Understanding How to Inform Blind and Low-Vision Users about Data Privacy through Privacy Question Answering Assistants

Yuanyuan Feng, *University of Vermont;* Abhilasha Ravichander, *Allen Institute for Artificial Intelligence;* Yaxing Yao, *Virginia Tech;* Shikun Zhang and Rex Chen, *Carnegie Mellon University;* Shomir Wilson, *Pennsylvania State University;* Norman Sadeh, *Carnegie Mellon University*

https://www.usenix.org/conference/usenixsecurity24/presentation/feng-yuanyuan

This paper is included in the Proceedings of the 33rd USENIX Security Symposium.

August 14–16, 2024 • Philadelphia, PA, USA

978-1-939133-44-1

Open access to the Proceedings of the 33rd USENIX Security Symposium is sponsored by USENIX.

Understanding How to Inform Blind and Low-Vision Users about Data Privacy through Privacy Question Answering Assistants

Yuanyuan Feng
University of Vermont

Abhilasha Ravichander
Allen Institute for AI

Yaxing Yao
Virginia Tech

Shikun Zhang
Carnegie Mellon University

Rex Chen
Carnegie Mellon University

Shomir Wilson
Pennsylvania State University

Norman Sadeh
Carnegie Mellon University

Abstract

Understanding and managing data privacy in the digital world can be challenging for sighted users, let alone blind and low-vision (BLV) users. There is limited research on how BLV users, who have special accessibility needs, navigate data privacy, and how potential privacy tools could assist them. We conducted an in-depth qualitative study with 21 US BLV participants to understand their data privacy risk perception and mitigation, as well as their information behaviors related to data privacy. We also explored BLV users' attitudes towards potential privacy question answering (Q&A) assistants that enable them to better navigate data privacy information. We found that BLV users face heightened security and privacy risks, but their risk mitigation is often insufficient. They do not necessarily seek data privacy information but clearly recognize the benefits of a potential privacy Q&A assistant. They also expect privacy Q&A assistants to possess cross-platform compatibility, support multi-modality, and demonstrate robust functionality. Our study sheds light on BLV users' expectations when it comes to usability, accessibility, trust and equity issues regarding digital data privacy.

1 Introduction

Navigating information about how websites, mobile applications, digital services, and Internet-connected devices ("digital technologies" thereafter) collect, use, and share personal data is challenging. First, it is difficult to find privacy policies – the legal documents mandated by many privacy regulations that disclose data privacy practices [69]. Then, understanding the data practices disclosed in these privacy policies can be even more challenging due to their length, legal jargon, and vague language [42, 48]. Given how difficult it is in general for people to find and understand data privacy information, it is natural to wonder how accessible these tasks are for people who are blind or have low vision. According to the Cornell University Employment and Disability Institute's interpretation of the 2016 American Community Survey, more than seven million US adults (2.4%) have a visual disability [53]. This population often relies on assistive technology, such as screen readers and magnifiers, to access digital information. In this paper, we use positive affirming adjectives recommended by the US National Federation of the Blind [1] —"blind" and "low-vision" (instead of "visually-impaired")—to describe this population ("BLV people/users" thereafter).

Prior research shows that BLV people are particularly vulnerable to online security and security threats due to the lack of visual cues [6, 34]. Generally, digital technologies tend to have poor accessibility features to support their general information needs in the digital world [17, 40]. Therefore, it is critical to improve not only the usability but also the accessibility of security and privacy (S&P) tools. We seek to understand BLV people's needs in navigating the data privacy information regarding the digital technologies with which they interact. We also aim to inform the design of accessible privacy tools, enabling BLV users to have the same level of access to privacy information as sighted users [83].

Specifically, we want to explore BLV users' attitudes towards "privacy assistants", broadly defined in this paper as tools designed to help users navigate and/or manage digital data privacy [12, 16, 45]. Recently, there has been considerable research on developing privacy assistants that can answer users' privacy questions based on the content of privacy policies using natural language processing (NLP) techniques [32, 60, 62]. Inspired by these recent advances in privacy question answering and the increasing public interest in applications such as ChatGPT[2] built on Large Language Models, we particularly investigate the NLP-based privacy question answering assistants ("privacy Q&A assistants" thereafter) as a promising approach to improve accessibility and bridge potential privacy inequity for BLV users.

To this end, we conducted a qualitative interview study with 21 US BLV users of various digital technologies to investigate three research questions (RQs):

[1] The National Federation of the Blind. https://nfb.org/
[2] ChatGPT. https://chat.openai.com/

- *RQ1: How do BLV people perceive and mitigate data privacy risks associated with digital technologies?*
- *RQ2: What are BLV people's information (seeking) behaviors around data privacy?*
- *RQ3: What do BLV people expect from potential privacy Q&A tools for navigating data privacy information?*

By answering these RQs, this paper contributes:

- The first in-depth qualitative investigation into how BLV users perceive and mitigate data privacy risks and how they seek (if any) data privacy information.
- To the understanding of BLV users' expectations around functionality and accessibility for potential privacy Q&A assistants that inform them about data privacy.
- To the growing area of inclusive security and privacy (S&P), providing insights into designing accessible, usable, and equitable S&P tools for BLV users and beyond.

2 Background and Related Work

2.1 How BLV Users Access Information

BLV users primarily rely on screen readers (e.g., JAWS and NVDA software on computers, VoiceOver on Apple devices, Talkback on Android devices), a type of assistive technology that reads out aloud text on the device screen, to access digital information and utilize other digital technologies. Screen readers only function well when digital technologies and contents follow good accessibility practices. However, research [2, 3, 29] revealed that websites and mobile applications do not consistently adhere to accessibility standards such as the Web Content Accessibility Guidelines [35].

Recently, agent-based visual interpreter services, including Aira[3] and Be My Eyes[4] mobile apps, have gained traction among BLV users. These services connect BLV users with sighted human agents, who help them recognize objects and cope with everyday situations through the phone cameras. Similarly, artificial intelligence(AI)-based visual-aid apps like Seeing AI[5] that leverage device cameras to describe texts, objects, people, and environments are also on the rise, enabling BLV users to access certain information independently.

Meanwhile, computing researchers have explored novel technical solutions and interaction modalities to improve information access for BLV people. As smart home speakers and their built-in voice-based assistants (e.g., Amazon Echo devices with Alexa) gain popularity, voice user interfaces significantly improve information accessibility for BLV people [56]. Also, research advances in computer vision [27] and mixed reality [81, 82] have also opened new possibilities to convey visual information through other modalities. We broadly categorize these above-mentioned examples as assistive technology in this paper.

[3] Aira. https://aira.io

[4] Be My Eyes. https://www.bemyeyes.com

[5] Seeing AI. https://www.microsoft.com/en-us/ai/seeing-ai/

2.2 Privacy and Security for BLV Users

With growing recognition of the importance of accessibility and inclusiveness in privacy and security technologies [74], many research studies examined blind and low-vision users' privacy concerns and behaviors. Ahmed et al. [6]'s interview study revealed blind participants' unique privacy concerns in three environments: physical (e.g., eavesdropping), digital (e.g., privacy settings in social media), and the intersection of physical and digital (e.g., shoulder-surfing). To fulfill their privacy and security needs, Hayes et al. found that BLV users also rely heavily on their allies (e.g., family members, caregivers) to protect their privacy and security cooperatively [34]. Akter et al.'s survey study elaborated on BLV users' concerns regarding camera-based visual interpreter services, where volunteers answer their questions about photos or videos [7].

Though focusing on a different population, Hamidi et al.'s study revealed that the privacy and utility trade-offs of adaptive assistive technologies might be overlooked among older adults with pointing problems [31]. This suggests that users with accessibility needs may knowingly use assistive technologies or services that compromise their data privacy. Such divergence between privacy attitudes and behaviors is commonly known as the privacy paradox [25].

Another important body of work focuses on the usability and accessibility of security and privacy tools for BLV users. For example, Danoso et al. found that web authentication can be time-consuming to BLV users and pose significant challenges, such as accessing error messages [17, 18]. More importantly, these usability issues may result in BLV users' risky behaviors (e.g., not being able to identify phishing websites) or decisions that compromise security [52]. Such usability issues also impact how BLV users seek and access general information [9, 51, 70, 76]. A study investigating online information behaviors revealed the barriers for blind web users to assess the credibility of websites [1], highlighting the need to assist BLV users in verifying the credibility of online information. However, to our knowledge, there is no research focusing on BLV users' information behaviors for data privacy information, a type of information that is challenging even for sighted users to navigate and understand [55, 63, 69, 72]. Our study aims to explore this untrodden topic to understand the challenges faced by BLV users in navigating data privacy.

2.3 Privacy Question Answering (Q&A)

Data privacy information about digital technologies can be obtained through many channels from different sources. A primary source is the official privacy policies – the legal doc-

uments in which companies or organizations self-disclose the practices they engage in with user data. Despite being required in many regulatory regimes around the world, privacy policies are difficult to find [69] and even more difficult to understand due to their technical and legal jargon [50]. The time required to read them is also impractical for most users [54].

Thus a growing line of research has explored using natural language processing (NLP) techniques to extract salient information from privacy policies, which can empower users to take control of their data privacy in the digital world. Prior studies [8, 13, 14, 46, 57, 75] have successfully identified data practices within privacy policies to potentially enable easier navigation of privacy policy content. Kumar et al. [43] extracted opt-out choices from website privacy policies and presented them to users in a web browser extension, *Opt-Out Easy*. Prior research efforts [39, 71, 78, 79] also developed techniques to summarize the most salient aspects of privacy policies to present to end-users. However, other research has indicated that summarization approaches that are not tailored to the needs of individual users are unlikely to be effective at meaningfully informing people about the information in privacy policies [26, 58]. Consequently, in recent years, there has been considerable interest in developing assistants that will answer users' privacy questions, which would enable people to flexibly access information within privacy policies that are most pertinent to them. Harkous et al. [33] created *PriBots*, chatbots for communicating privacy practices to users based on questions asked by users on Twitter. Ravichander et al. [60] constructed a benchmark for privacy question answering systems, where they source answers from experts with legal training and provide systems based on pretrained language models to identify answers to these questions. Ahmad et al. [5] extracted segments from policies in the OPP-115 Corpus [75], recruited "skilled annotators" to construct questions based on these segments, and explored transfer learning-based approaches to provide answers to these questions.

