Making Lecture Videos Accessible for Students who are Blind or have Low Vision through AI-Assisted Navigation and Visual Question Answering

Katharina Anderer
Karlsruhe Institute of Technology
Karlsruhe, Germany
Karlsruhe University of Applied Sciences
Karlsruhe, Germany
katharina.anderer@kit.edu

Karin Müller
Karlsruhe Institute of Technology (KIT)
Karlsruhe, Germany
karin.e.mueller@kit.edu

Lukas Strobel
Karlsruhe Institute of Technology
Karlsruhe, Germany
lukas.strobel@kit.edu

Matthias Wölfel
University of Applied Sciences Karlsruhe
Karlsruhe, Germany
matthias.woelfel@h-ka.de

Jan Niehues
Karlsruhe Institute of Technology
Karlsruhe, Germany
jan.niehues@kit.edu

Kathrin Gerling
Karlsruhe Institute of Technology
Karlsruhe, Germany
kathrin.gerling@kit.edu

Abstract

Designing accessible lectures and lecture materials is crucial to promote inclusive higher education. We conducted need-finding interviews with 12 students who are blind or have low vision to learn their perspectives on how lectures and lecture material could become more accessible through Artificial Intelligence (AI) technologies. Key insights from the interviews reveal that students envision AI to automatically customize lecture material, connect disparate information sources, for example, to better keep track of the current lecture slide, and enhance interaction and engagement with lecture material. Based on these insights, we developed the LectureAssistant prototype, employing an iterative design process with visually impaired users that features AI-assisted video navigation and chatbot interaction. In a final evaluation with seven students, the participants expressed enthusiasm for features such as AI-powered video search and the possibility of asking questions about visual content in the current video frame. They provided valuable suggestions for future improvements, including notifications for lecture slide transitions and the provision of a short overview function for a slide. Insights from the study indicate great potential of the prototype to improve accessibility of lecture videos for students with visual impairments, although they also point to crucial areas for improvement, such as more reliable and personalized image descriptions.

ASSETS '25, Denver, CO, USA

ACM ISBN 979-8-4007-0676-9/25/10
https://doi.org/10.1145/3663547.3746349

CCS Concepts

• **Human-centered computing** → **Accessibility systems and tools**; **Empirical studies in HCI**; • **Applied computing** → *Interactive learning environments*; • **Computing methodologies** → *Natural language generation*.

Keywords

Assistive technology, blind and low vision, large-language models, vision-language models

ACM Reference Format:
Katharina Anderer, Karin Müller, Lukas Strobel, Matthias Wölfel, Jan Niehues, and Kathrin Gerling. 2025. Making Lecture Videos Accessible for Students who are Blind or have Low Vision through AI-Assisted Navigation and Visual Question Answering. In *The 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25), October 26–29, 2025, Denver, CO, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3663547.3746349

1 Introduction

Digital formats in education, such as online lectures, hybrid formats, or digital material in the form of lecture slides, PDFs, or video uploads, have become indispensable since the coronavirus pandemic in 2020 at the latest. For students who are blind or have low vision, this has brought new challenges in terms of accessibility because there is an increasing need to transform digital material into formats that are accessible through screen readers or braille displays [41]. However, students who are blind or have low vision often lack access to structural support [4, 31], they need to put a lot of effort into obtaining accessible material and more time than the average student to prepare for or comprehend lecture material [6, 18, 32]. Manually preparing material such that it is accessible to students with blindness or low vision usually costs a lot of time and effort, for example, when alternative text descriptions must be crafted for images or charts [40, 43, 56, 57, 62]. Artificial intelligence (AI) has the potential here to make this process more efficient [36], which is, in the case of alternative text descriptions for images, facilitated due to the rapid development of vision language models (VLM) during the last few years. VLMs are now capable of giving detailed descriptions of images or even extracting text from a photo or scan, creating new opportunities for an effortless and efficient transfer of visual information into textual or auditive information. Furthermore, the development of large language models (LLMs) in general has great potential to individualize learning and adapt to individual needs [66, 71, 78], making it possible, for example, to generate summaries, transfer an explanation into easier language or go deeper for a specific topic.

In our need-finding interview study, there was strong agreement among the participants that current practice in universities discriminates students who are blind or have low vision due to a lack of resources and support infrastructure or by social structures and stereotypes. A common example provided by participants was that lecturers often do not understand their needs or do not feel responsible for offering accessible lectures or lecture materials. Although this work does not address these structural and social issues directly, our objective is to develop an accessible AI-assisted lecture video platform for students to help them become more independent of these structural and social barriers.

Although there is a lot of work on how to make lecture material more accessible by developing automatic conversion approaches [11, 36, 81, 84], or other work that discusses the potential of AI to make lectures more accessible [7], there are few studies that approach AI solutions for education in a human-centered approach, integrating potential users as a central part of the development process [20, 59]. However, a human-centered approach is crucial in order to avoid the 'Disability Dongle' [16],which describes the phenomenon in which developers or designers create technological solutions for people with a disability that they think are useful and innovative, without knowing the needs and preferences of end users [16]. In the book Design Justice [23] it is further argued that neglecting the inclusion of marginalized groups in the design process can lead to systems that inherently favor privileged individuals. This bias arises because developers tend to belong to privileged groups, which influences their perspectives and design decisions [23]. Recognizing the importance of user-centered design for accessibility, this study integrates potential users from the beginning with the goal of designing a prototype that makes lectures more accessible to students who are blind or have low vision. We will henceforth refer to this prototype as LectureAssistant. Our research questions are as follows.

- RQ1: How do students who are blind or have low vision envision using AI technology to support their studies?
- RQ2: How do students who are blind or have low vision perceive the LectureAssistant prototype?
  In particular, we wanted to investigate:
  - RQ2-A: How do students assess the usefulness of the prototype?,
  - RQ2-B: How accessible is the prototype for them?, and
  - RQ2-C: What is the perceived ease of use of the prototype among students?

We address RQ1 by conducting semi-structured interviews that include questions about students' challenges in their study life, their suggestions for possible improvements, and where they see potential to use AI-assisted technology for their studies. Based on the interviews, we collected concrete implementation requirements for the LectureAssistant prototype. The second part of this paper focuses on the iterative design process of the prototype with potential users. Finally, we evaluated the prototype to answer RQ2. Therefore, we included open questions about ease of use, accessibility, and usefulness.

The contributions of this paper are the following.

- A thematic synthesis that illustrates students' main challenges with inaccessible lecture material along with their perspectives on AI assistance to mitigate these barriers.
- LectureAssistant, an AI-based prototype that features lecture video navigation through an AI chatbot and visual question answering through a vision-language model.
- Based on the evaluation of LectureAssistant, we provide insight into challenges and opportunities for implementing future accessible lecture assistance tools.

2 Related Work

Assistive technology can improve access to educational material [42]. Depending on the degree of vision loss, blind people or people with low vision use optical or electronic magnifiers, large keyboards or large printed books [74], screen readers that read digital text aloud [9], braille displays or tactile devices [72], or a combination of them. The main barriers to these tools include that many websites or learning resources are not accessible to the screen reader [9] and the high cost of the tools [74]. During the last few years, AI has significantly impacted assistive technology also within the educational sector. Although there have been great advances in intelligent tutoring systems for education before the rise of LLMs [44, 86], LLMs have tremendously accelerated the pace of assistive systems for education [58, 76]. An overview of AI technologies used particularly to support people who are blind or have low vision is given by [80], highlighting the potential of AI to facilitate learning and help foster autonomy and independence. The following subsections discuss different fields of assistive AI technologies that can be used to make lecture material more accessible.

2.1 Automatic Text Conversion Approaches

Multiple approaches have been proposed to convert educational material into other formats that are accessible to students who are blind or have low vision. A semi-automated approach to convert the exams into a more accessible format has been proposed by [84]. They focused on how to make the hierarchy of an exam document more screen-reader friendly, applying Yolov8 [67] to split the document into content blocks and then applying heuristic rules to define the order of the contents [84]. Another approach of [81], called SciA11y, is specifically tailored to convert scientific articles into HTML code that generates hyperlinks to improve navigation for blind users. Their tool integrates several different algorithms to detect bounding boxes for figures and tables or to identify and extract textual elements. Another example of automatic conversion of PDFs into markup language is the transformer-based algorithm, Nougat, proposed by [11]. These approaches focus on conversion, without taking on a human-centered perspective or exploring how interactive engagement with the material is possible.

2.2 Alternative Text Description with VLMs

There was a rapid improvement in the ability of VLMs to generate high-quality image descriptions [34]. VLMs combine a pre-trained LLM with a vision encoder to generate visual descriptions in natural language [34]. Examples of open source models that have evolved during the last years are Blip-2 [46], Flamingo [3], MiniGPT-4 [88], LLaVA [48], Minicpm-v [87], LLaMA 3.2-vision [26], DeepSeek-VL [51] or internVL [21].

Although there is a growing effort to build accessible applications with open-source models, app solutions from big tech companies like SeeingAI [73] from Microsoft or Be My AI from Be My Eyes [28] that work with gpt-4v from OpenAI [2] are still predominant, as they can usually offer more computational resources and thus more accurate image descriptions. However, these apps work through an Application Programming Interface (API) to external servers and data privacy cannot be guaranteed. In addition, the number of user requests is often limited, not free of charge, or both. Another example is TapTapSee [79] that uses the CloudSight Image Recognition API, is free, but explicitly collects user data and shares this with external partners.

2.3 Making Video or Slide Information Accessible

An approach to make video information accessible to people who are blind or have low vision was given in 2021 by [12], who generated video commentary by combining methods such as optical character recognition (OCR) to detect text within a video frame, automatic object detection with Yolov3 [29], as well as convolutional neural network (CNN) architectures to detect objects and bounding boxes. However, their approach does not allow us to interactively engage with the video or ask detailed questions about the content. The rapid development of AI algorithms now provides new state-of-the-art approaches. An example of a real-time object detection algorithm is YOLOv8 [67].

A more recent approach that is more specifically tailored for lectures is DiagramVoice, a tool to generate video commentary for images within a lecture video [27]. Although it possesses some degree of adaptability by generating short or long commentary, it still lacks the capability for question and answering (Q&A) and is therefore limited to providing a broad overview of what is happening in a lecture video. As investigated in depth by [39], it depends on context, scenario, and individual preferences what and how users want video commentary. Therefore, it might be difficult to efficiently provide the information the user is looking for in such a generic approach.

