Q&A or Document-Based? The Effects of Interface Type on How Screen Reader Users Access Interconnected Documents

Colleen F. Cipriano
Department of Computer Science
University of Victoria
Victoria, British Columbia, Canada
colleencipriano@uvic.ca

Yichun Zhao
Department of Computer Science
University of Victoria
Victoria, British Columbia, Canada
yichunzhao@uvic.ca

Miguel A. Nacenta
Department of Computer Science
University of Victoria
Victoria, British Columbia, Canada
nacenta@uvic.ca

Kotaro Hara
School of Computing and Information Systems
Singapore Management University
Singapore, Singapore
kotarohara@smu.edu.sg

Jaylee Soh
Lee Kong Chian School of Business
Singapore Management University
Singapore, Singapore
jaylee.soh.2022@business.smu.edu.sg

Abstract

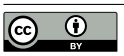
Blind and low-vision (BLV) users are increasingly engaging with large language model (LLM) interfaces to access documents, but it is unclear how such systems support or hinder their ability to build interconnected knowledge. To examine this gap, we compared a Question–Answer Interface (QAI) that supports open-ended conversational inquiry, with a Document Interface (DI) based mostly on traditional structured text document navigation. We recruited 16 BLV screen reader users where they used both interfaces to explore two fictional worlds. Data from interaction logs, concept maps, decision-based tasks, and semi-structured interviews provide comparative insights into how interface design supports knowledge construction. Findings show that participants visited more distinct documents with the DI and formed larger and more correct mental models with the DI than with the QAI. They were also more able to apply the knowledge they had gained. Simultaneously, many still preferred the QAI and often estimated that they had explored more, formed better mental models and applied their models better when acquiring the information with the QAI, despite this not being the case. Our analysis suggests possible interface design reasons for these differences and highlights some of the risks introduced by using question-answer interfaces to access information spaces.

CCS Concepts

• **Human-centered computing** → **Accessibility**; **Empirical studies in accessibility**; *Accessibility technologies.*

Keywords

Document Accessibility, Mental Models, Conversational User Interface, Exploratory Search, Information Seeking, Blind People, Screen-Readers



ASSETS '26, Vila Nova de Gaia, Portugal

ACM ISBN 979-8-4007-2521-0/2026/10
https://doi.org/10.1145/3797867.3829036

ACM Reference Format:
Colleen F. Cipriano, Yichun Zhao, Miguel A. Nacenta, Kotaro Hara, and Jaylee Soh. 2026. Q&A or Document-Based? The Effects of Interface Type on How Screen Reader Users Access Interconnected Documents. In *The 28th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '26), October 25–28, 2026, Vila Nova de Gaia, Portugal.* ACM, New York, NY, USA, 17 pages. https://doi.org/10.1145/3797867.3829036

1 Introduction

The conversational approach to browsing information enabled by generative AI (GenAI) has transformed how people can explore document collections. This emergent tool could benefit those who use screen readers (SRs), most often those coming from the Blind and Low Vision (BLV) community. Traditionally, blind users have relied on mainstream browsing technology and assistive technologies like SRs to access information [9, 69, 90]. Despite extensive efforts to establish accessibility guidelines and develop automated accessibility checkers, many documents remain inaccessible [53]. Even technically accessible documents can pose challenges when visual information lacks proper context or description, leaving SR users unaware of content on the page [17]. Some expect that GenAI tools could eliminate these barriers [35]; poor alt text becomes less problematic when AI can describe images directly, and one could find information quickly just by prompting, without having to search through lengthy reports.

However, current AI-powered tools commonly used by BLV users (*e.g.,* ChatGPT [5, 10]) often provide targeted answers to pointed questions. This approach risks creating biased access patterns, systematically directing users toward certain types of content while obscuring others [5, 101]. This, in turn, could limit users to partial understanding of complex information in a given collection of documents [110]. That is, we risk creating new accessibility barriers, much like how missing accessibility tags or image descriptions currently hide information for BLV users [33, 43], the design of information access technology could require additional effort into achieving comprehensive document understanding [5]. Better understanding of how GenAI tools influence BLV people's information exploration and consumption behavior is crucial for the future design of accessible technologies [42, 110].

The overarching goal of this paper is to compare traditional document browsing tools with LLM-driven conversational access tools and understand how SR users explore documents, comprehend the information they contain, and assess which aspects of each tool influence preference. We call the former a *document-based interface* (DI) and the latter a *question and answering interface* (QAI). We pose three research questions around these interface types.

(RQ1) *How does the difference in interface type affect the patterns by which SR users browse documents?*

(RQ2) *How does the difference in interface type influence how SR users integrate and apply knowledge from documents?*

(RQ3) *What aspects of the interface influence SR users' preference over the other?*

To answer these questions, we conducted a study with 16 SR users to examine how they interact with a DI and a QAI when exploring multimodal documents about complex, information spaces. We created two fictional worlds, *Solana* and *Dominion*, with corresponding document sets. Our mixed-methods study comprised of four phases. In Phase 1, we logged participants' document browsing patterns with both interfaces. In Phase 2, we asked participants to engage in concept mapping activities to examine how they construct mental models using different interfaces. In Phase 3, participants applied their mental models by answering a decision-based scenario within the worlds. Lastly, in Phase 4, we conducted interviews to record their overall experience and preferences.

Our results showed that participants covered a wider range of documents with the DI, whereas the QAI supported more broad connections transitions within a smaller set of documents. Participants produced larger and broader concept maps with fewer errors with the DI, although the QAI concept maps were more densely linked. Interestingly, most participants did not detect these differences, as some preferred DI's explicit structure and QAI for its conversational style. The subjective perceptions of the two interfaces were divided, suggesting divergence in exploration behaviors, strategies, and experiences.

This study primarily contributes: (i) an empirical examination of how SR users browse and synthesize multimodal information using a DI and a QAI; and (ii) insights on how these information access interface types shape knowledge construction and application. Given how long SR users have depended on mainstream browsing tools, we seek to advance understanding of how GenAI can influence and change the future of accessible information access.

2 Background and Related Work

We first introduce the concept of mental model, which we extensively leverage in this paper to conceptualize how people comprehend documents and to operationalize our measurement of this comprehension. Then we review document-based interfaces and conversational user interfaces, and the efforts in making these interfaces accessible for SR users. Finally, we introduce how GenAI technologies have enabled the conversational user interfaces.

2.1 Mental Models and Mental Model Elicitation

We define mental model (based on [40, 82, 103]) as *an internal representation that allows a person to organize knowledge gained about the world and act in consequence to this knowledge.* Mental models are commonly used to examine cognition in applied and theoretical areas (*e.g.,* [52, 99]). A mental model offers a lens through which we can examine cognitive activities that are otherwise hard to observe. For example, learners construct new mental models "when confronted with new learning tasks" [22, 78]. Methods for extracting these models include concept mapping and card sorting, each appropriate for different scenarios [48]. Prior work on blind users' mental models has established that non-visual interaction is shaped by the structure of the internal representation of a system, but has mostly examined desktop environments and screen readers through verbal protocols, observations, and performance data, rather than externalized representations [2, 63]. Moreover, work in this area emphasizes the difficulty of eliciting mental models directly [95].

In this paper, we use concept mapping as a way to assess the cognitive process of SR users as they engage with a corpus of documents and develop comprehension. Concept mapping is a way to externalize how people organize, connect, and build comprehension by representing concepts as nodes and their relationships as links [16, 41, 96]. Adopting on earlier critiques that verbal protocol alone is incomplete [2], concept mapping allowed us to examine construction of internal representations, beyond just measuring retrieval from memory.

2.2 Accessing Documents and Information Access Technologies

In traditional information retrieval literature, users' tasks in accessing a corpus of documents are broadly divided into search and browsing [6, 13, 29, 97]. In search, the goal is to find relevant information or a list of documents based on a search query [59, 92]. In its most basic form, search prompts a user to enter a query, and the information access technology (*e.g.,* a search engine) returns the results [66]. While search has evolved to accommodate more natural language queries to handle greater ambiguity [75], and support iterative refinement conversationally with the use of *conversational user interfaces* (CUIs) [31, 46, 56, 59, 97], the fundamental interaction type of "query-and-get-results" paradigm remained essentially unchanged until recently. Browsing represents a more exploratory approach to information access, which is associated with learning and investigative sense-making [71]—activities that are more aligned with the focus of our work. Users typically lack a specific query in browsing; instead, they may seek to understand general concepts within documents or gain an overview of available information before conducting more targeted searches [11, 71]. Browsing occurs through *document-based interaction.* When navigating the Web, for instance, sighted people use browsers through direct manipulation, while SR users additionally rely on assistive technologies [9, 69].

The emergence of GenAI-based tools follows a broader shift in information retrieval toward treating search as a text-to-text transformation problem [50, 74, 92, 105, 108, 118]. Compared with traditional CUIs, these systems can interpret nuanced input, generate human-like responses, and—crucially for document access—–synthesize text across multiple documents and maintain conversational context [7, 76]. Furthermore, tools like NotebookLM[1] and

[1] https://notebooklm.google.com

Elicit[2] enable information exploration within document corpora through iterative question-answering, and prior work shows that people do ask open-ended questions to navigate large collections without explicit queries [5, 10]. However, we still know little about what happens when an LLM becomes the intermediate layer between users and documents, especially in accessibility-centered exploration of interlinked documents.

This gap matters because the same properties that make GenAI tools appealing may also reshape access in limiting ways. Beyond general concerns like GenAI inducing overreliance [14, 88], and difficulty and cognitive burden associated with prompt-based interaction [30, 65, 107, 111], the dominant "prompt-and-get-summary" interaction encourages targeted information extraction and may reduce exposure to a wider range of documents [12, 39, 61, 93]. Prior work has linked this to weaker exploration, comparison, and continuity of access [12, 37, 114]. Using the stratified model of information retrieval as an analytical lens, we expect that effort is made toward prompt formulation and summary interpretation at the affective layer [97], while reducing direct interaction with the document structure at the interface layer.

If we unknowingly design GenAI-powered exploration tools that could bias information access, we risk creating new barriers for SR users; just as lack of appropriate document tags limits access to information, the design of information access technology may obscure parts of a collection and require additional effort into achieving comprehensive understanding. Therefore, in this work, we conduct a study comparing two interfaces—document-based interface and CUI-based question and answering interface powered by GenAI—that manifest different types of interaction.

2.3 Document Accessibility

Our goal is to understand the effects of interaction types on how SR users explore and comprehend information in documents. This requires us to isolate the accessibility of documents and interfaces by making them as accessible as possible.

The development of accessible document-based interfaces and study of information access using these tools has been a significant focus for the HCI and accessibility research communities. For example, past work has explored ways to present and extend alt-text or image descriptions to enrich screen reader-friendly interaction with static content [27, 54, 113, 123, 125, 126]. Tactile graphics offer another means of access: raised-line representations of visual materials that can be explored through touch [44, 77]. For tactile graphics there are guidelines that can aid in their creation [23, 36, 115]. In our study, we created the document-based interface by learning from past interface designs and following existing guidelines. To make our document-based interface accessible, we provided sufficient annotations for screen-reading access and used a tactile overlay on a touchscreen device for graphical access.

Early conversational assistants like Alexa and Siri opened the door to voice-based, hands-free information access for SR users. These assistants were appreciated for quick and simple tasks, which lowered barriers to accessing digital content [1, 84, 91]. Today's LLM-based interfaces can further facilitate access to documents [90]. Rather than navigating documents line-by-line, SR users can query LLMs for descriptions, summaries, or clarifications, enabling them to extract and synthesize information more efficiently. This shift in information access allows SR users to direct their navigation path instead of adhering to rigid document structures [5]. At the same time, LLMs may also introduce problems by limiting access to a subset of content, making it difficult to ensure complete rather than fragmented understanding [12, 39, 61, 93].

[2]https://elicit.com

3 Study Methodology

To address the RQs from the Introduction, we designed an empirical study to observe SR users navigating unfamiliar information spaces, building mental models of those spaces, and applying the gained knowledge. The study took place over two laboratory sessions in which SR users worked with two interfaces, one based on traditional hyperlinked documents (the Document Interface—DI), and one based on a question-answer paradigm powered by an LLM—QAI). The laboratory setting allowed us to gather detailed data and control our observations. Participants experienced both interfaces (a within-subjects design), allowing explicit comparison.

3.1 Study Materials: Fictional Worlds

An important concern in this kind of investigation is the large potential effect of familiarity on people's behavior and understanding. For this reason, we created two information spaces from scratch, guaranteeing no familiarity. These take the form of *interconnected documents*, similar to Wikipedia's hyperlinked pages but describing fictional worlds. The document sets followed the following criteria: (i) minimize the influence of participants' prior knowledge, (ii) encompass sufficient diversity of information and of document types to reproduce real-life complexity, and (iii) provide sufficient depth for extended exploration. We iteratively composed the *Solana* and *Dominion* worlds in a research approach related to Kieras et al.'s [58], informed by existing methodologies in fictional world building [38].

