# Enhancing Accessibility and User Experience through Voice-and-Text Multimodal AI: The Mijke Chatbot

Sichen Guo[1 [0009−0005−2163−4218]], Jun Hu[1 [0000−0003−2714−6264]], and Walter Baets[2 [0000−0003−1860−2483]]

[1] Eindhoven University of Technology, Groene Loper 3, 5612 AE Eindhoven, The Netherlands
info@tue.nl
https://www.tue.nl/en/
[2] Eindhoven Engine, Eindhoven, The Netherlands
https://eindhovenengine.nl/

**Abstract.** Mijke is an AI-powered chatbot designed to connect users, especially those with limited basic skills, to appropriate organizations and services in the Eindhoven. This work-in-progress project addresses two key challenges: improving accessibility for users with lower literacy skills and enhancing the accuracy of service matching. Mijke enables natural, voice-based interaction through WhatsApp by integrating speech-to-text and text-to-speech technologies, and a match validation process is under development to evaluate how well the chatbot aligns with user needs. The system also explores how tone of voice affects user trust and comfort. Informed by user testing and designed with multimodal interaction in mind, the project aims to reduce barriers to essential services and foster inclusive, adaptive digital support systems.

**Keywords:** Conversational AI · Low Literacy· Multimodal Interaction

## 1 Introduction

Mijke is the chatbot prototype from the Met Mij project [1], aiming to create an AI ecosystem that connects individuals with limited skills to local social support services through easy-to-use technology. People lacking basic skills—like literacy, numeracy, and problem-solving in tech-rich environments—face barriers in accessing public services. These "key competencies" from the PIAAC 2023 survey are essential for societal participation and daily support. PIAAC categorizes these competencies into five proficiency levels, with Level 1 reflecting basic tasks like finding specific information in a short text, while higher levels involve more complex information evaluation [2].

In the Netherlands, the PIAAC 2023 survey shows that 16% of adults aged 16 to 65 score at Level 1 or below in literacy, totaling about 2.2 million individuals with limited writing skills. Including those up to age 75, nearly 3 million adults

are low-skilled in key competencies. Additionally, eight in ten adults with low literacy also have low numeracy, indicating a substantial overlap [2].

To address these gaps, conversational agents and AI chatbots are increasingly adopted in digital service delivery. Prior studies show that multimodal AI—combining text, voice, and interface cues—improves accessibility, engagement, and trust, especially among underserved populations [3,4]. Saraswat et al. [5] emphasize the need for well-structured conversational flows and design strategies to guide users with minimal cognitive friction. In healthcare and senior support, multimodal systems help manage complexity and enhance the perceived helpfulness of digital assistants [6,7].

Relevant examples like Lukas et al. [8] and Candello et al. [9] highlight the effectiveness of WhatsApp-based conversational AI in multilingual, low-literacy contexts, where voice input and rich media enhance data quality and user comfort. From a design perspective, Ghosh et al. [10] and Tseng et al. [11] advocate for value-sensitive features such as trust-building responses, easy onboarding, and clear conversation scaffolds for misinterpretation or fallback scenarios.

The role of prosodic elements in voice interaction, including intonation, pitch, and speech rate, influences voice attractiveness and user engagement. Wang et al. [12] built on Zuckerman & Driver's framework [13] to evaluate AI-generated voices. Findings reveal that high speech rates and unnatural pitch reduce perceived trustworthiness and warmth. Although the original study didn't target users lacking basic skills, we adapted these prosodic insights into a simplified tone-of-voice questionnaire in Dutch and English to evaluate perceived voice quality for our target group.

The Mijke chatbot builds on existing work by adding voice interaction and a match validation system for users with limited reading and writing skills in the Dutch context. This work in progress study outlines the technical and methodological foundation of Mijke's ongoing development, contextualizes it within broader efforts toward digital inclusion, and shares insights from early implementation.

## 2 Method

The upcoming evaluation phase will adopt a mixed-methods approach to collect both quantitative and qualitative data. This includes a short version user experience questionnaire (UEQ), a tone-of-voice assessment, and semi-structured interviews. These instruments are designed to capture user perceptions of clarity, trust, and comfort in multimodal interaction.

