

Exploring the Advantages of Personalized LLM-Generated Responses for Visually Impaired Users

Master's Thesis at the
Media Computing Group
Prof. Dr. Jan Borchers
Computer Science Department
RWTH Aachen University

by
Aleksandra Kukawka

Thesis advisor:
Prof. Dr. Jan Borchers

Second examiner:
Prof. Dr. Ulrik Schroeder

Registration date: 07.06.2024
Submission date: 18.11.2024

Eidesstattliche Versicherung

Declaration of Academic Integrity

Kukawka, Aleksandra

Name, Vorname/Last Name, First Name

448975

Matrikelnummer (freiwillige Angabe)
Student ID Number (optional)

Ich versichere hiermit an Eides Statt, dass ich die vorliegende ~~Arbeit/Bachelorarbeit/~~
Masterarbeit* mit dem Titel

I hereby declare under penalty of perjury that I have completed the present ~~paper/bachelor's thesis/~~master's thesis* entitled

Exploring the Advantages of Personalized LLM-Generated Responses for Visually Impaired Users

selbstständig und ohne unzulässige fremde Hilfe (insbes. akademisches Ghostwriting) erbracht habe. Ich habe keine anderen als die angegebenen Quellen und Hilfsmittel benutzt; dies umfasst insbesondere auch Software und Dienste zur Sprach-, Text- und Medienproduktion. Ich erkläre, dass für den Fall, dass die Arbeit in unterschiedlichen Formen eingereicht wird (z.B. elektronisch, gedruckt, geplottet, auf einem Datenträger) alle eingereichten Versionen vollständig übereinstimmen. Die Arbeit hat in gleicher oder ähnlicher Form noch keiner Prüfungsbehörde vorgelegen.

independently and without unauthorized assistance from third parties (in particular academic ghostwriting). I have not used any other sources or aids than those indicated; this includes in particular software and services for language, text, and media production. In the event that the work is submitted in different formats (e.g. electronically, printed, plotted, on a data carrier), I declare that all the submitted versions are fully identical. I have not previously submitted this work, either in the same or a similar form to an examination body.

Aachen, 18.11.2024

Ort, Datum/City, Date

Unterschrift/Signature

*Nichtzutreffendes bitte streichen/Please delete as appropriate

Belehrung:

Official Notification:

§ 156 StGB: Falsche Versicherung an Eides Statt

Wer vor einer zur Abnahme einer Versicherung an Eides Statt zuständigen Behörde eine solche Versicherung falsch abgibt oder unter Berufung auf eine solche Versicherung falsch aussagt, wird mit Freiheitsstrafe bis zu drei Jahren oder mit Geldstrafe bestraft.

§ 156 StGB (German Criminal Code): False Unsworn Declarations

Whosoever before a public authority competent to administer unsworn declarations (including Declarations of Academic Integrity) falsely submits such a declaration or falsely testifies while referring to such a declaration shall be liable to imprisonment for a term not exceeding three years or to a fine.

§ 161 StGB: Fahrlässiger Falscheid; fahrlässige falsche Versicherung an Eides Statt

(1) Wenn eine der in den §§ 154 bis 156 bezeichneten Handlungen aus Fahrlässigkeit begangen worden ist, so tritt Freiheitsstrafe bis zu einem Jahr oder Geldstrafe ein.

(2) Straflosigkeit tritt ein, wenn der Täter die falsche Angabe rechtzeitig berichtigt. Die Vorschriften des § 158 Abs. 2 und 3 gelten entsprechend.

§ 161 StGB (German Criminal Code): False Unsworn Declarations Due to Negligence

(1) If an individual commits one of the offenses listed in §§ 154 to 156 due to negligence, they are liable to imprisonment for a term not exceeding one year or to a fine.

(2) The offender shall be exempt from liability if they correct their false testimony in time. The provisions of § 158 (2) and (3) shall apply accordingly.

Die vorstehende Belehrung habe ich zur Kenntnis genommen:

I have read and understood the above official notification:

Aachen, 18.11.2024

Ort, Datum/City, Date

Unterschrift/Signature

Contents

Abstract	ix
Überblick	xi
Acknowledgments	xiii
Conventions	xv
1 Introduction	1
1.1 Accessibility and Visual Impairment	1
1.2 Challenges Faced by Blind and Visually Impaired Individuals	2
1.3 Overview of Assistive Technologies	3
1.4 Motivation	4
1.5 Outline	5
2 Related Work	7
2.1 Assistive Technologies and Their Limitations	7
2.1.1 Screen Readers	7

2.1.2	Voice Assistants	8
2.2	Addressing the Challenges	9
2.2.1	Combining Screen Readers and Voice Assistants	10
2.2.2	Large Language Models	11
2.2.3	Actions	13
2.3	Information Retrieval	14
2.3.1	Classifying Search Activities	14
3	Methodology	17
3.1	User Study	17
3.1.1	Grounded Theory Method	17
3.1.2	Data Collection Methods and Sources	19
3.1.3	Interactive Interviews	20
	App Prototype	21
	Study Procedure	24
	Ethical Considerations	26
	Participants	27
3.1.4	Data Analysis	27
3.2	Results	28
3.2.1	Assistive Technologies and Challenges	28
3.2.2	AI Voice Assistants With LLM-Generated Responses	31
	Use Cases	31

Advantages	34
Comparison to Other Assistive Technologies	38
Limitations and Disadvantages	40
Design Recommendations From Participants	46
4 Discussion	53
4.1 Theories	53
4.1.1 No matter how good the AI is, screen readers remain irreplaceable	53
4.1.2 The most crucial feature sought by BVI users is the ability to describe images and ask follow-up questions	55
4.1.3 AI assistants can be effective information retrieval tools for BVI users	56
4.1.4 BVI users expect AI systems to be flexible	57
4.1.5 BVI users' big frustration with voice assistants is the lack of understanding	59
4.1.6 Inaccurate LLM-generated responses lead to BVI users' mistrust of AI systems	60
4.1.7 An AI system that performs actions on behalf of users could be beneficial for BVI individuals	61
4.1.8 BVI users seek safe AI systems that do not share their data	62
4.1.9 Simplicity in AI systems is the key	63
4.1.10 BVI users do not see much difference between an AI-based voice assistant and a traditional one	65
4.1.11 Users take sustainability into account when using AI	65
4.2 Design Considerations	66

4.3	Limitations	68
5	Summary and Future Work	69
5.1	Summary and Contributions	69
5.2	Future Work	71
A	List of Most Important Interview Questions for User Study	73
	Bibliography	75
	Index	81

List of Figures and Tables

2.1	<i>Classified Search Actions</i> A diagram showing search actions classified into three categories: lookup, learn and investigate [Marchionini, 2006].	15
3.1	<i>Blind and Low-Vision Assistant</i> An app prototype used in the study sessions.	21
3.2	<i>Blind and Low-Vision Assistant With a Text Input</i> A backup app prototype for the study sessions.	24
4.1	<i>Visualization of Theory 4.1.4</i> BVI users expect AI systems to be flexible.	58
4.2	<i>Visualization of Theory 4.1.9</i> Simplicity in AI systems is the key.	64
4.3	<i>Design Considerations for AI System With LLM-Generated Responses</i> Recommendations based on the study results and literature review.	67

Abstract

Blind and visually impaired (BVI) people face many challenges in various areas of life, particularly in accessing digital content. They rely on assistive technologies like screen readers and voice assistants, but these tools have limitations. With the growth of artificial intelligence (AI), there is an opportunity to enhance the existing assistive technologies and develop new ones. This thesis investigates the potential of AI and responses generated by large language models (LLMs) to help BVI people in their daily lives. Using a combination of prototype testing, user interviews, and online surveys, we employ the grounded theory approach to understand the needs and preferences of BVI people regarding AI-based systems. We answer the following research question: What are the advantages and disadvantages of AI voice assistants with LLM-generated responses for BVI users, and how can they be improved? The findings suggest that AI can help BVI people in their daily lives. They value the personalisation and natural interactions such a system can provide. However, due to existing limitations of AI technologies, such as privacy concerns, efficiency, or inaccuracy, they still prefer to use traditional tools like screen readers over entirely replacing them with AI. Our work contributes to understanding the potential of AI for BVI users and provides insights into how AI-based systems can be designed to meet their needs better.

Überblick

Blinde und sehbehinderte Menschen sehen sich in verschiedenen Lebensbereichen mit vielen Herausforderungen konfrontiert, etwa beim Zugang zu digitalen Inhalten. Sie sind auf unterstützende Technologien wie Bildschirmlesegeräte und Sprachassistenten angewiesen, aber diese Hilfsmittel haben ihre Grenzen. Mit dem Aufkommen der künstlichen Intelligenz (KI) besteht die Möglichkeit, die bestehenden Hilfstechnologien zu verbessern und neue zu entwickeln. In dieser Arbeit wird das Potenzial der KI und der von Large Language Models (LLMs) generierten Antworten untersucht, um blinde und sehbehinderte Menschen in ihrem täglichen Leben zu unterstützen. Mithilfe einer Kombination aus Prototypentests, Nutzerinterviews und Online-Umfragen wenden wir den Ansatz der Grounded Theory an, um die Bedürfnisse und Präferenzen von blinden und sehbehinderten Menschen in Bezug auf KI-basierte Systeme zu verstehen. Wir beantworten die folgende Forschungsfrage: Was sind die Vor- und Nachteile von KI-Sprachassistenten mit LLM-generierten Antworten für blinde und sehbehinderte Menschen, und wie können sie verbessert werden? Die Ergebnisse deuten darauf hin, dass KI den blinden und sehbehinderten Menschen in ihrem täglichen Leben helfen kann. Sie schätzen die Personalisierung und die natürlichen Interaktionen, die ein solches System bieten kann. Aufgrund bestehender Einschränkungen von KI-Technologien, wie z.B. Bedenken hinsichtlich des Datenschutzes, der Effizienz oder der Ungenauigkeit, ziehen sie es jedoch immer noch vor, herkömmliche Hilfsmittel wie Bildschirmlesegeräte zu verwenden, anstatt sie vollständig durch KI zu ersetzen. Unsere Arbeit trägt dazu bei, das Potenzial von KI für blinde und sehbehinderte Nutzende zu verstehen, und gibt Aufschluss darüber, wie KI-basierte Systeme so gestaltet werden können, dass sie ihre Bedürfnisse besser erfüllen.

Acknowledgments

Firstly, I would like to thank Prof. Dr. Jan Borchers and Prof. Dr. Ulrik Schroeder for examining my thesis.

Secondly, I would like to thank my advisor - Sarah Sahabi. Thank you for guiding me through this journey. Working and learning from you was an honour, and I am grateful for your support, patience, and understanding. Your responses to my questions, even during late hours, were always invaluable. Thank you for believing in me and my work.

To all my study participants, thank you for your time and effort. Your feedback was crucial for the success of this research. I am grateful for your willingness to share your experiences and thoughts.

Special thanks to my husband, who has been my rock and my biggest supporter since the beginning of my whole academic journey. I would not do it without you.

Finally, many thanks to my parents, my sister and my grandma for their support and encouragement throughout my studies. You have always been there for me, and I am grateful for your love and understanding.

Conventions

Throughout this thesis, we use the following conventions:

- The thesis is written in British English.
- The first person is written in plural form.
- Unidentified third persons are described in plural form.

Where appropriate, paragraphs are summarized by one or two sentences positioned at the margin of the page.

This is a summary of a paragraph.

Source code and implementation symbols are written in typewriter-style text.

Chapter 1

Introduction

The World Health Organization defines disability as a condition that reflects the interaction between a person's health issues and factors such as the environment and personal characteristics [WHO, 2023]. According to them, 16% of the global population live with disabilities. WHO expects this number to increase due to higher life expectancy and the rise of chronic health conditions. Moreover, they confirm that people with disabilities experience more limitations in everyday life than others.

Definition and statistics on disability

1.1 Accessibility and Visual Impairment

Even in 2024, people with disabilities are still facing issues with accessibility because their environments create barriers to equal participation in society [WHO, 2023]. The organisation emphasises that inclusion is essential for achieving the Sustainable Development Goals¹ and global health priorities². Improving accessibility can be accomplished by designing inclusive products that address the challenges faced by people with disabilities in their daily lives³.

Accessibility barriers persist for people with disabilities

¹ <https://www.globalgoals.org/goals/>

² <https://www.who.int/europe/about-us/our-work/core-priorities>

³ <https://www.who.int/health-topics/disability>

Disabilities enclose
diverse challenges and
needs

WHO explains that the term "people with disabilities" represents a wide range of individuals. They call this group diverse and clarify that it includes people with different needs - there are many types of disabilities, and each of them poses various challenges. Some of the most common types of disabilities affect vision, movement, thinking, remembering, learning, communicating, hearing, mental health or social relationships [WHO, 2023].

Global statistics on
vision impairments

Our research focuses on people with vision impairments. In 2020, 49.1 million people globally were blind, and 225 million had moderate to severe vision impairment [Bourne et al., 2020]. The number of blind and visually impaired (BVI) people is expected to increase due to the ageing population [Bourne et al., 2020].

1.2 Challenges Faced by Blind and Visually Impaired Individuals

Visually impaired
people encounter
challenges

BVI individuals face many challenges in various areas of life, such as education and employment. According to the National Federation of the Blind in the United States⁴, only around 10% of the people who were legally blind in the US were Braille readers in 2009. Around 70% of blind adults were unemployed, and 50% of blind high school students dropped out of school.

Daily difficulties faced
by BVI individuals

Daily tasks such as identifying objects, reading or navigating the environment can be challenging for BVI people and impact their independence. Reliance on visual information often makes it difficult to perform activities of daily living such as cooking, cleaning, or shopping [Bhosale et al., 2023]. Moreover, BVI people often face social isolation and frustration due to a lack of visual cues, which may complicate social interactions and cause mental health issues [Rajendran et al., 2024; Filippini et al., 2024]. Mobility is another challenge for BVI people because they must rely on alternative senses or assistive technologies to navi-

⁴ https://nfb.org/images/nfb/documents/pdf/braille_literacy_report_web.pdf

gate the environment [Saleem and Sivakumar, 2024]. In the digital world, BVI people face challenges when accessing digital content such as dynamic web pages that are hard to browse via a screen reader, inaccessible images or complex interfaces [Stangl et al., 2020; Phutane et al., 2023].

1.3 Overview of Assistive Technologies

Assistive technologies are designed to help people with disabilities perform tasks that they would otherwise find difficult or impossible. Non-visual access to digital content is crucial for BVI individuals as they rely on assistive technologies to interact with interfaces [Phutane et al., 2023].

Assistive technologies aid disabled individuals in challenging tasks

Screen readers are the most common assistive technology BVI people use to access interfaces [Phutane et al., 2023]. A screen reader is software that renders the screen's content into speech or Braille output. Users can navigate through the content using keyboard shortcuts or gestures. Popular screen readers are JAWS⁵, NVDA⁶, Apple's VoiceOver⁷ and Google's TalkBack⁸.

Screen readers assist BVI users

Another standard technology BVI people use is voice assistants, which offer a hands-free audio-based interaction with the device. Examples of popular voice assistants include Amazon's Alexa⁹, Google Assistant¹⁰, and Apple's Siri¹¹. Although they were designed for the general audience, BVI people are their heavy users. They would benefit from more advanced features tailored to their needs, such as synchronisation between visual and non-visual cues, access to third-party devices, or memory support [Abdolrahmani et al., 2018; Pradhan et al., 2018].

Voice assistants provide support for BVI users

⁵ <https://www.freedomscientific.com/products/software/jaws/>

⁶ <https://www.nvaccess.org/>

⁷ https://www.apple.com/voiceover/info/guide/_1121.html

⁸ <https://support.google.com/accessibility/android/answer/6007100>

⁹ <https://www.alexa.com/>

¹⁰ <https://assistant.google.com/>

¹¹ <https://www.apple.com/siri/>

Other tools that help
BVI individuals

There are also other assistive technologies available. The examples include:

- Braille displays - devices that convert digital text into Braille output,
- magnifiers - software that enlarges the screen content,
- optical character recognition (OCR) - software that converts images of text into digital text and
- apps dedicated specifically for BVI people (navigation apps, object recognition apps, etc.)