Each document set included 25 documents, including an index. Out of the 25 documents of each world, 8 (Solana) and 6 (Dominion) were diagrams representing spatial or conceptual aspects of the fictional worlds (*e.g.*, a map of the territory and a diagram of family group configurations). The textual representations resemble wikipedia articles, with headings and meaningful inter-document links. Each document describes an aspect of the world, such as history, governance, and culture.

We included spatial representations because they are common in real documents and are recognized as important to the BLV community [34, 123]. To make spatial documents accessible to our participants, we designed 3D-printed tactile overlays following best practices [23, 45], considering BANA guidelines [18] and current research on tactile representations [23, 45]. We iterated over the designs three times to make them representative of the best accessible spatial representations currently available to the community [49]. Although interactive printable tactile overlays are not currently widespread, they are not very different from printed materials that many in the BLV community are familiar with (e.g., embossed tactile prints) and may soon become practical and affordable [15]. Excluding spatial information from the DI would have artificially constrained the DI from presenting spatial structure in general. The

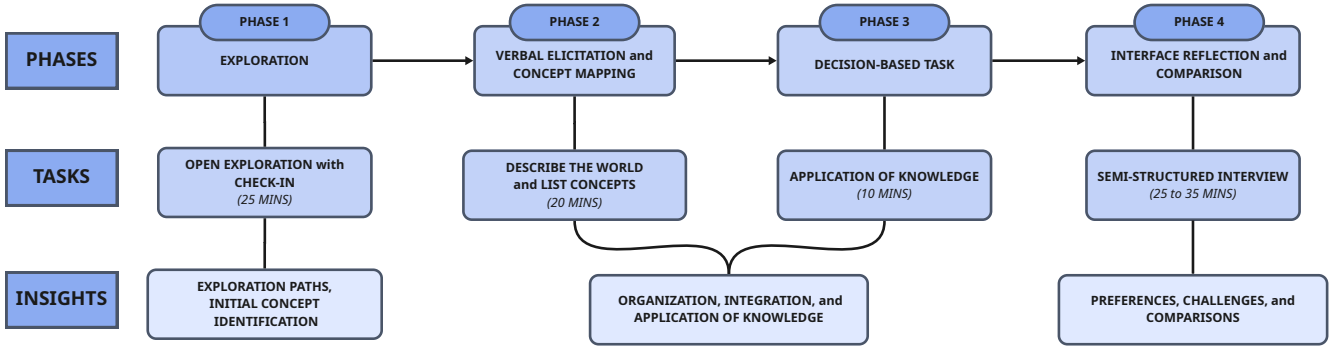


Figure 1: Overview of the study procedure and measurements. Participants completed all phases for each of the DI and QAI conditions in a counterbalanced order. At the end of the second session in Phase 4, participants completed an additional set of interview questions comparing the interfaces.

Figure 2: Experimental setup for the DI condition showing the tactile overlay on top of the tablet (left) on a non-slip mat, with the laptop showing the Dominion interface (right).

implications of this choice are discussed in the Limitations Section below. The documents that form the information spaces of both worlds, including the visual diagrams and their 3D-printed versions are open sourced for use in subsequent research and shared in the Supplementary Materials.

3.2 Interfaces (Main Condition)

The two interfaces (DI and QAI, shown in Figures 3 and 5, provided access to the fictional worlds with the interaction paradigms that we want to compare: direct corpus exploration and LLM-mediated question answering. We designed the conditions to isolate these access paradigms so that the observed differences could be attributed to the interaction approach. We strove to implement state-of-the-art interfaces that are the best reasonable alternatives available or soon to be available.

3.2.1 Document Interface (DI). To minimize the need to train participants and maximize ecological validity, participants used their own device and screen reader (SR) to access textual documents. These were web pages meeting the Web Content Accessibility Guidelines—WCAG 2.1 [116] (see Figure 3). The DI supported index and in-page search with the participants' own SRs. The spatial documents were accessed as the 3D-printed paper diagrams discussed above, overlaid over a tablet (Samsung Galaxy Tab A9+— see Figures 4 and 2). Moving the finger over the overlay produced verbal descriptions of manually labeled areas, in addition to the static relief of the 3D-printed overlay. When a participant followed a hyperlink in a textual document connected to a spatial document, they received a verbal notification to switch to the tablet. If the spatial document had changed, the experimenter manually placed the corresponding document's overlay. The spatial documents also linked to text documents: when tapped, the browser on the main computer jumped to the corresponding document.

The tactile representations preserve spatial relationships that are difficult to communicate through SR navigation alone. Their inclusion reflects our goal of implementing DI as a realistic form of document access. We considered several alternatives for spatial document display such as dynamic pin displays, but decided against because of their low resolution and reliance on Braille, which is not universal among BLV users. Although printable tactile overlays are not currently widespread, they are not very different from printed materials that many in the BLV community are already familiar with (e.g., embossed tactile prints) and may soon become practical and affordable [15].

3.2.2 Question-Answer Interface (QAI). The QAI was a WCAG 2.1-compliant web-based [116] chatbot accessible through participants' devices and screen readers. Participants could enter questions by typing or using speech input. The system returned information in multiple formats: text, structured bullet points, tables, detailed image descriptions when relevant, and direct quotations from the source documents (Figure 5). The interface was designed to retrieve and summarize information only from the existing knowledge base.

The underlying conversational system was built using Python with Flask, integrated with Google's Gemini API (Gemini 2.5 Pro). To constrain conversations to information in the documents we used a custom prompt with cache-augmented generation (CAG) [26], loading the corpus of 25 documents to the context window (including the spatial documents). CAG is well-suited to closed knowledge bases because it reduces retrieval latency and errors. When direct

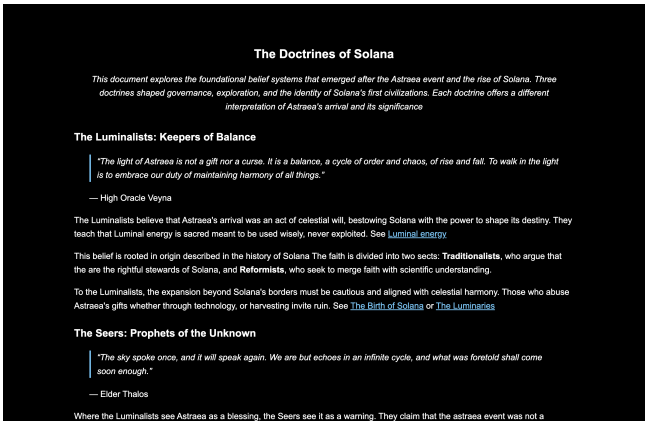


Figure 3: An example of one of the documents in the DI interface for the world of Solana.

Figure 4: Examples of tactile overlays used in the study: a map (left), a diagram of relationships (center), and a set of icons (right).

answers were unavailable, the QAI suggested related documents to stay within content boundaries. For out-of-scope questions (*e.g.*, "Do you have Oreo cookies in Solana?"), it explicitly stated that the information was unavailable, offering relevant alternatives such as "The archives do not mention that topic. If you're interested, there is detailed information on Solana's economy or its annual festivals." We did not observe any instances of hallucination or erroneous responses.

3.3 Participants

We recruited 16 participants (age ≥ 19) who reported regular use of at least one screen reader (*e.g.*, JAWS, NVDA, VoiceOver, TalkBack). Recruitment took place through community organizations, social media, and snowball sampling. Participants reported varied GenAI uses (Table 1), including accessibility features ($n = 11$), work or productivity ($n = 8$), technical assistance ($n = 7$), writing or content generation ($n = 5$), and information seeking ($n = 4$). Participants received $40 per session. The study was approved by the local ethics board.

3.4 Procedure

Each participant completed two 90-minute sessions on separate days at a location of their choice (*e.g.*, public library, participant home, community center). We counterbalanced interface order (DI vs. QAI) and world order (Solana vs. Dominion) across participants. Given the study duration (Figure 1), sessions were separated to reduce cognitive load, fatigue, and carryover effects, consistent with prior cognitive work [51, 57]. To preserve comparative judgments,

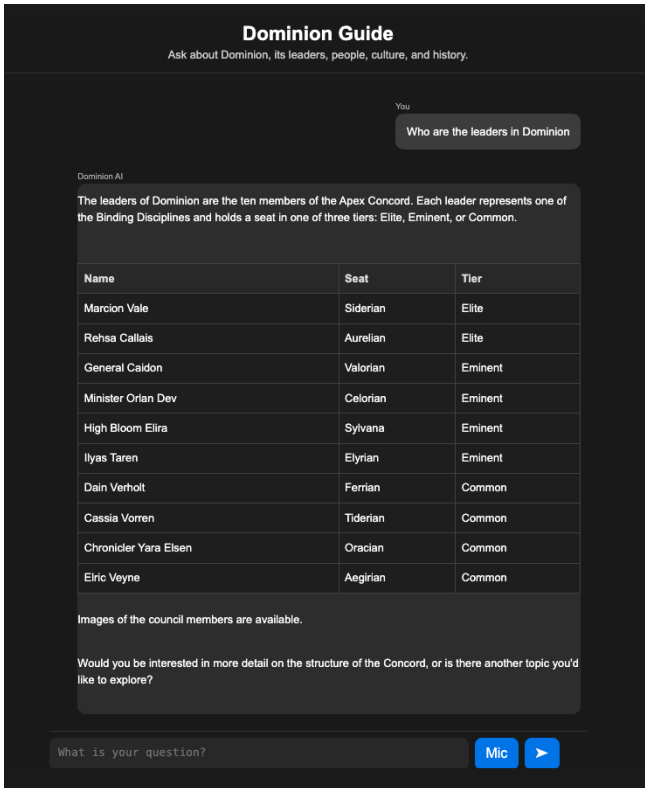


Figure 5: The QAI interface for the world of Dominion showing a user query and a structured tabular response.

we capped the interval between sessions at five days [25]. In practice, participants completed the second session one to three days later, depending on scheduling and location constraints. This also reflects best practices for working with BLV individuals by supporting flexible scheduling, agency in choosing comfortable locations, and use of their preferred device and screen readers. Each session proceeded with four phases in Figure 1.

Phase 1. Participants had to follow their interests and openly explore the world using the assigned interface, while developing an overall understanding of the world. We explicitly told them that recall was not the goal and asked them to prioritize acquiring domain knowledge over constrained task performance [71, 98, 104]. Before starting their QAI session, participants were informed that the system has an underlying content of 25 documents. Participants could take notes, but none chose to.

Phase 2. Participants had to describe the world as if to a friend, highlighting what they found most important. From their description, they listed important concepts and explained how concepts relate to each other. Using these responses, the researcher constructed a concept map and asked probing questions to address gaps or overlooked relationships. The map was then validated with the participant and refined as needed to ensure that it accurately reflected their understanding. The researcher followed a structured script and added or connected only concepts and relationships introduced by the participant in order to prevent bias in the analysis.

Table 1: Self-reported demographics of the study participants.

ID	Age	Visual ability	Occupation	GenAI tools used	GenAI usage
1	40–44	Totally blind	Program coordinator	ChatGPT, Copilot, Be My AI	Daily
2	19–24	Totally blind	Voiceover artist	ChatGPT, Copilot, Gemini, Be My AI	Daily
3	50–54	Low vision	Unemployed	ChatGPT, Be My AI	Daily
4	55–59	Totally blind	Administrative work	ChatGPT, Gemini, Be My AI	2–3 times a week
5	55–59	Totally blind	Retired	Be My AI	Daily
6	25–29	Totally blind	Unemployed	ChatGPT, Be My AI	Daily
7	70–74	Totally blind	Instructor	Be My AI	Daily
8	65–69	Totally blind	Bookkeeper	None	N/A
9	55–59	Low vision	Retired	None	N/A
10	19–24	Totally blind	Accessibility tester	ChatGPT, Copilot, Gemini, Be My AI	Daily
11	30–34	Low vision	Marketer	None	N/A
12	50–54	Totally blind	Trades	ChatGPT, Gemini, Be My AI	Daily
13	65–69	Totally blind	Administrative work	None	N/A
14	55–59	Totally blind	International relations	Copilot	Daily
15	55–59	Totally blind	Client scout	Gemini, Copilot	Daily
16	45–49	Totally blind	Administrative work	ChatGPT, Gemini, Be My AI	Daily

Participants did not have access to the assigned interface because doing so would have shifted the evaluation from their mental models to assisted performance. This restricted interface access means that the experimental design aligns with our formulation of RQ2. This phase implements a mental model elicitation technique similar to interview-based concept mapping method [70]. The assisted concept mapping exercise, in which the experimenter built a graph of concepts and connections is designed to capture mental model organization beyond recall measures [16, 41].