Building on existing research efforts to automatically answer users' data privacy questions [59, 61], our study explores BLV users' expectations for similar privacy Q&A tools to assist them in navigating data privacy information.

3 Methods

3.1 Assumption and Method Justification

Informed by literature review and guidance from two blind consultants, this study assumes that BLV people face challenges in navigating data privacy information and they can benefit from accessible privacy tools like the privacy Q&A assistant. To account for the potential inaccuracy of this assumption, we carefully formulated our research questions (RQs) in an open-ended manner, as shown in Section 1).

We chose the qualitative interviewing method for its strengths in obtaining a deep understanding of participants' perceptions, attitudes, and experiences [10], which is well-suited for RQ1 and RQ2. Alternative methods to address RQ3 are prototype testing and evaluation with users, which are more beneficial during the late stages of software development [64]. Because there was little understanding of BLV users' needs around privacy Q&A, we decided not to present researcher-derived prototypes to avoid inaccurate assumptions about BLV users' preferences. Instead, we decided to ask participants to freely imagine a hypothetical "digital assistant", a technique used in prior privacy and security interview studies [12, 68, 77] to elicit their requirements and expectations. This technique is particularly advantageous for requirements gathering in the early stages of tool design [37].

3.2 Study Design and Research Ethics

We used the university-licensed Zoom software to conduct remote interviews in 2021 during the COVID pandemic. We sought advice from two blind consultants in our academic network to ensure all study procedures are culturally responsible. The first consultant provided us with best practices for recruiting BLV participants into research studies. The second consultant completed a mock-up interview and provided suggestions to improve the study materials. For example, we revised our study materials to accommodate varying technology literacy and the high unemployment rate among BLV people in the US. Then, we conducted a pilot interview with a blind friend to refine the wording of some interview questions and finalize all study materials.

Our study protocol was approved by the Institution Review Board (IRB) at Carnegie Mellon University with the permission to obtain informed consent verbally because our participants may have difficulty signing electronic documents. We emailed participants the IRB-approved consent form when we scheduled interviews because many BLV users prefer reading text via screen readers at a fast speed. When they joined the interview remotely, we asked if they had read the consent form. If yes, we read the study summary section of the consent form; if not, one researcher read the full consent form. Then we obtained their verbal consent before starting the audio recording. We also completed a research data sharing agreement after the first author joined the University of Vermont.

3.3 Interview Questions

We started each interview by asking the participant's preferred terminology to describe their vision status so that we could use their preferred terminology throughout the interview. We structured our interview questions based on the RQs, with additional baseline questions at the start and optional demographic questions at the end. We describe interview question design rationale below and provide them in Appendix A.

Baseline questions: We started with questions to establish a baseline of participants' technology use, understanding of

data practices, and attitude towards data privacy. During baseline questions, we provided them with plain-language definitions for "digital technologies", "data privacy", and "data practices" in the context of this study to ensure a shared understanding of these terminologies (see Appendix B). After providing the definitions, we asked about their general thoughts around data privacy because people's privacy attitudes are shown to impact their perceived risks [41].

RQ1: Risk perception and mitigation. We asked participants to describe the potential risks associated with the digital technologies that they use and optionally to compare the risk levels between general and assistive technology. We employed the critical incident technique [23] by asking participants to share their most memorable experiences (i.e., critical incidents) around data privacy in the past few months. Then, we asked a hypothetical dilemma question to understand how they would mitigate privacy risks if assistive technology engaged in data practices that they were uncomfortable with.

RQ2: Information (seeking) behaviors. Not assuming all participants actively seek data privacy information, we used "information behaviors" [9] to broadly describe how participants navigate, access, seek, and understand data privacy information. We first asked about the sources from which they obtain data privacy information and their perceived credibility of these sources. Then, we focused on their prior experience seeking data privacy information, if any. We noted if participants naturally mentioned privacy policies in their responses. If not, we prompted privacy policies as a source.

RQ3: Expectations for privacy Q&A tools. To minimize the bias towards privacy Q&A assistants as the researcher-introduced solution, we crafted the interview questions to emphasize participants' needs for privacy Q&A. We first introduced an imaginary privacy expert who can answer their data privacy questions for any digital technologies they use. Then, we let them assume this expert was a "digital assistant" and asked them to freely describe their expectations for this digital assistant without priming them with any information about how the assistant might function. After that, we probed further into detailed aspects of this hypothetical digital assistant, including modalities, developers, and information sources. Finally, we asked about their perceived benefits, concerns, and potential use cases.

Demographic questions: The interview ended with optional demographic questions. All participants answered these questions to provide us with data on sample diversity.

3.4 Recruitment and Data Collection

Recruiting BLV people into research studies requires trust-building, so we recruited participants through our personal contacts and the National Federation of the Blind. Our consultants cautioned us about the sampling bias towards BLV users who are already comfortable using digital technologies. To mitigate this, we included a phone number in the recruitment materials and provided accessible instructions for joining Zoom via the Internet or phone call.

Aiming to increase sample diversity, we included the following statement in all study recruitment materials: "We particularly welcome diverse perspectives from individuals who are less familiar with technology and who also belong to other underrepresented groups." When replying to study inquiries, we politely explain the rationale prioritizing participants from underrepresented groups. Many voluntarily shared their demographic information with us during inquiry, and those who were not chosen expressed their understanding. Besides the above-mentioned diversity consideration, we responded to potential participants primarily by their time of inquiry. We aimed for a sample size of around 20 by referencing prior qualitative studies with BLV participants [6, 17, 34, 56].

We recruited 21 self-identified blind or low-vision participants in the US: 2 from our personal connections with snowball sampling and 19 through the approved email solicitation of the National Federation of the Blind. To avoid the potential discomfort some people feel when realizing they are the first few participants in a study, we assigned the number P10 to the pilot participant and numbered the full study participants from P11 to P31. Note that we did not use data saturation [24] to determine the sample size due to the inherent difficulty recruiting BLV users. Instead, we set our sample size goal by referencing published qualitative studies with this population.

The first and second authors conducted the first four interviews together and slightly adjusted interview question wording through discussing the initial results with the research team. All remaining interviews were led by either the first or second author, with an optional secondary interviewer from the research team. We audio-recorded all interviews using Zoom after requesting participants to turn off their cameras during the interviews to ensure only audio was recorded. We compensated each participant with an accessible electronic gift card of 25 US dollars after the interview.

3.5 Qualitative Data Analysis

We collected audio recordings from 21 interviews ranging from 40 to 92 minutes in length (average was 65 minutes). We first leveraged Zoom's automatic transcription to generate initial transcripts and then manually reviewed all auto-generated transcripts for correctness according to the audio recordings. Then, the first author structured all transcripts with multi-level section headings according to RQs to ensure effective navigation of transcripts in the coding process. To rigorously assess the qualitative data, we combined both inductive and deductive coding approaches in our thematic analysis [11, 22]. Four team members with prior qualitative data analysis experience were involved as coders to ensure internal reliability. Our coding process included five steps: (1) The first author conducted the first round of inductive open coding and summarized emerging themes in the data; (2) The research team

discussed these themes with examples from the data and then create an initial codebook based on study RQs and themes identified in the first round coding. (3) The research team used the codebook to perform the second round of coding, where the first author and three co-authors (secondary coders) double-coded all transcripts; (4) The first authors discussed with three secondary coders individually to resolved all coding conflicts, which eliminated coder errors and further clarified a few codes. Inter-coder reliability measures are not necessary when coders reach full consensus [49]; (5) The first author performed additional axial coding and selective coding [73] to synthesize high-level meta-themes.

There is a growing recognition in the research community that the frequency of themes in qualitative research should not be interpreted quantitatively for generalization [20, 28]. Rather, the contribution lies in the identification of these themes and in-depth discussion of their implications. We report frequencies of themes and codes when appropriate, but also adopt a consistent terminology used by Zhang et al. [80] (Figure 1) to convey the relative prominence of themes. We archive our coded data including exact frequencies in Open Science Framework (see Appendix C).

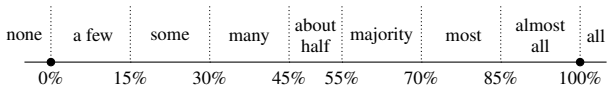

Figure 1: Our terminology to describe theme frequencies

4 Participants and Baseline Questions

4.1 Sample and Demographics

19 of 21 participants self-reported to be blind or legally blind, and two had low vision. Our sample is balanced across genders, age groups, and employment statuses (see Appendix D). We over-sampled non-white participants (N=8) compared to the US demographic distribution, while two participants voluntarily reported as members of the LGBTQ+ community. Our sample is biased towards people with high education levels as all participants had at least some college education. Overall, our sample is relatively inclusive and mitigates common sampling bias with the BLV population.

4.2 Participants' Technology Use

Almost all participants reported having used computers (N=20) and phones or tablets (N=21). Overall, we noticed an overall heavier reliance on mobile phones than computers. Specifically, two participants no longer use computers, citing the steep learning curve of computer screen readers and affordability considerations. 18 participants reported using various smart home devices, including smart speakers and other smart devices controlled through smart speakers. Participants also mentioned using Victor Stream devices (N=4) and Braille Displays (N=5), both of which are assistive output devices with narrow functionality and thus excluded from our analysis due to their limited data privacy implications.