Assistive technologies
pose challenges related
to accessibility,
understanding, and
personalisation

None of these technologies are perfect and pose different challenges. Screen readers struggle with rendering complex web content despite applying different web accessibility standards such as Web Content Accessibility Guidelines (WCAG) [Caldwell et al., 2008]. Voice assistants cause social awkwardness and privacy concerns. That is why users often prefer to use their keyboards and computers. Moreover, voice assistants often misinterpret user intent, not providing the desired information [Abdolrahmani et al., 2018]. Both screen readers and voice assistants cannot learn over time and adapt to the user’s preferences. They are not able to provide personalised information and recommendations.

AI assistants for BVI
users are the next step

Phutane et al. [2023] addressed some of these challenges and conducted a study to understand BVI users’ current screen reader and voice assistant practices. They suggested that the next step is an artificial intelligent (AI) conversational assistant that acts as both a virtual and personal assistant. To our knowledge, no research has yet focused on the potential of AI-generated responses tailored specifically to BVI people.

1.4 Motivation

In this thesis, we aim to explore the potential of responses generated by large language models (LLMs) and their abil-

ity to help BVI people access digital content and retrieve information from the web. We explore the advantages and limitations of such a solution and compare it to other assistive technologies, such as screen readers. The focus is on the potential of LLMs to support BVI individuals in information retrieval and performing actions on the device on behalf of the user. The study is conducted according to the principles of the grounded theory method [Glaser and Strauss, 1967] to derive theories from the data. It includes an interactive part where we present the participants with a very simple app prototype, a GPT-enabled voice assistant, and an interview part.

We explore how LLMs can assist BVI individuals, especially in information retrieval and performing actions on behalf of the user

The findings reveal that BVI users are open to using AI-based systems and appreciate personalised responses generated by LLMs because they make the interaction feel natural. Despite advancements, AI technologies still have limitations and cannot fully replace traditional assistive tools like screen readers.

Results show that BVI users value AI-based systems, but there exist limitations

1.5 Outline

After introducing the topic in Chapter 1, we present the related work in Chapter 2. We discuss the assistive technologies focusing on screen readers and voice assistants. The challenges they pose and the current research on addressing them are described. The focus is on LLMs and their potential to help BVI people retrieve information and access digital content in general. Moreover, we go one step further and describe how LLMs can assist users in performing actions on their devices.

In Chapter 3, we present the methodology. The grounded theory approach and the data collection process are described, and the whole study procedure is explained. The app prototype development and the limitations of the app itself are reported. We also discuss the ethical considerations regarding conducting research with BVI people and the data analysis process. The results are presented in the last part of this chapter.

Chapter 4 includes the results of the study together with the discussion. We present the most relevant theories derived from the data and discuss them in the context of the related work. We also consider how the findings may impact the design of AI-based systems for BVI users, highlighting key design considerations. Limitations of the study are also discussed.

Chapter 5 concludes the thesis, summarising the most important findings and proposing future work.

Chapter 2

Related Work

The following sections provide an overview of assistive technologies and their limitations. We focus on current research addressing these limitations and discuss the implications of utilising AI in this context. Additionally, we examine existing solutions to perform actions on behalf of the user. Finally, we review the work related to information retrieval, including classifying search actions.

2.1 Assistive Technologies and Their Limitations

BVI individuals use various tools to interact with computers and the web. While screen readers are typically used as a primary tool, the use of voice assistants has noticeably increased in recent years [Phutane et al., 2023].

Screen readers and voice assistants are most common tools

2.1.1 Screen Readers

Screen readers read out loud the visual content of the screen to the user and enable them to interact with the computer using keyboard shortcuts [Phutane et al., 2023]. They

Screen readers facilitate interaction with computers

are available on various platforms like Windows, macOS, iOS, and Android.

Web content accessibility is often problematic for screen readers

Many studies have shown that web content is not always accessible to screen readers. It is not always a smooth experience due to dynamic content changes, making it difficult for screen readers to keep up with the screen [Kim et al., 2021]. The images on the web are also a challenge for a screen reader because of the lack of alternative text or explanatory image descriptions [Stangl et al., 2020; Guinness et al., 2018]. Many users must rely on the context of the surrounding text to understand what is on the image [Bigham et al., 2007]. What is also challenging is the navigation of the web pages. The lack of proper headings, landmarks or ARIA (Accessible Rich Internet Applications) roles makes it daunting to go to the desired element on the page [Borodin et al., 2010]. That is why users often employ custom strategies to navigate web pages, like using the back button [Joyner et al., 2022; Borodin et al., 2010]. What helps designers and developers overcome navigation issues are the official guidelines such as WCAG on designing different content types, including images, charts, graphs and many others [Caldwell et al., 2008; Joyner et al., 2022].

Other screen reader challenges

One challenge not arising from the web accessibility issues but from the device itself is that the screen reader's commands or touch gestures are not always intuitive and require memorising [Vtyurina et al., 2019]. Regarding search and information retrieval, Sahib et al. [2012] found that BVI users usually submit fewer queries, and their exploration of the search results is less investigational than sighted users.

2.1.2 Voice Assistants

Rise in voice assistant usage

Tudor Car et al. [2020] noticed that there has been a rise in the use of voice assistants like Alexa, Siri or Google Assistant in the last few years. 50% of the US population uses voice search daily¹.

¹ <https://upcity.com/experts/consumers-and-voice-search-study/>

Pradhan et al. [2018] mentioned that voice assistants allow users to ask for information such as the weather, control smart home devices and even perform online tasks like shopping. They conducted two studies to explore how BVI individuals can use them. The most common tasks were listening to music, looking up information, checking the weather, playing audiobooks and home automation. The authors found that the three main benefits of using voice assistants for BVI people are efficiency, user independence and the ability to replace various other technologies. Vtyurina et al. [2019] reported that voice assistants require minimal to no training and that their portability makes them always ready to use.

Voice assistants enhance the independence of BVI users and efficiency in completing various tasks

However, a few limitations exist, such as difficulty for users in discovering unknown features and a lack of more advanced ones [Pradhan et al., 2018]. Abdolrahmani et al. [2018] found that voice assistants often respond too detailed or vague and sometimes cause social awkwardness when used in public. Moreover, Vtyurina et al. [2019] reported that voice assistants provide a single answer for a simple question and thus do not allow for exploration of the search results. They noted that while speaking is generally faster than typing, it can sometimes lead to speech recognition errors, which can be frustrating.

Voice assistants have limitations

2.2 Addressing the Challenges

Phutane et al. [2023] emphasised the importance of addressing these challenges by discovering new methods for web interaction, ensuring that BVI users can access digital content without barriers.

The need for new ways of accessible interaction

One of the solutions presented by Ashok et al. [2019] is SuggestOmatic, a system designed to improve web browsing by auto-suggesting the following browsing action in screen readers. It utilises the action history to predict and suggest the next move, decreasing the shortcuts used to conduct the same task. Their user study showed that this system could reduce browsing task times by up to 29% compared to other web automation tools.

SuggestOmatic enhances web browsing by predicting user actions

2.2.1 Combining Screen Readers and Voice Assistants

VERSE combines screen readers with voice assistants

Vtyurina et al. [2019] explored the opportunities of combining screen readers with voice-based virtual assistants and created a prototype called VERSE. Users can interact with it mainly through voice commands, similar to popular voice assistants. The system allows gestures on a so-called "companion device", such as a smartphone or a smartwatch. It gives users a general answer and enables them to explore additional sources the search engine offers if desired. Quickly and easily, users can switch between different search results using either voice commands or gestures. The authors conducted a study with BVI individuals to evaluate the prototype. Participants found it easy to learn but mentioned that it could not replace a traditional screen reader. There arose a need to improve the system's ability to hold a more natural conversation and to improve the language understanding part. This finding led authors to suggest future work on using AI.

A different study found that users prefer voice-enabled screen readers because they are easier to use

In another study by Ashok et al. [2015], the use of a traditional screen reader enabled by speech was explored. Contrary to the previous one, the findings showed that users prefer to use a screen reader with a voice over the usual screen reader. Participants mentioned the advantages, such as no need to remember and use keyboard shortcuts and no need to spend time and effort locating the desired content.

The Firefox Voice extension was developed as a tool for voice command interaction with the browser

Cambre et al. [2021] proposed a similar solution but in the form of an open-source browser extension called Firefox Voice. This extension enables users to interact with the browser using voice commands. It was not specifically designed for BVI users, but the authors consulted the design with an expert in screen readers and accessibility. However, the extension is no longer maintained due to various usability issues, such as too many voice commands that were hard to remember.

Key needs for a system integrating screen readers and voice assistants

Phutane et al. [2023] focused on identifying critical needs and design considerations for such solutions. They conducted a formative study with 14 BVI users to understand

current screen reader and conversational assistant practices. They asked the participants, "What if you could talk to your screen reader?" The main identified key needs are:

- | | |
|--|---|
| <ol style="list-style-type: none"> 1. Conversational assistants should be adaptive like AI and customisable like screen readers. They should adapt to different types of searches and information-seeking processes, preserving the tasks' confidentiality. They should learn user's behaviour and adapt to their personalities and moods. On the other hand, they should be customisable to the user's needs and preferences. Users want to customise the assistant's speech rate, visual semantic settings, information order, and voice. | <p>It should be adaptive and customisable</p> |
| <ol style="list-style-type: none"> 2. Conversational assistants should afford different levels of control, from granular control for cursors and voice commands to high-level control for task assistance. | <p>It should allow different control levels</p> |
| <ol style="list-style-type: none"> 3. Conversational assistants should be able to serve in public settings while maintaining privacy. | <p>It should be confidential</p> |

Discussion revealed that AI-based conversational assistants could be a promising solution that addresses these needs, and recent advancements in LLMs, especially OpenAI's GPT models, are worth exploring. Due to the ability to hold natural conversations and answer follow-up questions, LLMs can be used to create conversational assistants that can help BVI users access web content. Further research is needed to understand how LLMs can help with information seeking.

AI assistants may fulfill these needs

LLMs can assist BVI users in accessing web content

2.2.2 Large Language Models

An LLM is a type of AI that generates human-like text. Such models are trained on enormous amounts of text data and can be fine-tuned to perform specific tasks. OpenAI's GPT models² have gained significant popularity. One of the

LLM generates human-like text and can browse the web to answer questions

² <https://openai.com/index/chatgpt/>

most interesting models is WebGPT [Hilton et al., 2021]. It was developed to allow users to browse the web and answer questions from the web pages. This variant can ask for search queries, click on links, read the web page’s content, and provide information sources. It was trained on human examples and learned from an algorithm analysing the types of answers users prefer.

LLMs aid web access

As mentioned in the 2.2.1, LLMs can be used to create conversational assistants that can help BVI users access web content.

An accessible mobile app for BVI users to interact with OpenAI’s ChatGPT

Kuzdeuov et al. [2024] made an artefact contribution and designed a mobile app to interact with OpenAI’s ChatGPT in an accessible way for BVI users and open-sourced the code. They used the keyword spotting model and the voice activity detection model. Like most popular voice assistants, the app detects a keyword and listens to the input. It uses automatic speech recognition and text-to-speech conversion. The user interface contains only four icons for each mode: listening, recording, processing and speaking.

The GPT-4o model enhances human-computer interactions through voice input and output, enabling human-like conversations

The authors no longer maintain the repository, but new techniques such as OpenAI’s GPT-4o model³ are yet emerging. As of November 2024, it is the newest model that allows for even more natural human-computer interaction. It enables voice input and offers a smooth, human-like interaction. It is enough to press the button once at the beginning of the conversation. The model identifies when the user has finished speaking and provides a voice output responding to the user’s question. After it finishes talking, it waits for the user’s response. The conversation continues until the user decides to stop. This approach could also be explored as a way to assist BVI individuals.

There is no further research on LLMs for BVI

To our knowledge, no more research has been conducted yet to explore how LLMs can help with information-seeking for BVI users.

³ <https://openai.com/index/hello-gpt-4o/>

2.2.3 Actions

The next step is performing actions on the device on behalf of the user. Examples of actions include booking a flight, adding a task to the calendar, ordering an Uber ride to some specific location, etc. In addition, poorly designed web layouts with financial details, CAPTCHAs, and other complex forms are challenging for BVI users and often require sighted assistance [Murphy et al., 2008]. Although it has not yet been researched explicitly in the context of BVI users, triggering actions on the user's behalf could make them more independent.

The ability to perform actions on devices for BVI users, such as booking flights or ordering rides, could enhance their independence

Recently introduced Rabbit R1 device⁴ addresses issues of smartphones' app-based operating systems and voice assistants. The authors argue that although OpenAI's GPT models are very good at understanding intentions, they cannot do actions. The goals of the Rabbit R1 device are to trigger actions on behalf of users across all environments and be as universal as possible, independent of the app or the web browser. The authors introduce the large action model (LAM), an LLM that executes human intentions on computers. It learns by demonstrations - users show it how to perform actions on the device, and it remembers the steps. Either voice commands and different gestures or a keyboard can control the device. To interact with third-party apps, users must log in through another device.

The Rabbit R1 device aims to address limitations of operating systems and voice assistants by learning through demonstrations and triggering actions

This approach could assist BVI users by enabling actions on their devices without having to navigate through the interface with a screen reader. Due to complex interfaces, BVI users spend too much time and effort performing simple day-to-day tasks on the device. This is a challenge, particularly for novices using screen readers and knowing only basic commands [Ashok et al., 2015].

This approach could help BVI users by reducing time and effort spent on daily tasks

⁴ <https://www.rabbit.tech/research>

2.3 Information Retrieval

Assistive technologies enable BVI users to access web content. However, obtaining information using online search engines is still challenging.

Screen readers struggle with dynamic web content

As Kim et al. [2021] pointed out, web pages have become more dynamic and complex, which makes it difficult for screen readers to keep up with the changes. Craven [2004] reported that BVI users do not like searching the search engine results using their screen readers.

Voice commands may simplify web searching

One of the solutions to these issues is a voice-enabled screen reader mentioned in 2.2.1. Ashok et al. [2015] and Zhong et al. [2014] showed that voice commands decrease the number of actions users need to take when searching for information online.

Summarised search results lead to visiting fewer external pages

Aqle et al. [2022] investigated another possible solution - summarising search results. They conducted a study where participants had to complete one task with the Google search engine and another with the prototype summarising the search results. Findings showed that participants benefit from clustering search results and that the prototype allowed them to explore fewer search results. It decreased the number of external pages visited.

2.3.1 Classifying Search Activities

Tasks in our study were based on categorised search actions

In our study, we used categorised search actions and investigated how users interact with our app prototype depending on the search task category. We employed the classification created by Marchionini [2006]. He classified search actions into three categories shown in Figure 2.1.

Lookup is a simple search for specific facts

The first category, **lookup**, is the most basic kind of search. It is often called "fact retrieval" or "question answering". This search returns a single answer in a well-defined format, such as numbers or names. It focuses on the questions "Who?", "What?" and "Where?" rather than "Why?".

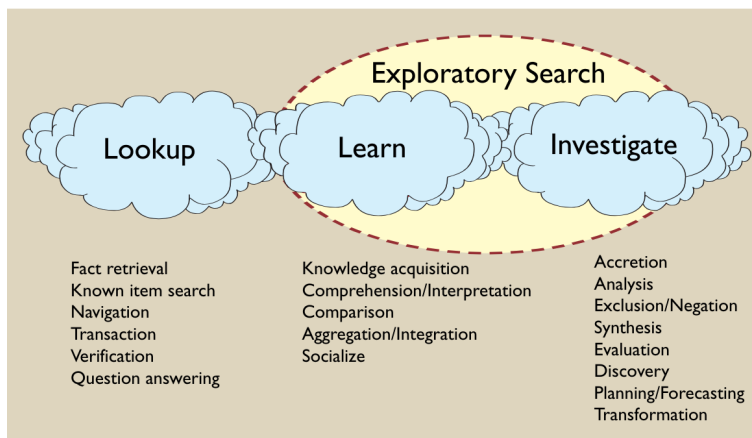


Figure 2.1: Search actions are classified into three categories: lookup, learn, and investigate. This figure is taken from Marchionini [2006].

The second category, **learn**, is more complex and involves understanding the information. It requires the user to read and synthesise the information to make qualitative judgments.

Learn requires understanding and synthesis

The third category, **investigate**, is the most complex and involves the user in the exploration process. It may help in supporting planning and forecasting. Both learn and investigate categories are more open-ended and require solid human judgment in a more exploratory process. Together, they belong to exploratory search.

Investigate involves deep exploration of the results

As we pointed out in 2.2.2, to our knowledge, there is no more research than described in this chapter on exploring how LLMs can help with information retrieval for BVI users. In this thesis, our goal is to address this research gap. In addition, we explore LLMs for performing actions on behalf of the user to find out if solutions like Rabbit R1 described in 2.2.3 could be helpful in BVI individuals' daily lives.