Phase 3. Participants applied their understanding by answering a question from a world-specific scenario (*e.g.,* "As Solana expands, kinship traditions are becoming difficult to maintain. The Luminaries are debating whether to modernize certain laws or preserve Kinship laws strictly. You've been asked to propose a policy on this") which tested how effectively they could use their gained knowledge. Decision-based scenarios complement the model elicitation procedure of Phase 2 by examining applied reasoning [24, 86].

Phase 4. Participants completed a semi-structured interview about their interactions. At the end of each session, we asked about ease of use, challenges, and overall experience. After the second session, participants additionally answered comparison questions, indicating overall preference, perceived support for exploration and learning, and confidence in answering, accompanied by free-form explanations.

3.5 Measurements and Analysis

The two sessions yielded quantitative and qualitative data to address our research questions. Figure 1 indicates which measurements correspond to which phases of the experiment.

3.5.1 Document Traversal Data. We logged the visited documents and transitions across the information space in Phase 1 of each session. We analyzed and plotted how participants moved within each interface. The quantitative measures extracted from this phase are:

- *Unique Document Visits:* A count of distinct documents counted as visited at least once.
- *Document Visits:* The total number of visits, including revisits.
- *Unique Document Transitions:* A count of distinct document-to-documenttransitions traversed at least once.
- *Document Transitions:* The sum of all transitions made between documents.
- *Transitions per Document:* The transitions normalized by the number of unique document visited.

Because the interfaces support different actions, visits and transitions, we necessarily had to operationalize their counts differently across interfaces maintaining equivalence as much possible. In DI, a *visit* occurs upon opening a document. In QAI, we manually mapped each response to the documents where its information appeared. We counted a source document as visited when the response included information corresponding to five or more contiguous sentences from that document; shorter references were treated as summaries and were not counted as visits. We chose this threshold to distinguish substantial exposure to a document's content from a brief mention or summary. It also reflects the DI structure, where each document is preceded by a short descriptive paragraph before the main content. The index document also contains a summary of the document, which therefore allows extracting information without actually visiting it. A *transition* represented movement between source documents. In DI, transitions are link traversals. In QAI, transitions occurred when the source-document mapping changed within a response or across follow-up responses. Although QAI transitions were mediated by the system's retrieval, they remained user-driven as they resulted from follow-up and compound questions. Self-loops were excluded for comparability, where rereading the currently selected document in DI does not generate a new transition.

With the visits data, we constructed graphs that describe the navigation pattern of each session. Topics are nodes and transitions are edges. From the graphs, which we plotted for visual qualitative analysis, we additionally extracted the following quantitative measurements (all common graph measurement described in [21, 83, 117, 120]:

- *Density:* Measures of overall interconnectedness. Higher values correspond to more tightly connected graphs.
- *Directed Diameter:* Measures the breadth of the graph. Lower values correspond to more tightly connected graphs.
- *Degree of Variance:* Measures the evenness of connectivity. Higher values correspond to more unevenly connected graphs.
- *Average shortest path length:* Measures the average number of steps needed to move between concepts. Lower values correspond to more tightly connected graphs.

3.5.2 Concept Maps. In Phase 2, we documented each topic and subtopic mentioned by the participant, and the explicit connections they made to form the concept maps. We collected this data as graphs, which we visualized and analyzed to examine the structure of gained knowledge and mental models following Paul [89] and Haque et al. [47].

Topics and Subtopics: A *topic* corresponds to overarching categories in the worlds represented in a document (*e.g.*, governance, geography (map), or the community structure). *Subtopics* refer to subsections within documents or facts that fit under a topic such as names, roles, dates, or landmarks. We counted a topic/subtopic when the participant made an unambiguous mention that could be mapped to the predefined documents. We did not count repeated mentions of the same topic/subtopic as new instances, only newly stated connections. References to information beyond the content were marked as inaccurate by the researcher, ensuring completeness since they could still link to new topics or subtopics.

Number of Connections: We counted the explicit connections from the topics/subtopics mentioned by participants and any additional connections were verified by the researcher through probing. We also calculated the number of connections per topic.

Error Metrics: We counted errors in the elicitation tasks to assess the alignment of participant responses with the actual content of the documents. Errors included statements that contradicted the source documents, unsupported inferences participants made without evidence, and incorrect recall of specific details. Repetitions of the same error were not counted as new, unless they added a new incorrect connection or distinct claim.

Graph Metrics: We calculated the same structural metrics as with the exploration graphs (Subsection 3.5.1) to assess the properties of participants' mental models.

3.5.3 Decision-based Tasks. We analyzed the decision-based task using the same metrics as the concept mapping task: topics and subtopics count, number of connections, and error metrics. We did not apply graph measures here because the graphs were substantially sparser and some participants were not able to produce responses. We discuss why in the Results section.

3.5.4 Qualitative Responses. The qualitative analysis drew on data sources across research questions. For RQ1, we used participants' interaction logs, queries, QAI responses, and graph visualizations, with relevant interview responses to understand exploration patterns, prompting strategies, and differences between interfaces.

For RQ2 and RQ3, we conducted thematic analysis [19, 20] of the interview reflections and comparison (Phase 4), adopting an inductive-deductive coding approach [3, 4, 32]. Coding was guided by the RQs while remaining open to emerging concepts. To develop the initial codebook, the first author began by open-coding four transcripts from two participants selected to capture variation in interface use and preference. Initial low-level codes were grouped into preliminary categories to code 12 transcripts more. We discovered new codes, and revisited transcripts to apply them. After this stage, the sample included a balanced set of participants who favored each interface. The research team met regularly to refine and develop themes across the categories of codes. The second round involved collectively collapsing redundancies and adding codes that aligned with the RQs. Specifically, we examined whether responses provided explanations for the quantitative results, such as insights on knowledge construction (RQ2). The third round involved revision and grouping codes into broader themes that reflected patterns in the data. We agreed on a final codebook of 93 codes that was applied deductively to the remaining 16 transcripts, with earlier transcripts revisited as needed to ensure consistency across the data.

Additionally, participants made direct comparisons by selecting which interface they preferred overall, which they believed supported greater exploration, more learning, and enabled more confident answers. These perceptions were followed by open-ended prompts asking to explain their choices.

3.6 Positionality Statement

None of the authors identify as BLV. Nevertheless, we intentionally ground our study design, analysis, and interpretation in participants' own accounts, practices, and perspectives. Our work centers around the lived realities of SR users, ensuring that RQs, metrics, and conclusions reflect what participants identify as meaningful and relevant. The second author has more than six years of sustained engagement with BLV communities. With the fourth author's experience in assistive technology design, these informed the study design, accessibility of the study materials and attention to the screen-reader practices of participants. The second author's qualitative research background guided the coding process. The third author's expertise in knowledge elicitation and interaction techniques informed our RQs, design and interpretation of elicitation tasks, and the analysis of how participants organized and applied information. The quantitative expertise of the third and fourth authors informed the selection and interpretation of our statistical analyses. We view technology as an active mediator that shapes how people access, explore, and make sense of information. Drawing on the relational model of disability [94, 112], while technology can reduce barriers, we believe restrictive design choices can also actively create disability and impose dysaffordances [28].

4 Results

This section presents the study evidence organized by the RQs. Within each subsection, we consider first analysis of quantitative measurements and then evidence from a qualitative analysis.

4.1 How do People Explore Information Spaces with the Different Interfaces? (RQ1)

4.1.1 Quantitative Results. The quantitative evidence for RQ1 is based on statistical comparisons of key measurements represented in Figure 6 and Figure 7. When counting the unique visited documents, we see that participants tended to visit more documents with the DI ($\mu_{documents,DI} = 11.44$, $\sigma_{documents,DI} = 2.58$) than with the QAI ($\mu_{documents,QAI} = 9.31$, $\sigma_{documents,QAI} = 2.68$), a 22.9% increase, ($t(15) = 2.34$, $p = 0.03 < 0.5$,$\eta^2 = 0.27$). Because documents correspond to topics, this generally supports that exploration was broader with the DI than with the QAI. This is despite the fact that our measurements of total document visits were similar (participants did roughly the same number of visits on average).

We expected that the QAI's nature would lead to more transitions between documents, or at least, more different instances of transitions between documents/topics. This is a property inherent to the QAI: when answering a query, the LLM can output material from multiple documents within the same answer (which we counted as transitions), whereas in the DI transitions can only be done in pre-established ways, such as navigating using hyperlinks between the documents or through the index document. Despite this, the counts of transitions between documents were not conclusively different (all $p > 0.05$ in Figure 6 b, d and e). However, a second set of measurements focuses on whether differences in the interface resulted in a tighter interconnection of more distant topics (Figure 7). This time the answer is positive. When measuring the density, directed diameter and average shortest path length of the navigation graphs (nodes are topics/documents and links are transitions between those topics) we see that the QAI enabled tighter connections between different topics as measured by a higher density of the navigation graph for QAI ($\mu_{density,QAI} = 0.21$, $\mu_{density,DI} = 0.13$), a lower directed diameter ($\mu_{diameter,QAI} = 4.94$, $\mu_{diameter,DI} = 7.75$), and a lower average shortest path length between two topics ($\mu_{aspl,QAI} = 2.44$, $\mu_{aspl,DI} = 3.77$).

4.1.2 Qualitative Results. We used answers from the interviews, visualizations of the exploration patterns, and analysis of the participants' questions and the QAI's answers to make sense of how the two interfaces differ and to contextualize the quantitative results from the previous section. We report counts (e.g., $n = 5$, where n denotes the number of distinct participants that fit the finding) to make the distribution of coded observations transparent, facilitate comparison across findings, and strengthen analytical credibility [80, 81]. We assessed the importance of each theme based on its relevance to our research questions and its relationship to the observed differences.

Our qualitative analysis of participants' explorations revealed three recurring subpatterns: *overview* (learning the overall structure of the information space), *depth* (learning in increasing detail about a particular topic), and *fact finding* (finding answers to specific questions). These align with existing frameworks in exploratory search, information foraging, and sensemaking literature [71, 72, 102] where similar knowledge-oriented intents have been documented as fundamental for complex information spaces. While these behaviors represent general patterns of engagement, we observed that for both *overview* and *depth* it did matter whether participants were content with the *predetermined* structure provided in the documents (i.e., the hierarchical organization and sequential order provided by the document author) or wanted to access the information in a *self-defined* organization. For fact finding this does not seem to matter that much.

	Overview Predetermined	Overview Self-Defined	Depth Predetermined	Depth Self-Defined	Fact Finding
DI	Easy	Hard	Easy	Hard	Hard
QAI	Hard	Easy	Hard	Easy	Easy

Table 2: Summary of the difficulty of using the DI and QAI across the exploration subpatterns, based on qualitative data.

For example, in our worlds, the documents are structured according to a traditional textbook-like set of topics such as history, politics, or culture. However, there are other ways to slice and index the same information space, such as organizing the descriptions as a storyline, or focusing on the influence of specific historical figures. Although both interfaces can support any of the subpatterns (overview, depth, fact finding), the data suggests that the effectiveness of the interface depends largely on whether the participant is ready to leverage the pre-defined structure of the documents or seeks a different one, as reflected in Table 2. The following paragraphs discuss evidence for each of the Table's cells.

Overview-Predetermined. Participants ($n = 9$) identified advantages of the DI when they were willing to accept the existing structure of the documents. P2 observed that they *"could look at a bunch of categories and decide what to read,"* reflecting the traditional way to familiarize oneself with the information space, akin to skimming the main subsections of a Wikipedia page. P12 echoes this: *"The information was there. It told me the links, told me 25 in a list, gave me an idea of what I was getting into."* In contrast, the QAI presents some difficulty, since a predetermined organization is not immediately apparent, and participants ($n = 9$) did not necessarily know how to obtain an overview. P4 observes: *"I didn't know what questions to ask it because I didn't know what information would be available to me."* Some participants did not even realize that an overview could be useful ($n = 6$), as P7 reflects, *"I have no idea what the parameters are. I guess I could have asked it a question like, give me the main topics."*

Overview-Self-Defined. When participants ($n = 5$) are, instead, interested in their particular overview of the information, DI can be frustrating because it requires them to define their own categories: *"It felt like reading a book and trying to navigate through the different pages." (P11)* This is much easier with the QAI, because participants ($n = 7$) can ask the interface to organize information based on their interest. P1 explains that they got *"to determine what I was going to look into and get those big concepts. The other interface [DI] was a bit overwhelming,"*. With QAI, they were able to request cross-document summaries based on their overview preferences without having to check every document.

