing the highest accuracy. The second study, conducted in a motion-tracked VICON environment, evaluated the system's ability to guide users along predefined paths. The final evaluation tested the integration of video-based navigation and LLM decision-making under three scenarios: no obstacles, fixed barriers, and moving objects. The system demonstrated high alignment with human decision-making, confirming its reliability in diverse conditions.

While these results highlight the system's potential, this work represents an early-stage implementation with opportunities for further refinement. Future research will focus on optimizing response times, particularly during rapid user movements, and evaluating performance in more complex environments, such as outdoor settings or varying lighting conditions. Additionally, enhancements to the haptic rendering mechanism and integration of advanced VLMs will improve obstacle detection and navigational precision.

With continued development, the system has the potential to become a comprehensive navigation aid for visually impaired individuals. By fostering greater independence, confidence, and accessibility, this technology can empower users to navigate complex environments, participate in social activities, and engage more fully in their communities.

REFERENCES

[1] A. M. Kappers, R. J. Holt, T. J. Junggeburth, M. F. S. Oen, B. J. van de Wetering, and M. A. Plaisier, "Hands-free haptic navigation devices for actual walking," *IEEE Transactions on Haptics*, pp. 1–18, 2024.

[2] S. Gallo, D. Chapuis, L. Santos-Carreras, Y. Kim, P. Retornaz, H. Bleuler, and R. Gassert, "Augmented white cane with multimodal haptic feedback," in *Proc. of the 3rd IEEE RAS & EMBS Int. Conf. on Biomedical Robotics and Biomechatronics*, 2010, pp. 149–155.

[3] F. Chinello, C. Pacchierotti, J. Bimbo, N. G. Tsagarakis, and D. Prattichizzo, "Design and evaluation of a wearable skin stretch device for haptic guidance," *IEEE Robotics and Automation Letters*, vol. 3, no. 1, pp. 524–531, 2018.

[4] W. Heuten, N. Henze, S. Boll, and M. Pielot, "Tactile wayfinder: a non-visual support system for wayfinding," in *Proc. of the 5th Nordic Conf. on Human-Computer Interaction: Building Bridges*, ser. NordiCHI '08, 2008, p. 172–181.

[5] B. Yang, L. He, K. Liu, and Z. Yan, "VIAssist: Adapting multi-modal large language models for users with visual impairments," 2024, arXiv:2404.02508.

[6] A. Hurst, A. Lerer, A. P. Goucher, A. Perelman, A. Ramesh, A. Clark, A. Ostrow, A. Welihinda, A. Hayes, A. Radford *et al.*, "GPT-4o system card," 2024, arxiv:2410.21276.

[7] H. Wang, J. Qin, A. Bastola, X. Chen, J. Suchanek, Z. Gong, and A. Razi, "VisionGPT: LLM-assisted real-time anomaly detection for safe visual navigation," 2024, arXiv:2403.12415.

[8] A. Kuzdeuov, O. Mukayev, S. Nurgaliyev, A. Kunbolsyn, and H. A. Varol, "ChatGPT for visually impaired and blind," in *Proc. of the Int. Conf. on Artificial Intelligence in Information and Communication (ICAIIC 2024)*, 2024, pp. 722–727.

[9] P.-A. Cabaret, C. Pacchierotti, M. Babel, and M. Marchal, "Design of haptic rendering techniques for navigating with a multi-actuator vibrotactile handle," in *Proc. of the 14th Int. Conf. on Human Haptic Sensing and Touch Enabled Computer Applications, (EuroHaptics 2024)*, 2024, p. 224–236.

[10] Y. H. Lee and G. Medioni, "RGB-D camera based wearable navigation system for the visually impaired," *Computer Vision and Image Understanding*, vol. 149, pp. 3–20, 2016.

[11] I. Lacôte, P.-A. Cabaret, C. Pacchierotti, M. Babel, D. Gueorguiev, and M. Marchal, "Do vibrotactile patterns on both hands improve guided navigation with a walker?" in *Proc. of the 14th Int. Conf. on Human Haptic Sensing and Touch Enabled Computer Applications, (EuroHaptics 2024)*, 2024, p. 391–404.

[12] M. Altamirano Cabrera, J. Tirado, J. Heredia, and D. Tsetserukou, "LinkGlide-S: A wearable multi-contact tactile display aimed at rendering object softness at the palm with impedance control in vr and telemanipulation," in *Proc. of the IEEE 18th Int. Conf. on Automation Science and Engineering (CASE 2022)*, 2022, pp. 647–652.

[13] A. M. L. Kappers, M. F. S. Oen, T. J. W. Junggeburth, and M. A. Plaisier, "Hand-held haptic navigation devices for actual walking," *IEEE Transactions on Haptics*, vol. 15, no. 4, pp. 655–666, 2022.

[14] S. Agrawal, M. E. West, and B. Hayes, "A novel perceptive robotic cane with haptic navigation for enabling vision-independent participation in the social dynamics of seat choice," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2022, pp. 9156–9163.

[15] Z. Liao, J. Salazar, and Y. Hirata, "Robotic guidance system for visually impaired users running outdoors using haptic feedback," in *Proc. of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, 2021, pp. 8325–8331.

[16] C. Bermejo and P. Hui, "A survey on haptic technologies for mobile augmented reality," *ACM Computing Surveys*, vol. 54, no. 9, 2021.

[17] S. Agrawal, "Assistive robotics for empowering humans with visual impairments to independently perform day-to-day tasks," in *Proc. of the Int. Conf. on Autonomous Agents and Multiagent Systems (AAMAS)*, 2023, p. 3023–3025.

[18] C. Jiang, E. Kuang, and M. Fan, "How can haptic feedback assist people with blind and low vision? a systematic review of 20 years of research," 2024, arxiv:2412.19105.

[19] J. Li, D. Li, C. Xiong, and S. Hoi, "BLIP: Bootstrapping language-image pre-training for unified vision-language understanding and generation," 2022, arXiv:2201.12086.

[20] G. Xu, P. Jin, H. Li, Y. Song, L. Sun, and L. Yuan, "LLaVA-CoT: Let vision language models reason step-by-step," 2025, arXiv:2411.10440.

[21] Q. Ren and T. Belpaeme, "Touched by ChatGPT: Using an LLM to drive affective tactile interaction," 2025, arxiv:2501.07224.

[22] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," 2016, arXiv:1506.02640.

[23] D. Tsetserukou, S. Hosokawa, and K. Terashima, "LinkTouch: A wearable haptic device with five-bar linkage mechanism for presentation of two-dof force feedback at the fingerpad," in *Proc. of the IEEE Haptics Symposium (HAPTICS 2014)*, 2014, pp. 307–312.

[24] Y. Kon, T. Nakamura, M. Sato, and H. Kajimoto, "Effect of hanger reflex on walking," in *Proc. of the IEEE Haptics Symposium (HAPTICS 2016)*, 2016, pp. 313–318.