In this thesis, we explore how LLMs can assist BVI users in daily tasks

Chapter 3

Methodology

This chapter contains a detailed description of our research methodology and user study. We conclude by presenting our results.

3.1 User Study

In this section, we describe our research approach - the grounded theory method and continue with data collection techniques. Next, we present the interactive interviews in detail and mention the data analysis procedure.

3.1.1 Grounded Theory Method

The grounded theory method is a research approach that generates theories from systematically collected and analysed data. One of the most critical aspects of this method is its inductive nature, meaning that theories are discovered, developed, and verified throughout the research process rather than a priori. It was designed by Glaser and Strauss [1967]. The grounded theory method is advantageous when the research question is broad, and there is a need to develop new theories in a given field. We chose

The grounded theory method generates new theories from data

It helps explore broad research questions

this method for our study to explore the potential of AI-based voice assistants for BVI users and develop new theories about their advantages, disadvantages, potential improvements, and daily use.

This method consists of several flexible steps

The grounded theory method consists of several essential steps and rules. Although they should be followed generally, the method is flexible and can be adapted to the specific research field. Several publications describe these steps, and we present shortly the ones by Pandit [1996] and Noble and Mitchell [2016].

1st step: specifying a balanced research question

The first step is identifying the research question, specifically the area of interest. The research question should be fundamental, not too narrow or too broad, to balance flexibility and focus.

We formulated the following research question:

RQ What are the advantages and disadvantages of AI voice assistants with LLM-generated responses for BVI users, and how can they be improved?

2nd step: data collection involving interviews, sampling and coding analysis

The second step of the grounded theory method involves data collection, including in-depth interviews supported by open-ended questions, gathering qualitative and quantitative data. The critical aspect of this step is theoretical sampling, where the researcher determines which details should be explored further. That is why we adjusted the questions asked during interviews for each participant as the study progressed. The data should be simultaneously analysed by coding and categorising it. The goal is to find the core category - the main phenomena around which the other categories are grouped.

3rd step: theoretical saturation

Interviews with participants should be concluded once theoretical saturation has been achieved. Theoretical saturation is when no additional information is obtained from the interviews. In our case, we decided to stop recruiting new participants after nine interviews.

There are many approaches on when to begin the literature review [Cutcliffe, 2000]. We followed the approach suggested by Pandit [1996] and conducted a shallow literature review in the beginning and an in-depth literature review after the data collection.

4th step: literature review

3.1.2 Data Collection Methods and Sources

Having multiple data sources is another critical feature of the grounded theory method. It ensures more validity and different perspectives on the research question [Pandit, 1996]. To achieve this, we used two main methods of data collection: interactive interviews and an online survey.

Multiple data sources enhance validity

The interactive interviews were conducted entirely online. Remote sessions were recommended by Trujillo Tanner et al. [2018], health organisations such as the National Health Service England¹ and various UX designers who posted online. We made our list of the pros and cons of online interviews. The pros included participants from various locations, eliminating the need for travel, allowing them to use their familiar devices, and providing the opportunity to remain as anonymous as possible due to the lack of a requirement to turn on the camera. The cons were the possibility of technical issues and the requirement to share the screen, which could be difficult for some participants. Finally, we concluded that the benefits outweighed the cons, so we continued with remote sessions.

Pros and cons of online interviews

We contacted 13 associations gathering BVI people in Germany and received responses from five. Four of these agreed to notify their members about the study. The one that denied the request gave a lack of compensation for every participant as the reason. We also posted the announcement in Online Participant Engagement Network for Vision Impairment Research². Through these two methods, we were able to gather nine participants. One did not want

We contacted associations for BVI individuals to gather study participants

¹ <https://digital.nhs.uk/blog/design-matters/2021/going-remote-how-we-adapted-our-accessibility-research>

² <https://sites.google.com/view/open-vi-research>

to participate in an online study but emailed us their experiences and opinions.

Online survey was the second source of data

The online survey contained more general questions than those asked during the interviews. The survey was created using Google Forms and was distributed among different Facebook groups from the list of 91 groups³. We joined all of them and posted the announcement about the study to those who allowed it and accepted our request to join. The online survey contained information that there is a possibility of participating in an interactive online interview and that the participants can reach out to us if they are interested. We received responses from 19 people, and five wanted to participate in the interactive interview.

We met with BVI people in person in Aachen

Apart from that, we also took part in the Association of the Blind and Visually Impaired meeting in Aachen, Germany, which takes place every three weeks. During the meeting, we talked to the association members about their general experiences with AI and assistive technologies. We also presented our research and invited them to participate in the interactive interview. However, nobody felt comfortable speaking English, which was a requirement for taking part in the study, so they suggested contacting other associations in Germany. Our discussion about assistive technologies and their overall knowledge of AI inspired interview questions and helped us connect with the community better before the study officially began.

3.1.3 Interactive Interviews

This section describes interactive interviews, including the app prototype used in the sessions, procedure, ethical considerations and participants.

³ <https://www.noisyvision.org/2019/04/22/all-the-facebook-groups-for-the-blind-and-visually-impaired/>

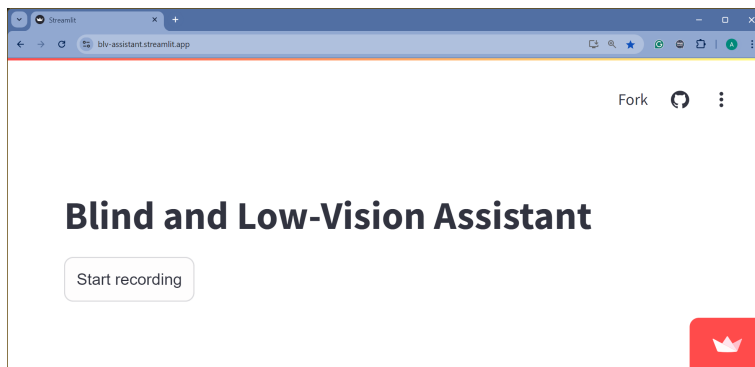


Figure 3.1: Blind and Low-Vision Assistant - an app prototype used in the study sessions. It contains one header, a labelled button to record voice, and a couple of buttons in the top right corner of the Streamlit framework.

App Prototype

The prototype participants used in the study session is a simple web application that acts as an AI-based voice assistant. It was developed using Streamlit⁴ - an open-source Python framework with simple components. We decided to use this framework because it allows for fast connection with the OpenAI API, which we wanted to utilise because of its high-quality LLMs.

We created the most straightforward possible interface with just one header and one button to maximise the app's accessibility for BVI users (see Figure 3.1). We used the custom component `streamlit-mic-recorder`⁵ to turn the basic Streamlit button component into a microphone that records the user's voice. The button can be clicked using either a mouse or a screen reader. We improved the application's accessibility for users unfamiliar with screen readers by adding custom JavaScript code that allows button activation using the space bar. This addition creates a more inclusive user experience without requiring navigation through a screen reader. When the user presses the button, the app starts recording their voice and plays an

Web AI voice assistant app prototype was implemented using Streamlit

We created a simple, accessible voice recording interface integrated with Google Web Speech recognition

⁴ <https://streamlit.io>

⁵ <https://github.com/B4PT0R/streamlit-mic-recorder>

increasing sound to indicate that it has begun the recording phase. When the user finishes speaking, they must press the button again to stop recording what a decreasing sound confirms. After the recording is processed, the app plays the voice, saying that the response is being generated. Additionally, the button is extended with a speech-to-text model to transcribe the user's speech into text, which is needed to obtain the OpenAI API response. The model chosen for speech recognition is the Google Web Speech⁶. We checked the quality of speech recognition and compared it to the other models, focusing primarily on the OpenAI Whisper. Other models, including the Whisper model, were not as accurate as the Google Web Speech, so we used the latter.

<p>The text is processed by GPT-4o, converted into speech and delivered as an MP3 audio file</p>	<p>The text is sent to the OpenAI API in this format, which returns the response using the GPT-4o model. As mentioned in 2.2.2, the GPT-4o model is the fastest available option in the OpenAI API, optimised for speed and quality. Then, the textual response is sent to a text-to-speech OpenAI model, which converts it into a natural-sounding voice and returns the MP3 audio file, which is played back to the user.</p>
<p>We simulated booking flights and hotels</p>	<p>We wanted to include the option for booking a flight or hotel to explore with participants the potential of an AI system to perform actions on behalf of the user. To achieve this, we used an initial prompt to inform the model that it was intended to simulate these capabilities.</p>
<p>App accessibility was evaluated with tools and user testing</p>	<p>The app's accessibility was checked using several methods. Firstly, free online accessibility checking tools were used. We tested the app using Silktide⁷ and Wave⁸. Although the first one found some issues, the second one confirmed that every component is accessible. Because the tools yielded different results, we used other methods. We installed a screen reader Google Chrome extension⁹ and tried it with the app. The button was successfully found under the label <code>aria-label= "record-button"</code> and could be clicked with-</p>

⁶ <https://cloud.google.com/speech-to-text>

⁷ <https://silktide.com/solutions/accessibility/>

⁸ <https://wave.webaim.org/>

⁹ <https://chromewebstore.google.com/detail/screen-reader/kgejglhpjiefppelpmljglcjbhoiplfn>

out issues. Finally, accessibility was checked and confirmed by two participants one week before the study officially started.

During development, we encountered some issues with the Streamlit framework. The main one was the long waiting time for the response from the OpenAI API. It was caused by several factors, such as the time needed to process the recording, transcribe the speech, and generate a response using the OpenAI text-to-speech model. We initially attempted to optimise the recording process by dividing the audio file into smaller chunks and sending them to the OpenAI API one at a time. However, this approach was ineffective because the OpenAI API requires the entire recording to be submitted at once. Then, we tried to change the speech-to-text model to a faster one. However, we found this was not the case - the model was not the bottleneck, and the time needed to transcribe the speech was acceptable. Finally, we tried to optimise the response generation using the OpenAI text-to-speech model. Indeed, it was a blockage, and the time needed to generate the response was too long because the responses were generally too detailed. We used the OpenAI API's streaming feature, allowing the response to be generated in chunks. This solution was ineffective, so we gave the model a detailed initial prompt. The initial prompt included the user's being BVI and the assistant's serving as a voice assistant. We instructed that responses should be concise to minimise generation time, and more details should be given only when the user asks for them. This solution worked surprisingly well; the responses were much shorter while still informative, decreasing the time needed to generate them.

During development, we faced delays due to the OpenAI API's response time, which we improved by optimising the initial prompt

We also found some limitations and encountered problems during the study sessions. First, the app required microphone access to be allowed each time the browser was reopened. Although we could not find a solution to this problem, we always informed the participants about this before the study started, and they had no issues with it. The second issue was that the app sometimes failed to respond after the user stopped recording. Although it happened only twice, we could not find the reason. It might have been the microphone that was used by the video conferencing

During the study, we faced issues with microphone access, app responsiveness, and smartphone compatibility

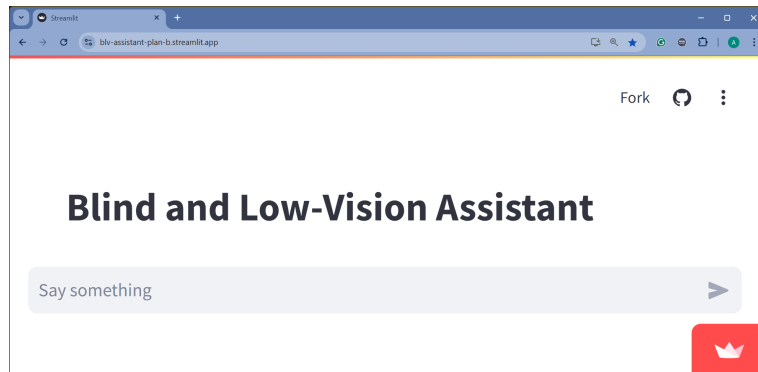


Figure 3.2: Blind and Low-Vision Assistant - a backup app prototype used in the study sessions with text input. It contains one header, a text input, and a couple of buttons from the framework in the top right corner.

system at the same time. Fortunately, we had an alternative app version that allowed participants to type their questions instead of speaking them (see Figure 3.2). The third limitation was that the app was incompatible with smartphones. It was designed to be used on a computer and not tested on smartphones. Before the study started, we informed the participants about this limitation, but two had to use their smartphones because they did not have a computer. One person dictated the questions to us, and we typed them into the app; another person made the app work on their smartphone. Although it was not intuitive, the participants completed all tasks successfully.

All participants completed their tasks despite inconveniences

Study Procedure

Study sessions were held on Zoom, lasting about 60 minutes, with communication in English

Study sessions were conducted using the Zoom platform¹⁰. We reached out to the first five participants to choose the most convenient and accessible platform for them, and they all chose Zoom, so we decided to use it for all of the sessions. We communicated in English throughout the study. Sessions lasted approximately 60 minutes, as our test study session took 45 minutes. We wanted to allow extra time

¹⁰ <https://zoom.us/>

for all participants to ask questions and share their experiences. The meetings were recorded with participants' verbal consent to ensure we could analyse the data later without missing any vital information. A single study session was divided into three parts.

In the first part, we gathered basic demographic data and information about the participants' experiences with AI and assistive technologies.

1st part: participant demographics and AI experiences

In the second part, we presented the app prototype and asked the participants to interact with it. We incorporated some interactivity into the study to observe the participants' reactions to such an app. Additionally, some individuals had never used an AI-based voice assistant or any application featuring LLM-generated responses, so it was crucial for us that everyone understood what we meant by an "AI-based voice assistant." Firstly, we asked participants to perform one or two warm-up tasks to get familiar with the app. After the warm-up tasks, we asked them to perform three main tasks. Each task was designed to fit into one of the three search categories identified by Marchionini [2006] described in 2.3.1. The first task required the user to ask about the recipe for a specific dessert. It fell into the category of lookup search. The second task asked users to compare different smartphones and choose the best one for them based on the information provided by the app. This task was designed to fit into the category of learn search. The third task was to plan a trip to a specific city. The user had to ask the app about flights, hotels and tourist attractions. It belonged to the category of investigate search, which is mainly concerned with supporting planning and forecasting. The critical aspect of the last task was that the user needed to concentrate on booking a flight or a hotel. This allowed us to observe the participant's reaction to the app's capability of performing actions on their behalf, as discussed in section 2.2.3. We could observe how the app handles the user's intent and discuss it with the participant in the last part of the study.

2nd part: interaction with app prototype

Three main tasks included finding a dessert recipe (lookup search), comparing smartphones (learn search), and planning a trip (investigate search)

In the third part, we asked the participants to share their thoughts about the app and their tasks. We also discussed their general experiences with AI and assistive technolo-

3rd part: interview

gies so we could identify the advantages and disadvantages of AI voice assistants. Due to the use of grounded theory methodology, we did not follow a strict list of questions for every participant, resulting in semi-structured interviews. We had a list of topics to discuss (see Appendix A) and chose the questions dynamically.

Ethical Considerations

As there is no Ethics Board of the Computer Science department at RWTH Aachen, we followed the guidelines provided by Moore [2002], Trujillo Tanner et al. [2018] and Nielsen Norman Group¹¹ described below.

Before: reach out to BVI associations to find participants, send the consent form via email

Consent form

Before the study, it was recommended to contact the associations gathering BVI people to find potential participants because people trust the associations and are more likely to participate if they recommend it. Secondly, it was recommended to send the consent form to participants before the study so they could read it and ask any questions they might have. Our consent form described the study's purpose and procedure. It also included information about potential risks such as mental fatigue, confidentiality, compensation, and the requirement to record the session without the need for a camera. Moreover, Trujillo Tanner et al. [2018] suggested that to make the study more accessible, it should be conducted online. It allows for more flexibility in scheduling and allows the participants to be in their familiar environment using their own devices.

During: give a lot of verbal feedback, offer breaks and update about the time

During the study, the etiquette was crucial. We followed a couple of rules to make the participants feel comfortable: introduced ourselves, gave a lot of verbal feedback ("I see", "I understand", "Yes"), offered breaks and updated the participants about the time left. At the beginning of each study session, we again read the consent form and asked for participants' verbal agreement.

After: offer a compensation

After completing all the study sessions, we randomly chose one participant who received a 25-euro Amazon gift

¹¹ <https://www.nngroup.com/>

card as compensation for their time and effort. The consent form clearly stated the compensation, ensuring all participants were informed.