Depth-Predetermined. When participants were looking for depth, opinions are parallel. DI is great if they do not have a particular view on how to dig deeper, and are comfortable following the document's structure ($n = 7$). Referring to what a document has to offer, P10 notes: *"I know there's a container full of information in each that I can just explore."* P12 echoes this, *"at least there was a container that I knew all the information was in, and I knew how big it was."* In contrast, this is harder to do with QAI ($n = 8$): *"What happens if I keep asking for more detail, it's eventually gonna run out of information about this one thing, so now what?" (P11)* Even when an answer felt detailed, the nature of the question-answer cycle makes it difficult to know when a topic has been exhausted.

Depth-Self-Defined. Nevertheless, if participants are not interested in the grouping (or "narrative"), then DI required readers to manually trace their path across documents and pursue a topic in detail based on what they deemed relevant ($n = 9$). P4 observes, *"...it led me to a whole different path of where I wanted to go to understand the world."* This made self-defined depth possible, but more effortful to keep track of. Conversely, the QAI strongly supports presenting information in whichever way the participant asks ($n = 7$). For example, participants asked questions such as *"How is life governed according to the full cycle?" (P4)* or *"Can you explain to me how is it from 12 to 50 years?" (P9).* P16 explicitly highlights this advantage, *"you can just keep asking questions and isolate on specific topics,"* without having to look for relevant pieces from each document.

Fact Finding. There is no distinction between whether participants are interested or not in the predetermined organization of the document. The QAI ($n = 9$) easily provides contextualized facts as answers (i.e., and advanced search feature) that the DI would require the reader to hunt throughout multiple documents ($n = 7$). This advantage with QAI over DI is acknowledged by P4, *"you're able to get your answers as opposed to having to read 30 pages to find out."*

4.1.3 RQ1 Synthesis. The quantitative data show that DI enables a wider coverage of the information space, whereas QAI supports a more tightly connected set of transitions between the different topics. The qualitative analysis helps us connect those results to the fundamental differences between the interfaces: there is value in having access to a pre-established structure that readers can just consume, which is much easier to do with DI. Participants struggle to access structure with QAI because they have to come up with appropriate queries about a structure and information space that they do not know much about in advance. However, when the user wants to access the information space in more sophisticated or custom ways, such as based on personal interest, then the QAI can significantly facilitate this process by crafting coherent answers that comprise information scattered around different areas of the document.

4.2 How do the Different Interfaces Influence Integration of Knowledge? (RQ2)

If RQ1's results address how people behaved with the interfaces, RQ2 focuses on the consequences. Were participants able to build, connect, and apply what they explored, integrating it into their

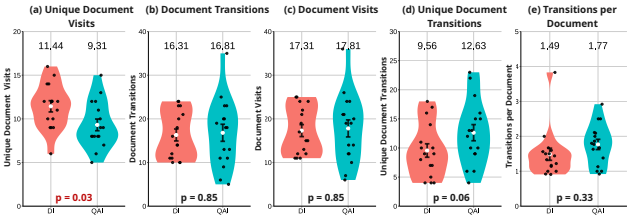


Figure 6: Exploration metrics (RQ1—Phase 1). DI is red, and QAI blue. Numbers above represent the averages. Error bars are standard error. Each black dot is a session.

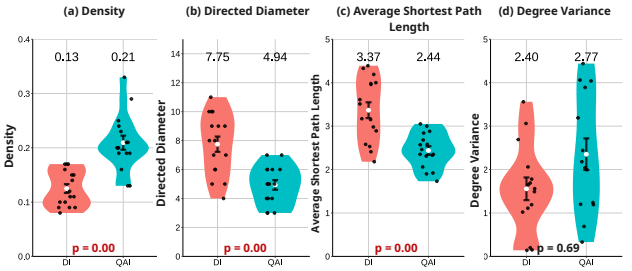


Figure 7: Exploration graph metrics (RQ1—Phase 1). See caption of Figure 6 for details.

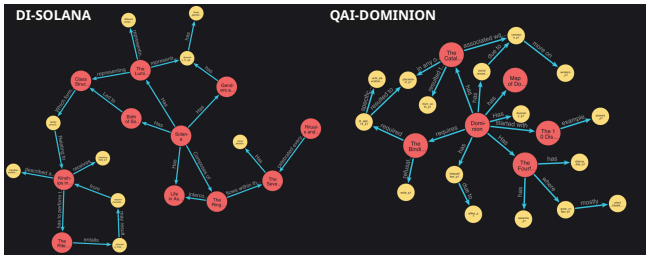


Figure 8: Participant 1's mental model graphs from the concept mapping. DI (left) and QAI (right). Nodes represent elicited topics (red) and subtopics (yellow). Edges represent explicitly described relationships.

mental models? (Phase 2) or when using it in scenario-based responses (Phase 3)? We look at each of the two phases quantitatively first, then discuss the qualitative evidence together.

4.2.1 Concept Maps (Phase 2). We asked participants to recreate their mental structures of the worlds they had explored (the elicitation process is described in Section 3.4). The result of this process is a mental model graph for each session and participant that reflects their understanding of the information space (Figure 8 displays two examples). The graphs are analyzed in three ways: straight counts of topics, subtopics, and their links (Figure 9), metrics of their connectedness (Figure 10), and the correctness of the remembered topics and links (Figure 12).

Overall, participants included 67% more topics on average in their elicited mental model graphs in the DI condition ($\mu_{topics,DI} = 9.44$)

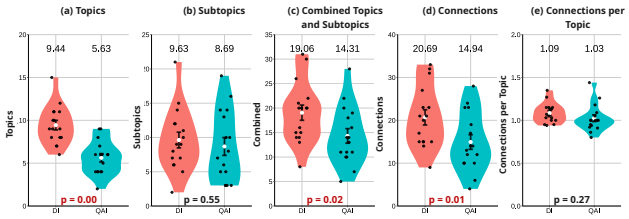


Figure 9: Concept map metrics (RQ2—Phase 2). See caption of Figure 6 for details.

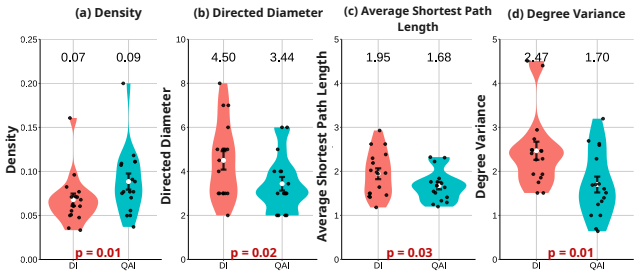


Figure 10: Concept map graph metrics (RQ2—Phase 2). See caption of Figure 6 for details.

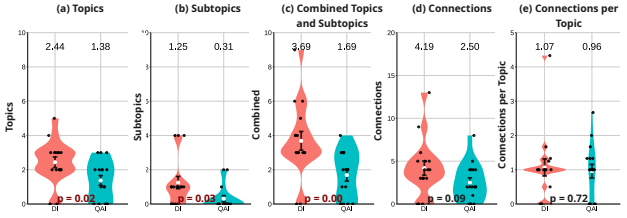


Figure 11: Decision-based metrics (RQ2—Phase 3). See caption of Figure 6 for details.

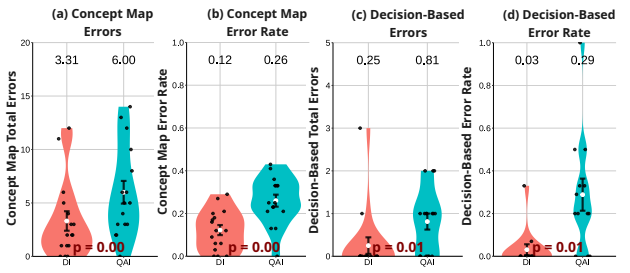


Figure 12: Concept map and Decision-based error metrics (RQ2—Phase 3). See caption of Figure 6 for details.

compared to the QAI condition ($\mu_{topics,QAI} = 5.63$, $t(7.25)$, $p = 0.02$, $\eta^2 = 0.78$) and 11% more subtopics ($\mu_{subtopics,DI} = 9.63$, $\mu_{subtopics,QAI} = 8.69$, $t(0.61)$, $p = 0.55$, $\eta^2 = 0.02$). Considering topics and subtopics together, this represents evidence that the models elicited in the DI condition were more comprehensive on average. With DI, participants made 38% more connections ($\mu_{connections,DI} = 20.69$, $\mu_{connections,QAI} = 14.94$, $t(2.92)$, $p = 0.01$, $\eta^2 = 0.36$), but this seems to be attributable to having more nodes to connect (the connections per topic measures are fairly similar and not statistically distinguishable: $\mu_{connections/topic,DI} = 1.09$, $\mu_{connections/topic,QAI} = 1.03$, $t(1.15)$, $p = 0.27$, $\eta^2 = 0.08$).

We also considered each node or link and classified it as correct or incorrect with respect to the knowledge contained in the documents. In terms of raw error counts, participants made 45% fewer errors with DI than with QAI ($\mu_{errors,DI} = 3.31$, $\mu_{errors,QAI} = 6.00$, $t(5)$, $p < 0.01$, $\eta^2 = 0.62$). Because the elicited models differed in size, we used error rate to show that the difference is slightly amplified (54%) ($\mu_{errorrate,DI} = 0.12$, $\mu_{errorrate,QAI} = 0.26$, $t(6.10)$, $p < 0.01$, $\eta^2 = 0.71$).

We then calculated graph metrics for the concept maps, after removing errors. The results show that the DI elicited mental model graphs were less interconnected than QAI's, with less average density, higher diameter, higher average shortest path length and higher degree variance (all $p < 0.04$, see Figure 10).

4.2.2 Decision-based Scenario (Phase 3). We performed a parallel analysis of the participant answers to the scenario phase, where we asked participants to apply their knowledge of the worlds to suggest solutions to conundrums. The results follow very similar patterns as the previous subsection, which instead of describing in text we leave summarized in Figures 11 and 12. However, there are a few differences between the two analyses. First, the scenario task was much harder than the elicitation task, which left 6 participants unable to provide coherent responses in one ($n = 5$, 4 of which were with the QAI) or both ($n = 1$) of the two sessions. Second, the pattern of higher connections in DI was not strong enough for the conditions to be distinguished statistically. We did not run the graph metrics for the resulting graphs because these graphs are much sparser than in Phase 2, hence noisier and less meaningful.

4.2.3 Qualitative Results. We used interview responses to analyze how participants made sense of the worlds using the two interfaces. Our qualitative analysis involved three processes: *building* understanding of the information space, *connecting* information across documents, and *applying* their gained knowledge to a scenario-based question.

Build. Participants described clear differences in how the two interfaces supported building an understanding of the created worlds. Participants appreciated that DI did not require them to construct their own structure since it is already visible ($n = 5$). P3 explains that DI *"gives you an understanding of what each chapter is and it shows you what you're looking, what you're headed into."* However, DI can be frustrating if readers want to build their own mental structure ($n = 4$). P10 notes, *"I basically have to figure out the structure based on the description of each thing,"* suggesting that they still had to determine where a given piece of information fit in their mental models. Conversely, the QAI supported participants in *"being able to curate the index in my own thought" (P3)* by building structure through their questions ($n = 7$). At the same time, this flexibility depended on participants having skills on what and how to ask ($n = 10$), *"it was a bit more tricky because I had to have a bit more imagination."(P6)*

Connect. When it comes to integrating disparate pieces of information across the information space, some participants perceived

the DI as providing more information volume which, in turn, allowed them to establish more connections between the different parts of the information space ($n = 6$). For example P3 explains that, with DI, *"I felt like I had more to work with and it was more that I could kind of jump off of, so it was helpful."* Yet for some, most connections with DI are implicit (i.e., not links), which forces them to make the connections mentally ($n = 3$). Even when the connection is explicit (i.e., a link), it might still be difficult to know why they are connected. P7 explains this: *"if you're going through it as link to here, link to here, you're not necessarily getting the connections between those."*.

In contrast, several participants recognized that the QAI directly supports integration of related topics, unconstrained by the structure ($n = 7$). In other words, participants appreciate that the QAI composes a coherent text using different bits of information: *"AI does a better job of linking stuff together than you having to pick a piece and go to another link. It pulls it together." (P7)* However, this is not for free, since participants recognized that asking the right questions to get the right integration is not trivial ($n = 8$): *"my questions were probably pretty simplistic and it sort of led you around in circles a couple of times. It's just like any AI"* (P8).