A system introduced by [64] called Slidecho allows users to extract text or image captions from a video or to receive notifications of undescribed elements. However, their tool also lacks the capability for interactive engagement with the video's content. Furthermore, despite evaluating the tool within a user study, the end-users were not directly involved in the design process.

[50] developed heuristics for measuring the accessibility of a video. An example is that accessibility is low when there are visual references without detailed explanations of the visual objects. However, a Q&A function with vision-language capabilities should also allow stopping the video and asking for more details to fill the information gap, which we address in our approach.

3 Part 1: Need-finding Interviews

To explore the challenges of students who are blind or have low vision and their perspectives on prospective AI solutions for accessibility, we conducted need-finding interviews, which are commonly used to understand user needs and problems [63]. The entire work of this paper was approved by the ethics committee of our home institution, as well as by the responsible data protection officer.

3.1 Method

In line with related work that collected requirements for the human-centered development of prototypes [37, 54], we use semi-structured interviews to explore the perspectives of the target group and applied a thematic analysis (TA) as a technique to synthesize these findings. Finally, on the basis of these results, we discuss implications for the development of LectureAssistant.

3.1.1 Interview Questions. We designed an interview guideline to explore the experiences and perspectives of students who are blind or have low vision. Questions about the students' experiences were detailed on their study routines, but also focused on challenges, their current use of technology, and how they envision further integration of AI. The complete list of questions is available in the Appendix 3.

3.1.2 Participants. Ensuring a variety of perspectives, we reached out to prospective participants by contacting representatives for accessibility at various German universities, who forwarded the invitation using mailing lists. To participate, individuals must be over 18 years old, have completed or enrolled in their studies in the last two years, and have a vision acuity equal to or less than 5%. 12 students from six different universities participated in the interviews. Due to Braun and Clark [15], the appropriate number of participants should be based on saturation. Observing that after feedback from several participants, the student's feedback did not introduce a lot of novel insights, we stopped recruiting further participants.

Two of the participants identified as women, one as non-binary, and nine as men. All of them were still enrolled in their studies, except one who participated in a program for several months to get to know university life. The study fields included Physics, Law, Economy, Philosophy, Computer Science, Education, Psychology, and Languages. Many of the interviewees work visually with magnification, whereas others completely or mainly work with audio via screen reader and sometimes in addition with a braille display. Please note that two of the participants did not strictly meet the criterion of having less than 5% visual acuity. One had a severe reduction in the visual field, the other had a vision acuity of 6%. We decided to include them in the study because we considered their experience to be comparable. The characteristics of the participants are detailed in Table 1.

3.1.3 Procedure. Prior to the interviews, participants received study information and data protection information. The interviews lasted between 30 minutes and 1 hour and were compensated with

Table 1: Participant Characteristics.

ID	Gender	Vision Acuity	Working Style	Interview	Evaluation
P1	m	2%	screen reader, braille display	Yes	Yes
P2	m	blind	screen reader, braille display	Yes	Yes
P3	m	≤ 5 % of vision field, 30% of vision acuity	works mostly visually, rarely screen reader	Yes	Yes
P4	m	≤ 2%	works mostly visually, screen reader for larger texts	Yes	No
P5	nb	≤ 2%	screen reader, a bit visually	Yes	Yes
P6	m	≤ 6%	works mostly visually, rarely screen reader	Yes	Yes
P7	m	blind	screen reader, braille display	Yes	No
P8	f	≤ 5%	screen reader and visually	Yes	No
P9	m	≤ 5%	screen reader and visually	Yes	Yes
P10	m	≤ 5%	screen reader and visually	Yes	No
P11	m	≤ 5%	mostly visually working, screen reader for longer texts	Yes	Yes
P12	f	≤ 2%	mostly screen reader, partially with magnification	Yes	No

10 euros. Ten interviews were conducted online through BigBlueButton [10], and two in person.

3.1.4 Data Analysis. We audio recorded the interviews and transcribed them afterwards. To synthesize the findings of the interviews, we applied a reflexive TA, based on the procedure outlined by Braun and Clarke [14, 22]. By ongoing reflection, reflexive analysis allows the researcher to engage with the data in-depth, uncovering nuanced interpretations, thereby providing a more profound understanding of it.

The first author, who also conducted the interviews, transcribed the interviews and crafted initial codes inductively. To ensure quality and a variety of perspectives, we collaboratively grouped the codes and crafted initial themes, which we then revised and reviewed in further discussions.

3.1.5 Positionality Statement. It is important to acknowledge subjectivity and privilege in qualitative research and reflect on how this can influence one's perspective. All authors have no visual impairment or have corrected vision and belong to Western, Educated, Industrialized, Rich, and Democratic (WEIRD) societies. Additionally, in the context of our research, it is important to underline that we, the authors, support the use of AI as one factor towards more accessibility, but we strongly argue for a carefully considered use of AI that is open to everyone.

The first author, who conducted the interviews and led the qualitative analysis, has a background in cognitive science, AI, and computer science. Reflecting on possible biases for the TA, we might have put more weight on participants' comments that matched our vision of a prototype as we already had the implementation of an AI-assisted prototype in mind before conducting the interviews.

3.2 Findings of the Thematic Analysis

We crafted three themes that address RQ1, illustrating the challenges of students during studies and their perspectives on how lectures can become more accessible using AI technology. The quotes were translated into English as all interviews were conducted in German. For the sake of brevity, we will refer to the participants as P1-P12, as detailed in Table 1.

3.2.1 Customizability can make current lecture materials more accessible. Customizability has the potential to better address the diverse needs of students who are blind or have low vision. Specifically, customizing the visual representation of lecture materials, but also the alternative presentation of visual information, can enhance accessibility.

Prospective AI assistants could help users visually adapt lecture material more efficiently, so that it is possible to *"feed the AI with how I want the lecture slides to be prepared. Basically, I need the texts to be bigger or with more contrast"* (P11). Other adaptation needs expressed for visual presentation of text included different font sizes or spacing.

Concerning alternative text descriptions of images or lecture slides, preferences ranged from interpretations to neutral descriptions of images, or with fewer details. Furthermore, there was a desire for more contextualized descriptions. This indicates that the assistive system may need to provide different levels of alternative descriptions to properly engage with the content and support understanding of visual content.

The interviews highlighted different preferences for input and output modalities for an assistance system depending on the context, such as whether it is used during lectures or individual learning. The input modality during lectures should be *"something keyboard-like or something with different short keys that generate different prompts, because you don't want to talk in the lecture"* (P1), or alternatively using touch gestures (P5). As an output modality, some participants prefer to combine audio and braille output simultaneously, or headphones during lectures to get descriptions in a situation where *"You have a graph and the lecturer says take a quick look at it. At that moment, you just put a pair of headphones in your ears or just have a set in parallel and then say [to the AI], describe it to me quickly"*.

These different preferences regarding visual descriptions or input and output modalities underscore the need for customizability within different contexts.

3.2.2 AI can empower interactive learning and engagement. The participants envisioned AI as a tool that supports interaction with the lecture material to make knowledge consolidation more accessible and efficient. AI assistants could provide clarification of inaccessible material, but could also assist in finding specific content.

Various participants had the idea of an interactive AI assistant that can be asked general questions about the current slide or to make inquiries about an image or graph on the slide, for instance, to *"ask the AI program what is currently displayed"* (P9), or to possibly implement multi-modal input, *"AI can then perhaps recognize where I am pointing at, so that it interprets the finger as an arrow and... shows... what text is there"* (P5). This provides an opportunity to compensate for information gaps that result from non-accessible visual information in study material. P9 said that the visual information gap could be compensated for as*"it is precisely this visual aspect, which you do not notice in the slides, that [AI] could take over this part"*. This is even crucial for study success, as P7 described the situation that *"in study groups or mock exams, I realized that I hadn't really grasped a lot of things because many visualizations are used in lectures.*

Furthermore, it was expressed that look-up functions would be useful, for example, to ask something general about the lecture, and the AI assistant would query the entire lecture content. P3 suggested to *"store the spoken word [of the lecturer], so that you can perhaps look for certain words and not have to look through 90 minutes to see when the word... came up"* (P3). Such an interaction with AI has potential for a significant increase in efficiency in engaging with a lecture. Individuals without blindness or low vision can navigate a video or a transcript by visually skimming. Although it benefits all students, a lookup function is more crucial for students who are blind or have low vision. In general, the participants' expressed desire for a Q&A function, especially for visual elements, highlights their need for deeper engagement with lecture material and the opportunity to fill information gaps resulting from non-accessible visual information.

3.2.3 Different sources of information can be synchronized and connected via AI. The interviews revealed a recurring emphasis of the participants on automatically coupled and synchronized lecture content. Participants expressed a desire for more integrated and accessible learning experiences "because this linking of slide and visual content and then in the auditory makes a huge difference" (P9). A common challenge mentioned by participants was to keep track of the current slide, e.g., P8 stating: *"When lecturers make a digression, I am often confused as to which slide we are actually on"*. P6 suggested *"a kind of combination of glasses that simultaneously see what the lecturer is doing in front and simultaneously combine the slides"* so that, for instance, a note could be added on the slide if the lecturer mentions something that is particularly relevant for exams. This parallel representation and synchronization of different content sources would help to address students' concerns not to *"have to lose track"* (P8).

Such connections could be facilitated through automated content summaries and note taking. For example, P7 further argued that a tool that was capable of making notes, but without using the exact word phrases of the lecturer, could help mitigate data protection issues. In addition, *"something that automatically converts blackboard images into text ... and then provides ... a suitable note"* (P7) could further enrich the original content with additional representations.

The participants' suggestions on automated note taking indicate a desire for tools that help them to concentrate more on following along with the lecture by integrating different knowledge sources for them.

Connecting different sources of information could also reduce the information gap that students experience when lecturers do not talk explicitly about visual content, and *"write things on the board and, in the worst case, don't say what they are writing, and perhaps point somewhere and refer to it"* (P11).

4 Part 2: Design and Implementation of LectureAssistant

This section details the basic features of LectureAssistant and its development process. The implementation requirements are based on need-finding interviews, as we describe next. Furthermore, we applied an iterative design that incorporates feedback from two people, one blind and one with low vision, which will be described below.