Participants used two major categories of digital technologies on their devices: (1) **Assistive technology** that enables them to access various digital information, products, or services. All participants reported using screen readers on both computers (e.g., JAWS, NVDA) and mobile devices (e.g., VoiceOver, TalkBack). 18 participants used agent-based visual interpreter (VI) services, while 15 participants used AI-based visual-aid apps (e.g., Seeing AI). All participants also used voice assistants (e.g., Siri, Alexa) on their devices for accessibility. (2) **General technology** that broadly refers to all devices, websites, apps, and digital products or services.

Additionally, our participants reported using computers mostly for formal activities such as work, banking, and information seeking. In contrast, they used their phones and/or tablets for broader online activities, including shopping, banking, communication, entertainment, and social media. 12 participants explicitly mentioned using websites or apps for their banking needs, but a few reported not using any online banking for security reasons. Those who have smart speakers primarily use them for content, quick life help, and controlling other smart home devices.

4.3 Data Privacy Awareness and Control

All participants identified at least one type of personal data being collected and knew advertising was a major purpose for data collection. 19 participants knew that digital technologies' data practices are pervasive. 17 participants were aware of the existence of data privacy control options made available by digital technologies, many of whom reported having configured cookie settings, privacy settings, and app permissions.

However, nine participants expressed that the existing control options are ineffective. Additionally, eight participants mentioned using some broadly defined privacy-enhancing technologies, including more private browsers or search engines and virtual private networks (VPNs). In summary, our participants are generally aware of data privacy and engage in some privacy control behaviors, which is similar to a recent sample with sighted Internet users in the US [67].

4.4 Data Privacy Attitudes

Our participants expressed multi-facet attitudes towards data privacy and we summarize the key themes below.

Pervasive data practices and the hope for change. Participants described digital technologies' data practices as "pervasive" (N=9) and even "intrusiveness" (N=9). Some participants believed such data practices should be limited or hoped for better privacy notice and choice mechanisms.

Not concerned and resigned. Only two participants expressed explicit concerns about personal digital data privacy, while four were not concerned at all due to perceived low risks. Four participants expressed privacy resignation [19], as P18 said: *"It's very difficult to contain [data practices] no matter what message you try."*

Low data privacy expectations online and in daily life. Seven participants felt their online privacy was limited and two participants explicitly commented on their low privacy expectations as blind individuals, said P11:

"I think privacy is a very important thing...for blind people, we just don't get that privacy. There's nothing in my whole life that's private because being blind...somebody has to know everything to help me out with something."

Neutral or positive attitudes towards data practices. Seven participants were okay with reasonable use of their personal data such as "providing or improving service" but felt differently about targeted advertising. Such a difference may derive from participants' opinions towards the data economy, the phenomenon that users' personal data is exchanged for access to a service and that data may subsequently be shared or sold to other entities. For example, three participants have neutral views about the data economy, as explained by P17: *"We are a capitalist society so companies need to make money."* Three participants mentioned personal benefits from targeted advertising, such as knowing about useful products.

Overall, the data reported in this section highlights the fact that BLV users must rely on assistive technology or sighted people (either from personal lives or interpreter services) to access online information, use other digital technologies, and manage their everyday lives, which invisibly impact their privacy perceptions and behaviors. Our participants' privacy awareness and their diverse attitudes towards digital data privacy share many similarities with sighted users [4, 38, 67].

5 BLV Users' Risk Perception and Mitigation

5.1 Perceived Security and Privacy Risks

Our participants reported a broad range of security and privacy (S&P) risks with digital technologies beyond the definition of data privacy presented to them (detailed in Appendix B).

General technology security risks. 15 participants mentioned the risks around personal financial data (e.g., social security numbers, credit card numbers, banking information). Eight participants also mentioned data breaches that leak their sensitive data collected by legitimate companies, and some cited the Equifax data breach as an example. Notably, the above-mentioned risks are security risks, not necessarily data privacy risks. Our analysis showed that only seven participants clearly distinguished security risks from data privacy risks associated with data practices. During the interview, we intentionally did not interrupt participants when they talked about security risks beyond data privacy. Though not asked in our interview, six participants voluntarily shared their experience falling victim to financial crimes, from personal financial information being used by a neighbor to credit card numbers being stolen online. Our participants' experiences revealed their vulnerability to financial crimes. It is only natural for many participants to be more concerned about financial security than data privacy.

Visual interpreter services: agents and companies. 14 participants identified the agents working for visual interpreter (VI) services as a risk. P20 expressed her concern: *"When you call in, if they [agents] can't see it [the content to be interpreted] properly, they're supposed to ask permission before they snap a picture of it, but 99% of agents will just snap a picture without asking you first now. It's your credit card sitting there; They've got your information."*

Two participants pointed out the risks of how VI service companies store their personal data, as P24 explained: *"There is the potential of my privacy being exposed because even though your information is stored in the cloud...those can also be compromised, and they may not know in time to inform you or to even make a solution for it."*

Compared to VI services, participants were less concerned about privacy risks associated with AI-based visual-aid apps, as P17 said: *"Seeing AI is not a problem because that's my phone reading it back to me. But Be My Eyes and Aira, I'm concerned about using them because it's a live person who may be reading sensitive information."*

Expanded physical-digital risks. Some participants mentioned physical-digital privacy risks, including shoulder surfing and eavesdropping by people nearby. We found such physical-digital risks expanded to assistive technology and other scenarios. For example, P31 was concerned about using voice assistants in public: *"One of the things I think is a real downside of the audio and the speech recognition style controls like Siri and Echo devices is that it forces you to speak information out loud...and that's a risk all by itself...if you forget other people [are] around sometimes."* Similarly, P11 mentioned a new category of physical-digital risk with VI services, she said *"Not only [the agents], there are whoever happens to be in their houses...now they're all working from home, so you hear other people in the background, and I wonder, is this really secure?"*

Risks around data practices. Participants mentioned risks associated with various data practices regarding both general technology (N=7) and assistive technology (N=2). A few participants reported physical safety concerns related to location tracking, said P21: *"If the data that's being collected could be shared with people that I haven't consented to, or could reveal things like my location and personal information about where I live, or might be expected to be in that create safety concerns."* Additionally, two raised concerns about unwanted surveillance, either by governments or companies. For example, P22 said, *"data privacy issues now are becoming a political thing...if the wrong people figure out with your*

data you're not in the same political camp as they are", and pointed out the risk to civil liberty if being censored by digital technologies.

More concerned about general technology. Mid-study, we added a follow-up question asking participants to compare their stated data privacy risks between general technology and assistive technology. 10 out of 13 participants who answered this question were more concerned about privacy risks with general technology. A few explained their rationale that assistive technology is a "small market" thus less likely to be targeted by attacks. Also, some participants already mitigated assistive technology use (see section 5.2) and thus were less concerned. However, within assistive technology, VI services draw the highest level of concern, while screen readers are considered low-risk.

Critical incidents around data privacy. Our critical incident technique probed into the participants' recent experiences when they felt surprised, uncomfortable, or suspicious of certain data practices. 19 participants recalled at least one critical incident. Ten of them were surprised by unexpected features of certain apps or services, including Facebook's friend recommendations, auto-fill features on some websites, and Apple devices asking to share Wi-Fi passwords with contacts. Ten participants feel surprised or uncomfortable with targeted advertising, particularly by the speed and accuracy of these ads. Five of them specifically mentioned ads that appeared to be cross-platform, as P22 explained: *"I'm on Amazon, and I searched for something, I would expect them to show me [the same] sort of things later on. I expect that. I don't expect Facebook to show it to me. And I've seen that happen."* Additionally, many shared their experiences receiving unexpected spam/scam emails (N=5) or security incidents online (N=3). Most of these reported critical incidents made participants realize how pervasively various digital technologies collect, use, and share their personal data. However, many participants' responses reveal misconceptions or gaps in understanding about online behavioral advertising and cookies.

5.2 Mitigation of Perceived Risks

Primary strategy: adjusted technology use Adjusting their usage patterns of digital technologies is the most commonly mentioned risk mitigation strategy. Four participants adjusted their general technology use, including limiting their social media usage to preserve online privacy (P22) and choosing mobile banking apps over web-based banking interfaces for better security (P24). For assistive technology, ten participants mitigated the risks with visual interpreter services by intentionally not sharing sensitive information with Be My Eyes volunteers, as P13 said: *"With Be My Eyes...you do not use them to read credit cards to you that you got in the mail because your other one expired. Because they're not background checked, they're just people that really want to help...But with Aira, the agents are extensively trained; they're also background checked, and I think they're bonded or something, so I have had them read credit cards..."*

Such usage adjustments were not financially viable for all BLV users. A few participants in our study mentioned that cost is a key consideration when using agent-based VI services, as P14 explained: *" I take advantage of it [Aira's promotion]. You don't have to pay for it for [participating] stores that you go shopping, you can get access [to Aira] and it doesn't cost in places like Walgreen [store]. But I also use Be My Eyes...That's a free one you don't have to pay for. They are volunteers from all over the world."*

Discontinued use and non-use. Only five participants stopped using certain digital technologies or switching to alternatives out of privacy or security reasons. These include deleting accounts on e-commerce sites due to security concerns (P11), stopping using social media apps out of privacy concerns (P12, P14), switching to new browsers because Internet Explorer is no longer maintained for security (P30), and adopting non-Google search engines to limit data exposure online (P27). However, no participants reported stopping using any assistive technologies for privacy and security reasons. In contrast, seven participants chose the non-use strategy to avoid risks with digital technologies: For general technology, a few did not use online or mobile banking to reduce financial risks. For assistive technology, a few mentioned that they avoided using agent-based VI services.