Apply. The decision-based scenario (Phase 3) interrogated whether participants could use the knowledge they had gained and connected through each interface. As we mentioned before, this is likely a much harder task that goes beyond accessing and remembering and requires integration, retrieval, and even creativity; in other words, it is a task higher in Bloom's taxonomy [62]. Some participants thought that the stable and well-organized structure of documents (DI) helped them answer the scenario ($n = 8$). P4 noted about the information to answer a question that *"it gave more, helped me get a visual representation first of what was happening with the various structures."* However, some participants also recognized that, to be successfully applied, they needed to actively process the information: *"It was much more of a lecture because you were just absorbing; I was filling the gaps,"* (P11) ($n = 3$).

With QAI the challenge is of a different nature. Although participants thought that they could ask for information that meaningfully supported their progressive understanding and ability to apply the knowledge ($n = 6$—"I was asking questions that were important to me." (P15)), they were never sure that what they had gathered through questions and answers was sufficient ($n = 8$): *"knowing what to ask and then questioning the output. Am I getting the full answer? Am I getting the correct information?"(P14).*

4.2.4 RQ2 Synthesis. The quantitative data show that participant's elicited mental models contained, on average, more topics and subtopics when using the DI, although less densely connected. Importantly, the elicited graphs also had many more errors. When asked to apply what they had learn, we also observed an advantage when using the DI, which resulted in substantially more topics and subtopics brought up, and also fewer errors. The qualitative data suggests that building mental models with the DI requires effort to connect what they read with their current understanding as they go, discovering implicit connections between the topics. In contrast, the QAI offers more flexibility to build the model based on the participant's own curiosity, and interest. The QAI answers provide custom-formulated answers that directly answer the participant's questions and connect the topics smoothly in the answer's text. In exchange, formulating the right questions might be challenging, and the participants were often not sure of whether they had gathered enough information or reached exhaustive cover of the available information.

	Perception		Measurement			Match	
	DI	QAI	DI	QAI	Equal	DI	QAI
Explore more	10	6	11	4	1	6	2
Learn more	6	10	11	5	0	5	4
Answer Confidence	7	9	12	1	3	5	0
Overall Preference	7	8	N/A	N/A	N/A	N/A	N/A

Table 3: Comparison of subjective judgments and measured outcomes for DI and QAI. Perception shows how many participants said each interface helped them explore more (RQ1), learn more and answer with greater confidence (RQ2). Measurement shows how many participants actually performed better with each interface. Overall preference is subjective only.

4.3 What Aspects of the Different Interfaces Affect Preferences? (RQ3)

At the end of Phase 4, we asked participants for comparative reflections on their experience with the interfaces. Table 3 summarizes these interview responses together with counts of their actual performance, and the counts of alignment. Overall, participants' perceptions did not fully align with the measured patterns. This discrepancy was evident for breadth of exploration and appeared even more strongly for perceived support for learning and confidence in answering the questions. When participants chose their preferred interface, 8 participants favored each. There is no obvious mapping between perceived advantages on either exploration, mental model building or scenario answer confidence and the final stated preference.

Nevertheless, two main themes came up in their interviews and feedback that can help us understand participant preferences: Agency/Control and Effort/Cognitive Load. Interestingly, arguments regarding each of those topics were used to support preferences for either of the two interfaces.

Agency and Control. Participants often evaluated the interfaces in terms of how independent they felt in steering the interaction and shaping control around their own interests. With DI, participants valued having the option to decide for themselves what documents to open and which links to pursue ($n = 9$) *"I felt really empowered. There's all these concepts and I'm going to go with what speaks to me," (P1).* P10 similarly describes *"I could just jump into it right away. I could be independent with it."* Still, for some participants, having the freedom to choose has drawbacks ($n = 3$), *"I lost control because of opening documents. I had to go back somewhere totally unrelated in my opinion." (P4)* In QAI, control meant having the ability to ask anything, which felt open-ended and self-directed ($n = 9$). P9 describes this, *"I liked that I could ask the questions and gave me answers. It acknowledged me."* Similarly, P14 mentions: *"I have more control over what I want to focus on,"* emphasizing the value of the question-answer cycle. However, control also depended

on knowing how to steer the conversation productively ($n = 6$), as P12 reflects: *"I don't think I'm asking the right questions to get the information that I wanted."*

Effort and Cognitive Load. Participants often mentioned the differences in effort involved in obtaining, tracking and working with the information with each interface. Some participants appreciated DI because, as P6 explains, *"there wasn't a lot of guesswork,"* and they could *"jump into it right away,"* ($n = 5$). P13 echoes this by describing DI as *"a straightforward situation,"* and P10 notes that the structure is *"pretty easy to understand."* However, the effort required with DI was often placed in repeated browsing and tracking of the documents ($n = 4$). P7 particularly dislikes *"I hate things that go on and on and send you off to other links. It's too much,"* with P11 also reflecting on the amount of material, *"I find this very exhausting and like a lot of information."* Some participants appreciated the QAI's ability to address several topics at once ($n = 9$), as P4 observes: *"I was able to ask three questions at once,"* and with P14 describing their experience as *"better, quicker, and clearer."* Yet the cost of this interaction relies heavily on the reader to probe the information effectively ($n = 5$), with P8 describing: *"it could be just insecurity. Trying to figure out the questions and do it in an intelligent way."* P15 also reflected that the interaction with QAI depends on the user, *"I think the limitation is your mind, right?"*.

5 Discussion

In this section we interpret the results above in light of the three research questions stated in the introduction, then connect the results to the emerging knowledge of GenAI-based interfaces for knowledge access, including accessibility. We finish by synthesizing recommendations for practitioners and users and qualifying the results based on the limitations of our study design choices.

5.1 Answers to RQ's and their connections

Our behavioral measures of how people chose to explore the information space exposed a trade-off between the two interface styles. Participants used the more traditional DI in a way that resulted in more unique documents visited than the QAI. However, the QAI enabled more flexible transitions from different parts of the information space (RQ1). These differences are not difficult to trace to the characteristics of the interface: the QAI—backed by a current LLM—is able to synthesize answers that connect disparate pieces of information into a smooth narrative, a hitherto difficult to achieve way to navigate information that many people liked and, importantly, allows the reader to personalize their knowledge acquisition.

Unfortunately, the QAI way of interacting with the information space did not seem to translate into more information remembered, better mental models, or an improved ability to apply the gained knowledge or models. In fact, the evidence that we collected supports the opposite: elicited mental models were smaller and contained more errors, although perhaps were a bit more tightly connected (RQ2). This pattern propagated to the application of their understanding, which was also poorer for QAI. We do not rely only on errors as the main or only measure of our analysis, in particular because task instructions de-emphasized recall. Nevertheless, the error measurements very clearly confirm results from the other measurements that models were overall poorer when using the QAI. Task instructions that explicitly emphasized memorization might have produced different outcomes, and this warrants further study.

We anticipate two plausible causal explanations for the connection between the behavior in access (RQ1) and our appraisals of the gained information and understanding (RQ2). First, the DI requires more effort (some participants brought this up), hence more elaboration which, in turn, results in deeper processing, better memory and more functional mental models [68]. Second, the QAI relinquishes the predetermined structure in the documents structure (a kind of ready-to-learn model curated by an expert in advance), forcing the reader to consider which mental structure to build and how to build it, then formulate the appropriate questions to fulfill this purpose. This might simply be too difficult and detract from the mental resources required to do the actual learning. This is, essentially, a Cognitive Load Theory argument about germane cognitive load adding to intrinsic cognitive load to overwhelm cognitive capacity [85, 109].

Interestingly, the challenges introduced by the QAI did not appear obvious to participants, who often rated it as best for the outcomes where we know they did worse (i.e., they were often wrong—see Table 3). In turn, their preferences were evenly split between the two interfaces (RQ3). This suggests that the nature of the interfaces somehow obscures their value to their users (a failure of metacognition, see the next subsection) or that preference was informed by other factors other than their ability to form more complete mental models and apply them. In any case, it is noteworthy and somewhat alarming that so many participants thought they were better with the interface that actually helped them least.

5.2 What does this mean for GenAI interfaces and non-visual information access?

Our findings extend work on traditional and GenAI search. Recent studies suggest that GenAI-based tools can support exploratory search, better user experience, and lower perceived effort [55, 59, 60, 67, 100, 118, 121]. Comparative work between chatbots and search tools afford different kinds of access, and that the strengths depend on the task, domain, and design [73, 87]. This is consistent with our study where the QAI was particularly effective for contextualized fact-finding across multiple documents, whereas the DI better supported broad exploration of the corpus. However, most prior evaluations focus primarily on search tasks and do not measure how people build mental models. Our results therefore contribute empirical evidence suggesting that the GenAI-enabled conversational interfaces that seem poised to replace traditional search as well as much direct document access, could be detrimental for tasks that are more open, exploratory or contributing to people's expertise and general knowledge. This is particularly important for the BLV community for two reasons. First, GenAI interfaces are being quickly adopted by the community for many tasks [5, 110], and traditional document access is often particularly tedious and time-consuming with non-visual input-output [64]. Both make it more likely that they will be used to replace document access. We also suspect that there might be important differences between BLV users and sighted users when using DI vs. QAI interfaces, but this will require further research.

Our results are also consistent with concerns about the metacognitive consequences of LLM use [111]. Experimental work has shown that despite how LLM-based search feels easier and faster, it can also yield superficial learning since users do less of synthesis [73, 106, 111, 119] This is consistent with our findings. First, we did observe that participants experienced the QAI as efficient, flexible, and responsive, yet these benefits did not translate into larger concept maps, or stronger performance in applying what they had learned. Second, participants were often poor at judging which interface actually supported their exploration or learning better. Moreover, we observed that the interfaces function not only as a tool for consuming information but also as a mechanism to help users determine what they have covered, what remains unseen, and how complete or reliable their current understanding is. In our study, the DI relies on users to build an internal representation of document architecture through system interaction at the interface/surface level of the stratified model of information access [97]. Users must grasp what documents exist and infer what information they contain, encouraging users to explore documents across various categories to construct this mental model. In contrast, QAI encourages users to focus on formulating their exploratory intent as prompts at the affective layer. Thus, the user's prompt and early document choices determine the subsequent course of exploration, making coverage and completeness harder to assess.

5.3 Implications for users and future directions for practitioners

Our findings have implications for the adoption of QAI technologies, for the design of future interactive systems and for future research.

First, we believe that users need to be aware that, despite their own initial appraisals of the value of LLM-based QA technologies for accessing documents and other information spaces, they incur the risk of processing information more superficially and being less able to apply it. This is particularly important for BLV individuals, who often seek technological solutions to avoid being perceived as less effective than their sighted peers [8, 79, 122, 124].Further confirmation should encourage wider dissemination of this issue, perhaps in educational contexts. These warnings should be qualified because readers are not always looking to form better models or remember more.

Second, some of the negative effects that we observed might be ameliorated or solved by better design. For example, when people are at a loss of where to start, or how to start structuring their own learning, the system might be able to, prompted or unprompted, offer help on how to carry out the process more efficiently. We did not observe this kind of behavior (asking for help about how to explore, learn or interpret), but it is not rare in current LLMs to offer different levels of hints to the user before or after prompting. Future document interfaces should also be able to incorporate notions of desired agency and effort that adapt to user preferences.

Third, designing new hybrid modes of access for document and information spaces might offer the advantages of DIs and QAIs without the disadvantages, depending on the task or the purpose of the interaction. Our study intentionally separated these modes of access to make differences easier to identify and attribute. Future systems, however, could combine document navigation, search, and conversational interaction to support different tasks and preferences. Interestingly, it might be of research interest to also consider whether documents and other current ways to structure information for human consumption can benefit from different forms and structures, especially if we accept that question-answer interfaces might become the dominant way to access them by the BLV community.

5.4 Limitations

We strove to design the QAI and DI interfaces in a way representative of current interfaces and as close to the state of the art as possible. Nevertheless, many different choices to the ones we made are justifiable, and we cannot discard that they could have influenced the results in different ways. One such choice was enabling touch interaction for spatial documents. This means that we cannot separate the effects of multimodality from other aspects of the DI interface. In other words, having different types of documents might have improved people's ability to remember and integrate the information. Similarly, we decided to include a document in each corpora that provides access to all other documents. This somewhat reduced the incentive for people to navigate between topics from other in-document links, but we also think it is more realistic. Although we still stand by our experimental design choices, we need further experiments to know whether either of these caused the improved performance in the DI condition. Specifically, future systems might include features of both DI and QAI, including both tactile spatial access, search, and presentation of literal information from the document corpus. Ecologically faithful evaluations of such system will require new experiments.

Additionally, we had to decide how to count visits and transitions between visits in Phase 1 in a way that would allow comparisons between the two interfaces. We chose before analysis and with our best judgment. However, we acknowledge that other choices were possible that might result in different balances. Nevertheless, this only affects some measurements in Phase 1; other measurements in other phases are less open to variation because the model elicitation and scenario processes are identical for the two conditions.