Participants

We recruited N=13 participants (seven females and six males) for the interactive interviews. After nine sessions, we reached data saturation and decided to stop recruiting additional participants. Participants aged between 19 and 65 years old (M=33.4, SD=14.3). We interviewed individuals from different countries: eight from Germany, two from the USA, two from the United Kingdom and one from Cambodia. Eight participants were completely blind, and five were partially visually impaired. Most of them used screen readers daily and had experience with at least one AI assistive technology. Eight participants considered themselves as advanced technology users, four as intermediate and one as a beginner.

We interviewed 13 participants of varying ages and backgrounds

Eleven participants took part in the whole online study sessions. One of the participants decided to continue answering the questions via email after we finished the interactive part of the study because they were not feeling comfortable during the online meeting. The second participant experienced issues with screen sharing during our session that we could not resolve. We conducted the demographics part during the online meeting but ran the rest of the study via email. We sent the tasks for the interactive session along with the interview questions. Both participants completed the whole study sessions, so we considered the results from all of the sessions.

Two participants faced challenges that led to completing the study via email

3.1.4 Data Analysis

The data collected during the study sessions was manually transcribed. We used the MAXQDA¹² software to code the data and identify the categories.

¹² <https://www.maxqda.com/>

We used open, axial, and selective coding to identify core categories

The data analysis was conducted according to the principles of the grounded theory. The coding process in this method should be iterative and consist of three types of coding: open coding, axial coding, and selective coding. **Open coding** is the first step in the coding process and involves identifying critical phrases in the data. In our case, we coded the data sentence by sentence and gave meaningful names to the codes. We were constantly comparing the codes and identifying relationships between them. Three rounds of open coding were conducted to ensure that the same name coded the same phenomena. After the open coding process, we moved on to **axial coding**, which involves finding relationships between categories. We grouped existing codes into categories and identified the relationships between them. Sometimes, we changed the names of the categories to reflect the data and relationships better. Finally, we conducted **selective coding** to identify the core categories and develop the theories based on them. Corbin and Strauss [1990] wrote: "The core category must be the sun, standing in orderly systematic relationships to its planets", and we followed this rule.

3.2 Results

This section provides a detailed presentation of the study's results. We demonstrate the pure data here and then analyse it to develop the theories in Chapter 4. We use the abbreviation "P" followed by the participant number (P1, P2, etc.) to refer to the interview participants. Participants in the online survey are referred to as "S," followed by their participant number (for example, S1, S2, etc.).

3.2.1 Assistive Technologies and Challenges

From both demographic data and the interviews, we found that the participants used a variety of assistive technologies in their daily lives.

Screen readers were the most commonly used assistive technology among the participants. P1 highlighted that screen readers were quick and easy to use, allowing them to work efficiently in the office. Other participants mentioned that screen readers were handy for reading documents and writing their texts.

Screen readers
enhance efficiency

Participants agreed that replacing the screen reader with any AI was impossible because it was the most efficient way to interact with the computer. However, they also mentioned that the screen reader was not always the perfect tool because some websites were incompatible. P5 encountered issues with some work applications because *"most of the time they have no labels."* S6 was often having *"difficulty in understanding which visual element corresponds to the textual information."* S14 was rarely using mobile apps because of the *"I/O interface issues with VoiceOver on iPhone."*

Screen readers are
essential, but have
some issues

To interact with the environment, participants were typically using smartphone apps, such as **Be My Eyes**¹³ - a free app that connects BVI people with sighted volunteers for visual assistance. A user takes a picture, and the volunteer describes what is on it. P1 encountered a challenge with this app: it did not allow follow-up questions. However, they were unaware that the app had recently introduced an AI feature that could continue the conversation about the picture. Other participants already knew Be My AI¹⁴ and were using it for various tasks. P3 mostly used it to describe pictures of people and places when they were travelling. They were delighted with the app's ability to ask follow-up questions about the picture. They also used the app to read printed documents and recognise specific information. P11 used it to identify expiration dates on food products or medicines and read the packaging instructions. P3 said that the app had greatly improved by introducing the AI feature, and they would like to see video support implemented. P8 found the app very helpful and claimed it had 99% accuracy, as they frequently checked with their sighted friends, who always confirmed its correctness. One crucial limitation of Be My AI was the camera's incorrect positioning and inability to move it in the right direction.

Participants use
smartphone apps like
Be My Eyes and Be My
AI for visual assistance

¹³ <https://www.bemyeyes.com/>

¹⁴ <https://www.bemyeyes.com/blog/introducing-be-my-ai>

Participants said that with Be My Eyes, the volunteer could tell them to move the camera in the right direction, but with Be My AI, it was less flexible. S19 suggested that the app should have *"sensor detection and voice-guided sensor to help to do better with taking photos."*

Voice assistants are helpful for various tasks but have issues with misunderstanding, accuracy, and app dependency

Another assistive technology the participants frequently used was various **voice assistants**. They used voice assistants for tasks like setting alarms, checking the weather, asking about current times, spelling words, tracking packages, playing music, asking basic questions, and controlling smart home devices. Participants were generally comfortable with using voice assistants in public. Many users reported misunderstandings or mentioned that their voices were not recognised correctly. P2 said that they hated Siri because *"it just makes up things"* and *"comes up with answers that make no sense."* Additionally, they always had to dictate the text to Siri with all the punctuation marks so it could understand them. P4 was sometimes frustrated because they were talking too quietly, and the voice assistant was missing their voice. P13 had a lot of Alexa devices at home, and sometimes, they were all responding at the same time. Other participants reported that they were often receiving inaccurate information about the weather. P11 said they had to be specific when asking about the weather because the voice assistant did not understand which city they asked about if there were two cities with similar names. Another issue one participant mentioned was *"not helpful visual feedback from Alexa."* They could not see the green circle around the device and were unsure if it was listening to them. Four participants mentioned apps' dependencies on voice assistants. Three of them treated it as a disadvantage because the voice assistant opened the app on their smartphone so they could do a specific task or access some particular information. The other participant said it was an advantage because they could access the desired app only by speaking to the voice assistant. P7 argued that voice assistants were *"not an assistive technology."* They mentioned that their *"use of those was more like everyone else using it."*

Other (assistive) technologies that the participants used were ChatGPT, Envision AI, Braille displays, Microsoft Copilot, Aira and Google Lookout.

Participants pointed out independence as the most important advantage of using assistive technologies. P11 mentioned it was a significant advantage because they did not require personal assistance. The general challenge pointed out by several participants was the lack of flexibility in the systems. Some individuals stated they did not understand how AI works, making it challenging to trust AI-based technologies.

Participants value independence provided by assistive technologies

3.2.2 AI Voice Assistants With LLM-Generated Responses

Most of the participants were satisfied with the app prototype and the interaction with it. Some of them had never used an AI voice assistant before, and they were surprised by the quality of the responses. In this section, we share their feedback on the app prototype and AI systems they might use.

Participants were satisfied with the app prototype

Use Cases

Many participants wanted to use AI voice assistant for **picture descriptions**. Some mentioned uploading a picture and getting a description like it can be done with Be My Eyes. Others would like to talk with the AI voice assistant about the picture they encounter while browsing the web without uploading it. Moreover, video support was mentioned as a significant feature.

Participants wanted AI voice assistants for image descriptions and videos

Some participants would like to have such an assistant for **navigating through the environment** in real-time. One mentioned that they loved running and would like to run without a guide. They could imagine running with an AI voice assistant app that tells them where to go or warns about obstacles such as curb stones or other people on the

AI voice assistant could assist with real-time navigation, public transportation and indoor location finding

way. P1 said they would love to have an AI voice assistant that announces the number of buses approaching the bus stop. P6 gave an idea to have an AI voice assistant to help them navigate inside buildings - tell them where the specific room number or toilet is. Others went even further and would like to have an AI voice assistant with a built-in GPS function: *"The app must provide the GPS function that can help blind and visually impaired people who navigate from one location to another location when not travelling with a sighted family or friends or stranger in public transportation"* (P8). S13 suggested a GPS explicitly created for BVI people: *"The AI in a GPS app would be useful. Instead of just saying 'turn left and cross the street', it can also say 'keep going straight', 'a little to the right', etc."* S6 would like an AI voice assistant that helps them use physical objects like a microwave or a washing machine: *"I struggle to understand which button corresponds to the action being described. For example, the AI might say: 'The start button is on the top right of the microwave.' However, the concept of 'top right' can vary depending on where I am standing in relation to the microwave. If I am standing to the left of the microwave, my 'top right' might be different from the AI's understanding of 'top right'."*

AI voice assistants could aid with daily tasks and planning

AI voice assistants could be used for **simple daily tasks**. Participants mentioned activities like checking the weather, reading mail and recipes or checking food labels: *"I have a few bottles with different labels on. I have not managed to Braille them yet, so I need to be able to tell what each bottle is"* (P9). Additionally, they would ask such an AI voice assistant spontaneous questions when they are too lazy to type them into the search engine. Some participants would use it for daily planning, for example, to check the morning calendar or set reminders.

AI systems may also help in smart home control

Other participants suggested use for **smart home control**. P4 said they would like to start the dishwasher or manage their home heating system using voice. P5 would like to have the possibility to check the state of every device in their home by voice, including the cat feeder. Similarly, S12 would like it to read text from *"electronic displays such as a heating pad, humidifier, dehumidifier, thermostat, and other electronic things that are not available through an app."* P13 men-

tioned that they already used Alexa for these things, but an AI voice assistant could be more efficient.

Summarising text from printed paper was also mentioned as a potential use case. Participants could imagine uploading a picture of the text to an AI voice assistant and getting a summary of it read out loud. Then, they could ask the AI voice assistant about the details of the text and extract relevant information.

AI could also summarise the text and extract details from it

A few participants would see such an AI voice assistant as a tool to **enhance the accessibility** of some websites when they are not fully accessible with the screen reader. LLM-generated information about the inaccessible parts of a website could be generated in real time, and users could ask the AI voice assistant for specific details. S6 would like AI voice assistant to *"recognise visual captchas, fill in the correct characters [and] relate contextual information to visual elements on the screen."* Participants also talked about using AI voice assistants to help fill out different forms: *"It would be interesting to use it for other things when you just work with the form, and the voice assistant helps you with interacting with it. But it has to be quite complicated and not accessible form"* (P3).

AI voice assistants could improve website accessibility

Emotional support was also a critical use case for three participants. They pointed out that sometimes they were just bored and would like to describe their feelings: *"I just consider the AI like my helpful assistant that can discuss all problems with me"* (P8). P12 said that they *"felt listened to"* by the prototyped AI voice assistant we had created.

Emotional support from AI was valued

Some people would use AI voice assistant as a **writing support** that helps them with their work. Not only for writing the whole text but also for providing text blocks that could be used in their texts. S6 was a writer, and they would like to have an AI voice assistant that could help them with visual descriptions for their storylines: *"I want the AI to be able to understand the context and details of a prompt and incorporate them into its responses in a meaningful way. I would like to be able to train the AI using paragraphs from authors and content that I enjoy, not to plagiarise but to use as guidelines for the AI to learn from."*

Writers sought AI assistance for creative writing tasks

Developers wished AI support for programming-related activities	Participants who were software developers would use it for programming : to debug the code, find the suitable icons or write the documentation. Other participants would learn new things using the AI voice assistant as a personal tutor .
Users wanted AI to control computer tasks like booking	Some participants would like to try an AI voice assistant that takes complete control over their computer : <i>"What is still missing is access to things controlling like booking flights. That would be a great option to have more possibilities on that side. Do and control things"</i> (P7).
Additional uses	Other use cases that participants mentioned were brainstorming, using for ideation, enhancing productivity, language translation, gaming and game development, travel planning, a search engine or just pure entertainment. One participant mentioned they would use it the same way as their current voice assistant.

Advantages

Participants appreciated the prototype's simplicity and usability	Almost every participant shared a positive opinion about the AI voice assistant prototyped by us. Everyone said that it was very easy to use : <i>"You open the web page, you press that button, and it works"</i> (P2), <i>"It could not be made easier than this"</i> (P3). P13 had no issues with using the app on their smartphone. Although P6 was not an advanced technology user, they said the app was intuitive and easy to learn. Participants agreed that there was no problem enabling microphone access at the beginning: <i>"I have to do it for everything on the computer, for some online conferencing platforms and voice search. I cannot avoid it. It is pretty straightforward because there are just two options, allow or block"</i> (P3). P10 liked the simplicity and the simple design of the app. They kept saying they would not add additional features because it would make the app more complex. Other participants also mentioned that there was no need to include new features because the app was already <i>"great the way it is"</i> (P13).
---	--

Many participants mentioned it might be great for **information retrieval and search**. It could *"help VI users conduct research almost as fast as sighted people instead of having to browse through search results one link at a time"* (S14). S18 already used ChatGPT for quick searches and was very satisfied with the results. P4 described the scenario where they would use the app to check something quickly on the internet. Some participants used Alexa and other voice assistants for this purpose but said such a voice assistant based on AI could be more efficient.

Participants valued AI for efficient information retrieval

The **accessible design** was also mentioned as an advantage. S9 said that such an AI-based system could be a great tool *"that would not have pop-ups wanting to sign up for this and that every few seconds."* Other participants liked that everything was labelled and could easily find the button to record their voice using their screen reader. They appreciated the possibility of using the app by pressing the spacebar without needing to search for the button with their screen reader. Several participants liked having the text form of the response because they could read it using their screen reader afterwards: *"My concern was, can it also generate a text response in addition to the voice-based response? You might need to interact with that text later or just copy the text. And it did it"* (P3). P5 considered himself a *"text person"*, and they liked having the text everywhere they could, so it was essential for them also to have the text form.

Accessible design was highly appreciated

Several participants said that the **voice-only interaction** was a great benefit. P10 said that although speech input was more sensitive, it had more potential than text input. P13, who had low vision, said that voice input enabled them to use the app without the need to use their magnifier, which could *"release the stress of having to do that."* P7 said that controlling it by voice was *"very handy and effortless."* P10 said the sound quality was excellent, even better than their screen reader's.

Voice-only interaction was seen as beneficial and efficient

Two participants pointed out the importance of **feedback that the response was being generated**. P4 said it was crucial because they knew the app was working. P13 told us it was beneficial because they did not have to check if the text was already generated using their screen reader.

Feedback on response generation confirmed as applicable

Participants valued accurate and precise responses

Participants appreciated the **accuracy of the responses**. P10 emphasised that the responses were *"very precise"* while the others said they got the exact information they requested. S3 mentioned that the most significant advantage of LLM-generated responses was the assurance of getting an accuracy of around 98%. P5 said, *"I am happy to have a system with 90% accuracy because I am getting like 120% speed increase on Google for example."* P7 told us that the main profit they saw was the generativeness of the responses. Other participants liked the details in the answers and that they did not have to ask for them. What also came as a great benefit was the AI's ability to learn over time through the user's interactions.

Participants appreciated the app's design for BVI users

Other participants liked that our app prototype was **designed specifically for BVI users**. P3 said that it was beneficial when the app mentioned the accessibility features of the phones in the second task regarding smartphone comparison in the second part of the study. S14 said: *"Having a VI-specific LLM [is] nice. [...] What I mean is training the LLM you use in the app to give every single response tailored for someone who is blind. This way, the user does not have to constantly provide that context or ask for specific instructions or helpful responses for blind people."*

Participants enjoyed natural interactions with the AI voice assistant

Participants also mentioned that one of the profits of using AI voice assistant was the possibility of **talking to it naturally**. P8 liked that the voice tone was not robotic, and the AI voice assistant sounded like an actual human. Two participants said that the voice was very clear and easy to understand. Others liked that the responses were read out loud very slowly and that the tone of voice was charming. Participants valued that the conversation felt natural and did not require specific phrases: *"It is great that you can give whole sentences and you do not have to think about keywords. In the third question, it was great that I did not have to name details. I can ask whatever I want, and it is not necessary to name the specific keywords; that is great. I could just say, please tell me the status of the dishwasher without telling the name of the app before"* (P4). P10 liked that they did not have to *"answer 100% exactly to the question it asked at the end."* P2 added: *"You do not even have to know much about how to talk to it be-*

cause it just tells you when it wants to know something or when something has not been understood."