In response to previous user testing, the voice-enabled interaction was also designed to support multilingual input. While not the primary focus of this study, Whisper's built-in multilingual capabilities enable users to speak in several languages beyond Dutch. This feature was particularly relevant during earlier pilots. Although Mijke's chatbot is primarily designed for NT1 users (Dutch native speakers) with limited literacy, due to recruitment challenges within this

group, some pilot testing was also conducted with NT2 users (Dutch as a second language), who expressed a preference for speaking in their native languages.

The development of the enhanced Mijke chatbot follows a three-part methodology:

***1. Voice-Enabled Interaction Layer*** A voice-enabled interaction layer was implemented to increase accessibility. The system uses OpenAI's Whisper API for speech-to-text transcription[3], enabling users to speak their requests through WhatsApp. Responses are returned in both text and audio formats, using OpenAI's text-to-speech API. This multimodal setup was informed by previous user testing, where participants expressed that receiving both spoken and written content simultaneously made information easier to understand and reduced stress. One participant noted that hearing the message while reading helped them process information faster and with greater confidence.

Although third-party APIs were used in this prototype, the development followed GDPR principles, with anonymized processing and secure message handling. Audio was streamed for transcription and synthesis but not stored, and no personal data was retained. Ethical approval was obtained from the Ethical Review Board at Eindhoven University of Technology. For future deployment, EU-hosted or on-device alternatives will be explored to enhance data sovereignty. An internship developer implemented this feature under the lead researcher's guidance, integrating user feedback to reduce barriers for text-only interaction.

***2. Multimodal Conversational Flow Design*** The conversational flow supports both voice-enabled and text-based input and output. It incorporates confirmation prompts and feedback checkpoints to enhance clarity and reduce confusion, especially in cases of transcription errors or ambiguous input. While the current evaluation focuses on Dutch-language interaction, the design accounts for language flexibility, improving onboarding instructions and inclusivity for diverse user groups. These design intentions are informed by best practices in multimodal chatbot design [5], particularly the emphasis on task-oriented flows and minimizing user uncertainty. As of writing, the updated version of the flow has not yet been evaluated with end users.

***3. Match Accuracy Evaluation Process & Tone of Voice Evaluation*** We are developing a process to evaluate Mijke's quality in matching users with services. This process focuses on intent recognition, referral relevance, and response clarity. It is based on service request patterns and conversation flows, refined through expert input and testing. The goal is to establish a practical baseline for assessing Mijke's service matching capabilities in real-world contexts.

In parallel, user perception of tone of voice is evaluated using a lightweight questionnaire adapted from prior MOS-related research [12,13]. The tool includes simplified language and visual cues to support users with low literacy,

[3] https://openai.com/index/whisper/

and it focuses on perceptual qualities including intelligibility, naturalness, voice preference, and tone appropriateness. It is designed to be low-barrier and accessible for the intended user group, without replicating full psychometric rigor. Given the exploratory nature of this research phase, its primary purpose was not rigorous psychometric validation but to guide practical adjustments to Mijke's tone of voice for improved user trust and comfort. These two evaluation tracks together offer insight into both functional precision and user experience quality.

## 3 Case Study: Pilot Implementation

An exploratory pilot session was conducted with a Dutch native speaker (NT1 user), who was recruited through the project partner organization. The participant previously experienced low literacy but no longer identifies as such to evaluate the naturalness and trustworthiness of the voice-interaction experience. The session combined task-based interactions with think-aloud reflection. The participant was encouraged to use voice messages. The participant ask for tax-related information and received spoken responses with accompanying text.

Key insights from the pilot study:

- **Voice felt more spontaneous but less controllable**: The participant appreciated the ability to speak directly, saying, *"I think, okay, I'm going to go for it now,"* but also noted they felt a bit more nervous than when typing.
- **Multimodal response improves comprehension**: They strongly valued receiving both voice and text, stating, *"While I listen, I can read along... I understand the information way faster."*
- **Tone of voice was generally pleasant but needed refinement**: The voice was described as *"very nice,"* but the participant suggested fewer accents and more natural breaks: *"She can also breathe."*
- **Desire for confirmation and reflection**: The participant wanted Mijke to confirm what was being asked, explaining, *"It would be really cool if she would still say this... that would show she understood my question."*
- **Potential confusion from incomplete or overly generic answers**: There was concern about the risk of following advice of the steps of preparing documents that might not fit one's situation: *"I might be frustrated if I followed her in the wrong... and I don't even need this folder."*
- **Social imagination of the chatbot**: Interestingly, the participant reflected on Mijke's behavior like a person: *"Maybe she's getting a cup of coffee... she's on the toilet."* This highlights the human-like expectations users may project onto AI.
- **Visual interface cues affect trust**: A small detail like WhatsApp's blue check mark gave the participant the impression that the chatbot was trustworthy and "*real*," suggesting that seemingly minor UI elements may play a surprisingly powerful role in shaping emotional connection and credibility.