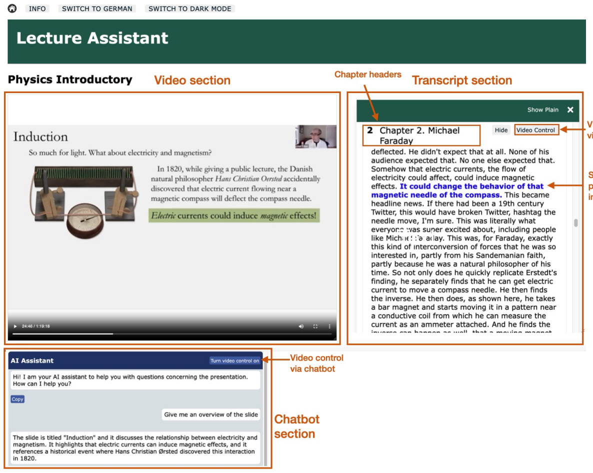
4.1 Requirements based on Need-Finding Interviews

The participants envisioned AI as a means to improve accessibility, having different ideas and visions for this. We decided to develop an AI-assisted prototype for lecture video recordings as videos, due to their inherently visual nature and dense information presentation, pose particular accessibility challenges that we want to address with our prototype. The features of LectureAssistant prototype address the following specific needs and visions that the students articulated in the interviews: To facilitate greater adaptability of lecture materials, we implemented a toggle to switch between dark and light mode, and adaptable font sizes. To account for a more interactive and engaging learning experience, we decided to implement a chatbot function to retrieve information from the lecture based on a large-language model and a chatbot function to answer visual questions about the current video frame with the help of a vision-language model. To address the theme of connecting different sources of information, we implemented a connection and synchronization of the transcript and the lecture video, so that a user can navigate from the transcript to the video and the other way, and we enabled navigation through the chatbot to video sections aligned with the user's question. Finally, we made the prototype more accessible to the screen reader and used contrasts that align with the WCAG 2.0 standards [19].

Ultimately, our prototype is built entirely on open source frameworks. We hope that this contributes to solutions independent of financial constraints and promotes widespread accessibility.

4.2 Implementation of the Basic Features

The architecture of LectureAssistant was built on the open source project of [38], who introduce a video platform called 'Lecture Translator' (LT) that allows students to watch videos in conjunction



Induction

So much for light. What about electricity and magnetism?

In 1820, while giving a public lecture, the Danish natural philosopher *Hans Christian Oersted* accidentally discovered that electric current flowing near a magnetic compass will deflect the compass needle.

Electric currents could induce *magnetic* effects!

2 Chapter 2. Michael Faraday

deflected. He didn't expect that at all. None of his audience expected that. No one else expected that. Somehow that electric currents, the flow of electricity could affect, could induce magnetic effects. **It could change the behavior of that magnetic needle of the compass.** This became headline news. If there had been a 19th century Twitter, this would have broken Twitter, hashtag the needle move, I'm sure. This was literally what everyone was super excited about, including people like Michael Faraday. This was, for Faraday, exactly this kind of interconversion of forces that he was so interested in, partly from his Sandemanian faith, partly because he was a natural philosopher of his time. So not only does he quickly replicate Erstedt's finding, he separately finds that he can get electric current to move a compass needle. He then finds the inverse. He then does, as shown here, he takes a bar magnet and starts moving it in a pattern near a conductive coil from which he can measure the current as an ammeter attached. And he finds the

Hi! I am your AI assistant to help you with questions concerning the presentation. How can I help you?

Give me an overview of the slide

The slide is titled "Induction" and it discusses the relationship between electricity and magnetism. It highlights that electric currents can induce magnetic effects, and it references a historical event where Hans Christian Ørsted discovered this interaction in 1820.

Figure 1: Application Interface: Lecture video player (i) is on the upper left, AI chatbot window (ii) on the lower left and transcript window (iii) on the upper right. The current position of the speaker is highlighted in the transcript.

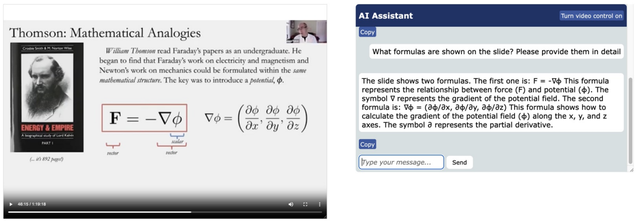
with the transcript of the lecture and to translate the transcript into different languages. The user interface of LectureAssistant contains three main windows with the following functionalities: (i) video player, (ii) AI chatbot, and (iii) transcript. A screenshot of these three windows is shown in Fig. 1. oh super, das

4.2.1 Technical details. Here, we briefly describe the technical framework of LT, which was also incorporated by LectureAssistant. For more detail, the reader can refer to, for example, [38]. LT is a component-based architecture with a central component, called 'mediator' that communicates between the different components. Through interaction with the application interface, the user can send requests to an API component that forwards the request to the mediator. When a video is uploaded, a component for automatic speech recognition gets the audio information from the video and can process the speech. The results are sent back to the mediator and from there other components can retrieve the information for further processing. There are multiple other components such as a chatbot component using the LlamaIndex framework [49], or a logging component that allows to store the user-chatbot interaction.

The entire system is hosted locally on servers of our home institution, such that the data of the users is only stored locally, and aligns with data protection rules of the institute's data protection department. In the following sections, we elaborate on the different features.

4.2.2 Video player. An HTML5 video player allows basic functions like setting the video speed, starting and stopping the video, or adding subtitles. It is possible to use shortcuts to start the video or search forward or backward in steps of 10 seconds.

4.2.3 Chatbot. The chatbot is based on the open source LLM Llama 3.1 [26] and allows the user to ask questions about the lecture transcript. This is made possible by retrieval-augmented generation (RAG), which was conceptually introduced by [45]. The transcript of the lecture is represented as vectors and is stored in a data store.



Thomson: Mathematical Analogies

William Thomson read Faraday's papers as an undergraduate. He began to find that Faraday's work on electricity and magnetism and Newton's work on mechanics could be formulated within the *same mathematical structure*. The key was to introduce a *potential*, ϕ.

$$\boxed{\mathbf{F} = -\nabla\phi} \qquad \nabla\phi = \left(\frac{\partial\phi}{\partial x}, \frac{\partial\phi}{\partial y}, \frac{\partial\phi}{\partial z}\right)$$

What formulas are shown on the slide? Please provide them in detail

The slide shows two formulas. The first one is: F = -∇ϕ This formula represents the relationship between force (F) and potential (ϕ). The symbol ∇ represents the gradient of the potential field. The second formula is: ∇ϕ = (∂ϕ/∂x, ∂ϕ/∂y, ∂ϕ/∂z) This formula shows how to calculate the gradient of the potential field (ϕ) along the x, y, and z axes. The symbol ∂ represents the partial derivative.

Figure 2: The figure displays a use case where a user is asking the AI chatbot about a detailed description of the formulas on the current slide. The chatbot gives back the formula in latex code and gives further context information about the nabla operator used as a tool to find the rate of change of a function.

When a user asks a question, the chatbot compares the question with the vectors in the data store, which represent small sections of the lecture transcript, and finds the most similar sections according to the user's question. These sections are then provided to the Llama 3.1 model along with the user's question. Furthermore, in order to allow for Q&A, the chatbot is enriched with a VLM called Minicpm-v [87]. This model is used when the Llama 3.1 agent determines that visual information is needed to answer a question. An agent in this context refers to the part of the chatbot that decides which tools or models to use to generate the best response. In such cases, a screenshot of the current video frame is sent to the VLM. The response generated by the VLM is then sent back to the agent, which creates the final answer. The final response can include additional context from the lecture, as the agent can use both the RAG and the VLM component. An exemplary use case is shown in Fig. 2 where a user asks the chatbot about the formulas on the slide.

4.2.4 Transcript and Post-Processing. The transcript of the lecture video is further processed by different components. One component automatically generates summaries, based on the algorithm presented by [69] and another generates chapters of the transcript, which are based on the algorithm presented by [68]. The transcript is synchronized with the video, and as soon as the user starts playing the video, the transcript window scrolls to the current speech so that the user can visually follow where the speech is currently. For screen reader users, we have implemented a focus on the current speech, which is discussed in further detail in Section 4.3.2.

The transcript can be displayed in two different modes on the user interface: (i) original transcript, and (ii) markup view with chapters and summaries. The user can toggle between these two modes via a button. The markup view presents the generated chapters and summaries of the video. Screen reader navigation is implemented by buttons with ARIA-labels and marked text regions.

4.2.5 Interaction with the Interface. We have prioritized accessibility for the screen reader by implementing several key features. For example, the system provides notification messages for updates and button state changes, establishes a logical focus order for all HTML elements, and ensures headers are correctly tagged. These measures enable users to receive audio output or even haptic feedback through a braille display. Beyond its main functionalities, the system also offers light- and dark-mode toggling and allows users to control interface text size via their browser's font settings, enhancing readability. Designing a system compatible with the assistive technologies users were already comfortable with seemed to be advantageous to us. The decision to focus on screen reader accessibility and, for instance, to not integrate a feature for voice input, was influenced by the need-finding interviews, as many participants reported using their own specialized speech synthesis software, and we wanted to avoid potential interference. Users can therefore interact with the chatbot through text or braille input. While direct voice input isn't integrated, some participants utilized personal speech-to-text software to interact with the LectureAssistant via voice.

4.3 Iterative Feedback Loop

After implementing the core features, we conducted user feedback sessions with two people to refine the prototype based on human-centered insights. We chose an iterative design protocol to implement and test the prototype with the features that stemmed from the need-finding interviews. In this phase, we had a concrete plan for what we were going to design, so open-ended co-creation was not seen as a suitable paradigm. The first participant, P13, is blind with no light perception and uses a screen reader and a braille display. The second participant, P14, has low vision with a visual acuity of approximately 10% and works visually with magnification software. These sessions aimed to collect feedback from users with diverse visual needs, ensuring screen reader accessibility on the one hand, and better visual readability on the other. We encouraged participants to think aloud during the exploration of the prototype.

4.3.1 First feedback round. Feedback from P13 was mainly related to the navigation of the screen reader. Despite the fact that most elements were already well accessible via screen reader, some screen reader notifications were still missing. We therefore added improved labeling to notify the user, for example, when the video jumped to a different chapter.