Security and privacy practices. Ten participants reported adopting good privacy and security practices for risk mitigation, such as using more private search engines or browsers, configuring privacy settings or mobile app permissions, using strong passwords, and being cautious wherever personal data is requested. We also observed **varying levels of security and privacy knowledge**. A few tech-savvy participants understood the benefits of using relatively effective security and privacy tools such as virtual private networks (VPNs). For example, P27 was confident about his security practices: *"The security is so locked down around my credit cards that I think it takes an actual data breach to get them. I don't think it's through the privacy settings."* In contrast, many participants **only followed generic S&P advice**, and a few reported struggling with online privacy, including difficulty understanding how cookies work (P11). A few participants also shared their frustration when security impeded accessibility, said P18: *"I was doing [mobile] banking. I had someone put the app on [my phone], I could check my balance and inquiry transactions, but then they [the bank] changed it. every time you go on the site, you have to put a new password. The major problem is that I can't type it in. On the Apple phone, you can dictate everything except your password."*

Hypothetical dilemma with assistive technology. Our hypothetical dilemma question forced participants to weigh up between accessibility and privacy: 17 participants reported that problematic data practices would affect their willingness to use certain assistive technology. However, the attitude may

not stop them from using the assistive technology in such a dilemma. Only eight firmly stated that they would stop using that assistive technology, most of whom felt confident in finding good alternatives. Nine participants said it would depend on the risk level, as P19 explained: *"No matter what you do, we're still going to have a risk. It all depends on to what extent the information is shared and how it's being shared. "* It also depended on how heavily they rely on the assistive technology, as P18 said: *"It has to do with how much you need this device (technology). I'm totally blind, I don't have any vision at all, and I live alone. So my need for the device is greater than someone else who maybe has some vision or lives with somebody."* Specifically, some participants mentioned giving up or switching screen readers would be the most difficult, and a few could not switch due to the limited available alternatives. We notice that less technology-proficient participants were less likely to stop using or switch, said P11: *"It's always a concern, and if I found out that they did have a privacy (issue), they weren't really secure, or they were leaking information. it's hard to say, because if you are blind, you don't really have a whole lot of options. So do you take a chance and do it, or can you do without? Some things you just can't do without."*

Participants' responses revealed a high level of trust in assistive technology, with the exception of agent-based VI services. For example, P25 said: *"Especially with assistive technology, I operate with a very high level of trust. If they would do something that would erode my trust level, I would seriously consider changing things or making a complete switch to something different, if need be."*

6 BLV Users' Information (Seeking) Behavior

6.1 Information Sources and Credibility

Data privacy information sources. Participants reported various sources that they obtained data privacy information from. Reputable news outlets (e.g., TV, radio, and print news) were the most reported source (N=13), followed by various online sources (N=11) and interpersonal sources (N=11) including tech-savvy friends and technical experts in the BLV communities. A third reported obtaining data privacy information from privacy policies, terms of use, or user agreements. Six participants realized the existence of data practices based on their empirical experiences, such as receiving targeted and various spam/scam emails, to rationalize that their personal data was collected or shared by companies.

Assessing information source credibility. Almost all participants reported certain criteria or preferences when assessing the credibility of information sources. Our analysis revealed that **trust** played a critical role in their credibility assessment process. 15 participants believed **their trusted entities** could provide relatively credible information, including reputable news outlets and BLV organizations. Nine participants placed trust in **the people** disseminating the information, including tech-savvy friends and influencers in the blind community. A third of the participants reported sophisticated credibility assessment strategies, such as cross-referencing multiple sources and seeking primary sources.

6.2 Privacy Policies as Information Source

Experience with privacy policies. Eight participants mentioned privacy policies in their responses without prompts, and all participants said they heard about privacy policies before after the prompt. 17 participants reported that they have read at least a few privacy policies before, often when signing up for a new service or receiving privacy updates from companies. Only two participants read privacy policies due to the perceived importance of the contents. This result and 6.1 together indicate that privacy policies are an underutilized source by BLV users.

Credibility and usability of privacy policies. 15 participants considered privacy policies a credible source for data privacy information because they are the official disclosure of companies' data practices. A few pointed out caveats with its credibility because privacy policies "may change without notice" and "do not prevent security problems". The remaining six participants thought privacy policies were not credible, because they "lack accountability" and credibility can "vary by the companies". 11 participants described difficulty reading privacy policies due to their length and vague languages, as P19 elaborated: *"It's a lengthy legal document so it's not like an exciting read to begin with...I feel that they're very vague...just very open to interpretation."*

6.3 Seeking and Non-seeking

13 participants reported that they sought information regarding the data practices of digital technologies and the remaining eight participants did not, as detailed below. **Seeking data privacy information** Among 13 participants, eight actively looked for data privacy information multiple times, while five only mentioned one or two examples and admitted that rarely sought data privacy information unless they had a concern. 10 out of the 13 participants successfully found the data privacy information they were looking for, including clarification about data practices, confirmation of data practices mentioned elsewhere, available privacy controls, and detailed information of known data breaches. To our surprise, only three participants mentioned that their search outcome were less than satisfactory, citing challenges in finding relevant information (P11), exercising privacy controls (P12), and understanding privacy policies (P30). Regarding information-seeking strategies, Google searches (N=6) and reputable news outlets (N=6) were the most commonly mentioned. Three participants reported using non-Google search engines for "more neutral results". A few participants also consulted privacy policies or terms of use (N=3), expert opinion (N=2), and trusted persons

(N=2). In summary, participants who sought information on data privacy generally succeeded in their search.

Not seeking data privacy information. Among the eight participants who did not seek data privacy information, four reported that privacy was not a major concern to motivate their information seeking. However, the lack of concern may derive from certain **misconceptions around data privacy**, as one marginalized participant said: *"(I didn't seek information) because it doesn't affect me. I don't have any information that is that important, that I would be upset about them collecting the information....I mean they can do anything with the little bit of money I got in the bank; I don't think it's important to them. I think I'm so glad that I'm down on the socioeconomic totem pole and that my information is not that important to them. But I think there are people whose information is."* This data suggests misconceptions may lead to an inaccurate assessment of data privacy risks and a failure to establish the appropriate level of concern. Two participants thought it was not necessary. For example, P29, who has high technology literacy, felt that "not much has changed in the data privacy landscape." Another participant admitted that "fear" prevented her from seeking such information, as she explained: *"Information, mainly because it's a scary topic, like the more I know the less I want to be on the web."* Only one participant admitted that they did not seek such information due to the perceived difficulty because "it's time-consuming and difficult" (P31). In summary, **the lack of data privacy concern** reduced the necessity for participants to seek data privacy information, which is consistent with findings in 5.1 that data privacy risks were not participants' primary concern.

7 Expectations for Privacy Q&A Tools

To mitigate participant response bias, we avoided priming participants with the idea of privacy Q&A assistants by phrasing it as a hypothetical "digital assitant" (detailed in Section 3).

7.1 Expectations without Priming

19 participants described at least one expectation for the privacy Q&A assistant without priming, as detailed below.

Good functionality. Most participants expected good Q&A functionality, which means the assistant should provide high-quality (N=9) and up-to-date (N=3) answers to their privacy questions in plain language (N6), as P15 described: *"It would give answers to the questions we ask...in straight-out answers, it's not what we think it's [the data] going to be used for...No, it's gonna be, not assumptions, just cold hard facts."*

Accessibility by default. Two participants explicitly expected the assistant to be accessible for BLV people. The other participants' follow-up comments confirmed that they implicitly assumed the assistant would be accessible by default.

Advanced features. Many participants elaborated on several advanced features, such as providing "links to references" in support of the answers(N=2), providing "a general overview" (N=2), and incorporating a mechanism to verify data privacy information. Seven participants expected other advanced features, including attractive accent (assuming the assistant is voice-based), the capability to naturally interact with users, and personalized reminders of potentially risky data practices.

Three prefer privacy experts over digital assistants. P20 commented on the poor accessibility of other digital assistants: *"I prefer it to be a human assistant because digital assistants can only give you answers that it's programmed to give. When you have something like Aira involved, it's person to person contact. Digital assistant is not enough to answer the concerns of a blind consumer."* Our analysis revealed that all three participants had negative user experiences with digital assistants. For example, P17 vividly shared a frustrating experience with us and commented: *"I'm done dealing with digital assistants on several different websites."* This data indicates that prior negative experiences could impede BLV users' acceptance of digital assistants in the future.

7.2 Expectations for Privacy Q&A Assistants

Cross-platform and multi-modality. While six participants preferred the privacy Q&A assistant to be available on one device type (i.e, smartphones, computers), 15 participants hoped the privacy Q&A assistant would be cross-platform and cross-device, as P13 commented: *"probably all of them [the devices]...in whatever form or shape... but I think the assistant should be on all devices."*

Regarding the interaction modality of the privacy Q&A assistant, 15 participants want it to support both textual and auditory Q&A experiences. P26 explained his understanding of accessibility: *"I've always been interested in accessibility for the most people, so I would say both. By voice for people that are interested in something like that...but maybe a lot of people are deaf and unable to speak, so to have an alternative method like being able to type would definitely bring accessibility up a lot, and maybe even getting a response back in text for somebody that cannot hear."*

Interesting, three participants strongly preferred the non-verbal interaction modality due to the limited accuracy of dictation on their devices, as P28 explained: *"I'm more comfortable doing research by typing. I haven't had much success by voice searching on any device, I always prefer to type...[For answers, I prefer] text format that can be accessed by anyone, but with me, it would be the screen reader."* Notably, two participants pointed out **modality consistency** during their Q&A interaction – they would prefer to receive answers in the same modality as they asked questions.

Preferred information sources. Participants reported their preferred information sources from which the privacy Q&A assistant should gather information. 13 participants preferred information from the first-party companies, about which the privacy questions were asked. 11 participants wanted infor-

mation from the official legal documents including privacy policies and terms of use. Many participants also mentioned reputable news outlets (N=6), other online sources (N=7), and organizations like privacy watch groups (N=6), while a few wanted the assistant to also look for actual regulations (N=2) as well as ratings of companies' privacy practices (N=2). Notably, four participants would like the assistant to employ some verification mechanisms to "constantly vet sources" and "compare what the company's saying to actually what has happened." This indicates participants valued both credible information sources and mechanisms (e.g., [47, 84] to verify the sources.