Most of the participants were positively surprised and impressed by the app's possibility of **performing actions** on behalf of the user, which we introduced by simulating booking a flight or a hotel:

Participants were impressed by the app's booking capabilities

P1: *"I do not want to interact with other websites, hotel websites, for example. It is really comfortable if the system [does it]."*

P3: *"If we can complete the process in one step, provide all the details and complete the booking, it would be much easier. If AI gets my information, it could pass it directly onto the app or website where you book the flight."*

P5: *"If I could just cut out all the clicks I have to do to make my screen reader navigate a website, that would make me a very happy bunny."*

P7: *"Booking things tend to be inaccessible at some point, so if an assistant was able to do this, that would be great."*

P10: *"It would ease it a lot like booking trains. That would be amazing. And give so much time back."*

P11: *"Most times I am planning it myself because I am comfortable using Google or other machines and websites. But sure, if it is such easy then why not?"*

Participants generally did not see any problems trusting the app with bank details for actions like booking a flight or a hotel. Some mentioned they would need only a secure connection or proper data protection to trust the app fully. The others said there was no problem with trust because they already trusted other apps with their bank details. P12 said the interaction felt so real that their mother asked them if they were booking a real trip.

Trust in the app for bank details was generally strong

Other advantages mentioned by participants included **real-time quick responses**: *"It puts the information literally at the sound of my voice"* (P13). Another noticed benefit was the **potential to replace human assistance**: *"In my private life,*

The app offered real-time responses and potential for independence

I use some apps that can describe objects and colours for which I would have to call somebody via video or ask some person in real life. And that can now be substituted by the app, and that is amazing” (P10).

Comparison to Other Assistive Technologies

Participants compared
AI voice assistants to
traditional assistive
technologies

We asked the participants to compare the potential of AI voice assistants with LLM-generated responses to other assistive technologies they use regarding efficiency, accuracy, and overall user satisfaction.

Participants analogised
AI voice assistants to
traditional ones

Most participants were analogising AI voice assistants to **traditional voice assistants** such as Siri, Alexa, or Google Assistant. P4 compared the ease of use of AI voice assistant and Alexa: *“It is rather easy because if you tell something your system does not understand, it will tell me that I have to ask again. Alexa also does it if she does not understand the question. It requests to repeat.”*

Some participants were
sceptical about needing
an AI assistant

P2 compared it to Siri and said they could do the same things with both of them, so they would not see why they should use AI voice assistant instead of the traditional one: *“It does the same things that Siri does. I would not use it because I already have Siri, and I use it sometimes for some tasks. I say something, it makes a speech to text stuff, so it can know what I want.”* On the other hand, P13 said *“I think it can do a lot more than Alexa can. So this is pretty amazing. I feel like this one is a lot more personalised and detailed. It gets right to the point. Alexa always adds a bunch of information that I did not ask about. I like this better than Alexa.”* P4 was excited that they did not have to name the specific keywords to get the information they wanted like they had to do with Alexa: *“It is great if we can only ask without telling which app you want to use.”*

Other participants liked
AI for personalisation
and efficiency

Issues with Siri as a
search engine

Some participants mentioned that Siri did not work well as a search engine and encountered more difficulties with it than with the app prototype we created: *“Your app might have the potential to be better pretty soon. It would even be better than Siri” (P2).*

Lastly, participants were concerned about the lack of conversation flow in our app prototype and said that it was handled better by other voice assistants: *"Every time I want to say something, or even when the AI is asking questions for me to respond, I have to hit the talk button which is not very realistic when you talk to artificial intelligence. A lot of artificial like Google, Siri, and Alexa, when they ask questions and they want a response from you, they switch on automatically, and you talk to it like you would talk to a real person. That is pretty much how the AI is operating at the moment. It is not very natural in terms of the flow of the conversation between myself and the AI. So rather than me hitting record every time, it should do it automatically when it wants responses from me"* (P9). On the contrary, P7 said that they could not talk to other non-AI voice assistants like a human and *"that was the main difference between such an AI assistant and these traditional voice assistants like Alexa and Siri."* However, they also compared the app to Google Assistant in terms of efficiency and speed and said the response time was much longer in our app prototype.

The conversation flow was criticised

Participants agreed that AI voice assistants could never replace a traditional **screen reader**: *"They are two different concepts. The screen reader reads the screen and basically anything that is on the screen, whereas the AI-based Voice Assistant can provide you with responses"* (P3). *"One of them is talking, the other one is reading"* (P1). P2 also said that these two devices did completely different things. P1 described the interaction with our app prototype to the screen reader's advantage: *"To use these AI systems, I have to press a button, and I have to speak. Then I get the response. I do not really talk with my screen reader."* Similarly, P5 said it would be frustrating *"if I had to say 'switch to VS code' and then start dictating my code."* P11 concluded that, in general, it was hard to say if the screen reader could be replaced because it depended on what they did. *"It would have to be able to control the whole computer or phone. As long as the AI cannot move my mouse and press buttons on my keyboard to control things, it cannot replace the screen reader at all because it cannot do things and control things"* (P7). However, P9 said there were things not reached by screen readers yet: *"Screen reader does not describe pictures on the screen. It only reads the text."*

Participants agreed that AI voice assistants and traditional screen readers serve entirely different purposes

Participants choose AI
for web searches

Nevertheless, participants saw the potential of AI voice assistants in **web searches**. When we were discussing the app's prototype response waiting time, some of the participants mentioned that despite not having immediate responses, they would still use the app because it was more efficient than using the screen reader to browse the web: *"I do not think it is too long. If you look in the web and you have difficult keywords, it needs more time"* (P4). P5 agreed: *"Even if the response time is longer, it was quicker than me going to Google to find all the stuff that it gave me."* Although P2 stated that both Siri and our app prototype could do the same things, they also recognised the potential to replace their voice assistant with the app for web searches because it was *"more efficient and smarter."*

Limitations and Disadvantages

Participants expressed
privacy concerns about
using AI systems

The main concerns participants usually raised were about **privacy**. Not only about booking details such as bank account or personal address but also about the whole content of the conversation. P9 said they always had privacy concerns while using any system with AI *"because it is an AI. I have provided the data to the AI, but AI is an Internet-based kind of system. It works off the cloud. I am providing my personal information to this piece of IT. How do I know that it is legit? How do I know that it puts it where I need it? [...]* While for some things I would rely on an AI, for some things I would not. For example where there is personal information involved. I would not take that risk in using AI because it is technology. Technology can always go wrong. [...] How will I know that it has not kept my details and put them elsewhere after I have completed the flight booking? How do I know that? If somebody hacks into the AI itself, they find my details because I have put them into that before."

Participants
communicated mistrust
in AI's reliability

Furthermore, participants were afraid of **unreliable responses**. P2 would not use an AI voice assistant to find real information but to *"play around and find out how it works, how it does things, how it answers."* P7 said that the fact that the responses were generative was a reason for AI voice assistants making up things, and that was why they would not

trust it 100%. P8 assumed that the AI systems were sometimes providing fake or old information, and they confirmed their assumption while testing our app prototype. On the other hand, S6 said that the problem was not about the fake information but about the AI repeating the same ideas and phrases all over again: *"I want the AI to be able to think outside the box and come up with new and innovative ideas, rather than relying on repetitive phrases and tropes."* P9 called all of the AI voice assistants *"useless things"*: *"It is difficult to put into words because even though it is meant to be artificial intelligence, I see no intelligence in that at all."*

Many participants agreed that such an AI voice assistant **could not be the only tool** they would use daily. Some of them mentioned that the devices they currently use and such AI voice assistants could be complementary *"because AI cannot perform some of the tasks that other tools can"* (P9). Others could not find any application: *"I will say it is not good at all for my daily activities"* (P8), *"I am not sure how I would use this in my daily life"* (P11). P10 said: *"The hugest problem for me is to remember to use it."* Additionally, participants said that AI voice assistants could never replace physical devices. Some of them mentioned the importance of physical devices themselves: *"If I am using, for example, a talking blood pressure monitor, I could not replace it with AI because I need the physical tool that I can use to check my blood pressure and I need the monitor to talk back to me and give me the results—the same with talking microwaves, talking cookers, etc. There are a lot of things that an AI could not do because there is a physical aspect of things as well as the intelligent side of things"* (P9). The others highlighted that *"It is more important to be able to use the device myself than it is needed to use an AI assistant"* (P11). Lastly, a few participants said that no matter how good the AI voice assistant was, it could not replace human assistance: *"The other thing about AI is that it can be the most advantageous piece of thing ever, but it can never replace the assistance of a sighted person, no matter how many improvements that tool could have"* (P9). S6 thought that *"AI lacks human expression and emotion in its responses."* Some participants always preferred to have someone look at what they do and if they do it right. S16 said that AI might not always be able to provide the same level of detail as a human assistant. Even though the potential solution could be con-

AI voice assistants
could not replace
physical devices

Human assistance was
preferred for specific
tasks

necting AI voice assistants with other assistive devices or apps, P4 raised some concerns: *"I am not sure if it is good to combine too many things into one system."*

Participants reported frustration with voice recognition systems due to misunderstandings

"I always have issues with things requiring me to speak because they never understand me" (P12). *"Sometimes it does not understand my questions. I need to repeat the same question again and again a few times to make it understand my query"* (P8). The next disadvantage that participants mentioned was **misunderstanding** of the users' voices and intents. P3 said that the results were not always what they expected because the answers were completely irrelevant to the questions. P11 would only use it for simple tasks because they feared the AI would not understand them. Others complained about voice recognition, which is still not as good as they would like, and mentioned that such systems often did not understand heavy accents from different countries. Lastly, two participants said that they always had to know specific phrases to get the correct response: *"We need to use specific words or phrases to trigger the desired response from the AI, but we are not aware of what those words or phrases are since we did not program the AI. For example, if I wanted to tell a smart washing machine to start, I might say 'start the washer' or 'begin the washing cycle.' However, if the AI is only programmed to respond to the specific phrase 'start the washing machine,' my attempts to communicate would be unsuccessful"* (S6). P4 had no problems with that but said that older people might have some issues with finding the correct phrases. S6, a writer said additionally: *"Sometimes, I provide too much detail, and the AI does not understand what I am asking for. Other times, I give specific instructions, but the AI fails to incorporate the details I want into the storyline."*

Internet dependency impacts usability and privacy

What was also noticed as a significant limitation was the AI voice assistant's **dependency on the internet**: *"It needs to send everything I do through the Internet"* (P2). Others said it was not very helpful if they could not use such an app when there was no internet at all. Moreover, this internet dependency meant they could not use it privately: *"It is always a question of how good and secure the Internet connection is"* (P11).

Some participants told us that they would **not use such an AI voice assistant in public**. P11 said using their voice assistant in public could sometimes annoy others. P2 would not want other people to know what they do, especially when they text someone: *"I like typing, because then only I can know what I am typing."* P5 said the only reason was that they did not like the feeling of walking around and talking on their phone.

Participants reported discomfort using voice assistants in public settings

Another concern that several participants had was the **sustainability** of the AI systems: *"Too much energy and too much time always to get it into the speech engine from the AI. It is just the thing I always think about when I use the computer and when I use any kind of technology. How much energy does it use, and is it necessary? Just a thought I always have in the back of my mind. I am kind of keen to save energy as much as I can"* (P2). *"It is clear that the data are sent around the world on many servers,"* they continued and highlighted that too much energy was used for too small tasks. *"One concern I have regarding all this AI hype is that it consumes a lot of energy and is very intense in terms of calculations. So, this is one aspect I always have in mind when people want to bring in more AI capabilities. It does not have to be used everywhere"* (P10). Others mentioned that storage also played a role, as the data was stored somewhere, and the location of the servers mattered for the environment.

There were concerns about AI sustainability and energy consumption

Other interesting limitations and disadvantages that participants mentioned were:

Other limitations

- **too general or too detailed LLM-generated responses:** while some participants liked the detailed responses, others would prefer to have more general information and the other way around,
- **voice assistant over-reliance on AI:** P8 said the AI voice assistant should not rely 100% on the AI but also on the pre-programmed responses,
- **blaming the user:** users often blamed themselves for the AI voice assistant's mistakes and thought they did something wrong: *"Maybe I should ask in more detail, I did not do that. Maybe it was my fault"* (P1 after getting

the wrong response from the app prototype), *"It is not because of the AI. It is because of me speaking not clear enough, or the words are not known to the AI, and it just does not understand it"* (P2),

- **performance dependency on the equipment:** participants mentioned there was the need to have a good microphone and a good sound system to get the most out of the AI voice assistant,
- **expensive software:** participants raised a concern that such software could be expensive, noting that more expensive software was often performing better: *"Some programs work better, but it is costly, so I just choose those without cost"* (S19), *"We really need [apps] for free at no cost"* (P8),
- **challenges faced by novice users of technology:** P2 was worried about the younger people who had just started using assistive technologies and the older people who were beginners in technology. They thought such an AI voice assistant could be too complicated for both groups.

Participants criticised the app prototype for its long response times, inefficient setup, and unnatural interaction flow

In addition to the general disadvantages, participants mentioned some specific aspects of the app prototype we created. One was the **long waiting time for the response**: *"It was a bit long. For AI, for computerised systems, it should not be that long. It is like asking a question to a real person, and they need to think before giving me a reply. But with a robotic, it should not be like that. If I am asking a question to a robot, that is the whole point of a robot: the answer should be instantly. I do not think the waiting time is appropriate at all"* (P9). Some of the participants mentioned that they were aware of this challenge: *"I know that AI takes a while because it is a lot to process"* (P10). P3 said they *"just needed efficiency and speed"* and that *"turning on the computer, opening the website, asking questions, waiting for the AI to generate the response"* was not ideal for them because they needed to get things done as fast as possible at their job. Similarly, P1 said they could not wait for it to generate the response because they were used to getting the information immediately from their screen reader when interacting with the computer. P11 told us it would be much easier to hit a shortcut in some situations

than to ask the AI voice assistant for a response. *"As long as I have the feeling it takes me more time to use an AI than to do it myself and just type it into a search engine, then I do not find it helpful in my everyday life because it takes longer most of the time"* (P2). *"If I always had to say 'read the next cell,' 'read the previous cell,' it would be way too much"* (P11). P2 said a few settings had to be made before they could use the app, which was inefficient if they wanted to use it quickly. P9 said that the conversation did not feel natural regarding the flow: *"Rather than me hitting record every time, it should do it automatically when it wants responses from me."*

Moreover, the participants could **not interrupt the app** while it was talking and had to wait until the end of the response, which was frustrating for some of them. They were also unsatisfied that our app prototype could **not understand other languages**. P2 spoke English very well but wanted to ask a question containing a German city name, which the app prototype could not understand.

Participants were discouraged by the inability to interrupt the app and the lack of language support

Furthermore, P3 raised a concern that the app was **not practical for everyone** but only for BVI individuals. In the second task in the interactive part of the study, they noticed that the app mentioned the accessibility features of the phones and that they were not helpful for sighted people. They said that although the app was helpful for them, it should be designed for a general audience and not only for BVI users.

Participants suggested that the app should be designed for all users, not just BVI

Some participants raised potential concerns about **performing actions on behalf of the user** in the third task of the interactive part of the study concerning booking. First, they saw some privacy issues: *"It depends on where the AI system is situated. Here in Germany, we have some more data protection than in the United States. I do not want any server or company to share my private details."* (P1). They also said they would not be concerned about minor things like online book shopping because they gave their bank details to the website anyway. Two other participants were satisfied with the app booking everything for them but would benefit from the step in between with all the details to confirm the booking. However, they would prefer to make the transfer manually for the bigger ones.

Participants expressed privacy concerns with booking actions

Some participants preferred reviewing details before confirming larger bookings

App mistakes caused
user distrust

Secondly, participants were afraid of making mistakes with the app. P1 said: *"I do not know if I would book something like that with this kind of app at the moment, because what if it books a flight which is not convenient for me, for example, 2:00 in the morning?"* P5 described themselves as a *"mistrustful person"*, and they would not trust the app with all the booking details. P9 was frustrated after the third task: *"Its job is to find me the first available date and say 'This is the date available. Are you alright with that, or should we look any further?' Instead, it just goes, 'Alright. I have booked your flight, hotel and everything.' without even telling me what it has done when my flight is, and what is going on."* P10 was surprised negatively because the app gave them *"concrete prices for a very unspecific connection. [...] I do not know how much I would have paid and when I would have to enter the train or if it is a day ticket."*

Participants desired
more information and
independent booking
confirmation

Lastly, participants would need to know more about the booking source. P1 suggested that the app should give them some alternatives to choose from. Several participants would prefer to receive the booking confirmation via email and complete the booking process independently to know the booking page exactly.