Our findings were not independently reviewed by the BLV community. We also focused on SR users to maintain a consistent access modality, as SR use imposes navigation demands central to how these interfaces are experienced, limiting generalizability to BLV users who use other access methods. Finally, we only tested participants in two sessions of ninety minutes. Prolonged use of interfaces for information access is likely to evolve, both in use patterns and in its outcomes.

6 Conclusion

Our study compared two interfaces, DI and QAI, to understand how BLV users explore and build knowledge of unfamiliar domains. From this comparison, we draw attention to a meaningful distinction between interfaces. The nature of interaction with QAI through conversation can feel deceptively expansive, offering a sense of coverage even though content is synthesized. This, in turn, did not translate into better learning. In contrast, DI may demand more effort or might be overwhelming, but they result in a more comprehensive mental model due to the provided structure. This

distinction is important because participants do not always recognize it. Many perceived the QAI as better even when performance suggested otherwise. For BLV users, who may be motivated to adopt these tools to reduce the burden of traditional document access, this introduces a risk of interfaces tha feel easier may also support weaker understanding. As LLM-based interfaces evolve closer to becoming intermediaries, accessible technologies should be evaluated beyond speed or convenience, but also accurate understanding and meaningful exploration.

Acknowledgments

We would like to thank the Canadian Council of the Blind for their support in connecting us with our participants. We are grateful to the participants who gave their time, insights, and trust to this work. We also acknowledge the broader network of practitioners, advocates, and community members who provided us support and guidance.

This research is supported by the University of Victoria, NSERC DG 2020-04401 and the Singapore Ministry of Education (MOE) Academic Research Fund (AcRF) Tier 1 grant (Project ID: 24-SIS-SMU-039).

References

[1] Ali Abdolrahmani, Ravi Kuber, and Stacy M. Branham. 2018. "Siri Talks at You": An Empirical Investigation of Voice-Activated Personal Assistant (VAPA) Usage by Individuals Who Are Blind. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '18)*. Association for Computing Machinery, New York, NY, USA, 249–258. doi:10.1145/3234695.3236344

[2] Ahmad Hisham Zainal Abidin, Hong Xie, and Kok Wai Wong. 2012. Blind users' mental model of web page using touch screen augmented with audio feedback. In *2012 International Conference on Computer & Information Science (ICCIS)*, Vol. 2. 1046–1051. doi:10.1109/ICCISci.2012.6297180

[3] Anne Adams, Peter Lunt, and Paul Cairns. 2008. A qualitative approach to HCI research. In *Research Methods for Human-Computer Interaction*, Anna L. Cox and Paul Cairns (Eds.). Cambridge University Press, Cambridge, 138–157. doi:10.1017/CBO9780511814570.008

[4] Anne Adams, Peter Lunt, and Paul Cairns. 2008. *A qualitative approach to HCI research* (1 ed.). Cambridge University Press, 138–157. doi:10.1017/CBO9780511814570.008

[5] Rudaiba Adnin and Maitraye Das. 2024. "I look at it as the king of knowledge": How Blind People Use and Understand Generative AI Tools. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3663548.3675631

[6] Maristella Agosti and W. Bruce Croft. 2008. *Information Access through Search Engines and Digital Libraries* (1 ed.). The Information Retrieval Series, Vol. 22. Springer Berlin Heidelberg, Berlin, Heidelberg. doi:10.1007/978-3-540-75134-2 ISSN: 1387-5264.

[7] Qingyao Ai, Jingtao Zhan, and Yiqun Liu. 2025. Foundations of Generative Information Retrieval. In *Information Access in the Era of Generative AI*, Ryen W. White and Chirag Shah (Eds.). Springer Nature Switzerland, Cham, 15–45. doi:10.1007/978-3-031-73147-1_2

[8] Tugba Kamali Arslantas and Abdulmenaf Gul. 2022. Digital literacy skills of university students with visual impairment: A mixed-methods analysis. *Education and Information Technologies* 27, 4 (2022), 5605–5625. doi:10.1007/s10639-021-10860-1

[9] Chieko Asakawa and Takashi Itoh. 1998. User interface of a Home Page Reader. In *Proceedings of the Third International ACM Conference on Assistive Technologies* (Marina del Rey, California, USA) *(Assets '98)*. Association for Computing Machinery, New York, NY, USA, 149–156. doi:10.1145/274497.274526

[10] Alex Atcheson, Omar Khan, Brian Siemann, Anika Jain, and Karrie Karahalios. 2025. "I'd Never Actually Realized How Big An Impact It Had Until Now": Perspectives of University Students with Disabilities on Generative Artificial Intelligence. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, Article 42, 22 pages. doi:10.1145/3706598.3714121

[11] Kumaripaba Athukorala, Dorota Głowacka, Giulio Jacucci, Antti Oulasvirta, and Jilles Vreeken. 2016. Is exploratory search different? A comparison of information search behavior for exploratory and lookup tasks. *J. Assoc. Inf. Sci. Technol.* 67, 11 (Nov. 2016), 2635–2651. doi:10.1002/asi.23617

[12] Marcos Baez, Claudia Maria Cutrupi, Maristella Matera, Isabella Possaghi, Emanuele Pucci, Gianluca Spadone, Cinzia Cappiello, and Antonella Pasquale. 2022. Exploring challenges for Conversational Web Browsing with Blind and Visually Impaired Users. In *Extended Abstracts of the 2022 CHI Conference on Human Factors in Computing Systems (CHI EA '22)*. Association for Computing Machinery, New York, NY, USA, 1–7. doi:10.1145/3491101.3519832

[13] Ricardo Baeza-Yates and Berthier Ribeiro-Neto. [n. d.]. *Modern Information Retrieval*. Pearson, Harlow.

[14] L. Bainbridge. 1983. IRONIES OF AUTOMATION. In *Analysis, Design and Evaluation of Man–Machine Systems*, G. Johannsen and J. E. Rijnsdorp (Eds.). Pergamon, 129–135. doi:10.1016/B978-0-08-029348-6.50026-9

[15] Gutenberg Barros, Walter Correia, and João Marcelo Teixeira. 2023. Towards the Effectiveness of 3D Printing on Tactile Content Creation for Visually Impaired Users. *Polymers* 15, 9 (Jan. 2023), 2180. doi:10.3390/polym15092180 Publisher: Multidisciplinary Digital Publishing Institute.

[16] Randolph G. Bias, Brian M. Moon, and Robert R. Hoffman. 2015. Concept Mapping Usability Evaluation: An Exploratory Study of a New Usability Inspection Method. *International Journal of Human–Computer Interaction* 31, 9 (Sept. 2015), 571–583. doi:10.1080/10447318.2015.1065692

[17] Jeffrey P. Bigham, Irene Lin, and Saiph Savage. 2017. The Effects of "Not Knowing What You Don't Know" on Web Accessibility for Blind Web Users. In *Proceedings of the 19th International ACM SIGACCESS Conference on Computers and Accessibility* (Baltimore, Maryland, USA) *(ASSETS '17)*. Association for Computing Machinery, New York, NY, USA, 101–109. doi:10.1145/3132525.3132533

[18] Braille Authority of North America. 2022. Guidelines and Standards for Tactile Graphics. https://www.brailleauthority.org/guidelines-and-standards-tactile-graphics

[19] Virginia Braun and Victoria Clarke. 2021. Thematic analysis: A practical guide. (2021).

[20] Virginia Braun and Victoria Clarke. 2023. Thematic analysis. In *APA handbook of research methods in psychology: Research designs: Quantitative, qualitative, neuropsychological, and biological, Vol. 2, 2nd ed.* American Psychological Association, Washington, DC, US, 65–81. doi:10.1037/0000319-004

[21] Andrei Broder, Ravi Kumar, Farzin Maghoul, Prabhakar Raghavan, Sridhar Rajagopalan, Raymie Stata, Andrew Tomkins, and Janet Wiener. 2000. Graph structure in the Web. *Computer Networks* 33, 1 (June 2000), 309–320. doi:10.1016/S1389-1286(00)00083-9

[22] Monica Bucciarelli and Ilaria Cutica. 2012. Mental Models in Improving Learning. In *Encyclopedia of the Sciences of Learning*. Springer, Boston, MA, 2213–2215. doi:10.1007/978-1-4419-1428-6_669

[23] Matthew Butler, Leona M Holloway, Samuel Reinders, Cagatay Goncu, and Kim Marriott. 2021. Technology Developments in Touch-Based Accessible Graphics: A Systematic Review of Research 2010-2020. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems (CHI '21)*. Association for Computing Machinery, New York, NY, USA, 1–15. doi:10.1145/3411764.3445207

[24] Katriina Byström and Kalervo Järvelin. 1995. Task complexity affects information seeking and use. *Information Processing & Management* 31, 2 (March 1995), 191–213. doi:10.1016/0306-4573(95)80035-R

[25] Nicholas J. Cepeda, Harold Pashler, Edward Vul, John T. Wixted, and Doug Rohrer. 2006. Distributed practice in verbal recall tasks: A review and quantitative synthesis. *Psychological Bulletin* 132, 3 (May 2006), 354–380. doi:10.1037/0033-2909.132.3.354

[26] Brian J. Chan, Chao-Ting Chen, Jui-Hung Cheng, and Hen-Hsen Huang. 2025. Don't Do RAG: When Cache-Augmented Generation is All You Need for Knowledge Tasks. In *Companion Proceedings of the ACM on Web Conference 2025 (WWW '25)*. Association for Computing Machinery, New York, NY, USA, 893–897. doi:10.1145/3701716.3715490

[27] Arnavi Chheda-Kothary, Ather Sharif, David Angel Rios, and Brian A. Smith. 2025. "It Brought Me Joy": Opportunities for Spatial Browsing in Desktop Screen Readers. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–18. doi:10.1145/3706598.3714125

[28] Sasha Costanza-Chock. 2020. *Design Justice: Community-Led Practices to Build the Worlds We Need*. The MIT Press. doi:10.7551/mitpress/12255.001.0001

[29] Bruce Croft, Donald Metzler, and Trevor Strohman. 2010. *Search Engines: Information Retrieval in Practice*. Pearson, Boston.

[30] Hai Dang, Sven Goller, Florian Lehmann, and Daniel Buschek. 2023. Choice Over Control: How Users Write with Large Language Models using Diegetic and Non-Diegetic Prompting. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3544548.3580969

[31] Bhuwan Dhingra, Lihong Li, Xiujun Li, Jianfeng Gao, Yun-Nung Chen, Faisal Ahmed, and Li Deng. 2017. Towards End-to-End Reinforcement Learning of Dialogue Agents for Information Access. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Regina

Barzilay and Min-Yen Kan (Eds.). Association for Computational Linguistics, Vancouver, Canada, 484–495. doi:10.18653/v1/P17-1045

[32] Alan Dix. 2020. *Statistics for HCI: Making Sense of Quantitative Data.* Morgan & Claypool Publishers.

[33] Stacy A. Doore, David Istrati, Chenchang Xu, Yixuan Qiu, Anais Sarrazin, and Nicholas A. Giudice. 2024. Images, Words, and Imagination: Accessible Descriptions to Support Blind and Low Vision Art Exploration and Engagement. *Journal of Imaging* 10, 1 (Jan. 2024), 26. doi:10.3390/jimaging10010026

[34] Niklas Elmqvist. 2023. Visualization for the Blind | IX Magazine Issue XXX.1 January - February 2023. *Interactions* 30, 1 (Jan. 2023). https://interactions.acm.org/archive/view/january-february-2023/visualization-for-the-blind

[35] Ernst & Young. 2024. *GenAI for Accessibility: More Human, Not Less – How Does Microsoft 365 Copilot Impact the Working Experience of People Living with Disability and/or Neurodivergence?* Research Report. Ernst & Young Global Limited. https://www.ey.com

[36] J. V. Erp. 2002. Guidelines for the use of vibro-tactile displays in human computer interaction. https://www.semanticscholar.org/paper/Guidelines-for-the-use-of-vibro-tactile-displays-in-Erp/cc553fb7289f993bd39c1ea67a93e4acfa8f5add

[37] Ethan Fast, Binbin Chen, Julia Mendelsohn, Jonathan Bassen, and Michael S. Bernstein. 2018. Iris: A Conversational Agent for Complex Tasks. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18).* Association for Computing Machinery, New York, NY, USA, 1–12. doi:10.1145/3173574.3174047

[38] Nele Fischer and Wenzel Mehnert. 2021. Building Possible Worlds: A Speculation Based Framework to Reflect on Images of the Future. https://www.semanticscholar.org/paper/Building-Possible-Worlds%3A-A-Speculation-Based-to-on-Fischer-Mehnert/7cd6bb89ce62b7e59dab4062153945719eb34805

[39] Marco Furini, Silvia Mirri, Manuela Montangero, and Catia Prandi. 2020. Do Conversational Interfaces Kill Web Accessibility? *2020 IEEE 17th Annual Consumer Communications & Networking Conference (CCNC)* (Jan. 2020), 1–6. doi:10.1109/CCNC46108.2020.9045477

[40] Dedre Gentner and Albert L. Stevens. 2014. *Mental Models.* Psychology Press.