Preferred developers. 16 participants would trust **nonprofit organizations** to develop the assistant because they are "neutral' and "have guiding purposes". Some cited Consumer Reports, the World Wide Web Consortium (W3C), and research universities as examples. 10 participants would trust **third-party companies** dedicated to developing the privacy Q&A assistant. Participants believed they had the technical capability but "no vested interest" in the data practices being questioned. Similarly, two participants preferred collaboration between two types of entities, as P31 said: *"Either a nonprofit and a third party working together, one with the tech and one with the ethics, or it could be a governmental entity working with a nonprofit."* Interestingly, two participants strongly preferred first-party companies (e.g., big tech companies). P19 specified that she would trust Apple to develop the assistant because she valued Apple's product quality and customer support, even if the company may not be completely neutral.

Participants also discussed the entities they would not trust to develop the privacy Q&A assistant. 12 participants distrusted **first-party companies** that also engage in data practices being questioned, citing that the assistant may not be neutral because the companies have vested interest in the data practices. **Governments** (N=10) were also unpopular due to participants' personal political opinions or their belief that governments cannot deliver good technology products.

7.3 Benefits, Concerns, and Use Cases

Assuming the privacy Q&A assistant was developed by their preferred developers, participants elaborated on the benefits, concerns, and use cases for the privacy Q&A assistant.

Benefits. 14 participants identified **easy access to trusted data privacy information** as a benefit. P13 explained: *"I don't have to be scurrying, like a little kitty cat over the Internet, to try to find this information as much. The assistant can read the boring information and then give me a snapshot of what it has found."* Three participants mentioned that using the assistant could increase their awareness of data practices, said P24: *" [Having] readily available, accurate information that was generated or produced by a group of people who have the knowledge...I, as a consumer, will be better informed about how my information is collected and shared. "* Some participants mentioned **benefits tied to their preferred developers**, including accessing neutral data privacy information (third-party companies or nonprofit organizations as developers) and enjoying highly-quality products (trusted first-party companies as developers). A few participants mentioned that the assistant can increase their confidence when dealing with businesses or making privacy decisions online. These results indicate that a privacy Q&A assistant would benefit BLV users in many ways, regardless of whether they sought data privacy information or not.

Concerns. Even with their preferred developers, participants still expressed concerns about the privacy Q&A assistant. Eight participants identified **the data practices of the assistant itself as a concern**, explained P24: *"The digital assistant is going to function like Siri or the smart speaker. Their constant availability accepting information via voice...because they're able to access more information and then that there might open up the channels for other people to have access to the information."* Note that our emphasis on data privacy in prior interview questions may have priming effect for this concern. A few participants worried about the **long-term neutrality** of the privacy Q&A assistant and the **quality of the answers** provided, said P14: *"There would definitely have to be checks and balances so that the digital assistant wouldn't morph into something that could be used against the purpose of its existence."* Only one participant explicitly voiced accessibility concerns from a blind user's perspective.

Use cases. 14 participants reported that they would most likely ask questions about data privacy **before starting to use a new or unfamiliar digital technologies**, including *"at the time of installing the app (P12)"* and *"when connecting with a merchant that I've not used before (P14)"*. Many participants said that they would like to use the privacy Q&A assistant to check the digital technologies they currently use (N=4), obtain proof for companies' data practices (N=4), and when there are updates to companies' data practices (N=3). Two participants said they would only use it **when there is a privacy concern**. Notably, two participants expressed strong enthusiasm towards using the privacy Q&A assistant, as P17 commented: *"[I would use it in] any place I have to put in something beyond my name. if I've got to put in my date of birth, my social security number, or any financial information linking any financial accounts to a certain company, I would definitely use that digital assistant to make sure that I know exactly what's happening with the data that I'm inputting."*

8 Discussion and Implications

8.1 Study Limitations

We acknowledge several limitations of this study. First, a purely self-reported interview study has its inherent methodological limitations. To increase data validity, we employed techniques [23] to help participants recall their experiences

and refined interview questions to evoke truthful answers.

Second, the qualitative interviewing method also restricted our sample. Though our sample size is on par or larger than prior security and privacy qualitative studies with BLV users, the findings may not generalize to all BLV users. Particularly, the US sample in this study cannot reflect the data privacy experience of BLV users worldwide, who may need to navigate more complicated consent procedures, such as the cookie banners mandated by the General Data Protection Regulation [15]. Nevertheless, compared to prior studies with BLV participants [6, 17, 34, 44, 56], we recruited a more diverse sample across age, gender, and race/ethnicity, which yielded more inclusive perspectives and uncovered disparities within this underrepresented user group.

Third, due to the study focus, our interview questions for RQ3 cannot escape the participant response bias towards NLP-based privacy Q&A assistants. This means our findings did not investigate other approaches that can also assist BLV users in navigating digital data privacy. To mitigate this researcher-induced bias, we did not explicitly mention "privacy Q&A assistant" or any NLP techniques in the interviews and carefully crafted our interview questions to focus on participants' needs and expectations for privacy question answering capabilities, as detailed in 3.3.

8.2 Mitigating Risks for BLV Users

Heightened risks for BLV users. Besides confirming prior research that BLV users face high security and privacy (S&P) risks when using digital technologies [6, 7, 52], this study revealed several **previously unstudied privacy risks**. Regarding VI services, we identify the privacy risks associated with VI companies' data practices and a new type of physical-digital risks caused by the VI agents' environment (5.1). BLV users are vulnerable to S&P risks because they de-prioritize S&P concerns when having to overcome accessibility challenges around digital technologies [52]. Our results further suggest participants' low levels of privacy concern (4.3) and certain misconceptions (5.1) may further expose them to heightened S&P risks.

BLV users' risk mitigation is insufficient. Our study is the first to articulate BLV users' risk mitigation strategies and behaviors. Our participants primarily mitigate perceived risks by adjusting how they use digital technologies (5.2). This is unique for assistive technology, including VI services, that some participants heavily relied on in their daily lives. With limited alternatives, BLV users *had to weigh between the essential utilities and risk mitigation*. We also found participants' risk perceptions affect their risk mitigation behaviors. For example, most participants considered AI-based visual-aid apps low-risk (5.1) because there seemed to be no human involvement. Such perception fails to consider S&P risks associated with these apps' data practices and their underlying algorithms. Our sample is skewed towards educated participants but their current risk mitigation strategies do not prevent them from many S&P risks.

Implications for mitigating S&P risks. Our findings indicate the importance of providing accessible information to BLV users to enable accurate assessment of S&P risks. This highlights the need for tools like privacy Q&A assistants. Specifically, compared to sighted users, BLV users tend to mitigate risks through behaviors, presenting both opportunities and challenges for S&P tools. The opportunities lie in the potential impact of significantly improving their S&P if they adopt effective tools. However, our results also identify several challenges: S&P tools should consider the unique risks faced by BLV users and their accessibility needs; these tools should be seamlessly integrated into their existing risk mitigation strategies without increased user burden.

8.3 Creating Inclusive Privacy Tools

This is the first study investigating BLV users' information behavior around data privacy information, which yielded novel findings on how they seek or do not seek such information (6.3) and how they assess information sources (6.1).

Generally successful information seeking outcomes. Different from prior research [51, 76], our participants who sought data privacy information were mostly successful and **reported little difficulty** during the process. Similarly, the main reason for not seeking such information was a lack of need or concern. However, it is impossible to know if those who claimed no need or concern fully understand the privacy risks or if they would feel the same way if they did. Regarding data privacy information sources, besides news outlets and privacy policies, interpersonal and BLV-specific channels are particularly important for BLV users.

Similar challenges but more patience with privacy policies. Our findings revealed how BLV participant felt and interacted with privacy policies (6.2), which is a major source for NLP-based privacy tools [33, 43, 61]. We found that BLV participants **face similar challenges as sighted users** in comprehending privacy policies [55, 63, 72] but generally consider privacy policies a credible source. Surprisingly different from sighted users, many participants **had the patience** to read through multiple privacy policies.

Implications for creating inclusive privacy tools. Our findings show that BLV users' information behaviors around data privacy exhibit both similarities and differences when compared to sighted users. Particularly, we did not find prevailing evidence that BLV users face more challenges when seeking data privacy information beyond accessibility considerations. This cautions researchers about inaccurate assumptions that they might bring into their research or tool development. Essentially, privacy tools should account for BLV users' information needs and align with their existing information behaviors. For instance, a search tool may not benefit those who seldom seek data privacy information but NLP-based privacy tools

capable of summarizing key information from privacy policies [39, 43, 78] can alleviate the usability frustrations for both sighted and BLV users. This calls for more research to understand underrepresented user groups' privacy perceptions and behaviors to design more inclusive privacy tools.

8.4 Strengthening Trust with BLV Users

Trust is a meta-theme that impacts BLV participants' technology selection and their approach to data privacy information.

Trust in entities and people. Participants' answers to the only trust-focused interview question (7.2) suggest that BLV users' trust in developers may impact their decision to adopt S&P tools. Our results showed that BLV users **distrust** big technology companies and governments to create transparent privacy tools. In contrast, they tend to trust companies that provide assistive technology and products with outstanding accessibility features (5.2 & 7.2). They also trust organizations and experts of the BLV community as well reliable friends and family members, because they believe these organizations and people have their best interests in mind.