Design Recommendations From Participants

We asked the participants what they would like to see in future AI voice assistants and what improvements they would suggest for the app prototype we created.

Some participants
preferred text input over
voice for convenience

Some participants preferred a **text input** option as an alternative to voice input. P5 said they were a *"real keyboard person"* and would always prefer to type than speak. P11 said they were much faster at typing than speaking, and interaction via text would be much simpler. S14 kept a wireless keyboard around their house, and they always used the browser versions of the apps on their laptop. P12 said they would decide about daily usefulness depending on existing text input options. Participants suggested adding a toggle between voice and text input so they could switch between them depending on the environment: *"For spelling things like a name or an address or for environments where I could not*

talk, it would be desirable to have a text input possibility" (P7). One participant thought about the separate input method for sensitive information: "[I could] just get a pop-up on my screen in the form of an edit field. And then just type the information, the data, the credentials or the card number. That is a better way of dealing with this. [Another] idea could be that you just type in some of the information you need, like the password for your PayPal account, so you do not need to say the password out loud" (P3).

Others would like a **text output** option, as it was already implemented in our app prototype. P10 said that text form was very beneficial "because then the user could also jump at the end of the results and look for the phrase the assistant is asking." Others mentioned that they would like to receive a text form before the voice output: "I would prefer to have the text output as soon as it starts. When the voice assistant is reading out the answer, I would love to read it in parallel" (P10), "I'd like to be able to review the text before it finished speaking" (P5). P9 said it was important to have "some quick way to go to the top of the text." Similarly to the text input, P5 suggested that such an AI voice assistant should have both options depending on the situation and the environment. Additionally, P5 would like to have the possibility to turn off the speech output entirely because they did not like human-sounding voices and any speech interactions with systems at all.

Other participants
desired simultaneous
text output with voice
responses

If it comes to the text output provided, participants mentioned some ideas to store the text for later use: "I would need that [cake recipe] in text form via email so I can use it later because I do not want to bake the cake now, but tomorrow and then I would have to go through the process again" (P2). "I would expect it to say things like 'Would you like to save this recipe? Would you like me to send the recipe by email?'" (P9). As an alternative, P2 suggested that the app should be able to copy the text into a file when the user asks and save it on the computer or download it in the browser. They also gave an idea for saving an audio file of the response. P9 said that it was an app that should propose to save the text in a file by saying "Would you like me to write this up for you and give you a link to download it?"

Users suggested
features for saving text
and audio responses

Smartphone voice assistant was suggested for efficiency and accessibility

The others would like AI voice assistant on their smartphones instead of the computer as a **mobile app**. P2 said it would be great, for example, when they were on the train and did not have a laptop. *"Your smartphone is always with you. Your laptop - you have to put it to start, and it takes a while. Sometimes, I am just too lazy to start my laptop, so I prefer to do things on my smartphone. Sometimes, it is easier to use. It would be OK if the same system would be on my smartphone"* (P1). Some participants proposed that AI voice assistant as a built-in feature might be more efficient and even more accessible: *"That might be easier for a lot of people because it is more accessible. You do not need to go over several steps to start it in the browser, to open the web page"* (P2). P3 subjectively evaluated that a built-in version *"would increase usability by 15 to 20%."*

Need for fully accessible app on all platforms

P10 summarised *"What I do not like about most assistance is that the navigation is relatively complex. You have an input field, and then another output field, and you have to jump from here to there and lose the overview over the results"* and recommended that the design of such voice assistant should be *"kept very easy."* Two participants raised the need for an app that is *"fully accessible on any platform and with any screen reader"* (S5).

AI voice assistant should connect to existing apps and devices

Many participants suggested that the AI voice assistant should be **connected to other apps**, devices or systems that they already use: *"Maybe an app that can link with things I already have"* (S9). Some of them proposed connecting the AI voice assistant to the screen reader: *"We can implement AI into the existing screen readers so that if there is any problem with accessing information, then we might make use of AI"* (P3). P4 said that good integration with the screen reader was the key because it should keep all its features. P5 suggested enhancing the current screen reader functions with AI capabilities. Still, they were unsure how it could be done: *"For example, on the ChatGPT website, when I type my response, there is no notification from the screen reader when the response comes in because they do not have ARIA set up. Nice if the screen goes 'OK, well, there is no text being off the screen, and it is scrolling.' Maybe it can make a scrolling noise, and then when it looks like it is done or it gets to know that website, play me a little song or something."* Others thought the other way around and sug-

gested creating an AI voice assistant with screen reader capabilities: *"The screen reader should be also implemented inside the AI-based app [...] We can extend [AI voice assistant] also to screen reading functionalities"* (P4). Instead of connecting the AI voice assistant to the screen reader, S6 suggested an AI voice assistant app which *"could automatically recognise when a screen reader is being used and assist. The AI [could] generate textual information through screen reader functionality, describing the different elements on the screen."*

Proposal for AI voice assistant to include screen reader features

"It should be possible to make some changes to these things because everybody has different preferences", said P2. Participants would like to have the possibility to **customise** the AI voice assistant to their needs as they do with their screen readers. They mainly mentioned settings related to the voice output, like the voice speed or tone. *"For me, it would not be needed because I like this voice and the speed is OK. It does not talk too quickly or too slowly, but it would be good if you could change that because some people like it more quickly, and some people like it more slowly. Some people like maybe a different voice"* (P2). Another participant, who was not entirely blind, raised a need to change the colour of the text in the app. Others mentioned that they would like to change labels by themselves. One of the participants said that it would be nice to change the words that response is being generated to *"sounds, some knock one after another so that you can hear all the time that it is working"* (P4). They said it was a necessary adjustment for the people who had problems with hearing because they might prefer to hear the sounds instead of the voice.

Participants desired customisable settings for AI voice assistants

What was also mentioned in terms of customisation was the possibility of providing feedback if the LLM-generated response was incorrect: *"Firstly, a feature that allows human correction and feedback to improve the AI's recognition and understanding capabilities would be incredibly beneficial. When I receive a description from artificial intelligence, I want to be able to correct it and provide feedback that would help the AI learn from its mistakes. This applies not just to images but also to descriptions for videos, documents, and overall context. This includes discussing the impact of lighting, image composition, and cultural references on image and video interpretation. I want to be able to speak to artificial intelligence regarding why it mis-*

Participants discussed feedback for improving AI responses

took one object for another, whether that was due to the angle of the picture, the context, the location, background elements, or the overall composition of the photo” (S6).

Participants want the AI to repeat answers

Participants wondered about the possibility of the AI voice assistant **repeating** answers: *“I should have asked if it is possible to repeat the last answer, especially in the recipes. It should be a possibility that it can be repeated because I cannot remember all steps of a recipe” (P4).* S19 said it was annoying that sometimes they had to repeat the process when the answer was unclear. They could ask the AI voice assistant to repeat the last answer if there was a repetition option.

Concerns about internet dependency; local hosting preferred

As mentioned in the limitations and disadvantages, participants were concerned about **internet dependency**. Some of them suggested that they could have the AI voice assistant on their local machine and not in the cloud: *“When it all takes place on my PC right in front of me, then it does not have to be sent through the world to be spoken to me” (P2).* P11, a software engineer, said that they would prefer to be able to host this app locally on their machine because of privacy concerns and hackers who may read the traffic. S16 said that the most beneficial feature for them would be *“if the AI was not an internet operating software, but rather a hardware that could provide visual assistance and sighted guiding.”*

Suggestion to upload files for enhanced interaction

Participants suggested that each AI voice assistant should have the option to **upload files** whether it is a text file, a picture, or a PDF document: *“Maybe it would be interesting to add more functionality such as uploading data, a picture, PDF document or Excel document” (P3).* This way, users could interact with the AI voice assistant more complexly by talking to it about the content of the uploaded files. Participants mentioned features similar to those within the Be My AI app, such as the possibility of asking follow-up questions about the picture.

AI should learn from the community to improve

Two participants mentioned the concept of tailoring the AI voice assistant specifically for BVI users by letting it **learn from the whole community**: *“The AI should also be able to learn from the community, gathering data from multiple users and incorporating it into its learning process. This would allow the AI to become more accurate and relevant over time and*

provide better responses to users. The AI should be able to suggest new parameters and ideas based on trending data and user feedback, allowing for continuous improvement” (S6). S14 suggested training the model on blind-specific forums: “Also training the LLM on blind-specific forums so it can learn from the collective knowledge of other blind people who have shared proven tips and tricks and strategies.” S6 also mentioned that it would be nice to subscribe to some packages, such as a writing package based on user feedback.

Training on
blind-specific forums'
data

Chapter 4

Discussion

In this chapter, we share the user study results and conduct a literature review to develop theories and design considerations. We also present the limitations of our study, pointing out any biases or challenges we faced along the way.

4.1 Theories

First, we examine the categories derived from the interactive study and online surveys. We present them below as a list of statements summarising the findings, and we discuss them in detail by comparing our results with the literature.

4.1.1 No matter how good the AI is, screen readers remain irreplaceable

Despite some imperfections and limitations such as dynamic content changes [Kim et al., 2021], image description [Guinness et al., 2018; Stangl et al., 2020] or numerous commands to remember [Vtyurina et al., 2019], almost all of our participants use screen readers daily and are very satisfied with them. Screen readers are quick and relatively easy to use once the user learns the shortcuts and gestures. BVI

BVI individuals use
screen readers daily
and are satisfied

people are used to screen readers and can do almost everything with them very efficiently. Vtyurina et al. [2019] recognised that screen readers are needed for in-depth exploration and can provide a significant number of navigation modes which operate at different granularities.

Screen readers and AI
serve different
purposes

When we asked to compare the AI-powered system with LLM-generated responses with the screen reader, our participants said these technologies were two completely different concepts serving distinct purposes. They found it frustrating to use the AI voice assistant for tasks they could complete with their screen reader because they would have to dictate every command to get the desired information. One of the findings of Phutane et al. [2023] is similar - their participants could not imagine talking to their screen reader. On the other hand, Ashok et al. [2015] found out that users prefer to use screen readers with a voice control feature to save time and effort. This might be because their research was conducted in 2015, and technology such as voice control was still new and exciting for the participants.

AI voice assistants
could complement
screen readers'
capabilities

However, the AI voice assistant could be a great addition to the screen reader. As per Phutane et al. [2023], AI-powered features could be an extension to any of the current users' ecosystems. AI might be used for tasks that cannot be done with the screen reader as an only tool, such as image description or quick information retrieval. We also discovered that users would appreciate the possibility of AI helping them browse inaccessible web pages. Kuzdeuov et al. [2024] noticed the same and called AI an "accessibility enhancer" in such cases.

Suggestions for AI
implementation

Such a feature could be implemented just as a single button that switches the AI voice assistant on and off or a feature that recognises in which situations the AI voice assistant could be helpful and switches it on automatically, as one of our participants proposed. Some suggested that this could be done the other way around, meaning that the AI voice assistant could be enhanced with screen reader features, such as navigation via shortcuts and commands.

4.1.2 The most crucial feature sought by BVI users is the ability to describe images and ask follow-up questions

In the previous theory 4.1.1, we mentioned that AI voice assistant could be an accessibility enhancer. As pointed out by our participants, some of the primary use cases of AI assistants associated with improving accessibility are image description and navigating the environment in real time. They include:

- manually uploading the image and getting its description,
- an immediate description of inaccessible images on the webpages (as a way to enhance the general accessibility of the website), and
- pointing the smartphone's camera to receive a real-time description.

Our participants need to have the opportunity to ask follow-up questions about images. Phutane et al. [2023] findings also suggest that BVI users' most common desire is to ask questions about images, which could be especially useful on social media pictures that often lack alt-text descriptions. Stangl et al. [2020] found out that news articles and social media sites are the ones that are the most inaccessible for BVI users. Their participants reported the need to have image descriptions that clarify the purpose of the image and the context in which it is used (for example, for images where people are present, participants wanted to know who was in the image and what they were doing).

The authors proposed future work on image description where users could specify the quantity of the content description, decide the level of precision and ask for descriptions only for demand, which is similar to what our participants noted and can be achieved together with the feature to ask follow-up questions about the image.

Although solutions such as Be My AI exist, they cannot serve other purposes, such as information retrieval

AI voice assistants enhance accessibility through image real-time descriptions

BVI users desire image context and clarification

Future image descriptions should allow user specifications

AI could enhance image description systems

GPS for BVI individuals
poses developmental
challenges but is
valuable

searches or setting reminders. The latest version of OpenAI's GPT model, GPT-4o (mentioned in 2.2.2), could be an excellent base for developing an AI system that describes pictures, navigates through the environment in real-time and performs other tasks, such as retrieving the information. What could be more challenging to achieve with this model right now is a GPS function specifically tailored to BVI individuals that would be a valuable addition for navigation purposes.

4.1.3 AI assistants can be effective information retrieval tools for BVI users

AI assistants aid BVI
users in online
searches

Another interesting use case is retrieving information from the internet. AI assistants can help BVI individuals conduct searches almost as efficiently as sighted users. Some of our participants already use traditional voice assistants or ChatGPT for quick information retrieval, but they all agree that AI-powered voice assistants could be faster. The need to review the search results individually and read the whole webpage to find particular information could be eliminated since the AI assistant can provide an immediate answer for the given query, including sources. Similarly, Murphy et al. [2008] found out that many web pages contain large numbers of outgoing links and going through them is time-consuming for BVI users. They confirmed that less experienced users tend to visit familiar or "blind-friendly" web pages so they do not experience information overload. Most of the time, they ask for help from a sighted person to find the information they seek. Moreover, Ashok et al. [2015] discussed that a serial narration of content from the web could cause irrelevant information to be read aloud.

WebChatGPT can be
improved for BVI
accessibility

There already exist Google Chrome extensions such as WebChatGPT¹ that provide an instant answer to the user's query provided in search engine together with web page sources. These could be redesigned to be more accessible for BVI users, for example, by serving as a voice assistant

¹ <https://tools.zmo.ai/webchatgpt>

that reads the answer aloud and provides an intuitive way to browse the sources.

However, there exist some limitations regarding the information retrieval capabilities of AI. Our participants pointed out that most of the time, responses are either too general or too detailed, and they would like more control over the amount of information received. It was also noticed by Phutane et al. [2023], who demonstrated that users prefer to have a system that has different roles for different tasks: for more general questions, they prefer to receive only a straightforward answer; for more complex questions - information about only specific things and for more exploratory questions - a list of keywords that they can further explore. Some of their participants viewed summarisation as a loss of user control, and they preferred to go over the web page themselves. Besides, Vtyurina et al. [2019] found that a single answer is appropriate for simple question answering but not for more exploratory search tasks, as it provides only limited information.

Users desire more control over AI response details

AI voice assistants are great as information retrieval tools or search engines. Still, they have to offer a more efficient and customisable way to do it than traditional voice assistants.

AI assistants need enhanced efficiency and customisation

4.1.4 BVI users expect AI systems to be flexible

The previous theory 4.1.3 raised the need for AI systems to be more customisable and flexible. We discussed the adaptability and personalisation of such AI tools with our participants and discovered that customisation is undoubtedly one of the most crucial aspects of such technologies. The core category and its subcategories are shown in Figure 4.1.

The need for customisable and flexible AI systems

A critical customisation parameter is switching between the text and voice modes for both input and output. Regarding input, it should be possible to change its type depending on the context or completely switch off the voice input. Phutane et al. [2023] and Vtyurina et al. [2019] found out that many of their participants were more effi-

Switching between text and voice input

Some users prefer typing

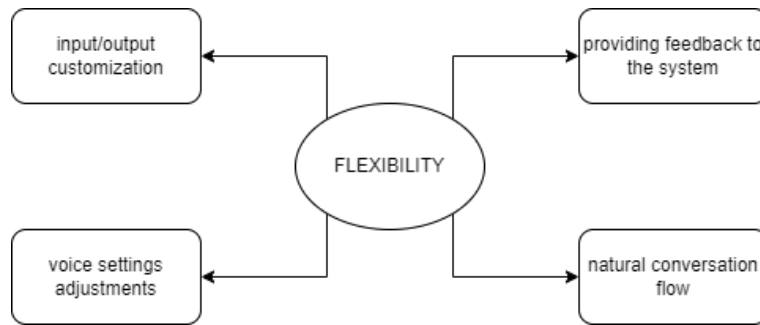


Figure 4.1: Theory 4.1.4: BVI users expect AI systems to be flexible. The figure shows the main category derived from the data we gathered: flexibility and groups of codes that belong to it - input/output customisation, voice settings adjustments, providing feedback to the system and natural conversation flow.

cient with a keyboard than their voice, so they would also appreciate the possibility of typing the question. Regarding output, our participants would like to have the possibility to read the response in addition to hearing it. Our app prototype already implements this, but the text appears after the voice output is finished, which is not always convenient.