[41] Jens Gerken, Hans-Christian Jetter, Michael Zöllner, Martin Mader, and Harald Reiterer. 2011. The concept maps method as a tool to evaluate the usability of APIs. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11).* Association for Computing Machinery, New York, NY, USA, 3373–3382. doi:10.1145/1978942.1979445

[42] Ricardo E. Gonzalez Penuela, Ruiying Hu, Sharon Lin, Tanisha Shende, and Shiri Azenkot. 2025. Towards Understanding the Use of MLLM-Enabled Applications for Visual Interpretation by Blind and Low Vision People. In *Proceedings of the Extended Abstracts of the CHI Conference on Human Factors in Computing Systems.* 1–8. doi:10.1145/3706599.3719714

[43] Ananya Gubbi Mohanbabu and Amy Pavel. 2024. Context-Aware Image Descriptions for Web Accessibility. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '24).* Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3663548.3675658

[44] Richa Gupta, M. Balakrishnan, and P.V.M. Rao. 2017. Tactile Diagrams for the Visually Impaired. *IEEE Potentials* 36, 1 (Jan. 2017), 14–18. doi:10.1109/MPOT.2016.2614754

[45] Richa Gupta, P. V. M. Rao, M. Balakrishnan, and S. Mannheimer. 2019. Evaluating the Use of Variable Height in Tactile Graphics. In *2019 IEEE World Haptics Conference (WHC).* 121–126. doi:10.1109/WHC.2019.8816083

[46] Erika Hall. 2018. *Conversational Design.* Mule Design, S.l.

[47] Sumaiya Haque, Hesam Mahmoudi, Navid Ghaffarzadegan, and Konstantinos Triantis. 2023. Mental models, cognitive maps, and the challenge of quantitative analysis of their network representations. *System Dynamics Review* 39, 2 (2023), 152–170. doi:10.1002/sdr.1729

[48] Samantha Harper and Stephen Dorton. 2019. A Context-Driven Framework for Selecting Mental Model Elicitation Methods. *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63, 1 (Nov. 2019), 367–371. doi:10.1177/1071181319631422

[49] Tingying He, Maggie McCracken, Daniel Hajas, Sarah Creem-Regehr, and Alexander Lex. 2025. Using Tactile Charts to Support Comprehension and Learning of Complex Visualizations for Blind and Low-Vision Individuals. doi:10.48550/arXiv.2507.21462 arXiv:2507.21462 [cs].

[50] Niklas Holtz, Sven Wittfoth, and Jorge Marx Gómez. 2024. The New Era of Knowledge Retrieval: Multi-Agent Systems Meet Generative AI. In *2024 Portland International Conference on Management of Engineering and Technology (PICMET).* 1–10. doi:10.23919/PICMET64035.2024.10653018 ISSN: 2159-5100.

[51] Kasper Hornbæk. 2013. Some Whys and Hows of Experiments in Human–Computer Interaction. *Found. Trends Hum.-Comput. Interact.* 5, 4 (June 2013), 299–373. doi:10.1561/1100000043

[52] Natalie Jones, Helen Ross, Timothy Lynam, Pascal Perez, and Anne Leitch. 2011. Mental models: an interdisciplinary synthesis of theory and methods. *SMART Infrastructure Facility - Papers* (Jan. 2011). http://ro.uow.edu.au/smartpapers/81

[53] J. Bern Jordan, Victoria Van Hyning, Mason A. Jones, Rachael Bradley Montgomery, Elizabeth Bottner, and Evan Tansil. 2024. Information Wayfinding of Screen Reader Users: Five Personas to Expand Conceptualizations of User Experiences. In *Proceedings of the 26th International ACM SIGACCESS Conference on Computers and Accessibility* (St. John's, NL, Canada) *(ASSETS '24).* Association for Computing Machinery, New York, NY, USA, Article 47, 7 pages. doi:10.1145/3663548.3688539

[54] Crescentia Jung, Shubham Mehta, Atharva Kulkarni, Yuhang Zhao, and Yea-Seul Kim. 2022. Communicating Visualizations without Visuals: Investigation of Visualization Alternative Text for People with Visual Impairments. *IEEE Transactions on Visualization and Computer Graphics* 28, 1 (Jan. 2022), 1095–1105. doi:10.1109/TVCG.2021.3114846

[55] Abhishek Kaushik and Gareth J. F. Jones. 2025. Comparing Conventional and Conversational Search Interaction Using Implicit Evaluation Methods. 292–304. https://www.scitepress.org/Link.aspx?doi=10.5220/0011798500003417

[56] Diane Kelly. 2009. Methods for Evaluating Interactive Information Retrieval Systems with Users. *Foundations and Trends® in Information Retrieval* 3, 1–2 (April 2009), 1–224. doi:10.1561/1500000012 Publisher: Now Publishers, Inc..

[57] Geoffrey Keppel. 1991. *Design and analysis: A researcher's handbook, 3rd ed.* Prentice-Hall, Inc, Englewood Cliffs, NJ, US.

[58] David E. Kieras and Susan Bovair. 1984. The role of a mental model in learning to operate a device. *Cognitive Science* 8, 3 (July 1984), 255–273. doi:10.1016/S0364-0213(84)80003-8

[59] Hannah Kim, Sergei L. Kosakovsky Pond, and Stephen MacNeil. 2025. Conversations over Clicks: Impact of Chatbots on Information Search in Interdisciplinary Learning. doi:10.48550/arXiv.2507.21490 arXiv:2507.21490 [cs].

[60] Nakyung Kim and Yong Gu Ji. 2026. Exploratory search with generative AI: An empirical study on the impact of interaction design strategies on information exploration and cognitive load. *International Journal of Human-Computer Studies* 210 (March 2026), 103771. doi:10.1016/j.ijhcs.2026.103771

[61] Satwik Ram Kodandaram, Utku Uckun, Xiaojun Bi, Iv Ramakrishnan, and Vikas Ashok. 2024. Enabling Uniform Computer Interaction Experience for Blind Users through Large Language Models. In *The 26th International ACM SIGACCESS Conference on Computers and Accessibility.* ACM, St. John's NL Canada, 1–14. doi:10.1145/3663548.3675605

[62] David R. Krathwohl. 2002. A Revision of Bloom's Taxonomy: An Overview. *Theory Into Practice* 41, 4 (Nov. 2002), 212–218. doi:10.1207/s15430421tip4104_2 _eprint: https://doi.org/10.1207/s15430421tip4104_2.

[63] Sri Hastuti Kurniawan and Alistair Sutcliffe. 2002. Mental Models of Blind Users in the Windows Environment. In *Computers Helping People with Special Needs*, Klaus Miesenberger, Joachim Klaus, and Wolfgang Zagler (Eds.). Springer, Berlin, Heidelberg, 568–574. doi:10.1007/3-540-45491-8_109

[64] Hae-Na Lee, Vikas Ashok, and I.V. Ramakrishnan. 2020. Repurposing Visual Input Modalities for Blind Users: A Case Study of Word Processors. *Conference proceedings. IEEE International Conference on Systems, Man, and Cybernetics* 2020 (Oct. 2020), 2714–2721. doi:10.1109/smc42975.2020.9283015

[65] Florian Lehmann and Daniel Buschek. 2024. Functional Flexibility in Generative AI Interfaces: Text Editing with LLMs through Conversations, Toolbars, and Prompts. doi:10.48550/arXiv.2410.10644 arXiv:2410.10644 [cs].

[66] David D. Lewis and Karen Spärck Jones. 1996. Natural language processing for information retrieval. *Commun. ACM* 39, 1 (Jan. 1996), 92–101. doi:10.1145/234173.234210

[67] Bulou Liu, Yueyue Wu, Yiqun Liu, Fan Zhang, Yunqiu Shao, Chenliang Li, Min Zhang, and Shaoping Ma. 2021. Conversational vs Traditional: Comparing Search Behavior and Outcome in Legal Case Retrieval. In *Proceedings of the 44th International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR '21).* Association for Computing Machinery, New York, NY, USA, 1622–1626. doi:10.1145/3404835.3463064

[68] Robert Lockhart and Fergus Craik. 1990. Levels of Processing: A Retrospective Commentary on a Framework for Memory Research. *Canadian Journal of Psychology* 44, 1 (March 1990), 87–112. insights.ovid.com

[69] Darren Lunn, Simon Harper, and Sean Bechhofer. 2011. Identifying Behavioral Strategies of Visually Impaired Users to Improve Access to Web Content. *ACM Trans. Access. Comput.* 3, 4, Article 13 (April 2011), 35 pages. doi:10.1145/1952388.1952390

[70] Heron M., Kinchin I.M., and Medland E. 2018. Interview talk and the co-construction of concept maps. *Educational Research* 60, 4 (Oct. 2018), 373–389. doi:10.1080/00131881.2018.1522963 _eprint: https://doi.org/10.1080/00131881.2018.1522963.

[71] Gary Marchionini. 2006. Exploratory search: from finding to understanding. *Commun. ACM* 49, 4 (April 2006), 41–46. doi:10.1145/1121949.1121979

[72] Gary Marchionini and Ben Shneiderman. 1988. Finding Facts vs. Browsing Knowledge in Hypertext Systems. *Computer* 21, 1 (Jan. 1988), 70–80. doi:10.1109/2.222119

[73] Shiri Melumad and Jin Ho Yun. 2025. Experimental evidence of the effects of large language models versus web search on depth of learning. *PNAS Nexus* 4, 10 (Oct. 2025), pgaf316. doi:10.1093/pnasnexus/pgaf316

[74] Fengran Mo, Kelong Mao, Ziliang Zhao, Hongjin Qian, Haonan Chen, Yiruo Cheng, Xiaoxi Li, Yutao Zhu, Zhicheng Dou, and Jian-Yun Nie. 2025. A Survey of Conversational Search. *ACM Trans. Inf. Syst.* (Aug. 2025). doi:10.1145/3759453 Just Accepted.

[75] Robert J. Moore. 2018. A Natural Conversation Framework for Conversational UX Design. In *Studies in Conversational UX Design*, Robert J. Moore, Margaret H. Szymanski, Raphael Arar, and Guang-Jie Ren (Eds.). Springer International Publishing, Cham, 181–204. doi:10.1007/978-3-319-95579-7_9

[76] M.P.Geetha, G. Thirukumaran, C.Pradakshana, B.Sudharsana, and T. Ashwin. 2024. Conversational AI Meets Documents Revolutionizing PDF Interaction with GenAI. In *2024 International Conference on Emerging Research in Computational Science (ICERCS)*. 1–6. doi:10.1109/ICERCS63125.2024.10895303

[77] 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 (2021), 1726. doi:10.3390/pr9101726

[78] Sucheta Nadkarni. 2003. Instructional Methods and Mental Models of Students: An Empirical Investigation. *Academy of Management Learning & Education* 2, 4 (Dec. 2003), 335–351. doi:10.5465/amle.2003.11901953

[79] Carmen del Rosario Navas-Bonilla, Julio Andrés Guerra-Arango, Daniel Alejandro Oviedo-Guado, and Daniel Eduardo Murillo-Noriega. 2025. Inclusive education through technology: a systematic review of types, tools and characteristics. *Frontiers in Education* 10 (Feb. 2025). doi:10.3389/feduc.2025.1527851

[80] Mitchell Nicmanis. 2024. Reflexive content analysis: An approach to qualitative data analysis, reduction, and description. *International Journal of Qualitative Methods* 23 (2024), 16094069241236603.

[81] Helen Noble and Joanna Smith. 2025. Ensuring validity and reliability in qualitative research. *Evidence-Based Nursing* 28, 4 (2025), 206–208.

[82] Donald A. Norman. 1983. Some observations on mental models. In *Mental models*. Hillsdale, USA, 7–14.