P13 also gave suggestions on useful shortcuts for video navigation. We implemented shortcuts in order to start and stop the video or search forward and backward, regardless of whether the video was in focus or not. In addition, it was suggested to implement a shortcut to switch between the three main windows, which we then realized with the shortcut "Ctrl + 1".

Furthermore, P13 expressed the wish for a copy function for the chatbot responses, which we therefore added to the prototype.

During testing the prototype, P13 expressed that allowing video navigation through the transcript chapters would be very beneficial for more efficient video navigation. We integrated this idea by a "Video Control" button to the right of the chapter header, which enables to navigate the video to the beginning of the corresponding chapter.

Feedback from P14 was mainly about simplifying the user interface. The three windows of chatbot, transcript and video, should not necessarily be displayed at the same time, but preferable only two windows appear next to each other for more clarity and simplicity in the interface. The user should be able to choose the displayed windows. In order to meet these criteria, we changed the layout of the user interface by arranging two columns next to each other, such that only two windows are displayed at a time and the third is displayed at the bottom. Switch positions of the windows, to bring a different window into focus, is enabled by dragging one window to the position of another and dropping it there.

Furthermore, P14 suggested that windows should be resizable so that magnification software can be well applied. If the window is, for example, already large and not resizable, some of the text will be out of focus when zooming in.

4.3.2 Second feedback loop. Prior to the final evaluation of the prototype, we made the web application available to P13 and P14 for a second feedback. P13 still identified some minor accessibility issues with the transcript's chapter headers, suggesting to mark them with appropriate HTML header tags. Additionally, P13 noted that the prototype lacked a mechanism to quickly navigate the transcript to the current speaker's position via screen reader. To address this, we implemented a focus mechanism for screen reader users, directing them to the current spoken words. However, the evaluation section will discuss the need for a more nuanced and cautious approach to this feature, as the forced focus interfered with the navigation of the application.

P14 was satisfied with the implemented version and had no further suggestions.

After the second feedback loop, main suggestions made by the two participants were addressed, such that we decided to proceed with the evaluation of the prototype.

5 Part 3: Evaluation of LectureAssistant

This study aims to address RQ2, that is, how LectureAssistant is perceived by students who are blind or have low vision. We were particularly interested in how they think about the prototype with respect to accessibility, usability, and ease of use. We start by describing the evaluation procedure and then move on to the analysis and findings of participants feedback. The evaluation should be seen as an initial assessment of the prototype in its early stage. Thus, we were especially interested in qualitative user feedback and did not incorporate any quantitative methods.

5.1 Participants

We contacted the participants of the need-finding interviews who had agreed to contact them again for evaluation. Seven of these participants (P1, P2, P3, P5, P6, P9, P11) participated in the evaluation (see Table 1). Two of the students use screen reader and braille display only, where the others work to some extent visually and with screen reader. Participants could participate in person or remotely using BigBlueButton [10]. Given our focus on qualitative feedback, we assume, in line with Braun and Clarke [15], that this sample size is appropriate, as we have observed that after some initial interviews, student feedback tends to converge and there are no significant new findings.

5.2 Procedure

For the prototype test, we made the web application available via a password-protected URL to make remote testing possible. A detailed testing protocol was prepared to ensure a standard testing procedure, which included the following phases: (1) pre-study preparation, (2) demographic questions and experience, (3) familiarization, (4) prototype testing, and (5) post-study interview.

5.2.1 *Pre-study preparation.* Participants received study information and data protection information prior to the test. They had the choice to select a lecture video that was used for the application or choose a topic with which they are familiar to avoid unfamiliar content.

5.2.2 *Demographic questions and experience.* First, we asked participants questions about their demographics, their use of assistive tools, and their experience with AI-based systems and video platforms. The full list of questions can be found in the Appendix B.

5.2.3 *Familiarization.* The application was then explained in detail, including the available shortcuts. The participants could explore the application freely for about 10 minutes and ask questions at any time. In addition, we encouraged participants to think aloud and share their thoughts at any time.

5.2.4 *Testing of the prototype.* After familiarization, we asked the participants to perform specific tasks to ensure that important functions were tested. The following tasks were included: (i) navigate the video roughly to the middle and start and stop the video there, (ii) ask the chatbot two or three questions about the current video frame, (iii) activate the 'Video Control' mode of the chatbot and ask the chatbot a question in order to test whether the video jumps to a position that fits to the question, (iv) navigate to the transcript window and navigate the video to an arbitrary chapter through the 'Video Control' of the chapter, and (v) test whether the text of the chapter is accessible.

5.2.5 *Post-study interview.* After the testing, we provided open-ended questions to obtain detailed feedback, to answer RQ2 about how participants perceive LectureAssistant in terms of usefulness, accessibility, and ease of use. The complete list of open feedback questions is shown in Table 2.

5.3 Analysis of Feedback

We conducted a content analysis [60] to systematically analyze user feedback, using the following deductive categories: (C1) Initial General Impressions, (C2) Experienced Barriers, (C3) Comparative Analysis, (C4) Perceived Utility, and (C5) Improvement Suggestions. The deductive categories were derived from open-ended feedback questions to capture user experiences relevant to RQ2. Table 4 shows how the categories relate to the questions.

5.3.1 *Coding process.* The interview transcripts were coded by the first author only due to limited project resources. The coding process and the categories are described in the following.

The first category, **Initial General Impressions (C1)**, acknowledges the significant impact of initial impressions on user perceptions, as highlighted, for example, in [47]. It is directly based on the first open feedback question. This category captures users' immediate reactions and general feelings about the tool's features.

The second category, **Experienced Barriers (C2)**, is about obstacles encountered by participants. Although open feedback questions already specifically include a question about navigational barriers, the other two subcategories Visibility and Perception were derived after going through the data. Navigation barriers refer to barriers in finding specific functions or moving through the interface (e.g., 'It did not work to navigate to the video'). Visibility barriers relate to issues in the visual presentation of the application (e.g., 'The boxes were too small'). Perception barriers refer to barriers in perceiving changes or notifications of the application (e.g., 'I did not get a notification when the answer had been generated').

The third category, **Comparative Analysis (C3)**, collects advantages and disadvantages compared to other video platforms that users already know. The Diffusion of Innovations Theory, introduced by Rogers in 1962, states that the relative advantage of an innovation influences its adoption [85]. Therefore, it is an important factor whether users think that the application offers any advantages compared to the status quo, which influences whether they will adopt the innovation [85].

The fourth category, **Perceived Utility (C4)** intends to help understand users' beliefs about the ability of the prototype to improve their productivity, the envisioned benefits, and the hypothetical

Table 2: Post-Study Interview: Open Feedback Questions for Evaluation of Prototype Study (translated from German) with the corresponding deductive categories for content analysis.

No.	Questions	Derived Category
1.	What are your thoughts on the features of the tool?	C1
2.	Have you experienced any specific barriers while using the tool? If so, which ones?	C2
3.	How easy or difficult did you find the navigation of the application?	C2
4.	Do you find the tool to be beneficial or hindering in terms of accessibility? How so?	C4
5.	How helpful do you find the application's features in relation to your studies?	C4
6.	Do you think the application helps to complete tasks or review faster?	C4
7.	How do you assess the range of functions and what additional functions would you like to see?	C4
8.	Do you think you would use the application more frequently?	C4
9.	Where do you see differences compared to other lecture platforms or video platforms? Are there advantages/disadvantages?	C3
10.	How did you like the image descriptions provided by the assistant?	C5 & C4
11.	How did you perceive the chatbot? How well did you find its responses?	C5 & C4
12.	Do you have any general improvement suggestions?	C5
13.	Do you have any other general comments?	C5

frequency of use. After going through the first three interviews, the categories were revised and questions 4-8, as well as aspects of the answers of questions 10 and 11 were merged into this category, similar to the construct of Perceived Usefulness of the Technology Acceptance Model (TAM), introduced by [24].

Finally, the fifth category, **Improvement Suggestions (C5)**, collects user suggestions for improvement. Consistent with the sub-categories of the Experienced Barriers category, we initially divided into the same sub-categories **Navigation**, **Visibility** and **Perception**, as before. Recognizing the specific importance of the AI chatbot within the prototype, we additionally included a sub-category for **AI Chatbot** specific suggestions, allowing for a focused analysis of feedback related to this key feature. After revising and finalizing of categories and subcategories, the remaining interviews were worked through.

The codes are finally analyzed in the next section.

5.4 Findings

In this section, we present the results of the content analysis of the evaluation.

5.4.1 Initial General Impressions. The participants generally had positive impressions of the application, stating, for example, that *"this is completely new. I think you can make a lot out of it"* (P2) or *"I would have really liked that in Corona times"* (P6). Four participants (P6, P2, P1, P5) highlighted the usefulness of the video control function to search within the video using the AI bot. P1 and P6 expressed to find the shortcuts intuitive and easy to use. Other comments included appreciation for automatic generation of chapters and chapter headings (P2, P1).

5.4.2 Barriers. Participants who work visually primarily mentioned visibility issues, those who use screen readers reported more navigation and perception barriers. Although not directly a barrier of the prototype, we note that there were also concerns about structural barriers such as data protection issues or the lack of lecturer's willingness to upload their videos.

Visibility issues. These included poor contrast (P11, P3), small box sizes (P3), difficulty finding video control elements (P5), and challenges to get an overview of the application (P9).

Navigation barriers. Screen reader users reported that a forced focus on the transcript blocked other functions (P1, P5, P9). Some participants suggested implementing a shortcut to jump to the current position in the transcript and removing forced focus. In addition, screen reader users mentioned unexpected jumps during navigation (P2, P9) and difficulties in navigating the video (P5). Furthermore, the transcript area was difficult to access (P2, P5) and the video slider was not easily usable via screen reader (P2, P9).

Perception barriers. Participants encountered the lack of audio or tactile feedback during AI responses (P2) and the lack of notification for slide switches (P1). LectureAssistant therefore needs further improvement to provide multi-modal feedback for all features.