Trust in risk and credibility assessment. BLV participants considered screen readers and AI-based visual-aid apps low-risk because these digital technologies do not directly expose data to strangers (human agents). Notably, our participants seemed to trust AI-based tools and did not express concerns about relying on these AI-based tools, potentially leading to overlooked AI-induced S&P risks. Additionally, in line with [34], we find that BLV participants' existing trust also impacts how they assess the credibility of data privacy information. The participants who did not use sophisticated credibility assessment criteria tended to believe information mostly from the people or channels they already trust(6.1).

Implications for building trust with BLV users. Trust is integral to BLV users' interactions with digital technologies, assessment of information credibility, and evaluation of S&P risks. S&P tools should establish trust with BLV users, where ensuring S&P tools' accessibility is an essential starting point. Collaborating with BLV organizations can further strengthen trust with BLV users. Moreover, under the backdrop that popular generative AI applications like ChatGPT often provide false answers [66], it is particularly crucial for tools like privacy Q&A assistants to build and strengthen trust with BLV and other marginalized users.

8.5 Expectations for Privacy Q&A Assistants

This is the first study exploring BLV participants' expectations for potential privacy Q&A assistants, enriching the growing research to design user-centered privacy assistants [12, 65, 68] with a focus on accessibility and the need of BLV users. While our participants may prioritize data security over data privacy(5.1), they clearly recognized the potential benefits of privacy Q&A assistants and expressed interest in them(7.3).

Rethinking accessibility. Our BLV participants expected the assistant to be **accessible by default**, meaning they could access the assistant across devices and platforms and via different modalities (7.2). Different from prior research showing that BLV users found voice-based apps particularly useful [56], our participants preferred interacting with the assistant in **multiple modalities**. Some also favored **textual input and output** so that they can ask more precise questions and obtain more comprehensive answers. This highlights the importance to extend existing NLP Q&A evaluations [21,30,62] to account for inputs from different modalities like speech.

The importance of quality and trustworthiness. Most participants implicitly assumed that the privacy Q&A assistant could provide high-quality answers to their privacy questions (7.1). Some also imagined advanced features to enhance answer credibility, including the ability to cross-reference sources and to verify data practices objectively. In addition, many participants generally emphasized the importance of having trustworthy assistants with their best interest in mind when providing answers. Given that three participants still prefer human experts due to prior negative experience with online chatbots, the assistants would need to provide high-quality and trustworthy answers to swing their perspectives.

Implications for privacy Q&A assistants. These findings underscore the significance of default accessibility for privacy Q&A assistants and broader S&P tools. To achieve optimal accessibility, developers should determine the best modality combination based on different user groups' accessibility needs [36], which typically supports multiple modalities. Moreover, BLV users' expectations for functionality drives their interest in privacy Q&A assistants. To meet such expectations, potential areas of research include applying NLP techniques to consolidate relevant information from both privacy policies and other credible sources (e.g., privacy regulations and news), and advancing verification mechanisms to validate the data practices disclosed in privacy policies.

8.6 Towards Equitable Privacy

Digital divide. The digital divide generally refers to unequal access to digital technology among populations. This was seen in several unemployed or racially minoritized participants. For example, one participant never learned to use computers due to the lack of affordable screen reader training (4.2) and another had to use the free but relatively risky VI services Be My Eyes due to financial considerations (5.2). We also observed that marginalized BLV participants in our sample heavily relied on mobile phones to access digital technologies. This directly increases the S&P risks faced by marginalized BLV users and limits their risk mitigation options.

Technology literacy. Our results reflected the inequality caused by disparities in technology literacy. A few participants enjoyed the security provided by paid privacy-enhancing technology like VPNs (4.2), while one marginal-

ized participant struggled with entering passwords on their iPhone that prevented them from using mobile banking(5.2).

Implications for S&P research. Our inclusive sample yielded perspectives on equity and inclusion even among BLV users. While technology alone cannot resolve the disparity in security and privacy resulting from societal factors, it is essential for the S&P community to acknowledge the presence of such inequality. We should actively work towards developing accessible, inclusive, and equitable S&P tools that can benefit the widest possible spectrum of users.

9 Conclusion

We presented an in-depth interview study aimed at understanding BLV users' data privacy risk perception and mitigation strategies, their information behaviors related to data privacy, and their expectations for privacy Q&A tools that could assist them in navigating data privacy information. This study yielded rich findings and implications around usability, accessibility, trust, equity, and S&P risk mitigation for BLV users and beyond. We want to conclude this paper with one participant's quote that uncovers the core value of accessibility:

> *"I know the survey [interview] is supposed to be for the blind...but really what you're talking about to me is a survey [interview] everybody should take...because I think what you're talking about [the privacy Q&A assistant]...will benefit everybody, and you know a lot of the cases in the community of people with disabilities, we try to find things that are going to benefit the whole populace."*

Acknowledgments

We want to thank our blind consultants Dr. Cynthia Bennett and Chancey Fleet for their invaluable advice and the National Federation of the Blind for their help with participant recruitment. This study is funded by the US National Science Foundation under the project "Automatically Answering People's Privacy Questions" (CNS-1914444 & CNS-1914486). The US government is authorized to reproduce and distribute reprints for government purposes not withstanding any copyright notices thereon. The views and conclusions contained herein are those of the authors and should not be interpreted as representing official policies or endorsements, either expressed or implied by NSF or the US government.

References

[1] Ali Abdolrahmani and Ravi Kuber. Should I trust it when I cannot see it? credibility assessment for blind web users. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '16, pages 191–199, 2016.

[2] Patricia Acosta-Vargas, Belén Salvador-Acosta, Luis Salvador-Ullauri, William Villegas-Ch., and Mario Gonzalez. Accessibility in native mobile applications for users with disabilities: A scoping review. *Applied Sciences*, 11(12):5707, 2021.

[3] Patricia Acosta-Vargas, Luis Salvador-Ullauri, Janio Jadán-Guerrero, César Guevara, Sandra Sanchez-Gordon, Tania Calle-Jimenez, Patricio Lara-Alvarez, Ana Medina, and Isabel L. Nunes. Accessibility assessment in mobile applications for Android. In *Proceedings of the 2019 International Conference on Applied Human Factors and Ergonomics*, AHFE '19, pages 279–288, 2019.

[4] Alessandro Acquisti and Jens Grossklags. Privacy and rationality in individual decision making. *IEEE security & privacy*, 3(1):26–33, 2005.

[5] Wasi Ahmad, Jianfeng Chi, Yuan Tian, and Kai-Wei Chang. PolicyQA: A reading comprehension dataset for privacy policies. In *Findings of the Association for Computational Linguistics: EMNLP 2020*, pages 743–749, Online, nov 2020. Association for Computational Linguistics.

[6] Tousif Ahmed, Roberto Hoyle, Kay Connelly, David Crandall, and Apu Kapadia. Privacy concerns and behaviors of people with visual impairments. In *Proceedings of the 2015 CHI Conference on Human Factors in Computing Systems*, CHI '15, pages 3523–3532, 2015.

[7] Taslima Akter, Bryan Dosono, Tousif Ahmed, Apu Kapadia, and Bryan Semaan. "i am uncomfortable sharing what i can't see": Privacy concerns of the visually impaired with camera based assistive applications. In *Proceedings of the 29th USENIX Security Symposium*, SEC '20, pages 1929–1948, 2020.

[8] Waleed Ammar, Shomir Wilson, Norman Sadeh, and Noah A Smith. Automatic categorization of privacy policies: A pilot study. *Technical Report CMU-LTI-12-019, Carnegie Mellon University*, 12 2012.

[9] Marcia J. Bates. Information behavior. In *Encyclopedia of Library and Information Sciences*, pages 2074–2085. Taylor & Francis, 4th edition, 2017.

[10] Glynis M. Breakwell. Interviewing methods. In *Research methods in psychology*, pages 232–253. SAGE, 3rd edition, 2006.

[11] Victoria Clarke, Virginia Braun, and Nikki Hayfield. Thematic analysis. In *Qualitative psychology: A practical guide to research methods*, page 248. SAGE, 3rd edition, 2015.

[12] Jessica Colnago, Yuanyuan Feng, Tharangini Palanivel, Sarah Pearman, Megan Ung, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. Informing the design of a personalized privacy assistant for the internet of things. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–13, 2020.

[13] Elisa Costante, Jerry den Hartog, and Milan Petković. What websites know about you. In *Data Privacy Management and Autonomous Spontaneous Security*, pages 146–159. Springer, 2012.

[14] Elisa Costante, Yuanhao Sun, Milan Petković, and Jerry den Hartog. A machine learning solution to assess privacy policy completeness: (short paper). In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society*, WPES '12, page 91–96, New York, NY, USA, 2012. Association for Computing Machinery.

[15] Council of European Union. General data protection regulation. https://gdpr-infor.eu, 2016.

[16] Anupam Das, Martin Degeling, Daniel Smullen, and Norman Sadeh. Personalized privacy assistants for the internet of things: Providing users with notice and choice. *IEEE Pervasive Computing*, 17(3):35–46, 2018.

[17] Bryan Dosono, Jordan Hayes, and Yang Wang. "I'm stuck!": A contextual inquiry of people with visual impairments in authentication. In *11th USENIX Conference on Usable Privacy and Security*, SOUPS '15, pages 151–168, 2015.

[18] Bryan Dosono, Jordan Hayes, and Yang Wang. Toward accessible authentication: Learning from people with visual impairments. *IEEE Internet Computing*, 22(2):62–70, 2018.

[19] Nora A. Draper. From privacy pragmatist to privacy resigned: Challenging narratives of rational choice in digital privacy debates. *Policy & Internet*, 9(2):232–251, 2017.

[20] Pardis Emami-Naeini, Henry Dixon, Yuvraj Agarwal, and Lorrie Faith Cranor. Exploring how privacy and security factor into IoT device purchase behavior. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–12, 2019.

[21] Fahim Faisal, Sharlina Keshava, Md Mahfuz Ibn Alam, and Antonios Anastasopoulos. SD-QA: Spoken dialectal question answering for the real world. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, EMNLP '21, pages 3296–3315, 2021.