[83] Maija Nousiainen and Ismo T. Koponen. 2010. Concept maps representing knowledge of physics: Connecting structure and content in the context of electricity and magnetism. *Nordic Studies in Science Education* 6, 2 (2010), 155–172. doi:10.5617/nordina.253

[84] Christina Oumard, Julian Kreimeier, and Timo Götzelmann. 2022. Pardon? An Overview of the Current State and Requirements of Voice User Interfaces for Blind and Visually Impaired Users. In *Computers Helping People with Special Needs: 18th International Conference, ICCHP-AAATE 2022, Lecco, Italy, July 11–15, 2022, Proceedings, Part I*. Springer-Verlag, Berlin, Heidelberg, 388–398. doi:10.1007/978-3-031-08648-9_45

[85] Fred G. W. C. Paas and Jeroen J. G. Van Merriënboer. 1994. Instructional control of cognitive load in the training of complex cognitive tasks. *Educational Psychology Review* 6, 4 (Dec. 1994), 351–371. doi:10.1007/BF02213420

[86] Bouquet Paolo and M. Warglien. 1999. Mental models and local semantics: the problem of information integration. https://www.semanticscholar.org/paper/Mental-models-and-local-semantics%3A-the-problem-of-Paolo-Warglien/74212e7ecac2dfc133c984375a68566061213663

[87] Leonardo Pasquarelli, Charles Koutcheme, and Arto Hellas. 2025. AI Chatbots vs. Traditional Search: A Comparative Study on Student Information Retrieval. In *2025 IEEE Frontiers in Education Conference (FIE)*. 1–8. doi:10.1109/FIE63693.2025.11328491 ISSN: 2377-634X.

[88] Samir Passi and Mihaela Vorvoreanu. 2022. Overreliance on AI literature review. *Microsoft Research* 339 (2022), 340. https://www.microsoft.com/en-us/research/wp-content/uploads/2022/06/Aether-Overreliance-on-AI-Review-Final-6.21.22.pdf

[89] Celeste Lyn Paul. 2014. Analyzing card-sorting data using graph visualization. *J. Usability Studies* 9, 3 (May 2014), 87–104.

[90] Minoli Perera, Swamy Ananthanarayan, Cagatay Goncu, and Kim Marriott. 2025. The Sky is the Limit: Understanding How Generative AI can Enhance Screen Reader Users' Experience with Productivity Applications. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3706598.3713634

[91] Alisha Pradhan, Kanika Mehta, and Leah Findlater. 2018. "Accessibility Came by Accident": Use of Voice-Controlled Intelligent Personal Assistants by People with Disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems (CHI '18)*. Association for Computing Machinery, New York, NY, USA, 1–13. doi:10.1145/3173574.3174033

[92] Anita M. Preininger, Bedda L. Rosario, Adam M. Buchold, Jeff Heiland, Nawshin Kutub, Bryan S. Bohanan, Brett South, and Gretchen P. Jackson. 2021. Differences in information accessed in a pharmacologic knowledge base using a conversational agent vs traditional search methods. *International Journal of Medical Informatics* 153 (Sept. 2021), 104530. doi:10.1016/j.ijmedinf.2021.104530

[93] Emanuele Pucci, Isabella Possaghi, Claudia Maria Cutrupi, Marcos Baez, Cinzia Cappiello, and Maristella Matera. 2023. Defining Patterns for a Conversational Web. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems (CHI '23)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3544548.3581145

[94] Solveig Magnus Reindal. 2008. A social relational model of disability: a theoretical framework for special needs education? *European Journal of Special Needs Education* 23, 2 (May 2008), 135–146. doi:10.1080/08856250801947812

[95] Siti Nur Syazana Mat Saei, Suziah Sulaiman, and Halabi Hasbullah. 2010. Mental model of blind users to assist designers in system development. In *2010 International Symposium on Information Technology*, Vol. 1. 1–5. doi:10.1109/ITSIM.2010.5561350 ISSN: 2155-899X.

[96] Jaime Sanchez and Hector Flores. 2010. Concept Mapping for Virtual Rehabilitation and Training of the Blind. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 18, 2 (April 2010), 210–219. doi:10.1109/TNSRE.2009.2032186

[97] Tefko Saracevic. 1997. The stratified model of information retrieval interaction: Extension and applications. In *Proceedings of the annual meeting-american society for information science*, Vol. 34. Learned Information (Europe) Ltd, 313–327.

[98] Bahareh Sarrafzadeh, Olga Vechtomova, and Vlado Jokic. 2014. Exploring knowledge graphs for exploratory search. In *Proceedings of the 5th Information Interaction in Context Symposium (IIiX '14)*. Association for Computing Machinery, New York, NY, USA, 135–144. doi:10.1145/2637002.2637019

[99] Camille J Saucier and Christopher M Dobmeier. 2025. A mental models approach to communication: integrating the features, functions, and mechanisms of mental modeling. *Communication Theory* (June 2025), qtaf012. doi:10.1093/ct/qtaf012

[100] Ted Selker and Yunzi Wu. 2024. Generative AI's aggregated knowledge versus web-based curated knowledge. doi:10.48550/arXiv.2410.12091 arXiv:2410.12091 [cs].

[101] Tanusree Sharma, Yu-Yun Tseng, Lotus Zhang, Ayae Ide, Kelly Avery Mack, Leah Findlater, Danna Gurari, and Yang Wang. 2025. "Before, I Asked My Mom, Now I Ask ChatGPT": Visual Privacy Management with Generative AI for Blind and Low-Vision People. In *Proceedings of the 27th International ACM SIGACCESS Conference on Computers and Accessibility (ASSETS '25)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3663547.3746335

[102] Ben Shneiderman. 1996. The Eyes Have It: A Task by Data Type Taxonomy for Information Visualizations. In *Proceedings of the 1996 IEEE Symposium on Visual Languages (VL '96)*. IEEE Computer Society, USA, 336.

[103] Herbert A. Simon. 1978. On the forms of mental representation. (1978). https://conservancy.umn.edu/bitstream/handle/11299/185337/9_01Simon.pdf?sequence=1

[104] Ayah Soufan. 2023. Towards Understanding and Supporting Exploratory Searches. In *Proceedings of the 2023 Conference on Human Information Interaction and Retrieval (CHIIR '23)*. Association for Computing Machinery, New York, NY, USA, 490–494. doi:10.1145/3576840.3578304

[105] Konrad Sowa and Aleksandra Przegalinska. 2025. From Expert Systems to Generative Artificial Experts: A New Concept for Human-AI Collaboration in Knowledge Work. *Journal of Artificial Intelligence Research* 82 (April 2025), 2101–2124. doi:10.1613/jair.1.17175

[106] Matthias Stadler, Maria Bannert, and Michael Sailer. 2024. Cognitive ease at a cost: LLMs reduce mental effort but compromise depth in student scientific inquiry. *Computers in Human Behavior* 160 (Nov. 2024), 108386. doi:10.1016/j.chb.2024.108386

[107] Hari Subramonyam, Roy Pea, Christopher Pondoc, Maneesh Agrawala, and Colleen Seifert. 2024. Bridging the Gulf of Envisioning: Cognitive Challenges in Prompt Based Interactions with LLMs. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–19. doi:10.1145/3613904.3642754

[108] Siddharth Suri, Scott Counts, Leijie Wang, Chacha Chen, Mengting Wan, Tara Safavi, Jennifer Neville, Chirag Shah, Ryen W. White, Reid Andersen, Georg Buscher, Sathish Manivannan, Nagu Rangan, and Longqi Yang. 2024. The Use of Generative Search Engines for Knowledge Work and Complex Tasks. *CoRR* (Jan. 2024). https://openreview.net/forum?id=qm5wHzFco2

[109] John Sweller. 1988. Cognitive Load During Problem Solving: Effects on Learning. *Cognitive Science* 12, 2 (1988), 257–285. doi:10.1207/s15516709cog1202_4

[110] Xinru Tang, Ali Abdolrahmani, Darren Gergle, and Anne Marie Piper. 2025. Everyday Uncertainty: How Blind People Use GenAI Tools for Information Access. In *Proceedings of the 2025 CHI Conference on Human Factors in Computing Systems (CHI '25)*. Association for Computing Machinery, New York, NY, USA, 1–17. doi:10.1145/3706598.3713433

[111] Lev Tankelevitch, Viktor Kewenig, Auste Simkute, Ava Elizabeth Scott, Advait Sarkar, Abigail Sellen, and Sean Rintel. 2024. The Metacognitive Demands and Opportunities of Generative AI. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems (CHI '24)*. Association for Computing Machinery, New York, NY, USA, 1–24. doi:10.1145/3613904.3642902

[112] Carol Thomas. 2004. Rescuing a social relational understanding of disability. *Scandinavian Journal of Disability Research* 6, 1 (2004), 22–36. arXiv:https://doi.org/10.1080/15017410409512637 doi:10.1080/15017410409512637

[113] John R Thompson, Jesse J Martinez, Alper Sarikaya, Edward Cutrell, and Bongshin Lee. 2023. Chart Reader: Accessible Visualization Experiences Designed with Screen Reader Users. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 802, 18 pages. doi:10.1145/3544548.3581186

[114] Cecília Torres, Walter Franklin, and Laura Martins. 2019. Accessibility in Chatbots: The State of the Art in Favor of Users with Visual Impairment. In *Advances in Usability, User Experience and Assistive Technology*. Springer, Cham, 623–635. doi:10.1007/978-3-319-94947-5_63

[115] Jakub Wabiński, Albina Mościcka, and Guillaume Touya. 2022. Guidelines for Standardizing the Design of Tactile Maps: A Review of Research and Best Practice. *The Cartographic Journal* 59, 3 (July 2022), 239–258. doi:10.1080/00087041.2022.2097760

[116] WCAG. 2025. Web Content Accessibility Guidelines (WCAG) 2.1. https://www.w3.org/TR/WCAG21/

[117] Christian von der Weth and Manfred Hauswirth. 2013. DOBBS: Towards a Comprehensive Dataset to Study the Browsing Behavior of Online Users. In *Proceedings of the 2013 IEEE/WIC/ACM International Joint Conferences on Web Intelligence (WI) and Intelligent Agent Technologies (IAT) - Volume 01 (WI-IAT '13, Vol. 01)*. IEEE Computer Society, USA, 51–56. doi:10.1109/WI-IAT.2013.8

[118] Yuyu Yang, Kelsey Urgo, Jaime Arguello, and Robert Capra. 2025. Search+Chat: Integrating Search and GenAI to Support Users with Learning-oriented Search Tasks. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '25)*. Association for Computing Machinery, New York, NY, USA, 57–70. doi:10.1145/3698204.3716446

[119] Quinton Yong, Anthony Estey, and Miguel Nacenta. 2026. Characterization and Effects of CS2 Learning with GenAI, Visualization, and Human Support. *arXiv preprint arXiv:2606.02933* (2026).

[120] Byung Sung Yoon and Antonie J. Jetter. 2016. Comparative analysis for Fuzzy Cognitive Mapping. In *2016 Portland International Conference on Management of Engineering and Technology (PICMET)*. 1897–1908. doi:10.1109/PICMET.2016.7806755

[121] Saber Zerhoudi and Michael Granitzer. 2025. SearchLab: Exploring Conversational and Traditional Search Interfaces in Information Retrieval. In *Proceedings of the 2025 ACM SIGIR Conference on Human Information Interaction and Retrieval (CHIIR '25)*. Association for Computing Machinery, New York, NY, USA, 382–389. doi:10.1145/3698204.3716475

[122] Yichun Zhao, Miguel A. Nacenta, Mahadeo A. Sukhai, and Sowmya Somanath. 2026. Accessibility-Driven Information Transformations in Mixed-Visual Ability Work Teams. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26)*. Association for Computing Machinery, New York, NY, USA, 1–14. doi:10.1145/3772318.3790872

[123] Yichun Zhao, Miguel A Nacenta, Mahadeo A. Sukhai, and Sowmya Somanath. 2024. TADA: Making Node-link Diagrams Accessible to Blind and Low-Vision People. In *Proceedings of the 2024 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '24)*. Association for Computing Machinery, New York, NY, USA, Article 45, 20 pages. doi:10.1145/3613904.3642222

[124] Yichun Zhao, Miguel A Nacenta, Mahadeo A Sukhai, and Sowmya Somanath. 2026. " If We Had the Information That We Need to Interpret the World Around Us, We Wouldn't Be Disabled:" Barriers and Opportunities in Information Work among Blind and Sighted Colleagues. In *Proceedings of the 5th Annual Symposium on Human-Computer Interaction for Work*. 1–15.

[125] Jonathan Zong, Crystal Lee, Alan Lundgard, JiWoong Jang, Daniel Hajas, and Arvind Satyanarayan. 2022. Rich Screen Reader Experiences for Accessible Data Visualization. *Computer Graphics Forum* 41, 3 (2022), 15–27. doi:10.1111/cgf.14519

[126] Jonathan Zong, Crystal Lee, Alan Lundgard, JiWoong Jang, Daniel Hajas, and Arvind Satyanarayan. 2022. Rich Screen Reader Experiences for Accessible Data Visualization. *Computer Graphics Forum* 41, 3 (2022), 15–27. doi:10.1111/cgf.14519