5.4.3 Comparative Analysis. All participants saw advantages of the application compared to common video platforms for lectures they have used. P6 and P11 stated that it is more clearly arranged. P2 stated that the AI chatbot with the video control function is a great advantage because *"YouTube is a good place to watch a video. But as soon as the video is longer than about 20 minutes, it becomes super difficult to search or navigate through it."*. Regarding YouTube, P1 stated that while it is well accessible, it is usually not chosen for lectures. Other points mentioned were that in contrast to other platforms visual elements are made accessible (P3, P5) because *"on other video platforms, you just don't have access to the visual content at all"* (P5). Furthermore, chapters are automatically generated and allow efficient navigation (P9), and finally, as P1 stated, *"the biggest difference is that it works for blind users."*.

This suggests that the application is generally considered an improvement in comparison to other video platforms for lectures, indicating that participants perceive the application as more accessible and perceive its features as helpful. As Diffusion of Innovations Theory suggests, this is an indicator of whether potential users are motivated to adopt a new innovation or not [85].

5.4.4 Perceived Utility. Although all participants stated that they saw the potential of LectureAssistant to be beneficial, some stated that the testing time was too short to evaluate this well (P11, P5, P9). For instance, P9 stated *"I don't think even a few hours are enough [for] test applications... And then somehow you realize that this seems great, but it is not. That is why I am perhaps a bit more technically critical. I honestly see a lot of potential there."*. Similarly, regarding the question whether the participant thinks to use this tool more frequently, all stated either they would clearly use it (P6, P11, P2, P3) or they would use it under some conditions such as that the AI chat should work more reliably and have an explicit function to reliably retrieve text on the current video frame (P2), or that it depends on whether the tool conforms to data privacy standards and is openly available (P5). P9 formulated the wish to further try the tool out. Most of the participants expressed that the application would be useful for their studies. P1 said *"I think it is really good that there is something like that now. And I would have liked that during my studies. Especially for the transcript and for jumping in the video. That would have been great. And if it worked really well, so that the AI would not use the wrong tool, then I would have used it a lot."* . Similarly P2 stated that *"as long as there exist videos for lectures, I would use this platform for it, I guess"*.

5.4.5 Improvement Suggestions. Four participants (P6, P11, P2, P5) suggested a feature to upload their own videos to be independent of lecturers or to work with openly available videos within the platform. Other feedback is divided into the subcategories AI Chatbot, Navigation, and Perception.

AI Chatbot. The AI chatbot was in general easy to understand (P9, P11) and the participant *"actually felt personally addressed somewhere"* (P9), it was perceived as *"not very casual or too complex* (P11), *"actually pretty good, somehow not artificial, or not extra complicated scientific"* (P9), or *"quite good"* (P3). However, others felt that it was limited in understanding the intent of the user or in language capabilities (P2, P6, P5). During testing, the chatbot occasionally misidentified user intents and called the wrong tool, leading to unexpected answers. To address this issue, P1 suggested introducing short keys to call agent tools more reliably. Furthermore, participants sometimes asked questions that the chatbot could not answer with its available tools, and it was not clear to users why this was the case. Understanding how the chatbot works could help users adapt their prompts and align them with the chatbot's capabilities, as stated by P1. These comments indicate that the chatbot's functionality should be made more transparent to users. Furthermore, it was suggested that the chatbot should have a specific optical character recognition (OCR) function to retrieve only the text of the video frame, without the danger that the LLM paraphrases it (P1). P5 suggested that the video control function of the chatbot should allow for more specific user requests allowing it, for example, to search a topic within the first half of the lecture or towards the end. This would contribute to a more efficient search. P11 and P2 wanted chat history storage to avoid the need to retype questions after browser reloads. Other wishes included chapter summaries by AI chatbot (P11), AI-generated exam study questions (P6), and automatic video frame overviews (P11). Many suggestions aim for a more personalized and efficient learning experience with an effective and reliable chatbot interaction.

Concerning specifically the image descriptions of the chatbot, P2 stated that the image descriptions were just right from the level of detail and just as a person would describe it. However, for many participants, they were not detailed enough and too condensed (P1, P9, P3). P11 and P5 stated that the descriptions should include more context-relevant information. Furthermore, the model sometimes described things that were actually not in the video frame, a phenomenon known as 'hallucination.' P5 and P6 stated that it should work more reliably, especially for handwritten board notes. In order to increase the perceived usefulness of the tool, improving the reliability and context-awareness for image descriptions would be important. In addition, the personalization of the image descriptions could deal with different user preferences.

Navigation. P1 expressed a desire to extend the video control function for the chapters, allowing jumps to specific video positions via text blocks within the chapters. Furthermore, optimizing navigation using arrow keys was suggested (P1). During testing, we focused on screen reader navigation using the 'Tab' key. However, P1 noted that the navigation order differs between 'Tab' and arrow keys. P5 recommended improving the navigation to the last chat message, as the screen reader consistently navigated to the first message instead. This did not happen in all tests, suggesting compatibility issues for different browsers and screen readers. Implementing these suggestions would primarily improve ease of use.

Perception. P1 suggested adding a signal tone when the slide in the video changes, addressing the challenge of keeping track of the current slide. Similarly, P11 recommended including a brief slide overview in parentheses within the transcript when the slide changes. Another suggestion was a tone or tactile feedback as long as the AI is generating its answer (P2). All of these suggestions are derived from a lack of adequate feedback on what is happening visually. This underscores that developers should always provide multimodal feedback to ensure that all users receive the necessary information. Some screen readers did not automatically read the new AI-generated answers, leading P2 to suggest implementing a more robust toast message or using an independent text-to-speech function. However, there was no consensus among participants on this, as a separate text-to-speech function could interfere with the screen reader, potentially causing irritation.

6 Discussion

In this paper, we explore the potential of an AI-assisted lecture tool to increase accessibility of lecture videos for students who are blind or have low vision. Here, we first answer research questions and then reflect on the challenges and opportunities of assistive technology in higher education.

6.1 RQ1: How Do Students Who Are Blind or Have Low Vision Envision Using AI Technology to Support Their Academic Life?

In summary, the participants of our study envisioned the use of AI technology to customize lecture content, interactive engagement with the lecture, as well as synchronization and connection of different information sources.

Customazibility of material. The study highlights the participants' wish for customizability of lecture content, where AI assistance could help to automatically adapt visual, but also alternative representations of lecture material to user needs. Lecture material should be made accessible via different modalities and tools, such as a screen reader, braille display, or magnification software. The participants envisioned that AI could personalize alternative text descriptions of images or automatically adapt lecture material. The personalization of image descriptions is not addressed yet by the current prototype. User preferences differ, which also depends on the context in which an image is provided as worked on by [52, 77] has analyzed. Users could potentially control both the level of detail and semantics through the application interface options or by specifying their general image description preferences to an AI chatbot during onboarding.

Interactive learning and engagement. AI can help create an interactive learning experience that allows the user to pose questions at any time and engage in discussions about the material. The participants envisioned using an AI chatbot to query the lecture or get information about visual material. As the evaluation of LectureAssistant made apparent, the participants wished for an even higher degree of interactivity, as currently implemented. This could be achieved with additional modes, such as a testing feature in which the AI generates exam questions, as suggested by a participant. Interactivity can increase the student's motivation as shown, for instance, by [8]. Therefore, all students would benefit from increased interactivity. In particular, for students who are blind or have low vision, it helps to mitigate visual information gaps.

Synchronizing and connecting different information sources. Furthermore, participants in the study envisioned using artificial intelligence to connect different sources of information, which could, for example, help them keep track of where the lecturer is pointing on the slide or when the slide is changing. They also saw potential to use AI for automated note taking, which could be used to connect the lecture slide's content with the lecturer's comments on it. The connection of information sources of LectureAssistant could still be extended as highlighted in the open feedback. For instance, one participant suggested using an auditory cue for slide changes, while another proposed incorporating annotations within the transcript to indicate changes in the video's visual content. A similar feature was implemented by [64], as discussed in the related work section. Integrating such a functionality into LectureAssistant appears beneficial to participants, related to Perception Barriers reported in Section 5.4.2.

6.2 RQ2: How do Participants Perceive the LectureAssistant Prototype?

In general, the prototype was perceived by participants as useful, accessible, and easy to use. However, they noted that issues such as lack of reliability and navigation barriers limited perceived utility and accessibility to some extent. We subsequently discuss RQ2 in more detail with regard to usefulness, accessibility, and ease of use.

Usefulness. Regarding the perceived usefulness of LectureAssistant, all participants reported that they found it beneficial or rather beneficial, with the limitation that for some testing time felt too short to evaluate the prototype in depth. During the evaluation of LectureAssistant, it became apparent that the current state of the prototype has limitations regarding reliability, context awareness, and level of detail for image descriptions. From the participants' perspective, a crucial factor for a regular use of the prototype would be the ability to reliably extract text from video frames and provide accurate qualitative image descriptions. We concur that achieving this reliability is a significant challenge, particularly for smaller open-source VLMs prone to hallucinations. Although research aims to address this issue [35, 82], hallucinations remain an open challenge. If users cannot independently verify the output of VLMs, their use should be carefully considered, especially within academic contexts.

Accessibility. Although participants generally perceived the prototype as accessible, there were still some barriers that participants reported. These included visibility issues for video control elements or buttons and navigation barriers, such as difficult navigation to the video or the transcript. Accessibility standards such as WCAG 2.0 [19] are dynamically evolving, and new technologies or applications can introduce novel barriers, as described by [13]. Although this study provides initial information, a larger and more diverse sample would be needed to comprehensively assess the general accessibility of the application. Furthermore, initial training for LectureAssistant could significantly enhance its accessibility. This recommendation emerged from the feedback of a participant and aligns with [19]'s observation that a lack of training for assistive technology often hinders accessibility.

Ease of use. Regarding the ease of use of LectureAssistant, most of the participants found it intuitive, particularly highlighting the helpful shortcuts and the clear layout of the user interface. However, several improvements should be made, specifically in relation to seamless screen reader navigation. In addition to human-centered development, further research could also look at the principle of affordance [61], which states that the functionality of interactive elements should be self-evident to the user, minimizing the need for explicit instruction and increasing ease of use.

6.3 Reflection on Challenges and Opportunities of AI Technology to Increase Accessibility of Higher Education

This section reflects on the potential and challenges associated with the introduction of artificial intelligence technology in higher education addressing students who are blind or have low vision. We consider the factors that influence the successful application of AI technology. Furthermore, we discuss the importance of embedding AI technology in a wider effort to provide more inclusive lectures. Finally, we reflect on the social implications that AI technology may have in this context.