These findings provide early validation of the multimodal setup while revealing actionable areas for refinement in voice interaction design, AI chat conversation flow optimization, confirmation mechanisms, and clarity of service guidance.

These highlights also provide the initiative validation for large-scale user testing procedure.

## 4 Future Work

The next project phase involves three main areas of focus:

1. **Comprehensive user testing** with users with lack of basic skills in the Eindhoven region to assess interaction preferences, trust levels, and overall accessibility.
2. **Establishment of a quantitative match accuracy baseline**, using the custom evaluation process to systematically assess and refine Mijke's service referral logic.
3. **Test and Improvement of the chatbot's tone of voice**, including refinements to speech pacing, prosody, and confirmation phrasing, guided by further iterations of the tone-of-voice evaluation tool.

These efforts will ensure the chatbot not only performs accurately but also communicates in a way that feels trustworthy, inclusive, and easy to engage with. Longer-term, we also anticipate expanding Mijke's capabilities to support multilingual and cross-domain referrals, aligned with broader goals of inclusive, AI-assisted public service navigation.

**Acknowledgments.** This project is part of a collaborative initiative supported by the Mijke development team and our organizational partners. We thank the internship developer for their contributions to the technical implementation of the voice-interaction feature. Special thanks also go to our pilot participant for their valuable feedback. This work was conducted within the context of an EngD research track focused on AI for social impact.

**About Sichen Guo**

Sichen Guo is an Engineering Doctorate (EngD) trainee at Eindhoven University of Technology, Netherlands, focusing on Human-System Interaction within socially impactful contexts. She is an active researcher at the Inclusive Society Lab at Eindhoven Engine, designing innovative technological solutions to bridge the gap between people with lack of basic skills and essential societal services. With degrees in Industrial Design (BSc, MSc), Sichen has explored inclusive design, multimodal user interaction, human-AI Interaction, virtual reality, and gamification. Her passion for user-centered innovation drives her to merge creativity and engineering, developing accessible and empathetic technologies that foster social inclusion.