[22] Jennifer Fereday and Eimear Muir-Cochrane. Demonstrating rigor using thematic analysis: A hybrid approach of inductive and deductive coding and theme development. *International Journal of Qualitative Methods*, 5(1):80–92, 2006.

[23] John C. Flanagan. The critical incident technique. *Psychological Bulletin*, 51(4):327, 1954.

[24] Patricia I Fusch and Lawrence R Ness. Are we there yet? data saturation in qualitative research. *The Qualitative Report*, 20(9):1408–1416, 2015.

[25] Nina Gerber, Paul Gerber, and Melanie Volkamer. Explaining the privacy paradox: A systematic review of literature investigating privacy attitude and behavior. *Computers & security*, 77:226–261, 2018.

[26] Joshua Gluck, Florian Schaub, Amy Friedman, Hana Habib, Norman Sadeh, Lorrie Faith Cranor, and Yuvraj Agarwal. How short is too short? implications of length and framing on the effectiveness of privacy notices. In *12th Symposium on Usable Privacy and Security*, SOUPS '16, pages 321–340, 2016.

[27] Danna Gurari, Qing Li, Abigale J. Stangl, Anhong Guo, Chi Lin, Kristen Grauman, Jiebo Luo, and Jeffrey P. Bigham. Vizwiz grand challenge: Answering visual questions from blind people. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, CVPR '18, pages 3608–3617, 2018.

[28] Hana Habib, Sarah Pearman, Jiamin Wang, Yixin Zou, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. "it's a scavenger hunt": Usability of websites' opt-out and data deletion choices. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–12, 2020.

[29] Stephanie Hackett, Bambang Parmanto, and Xiaoming Zeng. Accessibility of Internet websites through time. In *Proceedings of the 6th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '03, pages 32–39, 2003.

[30] Dilek Hakkani-Tur and Manaal Faruqui. A call for revisiting the boundary between asr and nlu in the age of conversational dialog systems. *Computational Linguistics*, 48(1):221–232, 2022.

[31] Foad Hamidi, Kellie Poneres, Aaron Massey, and Amy Hurst. Who should have access to my pointing data? privacy tradeoffs of adaptive assistive technologies. In *Proceedings of the 20th International ACM SIGACCESS Conference on Computers and Accessibility*, ASSETS '18, pages 203–216, 2018.

[32] Hamza Harkous, Kassem Fawaz, Rémi Lebret, Florian Schaub, Kang G Shin, and Karl Aberer. Polisis: Automated analysis and presentation of privacy policies using deep learning. In *27th USENIX Security Symposium (USENIX Security 18)*, pages 531–548, 2018.

[33] Hamza Harkous, Kassem Fawaz, Kang G. Shin, and Karl Aberer. PriBots: Conversational privacy with chatbots. In *Proceedings of the Workshop on the Future of Privacy Indicators, at the 12th Symposium on Usable Privacy and Security*, pages 1–6, 2016.

[34] Jordan Hayes, Smirity Kaushik, Charlotte Emily Price, and Yang Wang. Cooperative privacy and security: Learning from people with visual impairments and their allies. In *15th USENIX Symposium on Usable Privacy and Security*, SOUPS '19, pages 1–20, 2019.

[35] Shawn Lawton Henry. WCAG 2 overview. Web Accessibility Initiative (WAI) https://www.w3.org/WAI/standards-guidelines/wcag/, 2005.

[36] Julia Himmelsbach, Markus Garschall, Sebastian Egger, Susanne Steffek, and Manfred Tscheligi. Enabling accessibility through multimodality? interaction modality choices of older adults. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, MUM '15, pages 195–199, 2015.

[37] Karen Holtzblatt and Hugh R. Beyer. Requirements gathering: the human factor. *Communications of the ACM*, 38(5):31–32, 1995.

[38] Ruogu Kang, Stephanie Brown, Laura Dabbish, and Sara Kiesler. Privacy attitudes of mechanical turk workers and the US. public. In *10th Symposium On Usable Privacy and Security*, SOUPS '14, pages 37–49, 2014.

[39] Moniba Keymanesh, Micha Elsner, and Srinivasan Parthasarathy. Toward domain-guided controllable summarization of privacy policies. In *Natural Legal Language Processing Workshop*. KDD, 2020.

[40] Hyun K. Kim, Sung H. Han, Jaehyun Park, and Joohwan Park. The interaction experiences of visually impaired people with assistive technology: A case study of smartphones. *International Journal of Industrial Ergonomics*, 55:22–33, 2016.

[41] Spyros Kokolakis. Privacy attitudes and privacy behaviour: A review of current research on the privacy paradox phenomenon. *Computers & security*, 64:122–134, 2017.

[42] Barbara Krumay and Jennifer Klar. Readability of privacy policies. In *Proceedings of the 34th IFIP Annual Conference on Data and Applications Security and Privacy*, DBSec '20, pages 388–399, 2020.

[43] Vinayshekhar Bannihatti Kumar, Roger Iyengar, Namita Nisal, Yuanyuan Feng, Hana Habib, Peter Story, Sushain Cherivirala, Margaret Hagan, Lorrie Cranor, Shomir Wilson, Florian Schaub, and Norman Sadeh. Finding a choice in a haystack: Automatic extraction of opt-out statements from privacy policy text. In *Proceedings of The Web Conference 2020*, WWW '20, pages 1943–1954, 2020.

[44] Elaine Lau and Zachary Peterson. A research framework and initial study of browser security for the visually impaired. In *11th USENIX Symposium on Usable Privacy and Security*, SOUPS '15, pages 1–18, 2015.

[45] Bin Liu, Mads Schaarup Andersen, Florian Schaub, Hazim Almuhimedi, Shikun Aerin Zhang, Norman Sadeh, Yuvraj Agarwal, and Alessandro Acquisti. Follow my recommendations: A personalized privacy assistant for mobile app permissions. In *12th Symposium on Usable Privacy and Security*, SOUPS '16, pages 27–41, 2016.

[46] Fei Liu, Rohan Ramanath, Norman Sadeh, and Noah A. Smith. A step towards usable privacy policy: Automatic alignment of privacy statements. In *Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: Technical Papers*, pages 884–894, Dublin, Ireland, August 2014. Dublin City University and Association for Computational Linguistics.

[47] Shuang Liu, Baiyang Zhao, Renjie Guo, Guozhu Meng, Fan Zhang, and Meishan Zhang. Have you been properly notified? automatic compliance analysis of privacy policy text with gdpr article 13. In *Proceedings of the Web Conference 2021*, pages 2154–2164, 2021.

[48] Aleecia M. McDonald and Lorrie Faith Cranor. The cost of reading privacy policies. *I/S: A Journal of Law and Policy for the Information Society*, 4:543–568, 2008.

[49] Nora McDonald, Sarita Schoenebeck, and AndreaForte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for cscw and hci practice. *Proceedings of the ACM on Human-Computer Interaction*, 3(CSCW):1–23, 2019.

[50] Gabriele Meiselwitz. Readability assessment of policies and procedures of social networking sites. In *Proceedings of the 2013 International Conference on Online Communities and Social Computing*, OCSC '13, pages 67–75, 2013.

[51] Stephen Mutula and Rebecca M. Majinge. Information behaviour of students living with visual impairments in university libraries: A review of related literature. *The Journal of Academic Librarianship*, 42(5):522–528, 2016.

[52] Daniela Napoli, Khadija Baig, Sana Maqsood, and Sonia Chiasson. "i'm literally just hoping this will work": Obstacles blocking the online security and privacy of users with visual disabilities. In *17th Symposium on Usable Privacy and Security*, SOUPS '21, pages 263–280, 2021.

[53] National Federation of the Blind. Blindness statistics. https://nfb.org/resources/blindness-statistics, 2019.

[54] Jonathan A. Obar and Anne Oeldorf-Hirsch. The biggest lie on the internet: Ignoring the privacy policies and terms of service policies of social networking services. *Information, Communication & Society*, 23(1):128–147, 2020.

[55] Anne Oeldorf-Hirsch and Jonathan A. Obar. Overwhelming, important, irrelevant: Terms of service and privacy policy reading among older adults. In *Proceedings of the 10th International Conference on Social Media and Society*, SMSociety '19, pages 166–173, 2019.

[56] Alisha Pradhan, Kanika Mehta, and Leah Findlater. "accessibility came by accident": Use of voice-controlled intelligent personal assistants by people with disabilities. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pages 1–13, 2018.

[57] Rohan Ramanath, Fei Liu, Norman Sadeh, and Noah A. Smith. Unsupervised alignment of privacy policies using hidden markov models. In *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*, pages 605–610, Baltimore, Maryland, June 2014. Association for Computational Linguistics.

[58] Ashwini Rao, Florian Schaub, Norman Sadeh, Alessandro Acquisti, and Ruogu Kang. Expecting the unexpected: Understanding mismatched privacy expectations online. In *12th Symposium on Usable Privacy and Security*, SOUPS '16, pages 77–96, 2016.

[59] Abhilasha Ravichander, Alan W. Black, Thomas Norton, Shomir Wilson, and Norman Sadeh. Breaking down walls of text: How can NLP benefit consumer privacy? In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing*, ACL-IJCNLP '21, pages 4125–4140, 2021.

[60] Abhilasha Ravichander, Alan W Black, Shomir Wilson, Thomas Norton, and Norman Sadeh. Question answering for privacy policies: Combining computational and legal perspectives. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 4947–4958, Hong Kong, China, nov 2019. Association for Computational Linguistics.

[61] Abhilasha Ravichander, Alan W. Black, Shomir Wilson, Thomas Norton, and Norman Sadeh. Question answering for privacy policies: Combining computational and legal perspectives. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing*, EMNLP-IJCNLP '19, pages 4947–4958, 2019.