6.3.1 Sociotechnical factors influence the success of AI.. AI technology offers significant potential for automating the creation of accessible lecture material and to enable interactive engagement and efficient querying of lecture material using chatbots. Participants in our study were generally optimistic that AI technology can increase accessibility to lecture material and offer helpful assistance (see Section 3.2). However, while AI was seen as a promising tool for accessible learning, its practical impact was perceived as contingent on content availability. Here, participants raised concerns about the willingness of the lecturers to upload digital material (see Section 5.4.2). To address this, numerous participants emphasized that the value of LectureAssistant would be greatly improved by allowing users to upload their own video content, increasing user agency, and broadening the applicability of the system. Finally, we note that previous work has highlighted that AI literacy and technology skills are a key determinant of successful integration of AI technology [30, 70]. Thus, both students and lecturers need to be aware of potential issues and scope of a specific chatbot.

6.3.2 AI technology is not a replacement for inclusive lectures. There is a risk that AI technology is seen as an opportunity to delegate efforts to make lectures accessible. We want to note that the development of assistive AI technology should not be seen as a replacement for the creation of inclusive lectures, but should go hand in hand with general efforts to speak in an inclusive way, echoing previous findings on the use of assistive technology in classroom settings [17, 33, 55]. For example, to facilitate equitable access of students who are blind or have low vision to live lectures - which are a unique experience - lecturers should be trained in steps they can take to be more inclusive, for example by proactively offering verbal descriptions for images and explicitly reading formulas written on the board [65, 83]. An approach to automatically generate feedback for lecturers on how well they describe visual elements is presented by [65]. Having assistive AI technology should therefore not provide an excuse to not reflect on the diverse needs of students, and we encourage critical reflection on the fact that the heavy reliance of higher education on visual communication is inherently exclusionary [25]. Therefore, assistive AI technology should only be seen as a tool to mitigate barriers, but not as the only component to achieving inclusive higher education.

6.3.3 Social implications of AI technology in education. As a society, we should also carefully consider how assistive AI technology affects the isolation of students who are blind or have low vision. In particular, increased interaction with technology could reduce interpersonal interactions and participation in traditional lectures, and classroom use of such systems can be seen as stigmatizing. Here, work by [75] that addresses *social accessibility* of assistive technology argues that it should be 'built into mainstream technology' to increase acceptance and lower risk of social stigma. As LectureAssistant focuses on self-study before or after lectures for now, the risk of technology interfering in live interactions is not as high. However, the tool could be extended for real-time lecture assistance, a feature that was also of interest to some participants in our study. Here, future work must explore how to achieve suitable integration into classrooms and lecture halls.

7 Limitations and Future Work

There are a few limitations that we want to acknowledge. Although the prototype was developed and tested with VoiceOver [5] and NVDA [1] with Google Chrome as a browser, it was not optimized for the full spectrum of screen readers and browsers. We made a conscious decision to allow participants to use their own setups, including preferred browsers and laptops, to facilitate realistic remote testing and to reach more participants. Therefore, performance variations based on browser and screen reader type could have affected user study outcomes. Likewise, men were overrepresented in our sample, which needs to be addressed in future studies to include more women and non-binary people. We acknowledge that a self-selection bias cannot be excluded, such as some reason why male felt more motivated or comfortable to participate in the study. Overall, for stronger conclusions about usability and technology acceptance of our prototype, a longer testing phase would be required, ideally over a whole semester, where users can get used to LectureAssistant and use the tool with their own study materials. Furthermore, a larger sample could provide more diverse information on different challenges and needs with respect to different types or degrees of visual disability, various study subjects, or learning preferences. A larger sample would also give the opportunity to incorporate quantitative metrics about, for example, usability or technology acceptance [24].

Another potential limitation might arise from the repeated participation of participants. We invited participants from the interviews to the final study to assess how well our prototype addressed their visions, aligning with our iterative design approach. We cannot rule out the possibility that they were biased, for instance, by liking the prototype more because it originated from their ideas.

Finally, although AI assistance in the form of an LLM and VLM-based chatbot can help derive alternative text descriptions, it only offers a linear content representation, which may lack efficiency. Visual representations usually convey lots of information in parallel. To address this, research should not neglect to investigate and search for alternative representations such as haptic representations for graphs [53, 57]. This is also an interesting avenue for future extensions of LectureAssistant.

8 Conclusion

This paper presented a prototype design process centered on the needs of students who are blind or have low vision for an AI-assisted lecture video platform. Our initial need-finding interviews revealed that participants perceived significant potential in using AI technology to improve accessibility in education, offering concrete suggestions by students for its implementation. Subsequently, during the testing and evaluation of our prototype, participants expressed overall positive feedback, finding it easy to navigate and with a clear layout. In particular, the AI chatbot function for the in-video search yielded the most enthusiasm. Future development iterations should prioritize improving the quality and adaptability of AI-generated image descriptions and ensuring compatibility with diverse screen readers to maximize the usability and accessibility of such a system.

Acknowledgments

We acknowledge the use of ChatGPT and Gemini for rephrasing sentences and making grammatical corrections. Katharina Anderer is supported by the Ph.D. scholarship in the Cooperative Doctoral Program Accessibility through AI-based Assistive Technology (kooperatives Promotionskolleg Barrierefreiheit durch KI-Basierte Assistive Technologien (KATE)) provided by the Ministry of Science, Research and Arts Baden- Württemberg (MWK).

References

[1] NV Access. 2025. NVDA: The free software that makes that possible. Retrieved April 16, 2025 from https://www.nvaccess.org

[2] Josh Achiam, Steven Adler, Sandhini Agarwal, Lama Ahmad, Ilge Akkaya, Florencia Leoni Aleman, Diogo Almeida, Janko Altenschmidt, Sam Altman, Shyamal Anadkat, et al. 2023. Gpt-4 technical report. arXiv:2303.08774

[3] Jean-Baptiste Alayrac, Jeff Donahue, Pauline Luc, Antoine Miech, Iain Barr, Yana Hasson, Karel Lenc, Arthur Mensch, Katherine Millican, Malcolm Reynolds, Roman Ring, Eliza Rutherford, Serkan Cabi, Tengda Han, Zhitao Gong, Sina Samangooei, Marianne Monteiro, Jacob L Menick, Sebastian Borgeaud, Andy Brock, Aida Nematzadeh, Sahand Sharifzadeh, Mikoł aj Bińkowski, Ricardo Barreira, Oriol Vinyals, Andrew Zisserman, and Karén Simonyan. 2022. Flamingo: a Visual Language Model for Few-Shot Learning. In *Advances in Neural Information Processing Systems*, S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh (Eds.), Vol. 35. Curran Associates, Inc., 23716–23736. https://proceedings.neurips.cc/paper_files/paper/2022/file/960a172bc7fbf0177ccccbb411a7d800-Paper-Conference.pdf

[4] Aizan Sofia Amin, Norulhuda Sarnon, Noremy Md Akhir, SM Zakaria, and RNFRZ Badri. 2021. Main challenges of students with visual impairment at higher education institutions. *International Journal of Academic Research in Business and Social Sciences* 10, 1 (2021), 734–747.

[5] Apple. 2025. VoiceOver: Getting Started Guide. Retrieved April 16, 2025 from https://support.apple.com/en-gb/guide/voiceover-guide/welcome/web

[6] Helen Armstrong and Iain Murray. 2007. Remote and local delivery of cisco education for the vision-impaired. In *Proceedings of the 12th Annual SIGCSE Conference on Innovation and Technology in Computer Science Education* (Dundee, Scotland) *(ITiCSE '07)*. Association for Computing Machinery, New York, NY, USA, 78–81. https://doi.org/10.1145/1268784.1268809

[7] Pier Felice Balestrucci, Elisa Di Nuovo, Manuela Sanguinetti, Luca Anselma, Cristian Bernareggi, and Alessandro Mazzei. 2024. An Educational Dialogue System for Visually Impaired People. *IEEE Access* 12 (2024), 150502–150519. https://doi.org/10.1109/ACCESS.2024.3479883

[8] Esra Barut Tugtekin and Ozcan Ozgur Dursun. 2022. Effect of animated and interactive video variations on learners' motivation in distance Education. *Education and Information Technologies* 27, 3 (2022), 3247–3276. https://doi.org/10.1007/s10639-021-10735-5

[9] Vijesh J. Bhute, Ellen Player, and Deesha Chadha. 2023. Motivation and Evidence for Screen Reader Accessible Website as an Effective and Inclusive Delivery Method for Course Content in Higher Education. In *2023 ASEE Annual Conference & Exposition*. ASEE Conferences, Baltimore , Maryland. https://peer.asee.org/43941.

[10] BigBlueButton. [n. d.]. BigBlueButton: Virtual Classroom Software. Retrieved April 16, 2025 from https://bigbluebutton.org

[11] Lukas Blecher, Guillem Cucurull, Thomas Scialom, and Robert Stojnic. 2023. Nougat: Neural optical understanding for academic documents. arXiv:2308.13418

[12] Aditya Bodi, Pooyan Fazli, Shasta Ihorn, Yue-Ting Siu, Andrew T Scott, Lothar Narins, Yash Kant, Abhishek Das, and Ilmi Yoon. 2021. Automated Video Description for Blind and Low Vision Users. In *Extended Abstracts of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI EA*

'21). Association for Computing Machinery, New York, NY, USA, Article 230, 7 pages. https://doi.org/10.1145/3411763.3451810

[13] Fernando HF Botelho. 2021. Accessibility to digital technology: Virtual barriers, real opportunities. *Assistive Technology* 33, sup1 (2021), 27–34. https://doi.org/10.1080/10400435.2021.1945705

[14] Virginia Braun and Victoria Clarke. 2019. Reflecting on reflexive thematic analysis. *Qualitative research in sport, exercise and health* 11, 4 (2019), 589–597.

[15] Virginia Braun and Victoria Clarke. 2021. To saturate or not to saturate? Questioning data saturation as a useful concept for thematic analysis and sample-size rationales. *Qualitative research in sport, exercise and health* 13, 2 (2021), 201–216. https://doi.org/10.1080/2159676X.2019.1704846

[16] Scott Andrew Brown. 2023. From Assistive to Adaptive: Can We Bring a Strengths-Based Approach to Designing Disability Technology? In *Cultural Robotics: Social Robots and Their Emergent Cultural Ecologies*. Springer, 101–108.