Feedback feature
enables corrections and
system learning

Moreover, the possibility of providing feedback to the system is a helpful feature. The opportunity to correct the system when it makes a mistake and emphasise that it should learn from these corrections is very appreciated by our participants. Likewise, participants from a study conducted by Phutane et al. [2023] shared expectations from an AI to learn their behaviour and adapt to task, their personalities and their moods when using the system. They also mentioned that such systems should learn their navigation patterns.

Users desire voice
settings

Furthermore, our participants mentioned adjusting voice settings, like speed or tone. Participants from Phutane et al. [2023] and Abdolrahmani et al. [2018] also stressed that such an AI system should be customisable enough to allow users to set their preferences such as speech rate, the order in which information is read and the voice tone. We

found out that the option to choose the conversation's language should also be considered.

Lastly, the possibility of talking to the AI voice assistant naturally makes the system more flexible, and our participants appreciate it. They value not robotic human-like voice of our app prototype and the possibility to ask questions conversationally without using phrases as opposed to traditional voice assistants.

Natural conversation with AI improves flexibility and satisfaction

4.1.5 BVI users' big frustration with voice assistants is the lack of understanding

However, with natural conversation flow comes one of the most significant frustrations of BVI users with voice assistants - the lack of understanding. Some users do not like voice assistants at all because assistants often do not understand them, leading to frustration as users lack the patience to repeat questions multiple times. The voice recognition is not always perfect, and our participants complained about various systems being unable to understand their heavy accents. Phutane et al. [2023] also noticed the frustration of BVI users with voice assistants not understanding their speech. Likewise, Abdolrahmani et al. [2020] found out that misunderstanding caused participants to lose confidence in voice assistants' accuracy and utility. This led to significant trust issues.

Understanding difficulties frustrate BVI users of voice assistants

Misunderstandings cause mistrust in voice assistant accuracy

This theory does not refer only to misunderstanding the speech but also to misinterpreting the questions. Abdolrahmani et al. [2018] noticed that assistants often misinterpret users' questions, particularly in noisy public settings. Our participants would use voice assistants only for simple tasks because they fear they would not be well understood for more complex ones, such as writing an important email or scheduling a meeting in the calendar. Some suggested that the system should always ask for confirmation when it does more complicated things.

Misinterpretation of questions is a common issue

Moreover, users often need to memorise specific keywords, phrases, or voice commands to get the desired information,

Users must use specific commands

AI improvements could enhance natural conversation flow

for example, "start the washing machine whites program" instead of "switch on the washing machine and choose a program for white clothes." Voice assistants also often fail to catch colloquial phrases or names of the places or people [Abdolrahmani et al., 2018], which our participants also experienced. This is the room for improvement that AI-generated responses could fill in as they could provide more natural conversation without specific keywords and better understand the user's intent.

4.1.6 Inaccurate LLM-generated responses lead to BVI users' mistrust of AI systems

Users need reliable AI for daily practical use

Sources could enhance trustworthiness

Misunderstanding is not the only reason for BVI users' mistrust of AI systems. One of the most critical aspects of the AI tool with LLM-generated responses is the accuracy of the responses. We found that users need the tool to be reliable for daily use. Without trust, the usage will be somewhat playful and exploratory. The AI's generativeness is problematic due to its potential to yield false information. One of the suggested potential solutions to this problem is that the system could provide information from both web and pre-programmed responses. Furthermore, the system should be able to provide the sources of the information to increase trustworthiness. However, some of our participants mentioned that they would not use the system to find accurate information in any case, regardless of the information source, due to their general mistrust of this technology.

Inconsistent answers are as problematic as false information

On the other hand, participants noted that false information is as problematic as the system providing inconsistent answers to the same question. The same phenomenon was noticed by Wulf et al. [2014] but in the responses from traditional voice assistants, which negatively influence user behaviour.

4.1.7 An AI system that performs actions on behalf of users could be beneficial for BVI individuals

Most of our participants are very optimistic about the possibility of the AI system performing actions on their behalf. It is perceived as a great help in completing time-consuming or challenging tasks. Some tasks tend to be inaccessible at some point, as noted by Ashok et al. [2015]. They discovered that BVI users spend excessive time and effort on simple day-to-day tasks such as online shopping or flight reservations. This is primarily due to content-rich web pages filled with information and advertisements. The authors reported that it gets typically worse for novice users who are not familiar with the screen readers yet and use, on average, five different shortcuts.

Users are optimistic about AI assisting with tasks

A feature that automates tasks would ease a lot of online activities, allowing BVI people to be more independent and saving them a lot of time. Our participants appreciate putting the system in charge of specific tasks by providing personal information. In general, they do not see any privacy concerns in such a system because they give their personal information to other web pages anyway. Although some already use traditional voice assistants for less complex tasks, they sometimes get frustrated for various reasons, some of which we mentioned, in theory 4.1.5.

Automating tasks enhances independence

Similarly, Abdolrahmani et al. [2018] suggested that traditional voice assistants should support more apps, services, and complex tasks through voice interaction. In their other study, they noted that the biggest frustration is voice assistants having a finite time to listen to the user, which can create pressure to speak quickly before the time runs out [Abdolrahmani et al., 2020]. Moreover, they discovered that some apps may be inaccessible, so it is not always possible to interact with them using a traditional voice assistant. The AI system would be greatly appreciated for completing more complex tasks on any application without time constraints. Further enhancing inclusion is the fact that performing actions such as setting reminders or tracking calen-

Voice assistants need broader support and improved accessibility

dars could be helpful for people with memory loss [Vtyurina et al., 2019].

Our participants want more control and privacy with AI tasks

On the other hand, some of our participants are more cautious about the AI system performing actions on their behalf. They have privacy issues with the system's access to their personal information. Although they appreciate the help, they would prefer more control over providing the data to the system. On top of that, they would not trust the system with tasks containing detailed information, such as exact departure time while booking flights. They would prefer a confirmation step in between to prevent the system from making mistakes. Likewise, Phutane et al. [2023] found out that consent is crucial for users regarding AI systems performing actions.

Some are open to data access for efficiency

Other participants from our study are open to the system accessing their personal information if it would enhance task efficiency, as they already share their data with other websites. The only thing they would need is a secure internet connection.

Exploring Rabbit R1 as a service for BVI users

We suggest that Rabbit R1, described in section 2.2.3, might be one of the solutions to this problem as it could perform more complex tasks. This device could be redesigned so that the user does not have to buy the product but could use it as a service, for example, a simple smartphone app or a screen reader addition. However, the potential of this technology needs further exploration and testing with BVI users to determine its actual benefits for them.

4.1.8 BVI users seek safe AI systems that do not share their data

User data safety and privacy are critical concerns

Theory 4.1.7 leads us to another one - the safety of the users' data. As mentioned several times, one of the primary concerns expressed by our participants is the system's privacy. Similarly, the findings by Phutane et al. [2023] indicated that confidentiality and privacy were essential for new system users. Cambre et al. [2021] suggested that privacy should be vital to any web-based system. This refers

to the data collected by the system and the data that it shares. Users often seek knowledge about the system's privacy policy, especially the information about where the data is stored.

Moreover, the internet dependency is also considered as a privacy concern. The data is usually stored on the cloud, and the system must be connected to the internet to work. This raises issues about the system being hacked and data being stolen.

Internet dependency poses risks of data theft

Moreover, the use of the system in public is a danger to privacy. Apart from interrupting others with the system's voice output, which several of our participants mentioned, they see some privacy issues with others overhearing their personal information. As Abdolrahmani et al. [2018] noticed, participants prefer not to interact with Siri publicly because of privacy concerns and social awkwardness. However, there are not many privacy concerns when using them at home. According to Phutane et al. [2023], maintaining privacy in public was a concern for their participants, and each had different preferences.

Public use raises privacy and social awkwardness issues

It is crucial to provide the possibility of adjusting the privacy settings. Phutane et al. [2023] suggested three different levels of control for various kinds of tasks: for voice and cursors - granular control; for representing on-screen content - medium control; and for task assistance - high control. This way, users could always choose the level of privacy they are comfortable with.

Customisable privacy settings are essential for user control

4.1.9 Simplicity in AI systems is the key

However, the AI system should not be overloaded with features. In general, the simplicity is highly respected. Our participants appreciated the plainness of the app prototype. Figure 4.2 shows the theory's subcategories described in this section.

Simplicity is highly valued, and feature overload should be avoided

The participants found our app prototype easy to use. Even beginner users saw the usage of the app prototype

The system should be intuitive and user-friendly

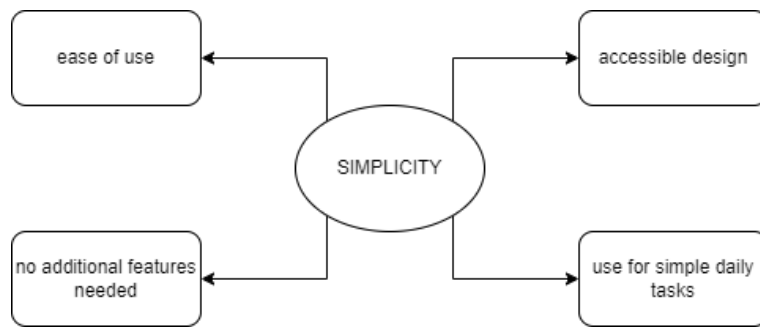


Figure 4.2: Theory 4.1.9: Simplicity in AI systems is the key. The figure shows the main category derived from the data we gathered: simplicity and groups of codes that belong to it - ease of use, no additional features needed, accessible design and use for simple daily tasks.

as intuitive and very easy to learn. The design was not overloaded with information and was easy to understand. Some said that the system did not need any additional features because it would cause it to be more complicated. They appreciated the fact that the responses were concise and detailed at the same time.

Accessible design aids
navigation and supports
simplicity

Moreover, the accessible design is often related to the system's simplicity. The fact that every element was clearly labelled made it easier to navigate through the app. Vtyurina et al. [2019] found out that users with memory loss sometimes encounter difficulties with the system's complexity due to the need to remember the commands. If the system is kept simple and easy to use, it can be more inclusive for BVI users with other additional impairments. Additionally, the AI system with LLM-generated responses could effectively assist with the simplest daily tasks, such as checking the weather or reading emails.

4.1.10 BVI users do not see much difference between an AI-based voice assistant and a traditional one

As mentioned several times, BVI people are heavy users of traditional voice assistants, such as Siri, Google Assistant or Alexa. They use them for simple daily tasks, such as setting reminders, checking the weather or reading mail. They generally appreciate voice-only interaction and the possibility to ask questions conversationally. However, they often get frustrated with them for various reasons, such as misunderstanding described in theory 4.1.5.

BVI users frequently utilise traditional voice assistants for daily tasks

We asked our participants to compare the AI-powered voice assistant with traditional ones. Some of them, particularly those with IT backgrounds, value the generative nature of the responses. In contrast, others perceive little difference because they could do the same tasks using both. Another point of view was that the AI-powered voice assistant could be better as a search engine, but the traditional voice assistant is better for daily tasks. The results might be biased because some participants were unaware that AI does not power traditional voice assistants. Likewise, Abdolrahmani et al. [2018] discovered that participants were surprised that Siri could not learn over time.

Our participants compare AI assistants to traditional ones, noting only minor differences

Some users are unaware traditional assistants lack AI capabilities

4.1.11 Users take sustainability into account when using AI

Last but not least, we also address the sustainability of AI. No matter how good the AI system is and how many features it has, some users may not use it daily because of the environmental impact this technology has.

Sustainability of AI may affect user adoption

Their worries are justified. Van Wynsberghe [2021] discussed that recent advancements in AI caused a significant increase in energy consumption and carbon emissions due to the computer power needed to train and use the models. For example, she presented that Google's AlphaGo model generated 96 tons of carbon dioxide over 40 training days, which equals 1000 hours of air travel. The author proposed

AI advancements increase energy use and emissions significantly

a definition of "sustainable AI" as AI that is applied in a way that addresses issues of sustainability. She highlighted the importance of being aware of the environmental costs associated with AI and emphasised the need to use it responsibly.

Users recognise sustainability issues and prioritise energy efficiency

Our participants are aware of the sustainability concerns related to AI systems and care deeply about these issues. They expressed worries about energy consumption and the environmental impact of these technologies. This awareness influences their approach to technology, aiming to save energy and minimise environmental impact whenever possible. They would appreciate an energy-efficient system that does not consume too much power. Others mentioned that AI does not have to be used everywhere and that, in some cases, it is more sustainable to do things without it.

Some believe AI should not be used everywhere

4.2 Design Considerations

Description of the table with design considerations

Table 4.3 gathers the design considerations for AI-based systems with LLM-generated responses for BVI users. We based them on the findings from the user study and the literature review. The first column of the table is the theme - one of the core categories derived from the data we gathered. The second column includes the design considerations that belong to the theme.

The design of the AI system should prioritise inclusivity to enhance accessibility and improve user comfort

These considerations may help to design AI systems for BVI users. However, they could also be applied to the general audience, as some appear not specific to BVI users only. By following these guidelines, we can create technology that is easier to use and understand for everyone. Good design for BVI users can lead to better experiences for all users, making technology more accessible. The system should be designed to ensure user comfort and safety. It should be intuitive, straightforward, customisable, and reliable. Moreover, the system should be energy-efficient and prioritise the user's privacy. Finally, it should be compatible with other assistive technologies and be trained on a diverse dataset that includes BVI users' data to ensure inclusivity.

Theme	Design Consideration
Simplicity	The system should be easy to use and understand even for beginner users. It should not be overloaded with features. The design should be accessible and easy to navigate.
Customization	Users should be able to change both the input and output types, choosing between voice and text. They should be able to adjust voice settings such as speed, tone, and the conversation's language. The system should learn the user's behaviour over time and adapt to their personalities and moods.
Natural Conversation Flow	The system should understand the user's intent and respond naturally. It should not require specific keywords or voice commands. It should indicate when the response is being generated. Users should also be able to interrupt the system's response if they do not wish to continue.
Information Retrieval	The system should serve as a search engine and provide immediate answers to the user's query. It should be possible to ask follow-up questions about the information provided.
Accuracy	Responses should be accurate and include sources to ensure reliability. Users should receive consistent information each time they ask the same question.
Internet Independence	The system should be able to work offline and not require a constant internet connection for simple tasks that do not need it.
Sustainability	The system should be energy-efficient and not consume too much power.
Privacy	There should exist a clear privacy policy. It should also be possible to adjust the privacy settings. The system must request consent before performing actions on behalf of the user.
Public Comfort	The user should feel comfortable using the system in public. The possibility of adjusting the privacy settings and input/output type should be provided.
Mobile Version	The system should either be available as a mobile app or have a mobile-friendly website version.
Compatibility	The system should be compatible with other assistive technologies, such as screen readers.
Appropriate Training Set	The system should be trained on a diverse dataset that includes data from BVI users, such as information from web forums.

Table 4.3: Design considerations for an AI system with LLM-generated responses based on the findings from interactive interviews and literature reviews. Some of them apply not only to AI voice assistants designed specifically for BVI people but also to the general audience.

4.3 Limitations

Study reveals limitations in interpreting results

Our study revealed several limitations that should be considered when interpreting the results. In section 3.1.3, we discussed the limitations of the app prototype, along with those identified by participants in section 3.2.2. In this section, we outline the limitations of the study itself.

Interactive study was fully online, causing technical issues

First of all, the interactive study was conducted entirely online. Several participants faced technical issues, as mentioned in section 3.1.2. Despite the problems with the video conferencing system being solved very quickly, some issues with the app prototype could not be fixed during the study because of time constraints (see 3.1.3). This might have influenced the participants' experience with the app prototype and their answers to the general questions about AI.

Limited participant diversity may affect generalisability

Secondly, the interactive study was conducted with a limited number of participants. Most considered themselves intermediate or advanced technology users who used an AI system at least once, so the beginner users' perspective might have been underrepresented. Although we reached data saturation, some findings might not be generalisable to the BVI population. Furthermore, participants focused primarily on the app prototype during the interview, even though we asked about AI in general. This focus may have biased the results toward the app prototype. Lastly, some participants were not advanced English speakers, and the language barrier might have affected the results because they might not have found the right words to express their thoughts.