[62] Abhilasha Ravichander, Siddharth Dalmia, Maria Ryskina, Florian Metze, Eduard Hovy, and Alan W. Black. NoiseQA: Challenge set evaluation for user-centric question answering. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics*, EACL '21, pages 2976–2992, 2021.

[63] Joel R Reidenberg, Travis Breaux, Lorrie Faith Cranor, Brian French, Amanda Grannis, James T Graves, Fei Liu, Aleecia McDonald, Thomas B Norton, Rohan Ramanath, N. Cameron Russell, Norman Sadeh, and Florian Schaub. Disagreeable privacy policies: Mismatches between meaning and users' understanding. *Berkeley Technology Law Journal*, 30:39–68, 2015.

[64] Antti Salovaara, Antti Oulasvirta, and Giulio Jacucci. Evaluation of prototypes and the problem of possible futures. In *Proceedings of the 2017 CHI conference on human factors in computing systems*, pages 2064–2077, 2017.

[65] William Seymour, Martin J. Kraemer, Reuben Binns, and Max Van Kleek. Informing the design of privacy-empowering tools for the connected home. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, pages 1–14, 2020.

[66] Chris Stokel-Walker and Richard Van Noorden. What ChatGPT and generative AI mean for science. *Nature*, 614(7947):214–216, 2023.

[67] Peter Story, Daniel Smullen, Yaxing Yao, Alessandro Acquisti, Lorrie Faith Cranor, Norman Sadeh, and Florian Schaub. Awareness, adoption, and misconceptions of web privacy tools. *Proceedings on Privacy Enhancing Technologies Symposium 2021*, pages 1–26, 2021.

[68] Alina Stöver, Sara Hahn, Felix Kretschmer, and Nina Gerber. Investigating how users imagine their personal privacy assistant. In *Proceedings on Privacy Enhancing Technologies Symposium 2023*, PETS '23, pages 384–402, 2023.

[69] Soundarya Nurani Sundareswara, Shomir Wilson, Mukund Srinath, and C. Lee Giles. Privacy not found: a study of the availability of privacy policies on the web. In *16th Symposium on Usable Privacy and Security*, SOUPS '20, pages 1–5, 2020.

[70] Shannon M. Tomlinson. Perceptions of accessibility and usability by blind or visually impaired persons: a pilot study. *Proceedings of the Association for Information Science and Technology*, 53(1):1–4, 2016.

[71] Noriko Tomuro, Steven Lytinen, and Kurt Hornsburg. Automatic summarization of privacy policies using ensemble learning. In *Proceedings of the Sixth ACM Conference on Data and Application Security and Privacy*, CODASPY '16, page 133–135, New York, NY, USA, 2016. Association for Computing Machinery.

[72] Matthew W. Vail, Julia B. Earp, and Annie I. Antón. An empirical study of consumer perceptions and comprehension of web site privacy policies. *IEEE Transactions on Engineering Management*, 55(3):442–454, 2008.

[73] Maike Vollstedt and Sebastian Rezat. An introduction to grounded theory with a special focus on axial coding and the coding paradigm. *Compendium for early career researchers in mathematics education*, 13(1):81–100, 2019.

[74] Junjue Wang, Brandon Amos, Anupam Das, Padmanabhan Pillai, Norman Sadeh, and Mahadev Satyanarayanan. Enabling live video analytics with a scalable and privacy-aware framework. *ACM Transactions on Multimedia Computing, Communications, and Applications*, 14(3s):1–24, 2018.

[75] Shomir Wilson, Florian Schaub, Aswarth Abhilash Dara, Frederick Liu, Sushain Cherivirala, Pedro Giovanni Leon, Mads Schaarup Andersen, Sebastian Zimmeck, Kanthashree Mysore Sathyendra, N Cameron Russell, et al. The creation and analysis of a website privacy policy corpus. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 1330–1340, 2016.

[76] Iris Xie, Shengang Wang, and Meredith Saba. Studies on blind and visually impaired users in LIS literature: A review of research methods. *Library & Information Science Research*, 43(3):101109, 2021.

[77] Yaxing Yao, Davide Lo Re, and Yang Wang. Folk models of online behavioral advertising. In *Proceedings of the 2017 ACM Conference on Computer Supported Cooperative Work and Social Computing*, pages 1957–1969, 2017.

[78] Razieh Nokhbeh Zaeem, Safa Anya, Alex Issa, Jake Nimergood, Isabelle Rogers, Vinay Shah, Ayush Srivastava, and K Suzanne Barber. Privacycheck v2: A tool that recaps privacy policies for you. In *29th ACM International Conference on Information and Knowledge Management (CIKM). ACM. To appear*, 2020.

[79] Razieh Nokhbeh Zaeem, Rachel L German, and K Suzanne Barber. Privacycheck: Automatic summarization of privacy policies using data mining. *ACM Transactions on Internet Technology (TOIT)*, 18(4):1–18, 2018.

[80] Shikun Zhang, Yuanyuan Feng, Yaxing Yao, Lorrie Faith Cranor, and Norman Sadeh. How usable are iOS app privacy labels? *Proceedings on Privacy Enhancing Technologies Symposium 2022*, pages 204–228, 2022.

[81] Yuhang Zhao, Edward Cutrell, Christian Holz, Meredith Ringel Morris, Eyal Ofek, and Andrew D. Wilson. SeeingVR: A set of tools to make virtual reality more accessible to people with low vision. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, CHI '19, pages 1–14, 2019.

[82] Yuhang Zhao, Sarit Szpiro, Jonathan Knighten, and Shiri Azenkot. CueSee: exploring visual cues for people with low vision to facilitate a visual search task. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp '16, pages 73–84, 2016.

[83] Yuhang Zhao, Yaxing Yao, Jiaru Fu, and Nihan Zhou. "if sighted people know, i should be able to know:" privacy perceptions of bystanders with visual impairments around camera-based technology. In *32nd USENIX Security Symposium (USENIX Security 23)*, pages 4661–4678, 2023.

[84] Sebastian Zimmeck, Peter Story, Daniel Smullen, Abhilasha Ravichander, Ziqi Wang, Joel R. Reidenberg, N. Russell, and N. Sadeh. Maps: Scaling privacy compliance analysis to a million apps. *Proceedings on Privacy Enhancing Technologies*, 2019:66 – 86, 2019.

A Interview Questions

We list all key interview questions and describe some (but not all) follow-up probing questions due to the page limit.

Baseline questions: (1) What electronic devices do you typically use to access digital information? (2) What assistive tools or technologies and other websites/apps/services do you use on this device? (for each type of device mentioned) (3) Do you know what types of personal data is collected when you

use these digital technologies on your devices? (4) Considering the extend of data practices discussed just now, what are your general thoughts about data privacy in regarding digital technologies?

RQ1: Risks perception and mitigation: (1) Please think about the digital tools/technologies that you use, can you think of some of the potential risks around your personal data privacy? (Follow-up questions about comparing risks between general technology and assistive technology). (2) *[Critical incident]* Please think about the past month when you used digital tools/technologies, were there any situations that you felt surprised, uncomfortable, or suspicious about how these tools/technologies use your personal data? (3) Have you ever stopped using a tool/technology or switched to an alternative tool/technology out of data privacy concerns? (4) *(Hypothetical dilemma question)* If you find out an assistive technology that you use handles your data in a way that you feel uncomfortable with, would it affect your willingness to use the technology?

RQ2: Information (seeking) behaviors: (1) How did you typically come across or get information about data privacy in the digital world? (follow-up questions: where, how, information sources , and perceived credibility of information sources) (2) Have you ever tried to find any information about data privacy? (If yes, follow up with where, how, and information sources; If not, ask why) (3) *(When participants did not mention privacy policies as a source)* Are you familiar with privacy policies? (follow up on perceived credibility for privacy policies)

RQ3: Expectations for Q&A tools: (1) You mentioned that you used [X]. Please imagine if there is an expert who can answer any questions around data privacy for [X]. What kinds of questions would you ask this expert about [X]? ([X] is a digital tool/technology mentioned by participants; up to 4 tools/technologies of different categories are asked here) — *Note: the data from this set of questions is not reported due to the scope and focus of this paper.* (2) Please imagine if the expert we discussed above is a digital assistant that can provide you with information around data privacy, how would you like the privacy assistant to be? (First, we let participants freely describe their imagined digital assistant without priming; then we asked follow up questions on preferred devices, modality, sources, developers, benefits, concerns, and use cases.)

B Definitions Given to Participants

Digital technologies: During this interview, I will use the general term "digital technologies" to refer to all the websites, apps, web services, and assistive tools/technologies that you mentioned. Does it sound okay to you?

Data practices: As you may know, when you use these digital tools/technologies on your devices, your personal data is often collected, used, or shared by these tools/technologies.
Data privacy: Data privacy concerns the handling of personal data by different entities, such as if the handling is appropriate and if it's in compliance with laws. The handling of personal data includes a variety of practices, such as how data is collected, used, or stored, whether data collectors share or sell the data to others. The focus of this interview study is data privacy with digital technologies, or digital data privacy.

C Coded Data

We archive the coded data in this Open Science Framework repository: https://doi.org/10.17605/OSF.IO/K9FV6

D Participants' Demographic Distribution

Gender		Education	
Female	52.4%	Some college	38.1%
Male	47.6%	College degree	33.3%
		Graduate degree	28.6%
Race/Ethnicity		**Age Group**	
White	61.9%	18–29	9.5%
Black	19.0%	30–39	19.0%
Asian	9.5%	40–49	19.0%
Hispanic	4.8%	50–59	28.6%
Mixed	4.8%	60–69	14.3%
		70+	9.5%
Employment Status			
Fully employed			23.8%
Partially employed			14.3%
Self-employed			9.5%
Unemployed			28.6%
Retired			19.0%
Homemaker			4.8%