[17] Theeraphong Bualar. 2018. Barriers to inclusive higher education in Thailand: voices of blind students. *Asia Pacific Education Review* 19, 4 (2018), 469–477.

[18] Matthew Butler, Leona Holloway, Kim Marriott, and Cagatay Goncu. 2017. Understanding the graphical challenges faced by vision-impaired students in Australian universities. *Higher Education Research & Development* 36, 1 (2017), 59–72.

[19] Ben Caldwell, Michael Cooper, Loretta Guarino Reid, Gregg Vanderheiden, Wendy Chisholm, John Slatin, and Jason White. 2008. Web content accessibility guidelines (WCAG) 2.0. *WWW Consortium (W3C)* 290, 1-34 (2008), 5–12.

[20] Zuzana Ceresnova, Lea Rollova, and Danica Koncekova. 2017. A HUMAN-CENTERED APPROACH IN AN EDUCATIONAL ENVIRONMENT. *Journal of Education Research* 11, 2 (2017).

[21] Zhe Chen, Jiannan Wu, Wenhai Wang, Weijie Su, Guo Chen, Sen Xing, Muyan Zhong, Qinglong Zhang, Xizhou Zhu, Lewei Lu, et al. 2024. Internvl: Scaling up vision foundation models and aligning for generic visual-linguistic tasks. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 24185–24198.

[22] Victoria Clarke and Virginia Braun. 2017. Thematic analysis. *The journal of positive psychology* 12, 3 (2017), 297–298.

[23] Sasha Costanza-Chock. 2020. *Design justice: Community-led practices to build the worlds we need*. The MIT Press.

[24] Fred D Davis et al. 1989. Technology acceptance model: TAM. *Al-Suqri, MN, Al-Aufi, AS: Information Seeking Behavior and Technology Adoption* 205, 219 (1989), 5.

[25] Michele dos Santos Soares, Cássio Andrade Furukawa, Maria Istela Cagnin, and Débora Maria Barroso Paiva. 2024. Accessibility Barriers for Blind Students in Teaching-learning Systems. *Journal of Universal Computer Science* 30, 10 (2024), 1342. https://doi.org/10.3897/jucs.106239

[26] Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Amy Yang, Angela Fan, et al. 2024. The llama 3 herd of models. arXiv:2407.21783

[27] Nahyun Eun and Jongwoo Lee. 2024. DiagramVoice: Automatic Lecture Video Commentator for Visually Impaired Students Supporting Diagram Commentary. In *International Congress on Information and Communication Technology*. Springer, 381–391.

[28] By My Eyes. [n. d.]. By My Eyes. Retrieved April 16, 2025 from https://www.bemyeyes.com

[29] Ali Farhadi and Joseph Redmon. 2018. Yolov3: An incremental improvement. In *Computer vision and pattern recognition*, Vol. 1804. Springer Berlin/Heidelberg, Germany, 1–6.

[30] José María Fernández-Batanero, Marta Montenegro-Rueda, José Fernández-Cerero, and Inmaculada García-Martínez. 2022. Assistive technology for the inclusion of students with disabilities: a systematic review. *Educational technology research and development* 70, 5 (2022), 1911–1930. https://doi.org/10.1007/s11423-022-10127-7

[31] Tahsin Firat. 2021. Experiences of students with visual impairments in higher education: barriers and facilitators. *British Journal of Special Education* 48, 3 (2021), 301–322.

[32] Helen Frank, Mike McLinden, and Graeme Douglas. 2014. Investigating the learning experiences of student physiotherapists with visual impairments: An exploratory study. *British Journal of Visual Impairment* 32, 3 (2014), 223–235.

[33] André P Freire, Flávia Linhalis, Sandro L Bianchini, Renata PM Fortes, and Maria da Graça C Pimentel. 2010. Revealing the whiteboard to blind students: An inclusive approach to provide mediation in synchronous e-learning activities. *Computers & Education* 54, 4 (2010), 866–876.

[34] Yunhao Ge, Xiaohui Zeng, Jacob Samuel Huffman, Tsung-Yi Lin, Ming-Yu Liu, and Yin Cui. 2024. Visual fact checker: enabling high-fidelity detailed caption generation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 14033–14042.

[35] Tianrui Guan, Fuxiao Liu, Xiyang Wu, Ruiqi Xian, Zongxia Li, Xiaoyu Liu, Xijun Wang, Lichang Chen, Furong Huang, Yaser Yacoob, Dinesh Manocha, and Tianyi Zhou. 2024. HallusionBench: An Advanced Diagnostic Suite for Entangled Language Hallucination and Visual Illusion in Large Vision-Language Models. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*. 14375–14385.

[36] Michaela Hanousková, Boris Janča, Lukáš Másilko, Karin Müller, Svatoslav Ondra, Radek Pavlicek, Petr Penaz, Andrea Petz, Thorsten Schwarz, and Rainer Stiefelhagen. 2024. STS New Methods for Creating Accessible Material in Higher Education: Introduction to the Special Thematic Session. In *International Conference on Computers Helping People with Special Needs*. Springer, 285–290.

[37] Raquel Hervás, Virginia Francisco, Gonzalo Méndez, and Susana Bautista. 2019. A user-centred methodology for the development of computer-based assistive technologies for individuals with autism. In *Human-Computer Interaction–INTERACT 2019: 17th IFIP TC 13 International Conference, Paphos, Cyprus, September 2–6, 2019, Proceedings, Part I 17*. Springer, 85–106.

[38] Christian Huber, Tu Anh Dinh, Carlos Mullov, Ngoc-Quan Pham, Thai Binh Nguyen, Fabian Retkowski, Stefan Constantin, Enes Ugan, Danni Liu, Zhaolin Li, Sai Koneru, Jan Niehues, and Alexander Waibel. 2023. End-to-End Evaluation for Low-Latency Simultaneous Speech Translation. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations. Ed.: Yansong Feng, Els Lefever*. Association for Computational Linguistics (ACL), 12–20. https://doi.org/10.18653/v1/2023.emnlp-demo.2

[39] Lucy Jiang, Crescentia Jung, Mahika Phutane, Abigale Stangl, and Shiri Azenkot. 2024. "It's Kind of Context Dependent": Understanding Blind and Low Vision People's Video Accessibility Preferences Across Viewing Scenarios. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–20.

[40] Crescentia Jung, Shubham Mehta, Atharva Kulkarni, Yuhang Zhao, and Yea-Seul Kim. 2021. Communicating visualizations without visuals: Investigation of visualization alternative text for people with visual impairments. *IEEE transactions on visualization and computer graphics* 28, 1 (2021), 1095–1105.

[41] Mohamed Koutheair Khribi. 2022. Toward accessible online learning for visually impaired and blind students. *Nafath* 6, 19 (2022).

[42] Sarah E Kisanga and Dalton H Kisanga. 2022. The role of assistive technology devices in fostering the participation and learning of students with visual impairment in higher education institutions in Tanzania. *Disability and Rehabilitation: Assistive Technology* 17, 7 (2022), 791–800.

[43] S. E. Krufka and K. E. Barner and. 2006. A user study on tactile graphic generation methods. *Behaviour & Information Technology* 25, 4 (2006), 297–311.

[44] Annabel Latham. 2022. Conversational intelligent tutoring systems: The state of the art. *Women in Computational Intelligence: Key Advances and Perspectives on Emerging Topics* (2022), 77–101.

[45] Patrick Lewis, Ethan Perez, Aleksandra Piktus, Fabio Petroni, Vladimir Karpukhin, Naman Goyal, Heinrich Küttler, Mike Lewis, Wen tau Yih, Tim Rocktäschel, Sebastian Riedel, and Douwe Kiela. 2021. Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks. arXiv:2005.11401

[46] Junnan Li, Dongxu Li, Silvio Savarese, and Steven Hoi. 2023. Blip-2: Bootstrapping language-image pre-training with frozen image encoders and large language models. In *International conference on machine learning*. PMLR, 19730–19742.

[47] Gitte Lindgaard, Gary Fernandes, Cathy Dudek, and Judith Brown. 2006. Attention web designers: You have 50 milliseconds to make a good first impression! *Behaviour & information technology* 25, 2 (2006), 115–126.

[48] Haotian Liu, Chunyuan Li, Yuheng Li, and Yong Jae Lee. 2024. Improved baselines with visual instruction tuning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*. 26296–26306. https://doi.org/10.48550/arXiv.2310.03744

[49] Jerry Liu. 2022. LlamaIndex. Retrieved April 16, 2025 from https://github.com/jerryjliu/llama_index

[50] Xingyu Liu, Patrick Carrington, Xiang'Anthony' Chen, and Amy Pavel. 2021. What makes videos accessible to blind and visually impaired people?. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–14.

[51] Haoyu Lu, Wen Liu, Bo Zhang, Bingxuan Wang, Kai Dong, Bo Liu, Jingxiang Sun, Tongzheng Ren, Zhuoshu Li, Hao Yang, et al. 2024. Deepseek-vl: towards real-world vision-language understanding. *arXiv preprint arXiv:2403.05525* (2024).

[52] Alan Lundgard and Arvind Satyanarayan. 2021. Accessible visualization via natural language descriptions: A four-level model of semantic content. *IEEE transactions on visualization and computer graphics* 28, 1 (2021), 1073–1083.

[53] Michał Maćkowski, Piotr Brzoza, Mateusz Kawulok, Rafał Meisel, and Dominik Spinczyk. 2023. Multimodal presentation of interactive audio-tactile graphics supporting the perception of visual information by blind people. *ACM Transactions on Multimedia Computing, Communications and Applications* 19, 5s (2023), 1–22.

[54] Conor McGinn, Michael F Cullinan, Mark Culleton, and Kevin Kelly. 2018. A human-oriented framework for developing assistive service robots. *Disability and rehabilitation: assistive technology* 13, 3 (2018), 293–304.