Grounded theory method relies on researcher interpretation

Last but not least, the grounded theory method has its limitations. It is a qualitative research method based on the researcher's interpretation of the data [Pandit, 1996; Noble and Mitchell, 2016]. Our background and experience with the topic might have impacted the findings. Moreover, this method requires several data sources to be used to ensure the reliability of the findings. We used the interactive study sessions, online surveys and the literature review to gather the data, but other sources may have provided more insights.

Chapter 5

Summary and Future Work

This chapter concludes our research by summarising the findings and presenting the contributions. We also discuss possible opportunities for future work.

5.1 Summary and Contributions

In this thesis, we explored the potential of personalised LLM-generated responses in the context of BVI users. We developed the simplest possible AI voice assistant with the ability to generate personalised responses. This prototype was evaluated with BVI users to understand the impact of personalised LLM-generated responses on their user experience. We conducted an interactive user study with 13 BVI participants to determine the advantages and limitations of AI-based systems. We also collected additional feedback through an online survey. The whole study design and data analysis followed the grounded theory approach, which led us to identify the key themes in the data without preconceived notions or overly detailed research questions and hypotheses.

We investigated the effectiveness of LLM-generated personalised responses for BVI users, using a grounded theory approach

We analysed the participants' experiences with AI-based systems and did a literature review to answer our wide research question:

RQ What are the advantages and disadvantages of AI voice assistants with LLM-generated responses for BVI users, and how can they be improved?

BVI users value personalised, simple, and safe AI systems

In general, BVI users are open to using AI-based systems. Some of our participants already used AI technologies, most of which were not designed specifically for BVI users. They showed a positive attitude towards the AI voice assistant prototype we created and found it useful. We concluded that BVI individuals could benefit from the personalised responses generated by the LLM in their daily lives. The most common use cases involve retrieving information, describing images, or performing actions on behalf of the user. Most importantly, the users appreciate personalised responses that make the interaction with the system feel more human-like and natural. AI-based systems should be simple, customisable, and safe - these are the key factors BVI users consider when adopting new technology.

They are concerned about privacy, reliability, and compatibility of AI systems

However, our study and the literature review highlighted several limitations and challenges with AI-based systems. Users are concerned about the privacy and security of their data and emphasise the need for a clear privacy policy. Additionally, there are worries about the reliability of AI systems as users recognise the potential for mistakes. Furthermore, many report that traditional assistive technologies, such as screen readers, outperform AI-based systems regarding efficiency and reliability. Any new AI technologies must be fully compatible with existing assistive tools. Most BVI users will hesitate to fully replace their current assistive means with a single AI-based system.

Recommendations for developing inclusive AI systems for BVI users are our main contribution

Our findings serve as design recommendations and guidelines for developers and designers of AI-based systems for BVI users. We gathered the main use cases and scenarios where AI-based systems can benefit BVI users, the key factors influencing the user experience, and the limitations and challenges that should be considered. Design consider-

ations included in 4.3 might help to create more accessible and inclusive AI-based systems for BVI users.

5.2 Future Work

Firstly, we suggest creating a more advanced AI-based prototype with more features and capabilities. It should include various customisation options and serve more purposes than just information retrieval. Performing actions on behalf of the user could be further investigated, as our findings revealed that it might be very beneficial for BVI users. It should be developed into a fully functional feature that triggers actions, allowing users to experience its complete potential. Such functionality should ensure safety and security, and the user should have complete control over the system. There is also a need for more research on image description, information retrieval and performing actions on behalf of the user for BVI users, as these are the most commonly mentioned use cases for AI.

Future work should include developing an advanced AI system for BVI users with enhanced features

Secondly, testing on a larger scale is recommended. Different people have different needs and preferences, so testing the system with a more diverse group of users is essential. Including more technology beginner users and those who have never used AI-based systems in the research is essential to ensure that these tools are easy to learn.

Larger, diverse user testing is essential for system evaluation

Finally, the most crucial aspect of future work is simply continuing to research AI-based systems for BVI users. The field of AI is rapidly evolving, and it is essential to keep up with the latest technologies and make them accessible and tailored to people with various disabilities.

Continuing research on accessible AI for BVI users is of utmost importance

Appendix A

List of Most Important Interview Questions for User Study

- Questions about the app prototype
 - Describe your overall experience.
 - How easy was it to use the app?
 - How did you feel about response time?
 - How could the app be improved to make the experience better?
 - Are there any additional features you would like to see?
 - What did you think about booking capabilities?
 - Do you see any privacy concerns?
- Questions about AI and non-AI assistive technologies
 - Share your experience with AI and non-AI assistive technologies. Which ones do you like the most and why? Which ones do you dislike and why?
 - What do you find most helpful about using systems with AI?
 - How could an AI voice assistant compare to other technologies you have used?

- Can an AI voice assistant replace other assistive technologies?
- Would you use an AI voice assistant in your daily life?
- Share some scenarios where you would use an AI voice assistant daily.

Bibliography

- [1] Ali Abdolrahmani, Ravi Kuber, and Stacy M Branham. "Siri Talks at You" An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th international ACM SIGACCESS conference on computers and accessibility*, pages 249–258, 2018. doi.org/10.1145/3234695.3236344.
- [2] Ali Abdolrahmani, Kevin M Storer, Antony Rishin Mukkath Roy, Ravi Kuber, and Stacy M Branham. Blind leading the sighted: drawing design insights from blind users towards more productivity-oriented voice interfaces. *ACM Transactions on Accessible Computing (TACCESS)*, 12(4):1–35, 2020. doi.org/10.1145/3368426.
- [3] Aboubakr Aqle, Dena Al-Thani, and Ali Jaoua. Can search result summaries enhance the web search efficiency and experiences of the visually impaired users? *Universal Access in the Information Society*, 21(1):171–192, 2022. doi.org/10.1007/s10209-020-00777-w.
- [4] Vikas Ashok, Yevgen Borodin, Yury Puzis, and I.V. Ramakrishnan. Capti-speak: a speech-enabled web screen reader. In *Proceedings of the 12th International Web for All Conference*, pages 1–10, 2015. doi.org/10.1145/2745555.2746660.
- [5] Vikas Ashok, Syed Masum Billah, Yevgen Borodin, and I.V. Ramakrishnan. Auto-suggesting browsing actions for personalized web screen reading. In *Proceedings of the 27th ACM Conference on User Modeling, Adaptation and Personalization*, pages 252–260, 2019. doi.org/10.1145/3320435.3320460.
- [6] Chaitanya Bhosale, Priyanka Joshi, Neha Dorage, Rutuja Lamkhade, and Sakshi Musale. Object Detection and Disease Prediction for Blind People. *Journal Of Technical Education*, page 307, 2023.
- [7] Jeffrey P. Bigham, Anna C. Cavender, Jeremy T. Brudvik, Jacob O. Wobbrock, and Richard E. Ladner. WebinSitu: a comparative analysis of blind and sighted browsing behavior. In *Proceedings of the 9th International ACM SIGACCESS*

- Conference on Computers and Accessibility*, page 51–58, 2007. doi.org/10.1145/1296843.1296854.
- [8] Yevgen Borodin, Jeffrey P. Bigham, Glenn Dausch, and I. V. Ramakrishnan. More than meets the eye: a survey of screen-reader browsing strategies. In *Proceedings of the 2010 International Cross Disciplinary Conference on Web Accessibility (W4A)*, 2010. doi.org/10.1145/1805986.1806005.
- [9] Rupert RA Bourne, Jaimie Adelson, Seth Flaxman, Paul Briant, Michele Bottono, Theo Vos, Kovin Naidoo, Tasanee Braithwaite, Maria Cicinelli, Jost Jonas, and others. Global Prevalence of Blindness and Distance and Near Vision Impairment in 2020: progress towards the Vision 2020 targets and what the future holds. *Investigative Ophthalmology & Visual Science*, 61(7):2317–2317, 2020.
- [10] Ben Caldwell, Michael Cooper, Loretta Guarino Reid, Gregg Vanderheiden, Wendy Chisholm, John Slatin, and Jason White. Web content accessibility guidelines (WCAG) 2.0. *WWW Consortium (W3C)*, 290(1-34):5–12, 2008.
- [11] Julia Cambre, Alex C Williams, Afsaneh Razi, Ian Bicking, Abraham Wallin, Janice Tsai, Chinmay Kulkarni, and Jofish Kaye. Firefox voice: An open and extensible voice assistant built upon the web. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–18, 2021. doi.org/10.1145/3411764.3445409.
- [12] Juliet M Corbin and Anselm Strauss. Grounded theory research: Procedures, canons, and evaluative criteria. *Qualitative sociology*, 13(1):3–21, 1990. doi.org/10.1007/BF00988593.
- [13] Jenny Craven. Linear searching in a non-linear environment: The information seeking behaviour of visually impaired people on the world wide web. In *Computers Helping People with Special Needs: 9th International Conference, ICCHP 2004, Paris, France. Proceedings 9*, pages 530–537, 2004. doi.org/10.1007/978-3-540-27817-7_77.
- [14] John R Cutcliffe. Methodological issues in grounded theory. *Journal of advanced nursing*, 31(6):1476–1484, 2000. doi.org/10.1046/j.1365-2648.2000.01430.x.
- [15] Chiara Filippini, Giulio E Lancioni, Gloria Alberti, Francesco Pezzuoli, and Patrizia Ceccarani. A Technology Aid to Help People with Blindness and Moderate Intellectual Disability Retrieve Common Objects from Storage Units: A Proof-of-Concept Study. *Sensors*, 24(14):4453, 2024. doi.org/10.3390/s24144453.
- [16] Barney Glaser and Anselm Strauss. *Discovery of grounded theory: Strategies for qualitative research*. Routledge, 1967. doi.org/10.4324/9780203793206.

- [17] Darren Guinness, Edward Cutrell, and Meredith Ringel Morris. Caption crawler: Enabling reusable alternative text descriptions using reverse image search. In *Proceedings of the 2018 CHI conference on human factors in computing systems*, pages 1–11, 2018. doi.org/10.1145/3173574.3174092.
- [18] Jacob Hilton, R Nakano, S Balaji, and John Schulman. WebGPT: Improving the factual accuracy of language models through web browsing. *OpenAI Blog*, December, 16, 2021.
- [19] Shakila Cherise S Joyner, Amalia Riegelhuth, Kathleen Garrity, Yea-Seul Kim, and Nam Wook Kim. Visualization accessibility in the wild: Challenges faced by visualization designers. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, pages 1–19, 2022. doi.org/10.1145/3491102.3517630.
- [20] Nam Wook Kim, Shakila Cherise Joyner, Amalia Riegelhuth, and Yeeun Kim. Accessible visualization: Design space, opportunities, and challenges. In *Computer graphics forum*, volume 40, pages 173–188, 2021. doi.org/10.1111/cgf.14298.
- [21] Askat Kuzdeuov, Olzhas Mukayev, Shakhizat Nurgaliyev, Alisher Kunbolsyn, and Huseyin Atakan Varol. ChatGPT for visually impaired and blind. In *2024 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC)*, pages 722–727, 2024. doi.org/10.1109/ICAIIIC60209.2024.10463430.
- [22] Gary Marchionini. Exploratory search: from finding to understanding. *Commun. ACM*, 49(4):41–46, apr 2006. doi.org/10.1145/1121949.1121979.
- [23] Linda Weaver Moore. Conducting research with visually impaired older adults. *Qualitative Health Research*, 12(4):559–565, 2002. doi.org/10.1177/104973202129120070.
- [24] Emma Murphy, Ravi Kuber, Graham McAllister, Philip Strain, and Wai Yu. An empirical investigation into the difficulties experienced by visually impaired Internet users. *Universal Access in the Information Society*, 7:79–91, 2008. doi.org/10.1007/s10209-007-0098-4.
- [25] Helen Noble and Gary Mitchell. What is grounded theory? *Evidence-based nursing*, 19(2):34–35, 2016. doi.org/10.1136/eb-2016-102306.
- [26] Naresh R Pandit. The creation of theory: A recent application of the grounded theory method. *The qualitative report*, 2(4):1–15, 1996.
- [27] Mahika Phutane, Crescentia Jung, Niu Chen, and Shiri Azenkot. Speaking with My Screen Reader: Using Audio Fictions to Explore Conversational Access to Interfaces. In *Proceedings of the 25th International ACM SIGACCESS*

- Conference on Computers and Accessibility, ASSETS '23, 2023. doi.org/10.1145/3597638.3608404.*
- [28] Alisha Pradhan, Kanika Mehta, and Leah Findlater. "Accessibility Came by Accident" Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, pages 1–13, 2018. doi.org/10.1145/3173574.3174033.
- [29] A Rajendran, C Dhanush, Krishna S Hari, Abul Hasil S Mohamed, and M Mo-hanaprasath. Visual Content to Spoken Word Transformation for Visually Impaired People. In *Emerging Trends in IoT and Computing Technologies*, pages 78–82. 2024.
- [30] Nuzhah Gooda Sahib, Anastasios Tombros, and Tony Stockman. A comparative analysis of the information-seeking behavior of visually impaired and sighted searchers. *Journal of the American Society for Information Science and Technology*, 63(2):377–391, 2012. doi.org/10.1002/asi.21696.
- [31] Talal Saleem and V Sivakumar. A Mobile Lens: Voice-Assisted Smartphone Solutions for the Sightless to Assist Indoor Object Identification. *EAI Endorsed Transactions on Internet of Things*, 10, 2024. doi.org/10.4108/eetiot.6450.
- [32] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. "Person, Shoes, Tree. Is the Person Naked?" What People with Vision Impairments Want in Image Descriptions. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems, CHI '20*, page 1–13, 2020. doi.org/10.1145/3313831.3376404.
- [33] Corinna Trujillo Tanner, Michael S Caserta, Julia J Kleinschmidt, Margaret S Clayton, Paul S Bernstein, and Jia Wen Guo. Conducting research with older adults with vision impairment: Lessons learned and recommended best practices. *Gerontology and Geriatric Medicine*, 4:2333721418812624, 2018. doi.org/10.1177/2333721418812624.
- [34] Lorainne Tudor Car, Dhakshenya Ardhithy Dhinakaran, Bhone Myint Kyaw, Tobias Kowatsch, Shafiq Joty, Yin-Leng Theng, and Rifat Atun. Conversational agents in health care: scoping review and conceptual analysis. *Journal of medical Internet research*, 22(8):e17158, 2020. doi.org/10.2196/17158.
- [35] Aimee Van Wynsberghe. Sustainable AI: AI for sustainability and the sustainability of AI. *AI and Ethics*, 1(3):213–218, 2021. doi.org/10.1007/s43681-021-00043-6.
- [36] Alexandra Vtyurina, Adam Fourney, Meredith Ringel Morris, Leah Findlater, and Ryen W White. Verse: Bridging screen readers and voice assistants for enhanced eyes-free web search. In *Proceedings of the 21st International ACM*

- SIGACCESS Conference on Computers and Accessibility*, pages 414–426, 2019. doi.org/10.1145/3308561.3353773.
- [37] WHO. Disability - World Health Organization. <https://www.who.int/news-room/fact-sheets/detail/disability-and-health>, 2023. [Accessed 15-09-2024].
- [38] Linda Wulf, Markus Garschall, Julia Himmelsbach, and Manfred Tscheligi. Hands free-care free: elderly people taking advantage of speech-only interaction. In *Proceedings of the 8th Nordic Conference on Human-Computer Interaction: Fun, Fast, Foundational*, pages 203–206, 2014. doi.org/10.1145/2639189.2639251.
- [39] Yu Zhong, TV Raman, Casey Burkhardt, Fadi Biadisy, and Jeffrey P Bigham. JustSpeak: enabling universal voice control on Android. In *Proceedings of the 11th Web for All Conference*, pages 1–4, 2014. doi.org/10.1145/2596695.2596720.

Index

Assistive Technology	3
Blind and Visually Impaired	2
BVI.....	<i>see</i> Blind and Visually Impaired
Core Category	18
Disability	1
Grounded Theory Method.....	17–19
Large Action Model.....	13
Large Language Model	11
LLM.....	<i>see</i> Large Language Model
Screen Reader	3
Theoretical Saturation.....	18
Voice Assistant.....	3
WCAG.....	<i>see</i> Web Content Accessibility Guidelines
Web Content Accessibility Guidelines	8