## References


1. Jéssica Messias Goss Dos Santos. *Met Mij: AI-powered tool for identifying and connecting people and local services.* Engd thesis, December 2024. EngD thesis.
2. M. Buisman, I. Bollen, B. Jacobs, T. Huijts, R. Cornelisse, N. Van Guilik, D. Elshof, and L. Van Griensven. *PIAAC 2023: Core Skills of Adults. Results of the Dutch Survey 2023.* Kohnstamm Institute, 2024.
3. Yue Huang, Lichao Sun, Haoran Wang, Siyuan Wu, Qihui Zhang, Yuan Li, Chujie Gao, Yixin Huang, Wenhan Lyu, Yixuan Zhang, Xiner Li, Zhengliang Liu, Yixin Liu, Yijue Wang, Zhikun Zhang, Bertie Vidgen, Bhavya Kailkhura, Caiming Xiong, Chaowei Xiao, Chunyuan Li, Eric Xing, Furong Huang, Hao Liu, Heng Ji, Hongyi Wang, Huan Zhang, Huaxiu Yao, Manolis Kellis, Marinka Zitnik, Meng Jiang, Mohit Bansal, James Zou, Jian Pei, Jian Liu, Jianfeng Gao, Jiawei Han, Jieyu Zhao, Jiliang Tang, Jindong Wang, Joaquin Vanschoren, John Mitchell, Kai Shu, Kaidi Xu, Kai-Wei Chang, Lifang He, Lifu Huang, Michael Backes, Neil Zhenqiang Gong, Philip S. Yu, Pin-Yu Chen, Quanquan Gu, Ran Xu, Rex Ying, Shuiwang Ji, Suman Jana, Tianlong Chen, Tianming Liu, Tianyi Zhou, William Wang, Xiang Li, Xiangliang Zhang, Xiao Wang, Xing Xie, Xun Chen, Xuyu Wang, Yan Liu, Yanfang Ye, Yinzhi Cao, Yong Chen, and Yue Zhao. Trustllm: Trustworthiness in large language models. https://arxiv.org/abs/2401.05561, 2024.
4. Gobinath A, Manjula Devi C, Suthan Raja S J, Prakash P, Anandan M, and Srinivasan A. Voice assistant with ai chat integration using openai. In *2024 Third International Conference on Intelligent Techniques in Control, Optimization and Signal Processing (INCOS)*, pages 1–6, 2024. https://doi.org/10.1109/INCOS59338.2024.10527726.
5. Pavi Saraswat, Bharat Bhardwaj, Prashant Naresh, Alaknanda Ashok, Raja Kumar, and Manish Kumar. Voice assistants and chatbots: Hands-on essentials of ui and feature design, development, and testing. In P. Kumar, A. Tomar, and R. Sharmila, editors, *Emerging Technologies in Computing: Theory, Practice, and Advances*, pages 217–240. Chapman and Hall/CRC, 1st edition, 2021. https://doi.org/10.1201/9781003121466-11.
6. Yuhao Chen, Jiahao Cai, Siyu Chen, Farhana Zulkernine, Nauman Jaffar, Amina Almarzouqi, Nabeel Al-Yateem, and Syed Aziz Rahman. Assist-bot: A voice-enabled assistant for seniors. In *2024 IEEE 48th Annual Computers, Software, and Applications Conference (COMPSAC)*, pages 1901–1906, Los Alamitos, CA, USA, July 2024. IEEE Computer Society. https://doi.ieeecomputersociety.org/10.1109/COMPSAC61105.2024.00301.
7. Divya Rao. The future of healthcare using multimodal ai: Technology that can read, see, hear and sense. *Oral Oncology Reports*, 10:100340, 2024. https://www.sciencedirect.com/science/article/pii/S2772906024001869.
8. Lukas Mueller, Jackson Mughuma, and Ulrich von Zadow. Exploratory study: How the usage of a whatsapp-based chatbot influences data collection in subsaharan africa. In *Proceedings of the 6th ACM Conference on Conversational User Interfaces*, New York, NY, USA, 2024. Association for Computing Machinery. https://doi.org/10.1145/3640794.3665584.
9. Heloisa Candello, Gabriel Soella, and Leandro Nascimento. Designing multi-model conversational ai financial systems: understanding sensitive values of women entrepreneurs in brazil. In *Proceedings of the 2024 ACM International Conference on Interactive Media Experiences Workshops*, pages 11–18, New York, NY, USA, 2024. Association for Computing Machinery. https://doi.org/10.1145/3672406.3672409.

10. Debmitra Ghosh, Sayani Ghatak, Hrithik Paul, et al. A proposed cognitive framework model for a student support voice based chatbot using xai. Preprint (Version 1) available at Research Square, May 2023. https://doi.org/10.21203/rs.3.rs-2888180/v1.
11. Yuan-Chi Tseng, Weerachaya Jarupreechachan, and Tuan-He Lee. Understanding the benefits and design of chatbots to meet the healthcare needs of migrant workers. *Proceedings of the ACM on Human-Computer Interaction*, 7(CSCW2), October 2023. https://doi.org/10.1145/3610106.
12. Yihui Wang, Haocheng Lu, and Gaowu Wang. A pilot study on the prosodic factors influencing voice attractiveness of ai speech. In Jia Jia, Zhenhua Ling, Xie Chen, Ya Li, and Zixing Zhang, editors, *Man-Machine Speech Communication*, pages 316–329, Singapore, 2024. Springer Nature Singapore. https://doi.org/10.1007/978-981-97-0601-3_27.
13. Miron Zuckerman and Robert E. Driver. What sounds beautiful is good: The vocal attractiveness stereotype. *Journal of Nonverbal Behavior*, 13(2):67–82, June 1989. https://doi.org/10.1007/BF00990791.