[55] Aoife McNicholl, Hannah Casey, Deirdre Desmond, and Pamela Gallagher. 2021. The impact of assistive technology use for students with disabilities in higher education: a systematic review. *Disability and rehabilitation: assistive Technology* 16, 2 (2021), 130–143. https://doi.org/10.1080/17483107.2019.1642395

[56] Omar Moured, Morris Baumgarten-Egemole, Karin Müller, Alina Roitberg, Thorsten Schwarz, and Rainer Stiefelhagen. 2024. Chart4blind: An intelligent interface for chart accessibility conversion. In *Proceedings of the 29th International Conference on Intelligent User Interfaces*. 504–514.

[57] Mukhriddin Mukhiddinov and Soon-Young Kim. 2021. A Systematic Literature Review on the Automatic Creation of Tactile Graphics for the Blind and Visually Impaired. *Processes* 9, 10 (2021).

[58] Sankalan Pal Chowdhury, Vilém Zouhar, and Mrinmaya Sachan. 2024. AutoTutor meets Large Language Models: A Language Model Tutor with Rich Pedagogy and Guardrails. In *Proceedings of the Eleventh ACM Conference on Learning @ Scale* (Atlanta, GA, USA) *(L@S '24)*. Association for Computing Machinery, New York, NY, USA, 5–15. https://doi.org/10.1145/3657604.3662041

[59] Sethuraman Panchanathan and Troy McDaniel. 2015. Person-centered accessible technologies and computing solutions through interdisciplinary and integrated perspectives from disability research. *Universal access in the Information Society* 14 (2015), 415–426. https://doi.org/10.1007/s10209-014-0369-9

[60] Jatin Pandey. 2019. Deductive approach to content analysis. In *Qualitative techniques for workplace data analysis*. IGI Global, 145–169. https://doi.org/10.1111/j.1365-2648.2007.04569.x

[61] Hyungjoo Park and Hae-Deok Song. 2015. Make e-learning effortless! Impact of a redesigned user interface on usability through the application of an affordance design approach. *Journal of Educational Technology & Society* 18, 3 (2015), 185–196.

[62] Jinseok Park and Sunggye Hong. 2023. Creating tactile graphics in school settings: A survey of training experience, competence, challenges, and future support needs. *British Journal of Visual Impairment* 41, 4 (2023), 864–875. https://doi.org/10.1177/02646196221109080

[63] Dev Patnaik and Robert Becker. 1999. Needfinding: the why and how of uncovering people's needs. *Design Management Journal (Former Series)* 10, 2 (1999), 37–43.

[64] Yi-Hao Peng, Jeffrey P Bigham, and Amy Pavel. 2021. Slidecho: Flexible non-visual exploration of presentation videos. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–12.

[65] Yi-Hao Peng, JiWoong Jang, Jeffrey P Bigham, and Amy Pavel. 2021. Say it all: Feedback for improving non-visual presentation accessibility. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*. 1–12.

[66] Nitin Rane, Saurabh Choudhary, and Jayesh Rane. 2023. Education 4.0 and 5.0: Integrating artificial intelligence (AI) for personalized and adaptive learning. *Journal of Artificial Intelligence and Robotics* 1 (Jan. 2023), 29–43. https://doi.org/10.61577/jaiar.2024.100006

[67] Dillon Reis, Jordan Kupec, Jacqueline Hong, and Ahmad Daoudi. 2023. Real-time flying object detection with YOLOv8. *arXiv preprint arXiv:2305.09972* (2023).

[68] Fabian Retkowski and Alexander Waibel. 2024. From Text Segmentation to Smart Chaptering: A Novel Benchmark for Structuring Video Transcriptions. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics, EACL 2024 - Volume 1: Long Papers*. Association for Computational Linguistics, St. Julian's, Malta, 406–419. https://aclanthology.org/2024.eacl-long.30

[69] Fabian Retkowski and Alexander Waibel. 2025. Zero-Shot Strategies for Length-Controllable Summarization. In *Findings of the Association for Computational Linguistics: NAACL 2025*, Luis Chiruzzo, Alan Ritter, and Lu Wang (Eds.). Association for Computational Linguistics, Albuquerque, New Mexico, 551–572. https://doi.org/10.18653/v1/2025.findings-naacl.34

[70] Carlie R Rhoads, Arielle M Silverman, and L Penny Rosenblum. 2022. Voices from Academia Providing Education to Students with Visual Impairments During the Pandemic. *Assistive Technology Outcomes and Benefits AT Services During & After the COVID-19 Pandemic* (2022).

[71] Lena Ivannova Ruiz-Rojas, Luis Salvador-Ullauri, and Patricia Acosta-Vargas. 2024. Collaborative working and critical thinking: Adoption of generative artificial intelligence tools in higher education. *Sustainability* 16, 13 (2024), 5367.

[72] Alexander Russomanno, Sile O'Modhrain, R Brent Gillespie, and Matthew WM Rodger. 2015. Refreshing refreshable braille displays. *IEEE transactions on haptics* 8, 3 (2015), 287–297.

[73] SeeingAI. [n. d.]. Seeing AI: Talking Camera for the Blind. Retrieved April 16, 2025 from https://www.seeingai.com

[74] Suraj Singh Senjam, Allen Foster, and Covadonga Bascaran. 2021. Barriers to using assistive technology among students with visual disability in schools for the blind in Delhi, India. *Disability and Rehabilitation: Assistive Technology* 16, 7 (2021), 802–806.

[75] Kristen Shinohara and Jacob O Wobbrock. 2011. In the shadow of misperception: assistive technology use and social interactions. In *Proceedings of the SIGCHI conference on human factors in computing systems*. 705–714. https://doi.org/10.1145/1978942.1979044

[76] John Stamper, Ruiwei Xiao, and Xinying Hou. 2024. Enhancing llm-based feedback: Insights from intelligent tutoring systems and the learning sciences. In *International Conference on Artificial Intelligence in Education*. Springer, 32–43.

[77] Abigale Stangl, Meredith Ringel Morris, and Danna Gurari. 2020. " 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*. 1–13.

[78] Jiahong Su and Weipeng Yang. 2023. Unlocking the power of ChatGPT: A framework for applying generative AI in education. *ECNU Review of Education* 6, 3 (2023), 355–366.

[79] TapTapSee. [n. d.]. TapTapSee: Assistive Technology for the Blind and Visually Impaired. Retrieved April 16, 2025 from https://taptapseeapp.com

[80] Aikaterini Tsouktakou, Angelos Hamouroudis, and Anastasia Chorti. 2024. The use of artificial intelligence in the education of people with visual impairment. *World Journal of Advanced Engineering Technology and Sciences* 13 (10 2024), 734–744. https://doi.org/10.30574/wjaets.2024.13.1.0481

[81] Lucy Lu Wang, Isabel Cachola, Jonathan Bragg, Evie Yu-Yen Cheng, Chelsea Haupt, Matt Latzke, Bailey Kuehl, Madeleine N van Zuylen, Linda Wagner, and Daniel Weld. 2021. Scia11y: Converting scientific papers to accessible html. In *Proceedings of the 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. 1–4.

[82] Zhecan Wang, Garrett Bingham, Adams Wei Yu, Quoc V Le, Thang Luong, and Golnaz Ghiasi. 2024. Haloquest: A visual hallucination dataset for advancing multimodal reasoning. In *European Conference on Computer Vision*. Springer, 288–304.

[83] Christof Wecker. 2012. Slide presentations as speech suppressors: When and why learners miss oral information. *Computers & Education* 59, 2 (2012), 260–273. https://doi.org/10.1016/j.compedu.2012.01.013

[84] David Wilkening, Omar Moured, Thorsten Schwarz, Karin Müller, and Rainer Stiefelhagen. 2024. ACCSAMS: Automatic Conversion of Exam Documents to Accessible Learning Material for Blind and Visually Impaired. In *International Conference on Computers Helping People with Special Needs*. Springer, 322–330.

[85] Cornelia Wolf. 2022. Diffusion of Innovations: von Everett M. Rogers (1962). In *Schlüsselwerke: Theorien (in) der Kommunikationswissenschaft*. Springer, 151–170.

[86] Matthias Wölfel. 2021. Towards the automatic generation of pedagogical conversational agents from lecture slides. In *International Conference on Multimedia Technology and Enhanced Learning*. Springer, 216–229.

[87] Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin Zhao, Zhihui He, et al. 2024. Minicpm-v: A gpt-4v level mllm on your phone. arXiv:2408.01800

[88] Deyao Zhu, Jun Chen, Xiaoqian Shen, Xiang Li, and Mohamed Elhoseiny. 2023. Minigpt-4: Enhancing vision-language understanding with advanced large language models. (2023). arXiv:2304.10592

A Semi-Structured Interview Questions

Table 3: Semi-Structured Interview Questions (Translated from German)

No.	Questions
1.	Could you please tell me what your current field of study is and how long you have been studying?
2.	What is your visual acuity?
3.	Can you please describe your study life? *Sub-questions*: Do you usually attend lectures in person, online, or a combination? What study formats do you usually have? What types of study material do you use? Are there any assistive tools that you regularly use?
4.	Do you face any challenges in your study life?
5.	What, in your perspective, should be done to make your study life more inclusive and accessible?
6.	Are alternative text descriptions for images usually provided to you? *Sub-questions:* If they are provided, how would you generally rate their quality and usefulness?
7.	Do you use large language models like ChatGPT for studying? *Sub-questions:* If yes, how helpful do you perceive the use of it? Have you encountered any accessibility barriers using it?
8.	Do lecturers usually present or teach in an inclusive way? How accessible do you find the lecture material that is provided to you?
9.	Considering an AI-assisted tool designed to support you during lectures, do you believe such a tool could be beneficial? In what specific contexts or aspects of lectures do you think AI assistance could be helpful? Do you have any concrete ideas or features you would wish for regarding an AI-assisted tool?
10.	We plan to implement a prototype based on this interview study and then evaluate it. Would you be open to participate in the evaluation study again? Could we contact you a second time for that?

B Survey Questions for Evaluation

Table 4: Pre-Survey of Prototype Evaluation (Translated from German)

No.	Questions
1.	What is your age?
2.	What is the gender you feel you belong to?
3.	What are you studying and for how long?
4.	What is your visual acuity?
5.	Are there video platforms for lectures that you use or are familiar with? If yes, which one? What is your typical workflow?
6.	Do you use large language models like ChatGPT? If yes, how often? For what?
7.	Do you work with a screen reader, visual aids, or both? If a screen reader, which one do you use and with which browser? If visual aids or both, do you use magnification tools? Which ones?