Sarah Pearman*, Ellie Young, and Lorrie Faith Cranor

User-friendly yet rarely read: A case study on the redesign of an online HIPAA authorization

Abstract: In this paper we describe the iterative evaluation and refinement of a consent flow for a chatbot being developed by a large U.S. health insurance company. This chatbot's use of a cloud service provider triggers a requirement for users to agree to a HIPAA authorization. We highlight remote usability study and online survey findings indicating that simplifying the interface and language of the consent flow can improve the user experience and help users who read the content understand how their data may be used. However, we observe that most users in our studies, even those using our improved consent flows, missed important information in the authorization until we asked them to review it again. We also show that many people are overconfident about the privacy and security of healthcare data and that many people believe HIPAA protects in far more contexts than it actually does. Given that our redesigns following best practices did not produce many meaningful improvements in informed consent, we argue for the need for research on alternate approaches to health data disclosures such as standardized disclosures; methods borrowed from clinical research contexts such as multimedia formats, quizzes, and conversational approaches; and automated privacy assistants.

Keywords: privacy, healthcare privacy, usable privacy, consent, HIPAA, notice and choice

DOI 10.56553/popets-2022-0086
Received 2021-11-30; revised 2022-03-15; accepted 2022-03-16.

1 Introduction

With the proliferation of apps and devices to help users harness and manage their own health data, consumers are encouraged to learn how their health information will be used before they disclose it to third parties who are not their healthcare providers [36]. However, privacy notices are notoriously long and time consuming to read [27], and it is unclear whether users are equipped to make informed choices. Some studies show that shortening and simplifying disclosures may help users understand the choices available [17, 21], while others have shown little effect from simplifying such disclosures [7].

In this paper we present a case study on the impact of simplifying, shortening, and clarifying authorizations related to the U.S. Health Insurance Portability and Accountability Act of 1996 (HIPAA). While we show that our iterative changes improve user experience and better inform those who take the time to read the authorization, it is not clear that our changes result in users being significantly more informed in the typical use case in which users rush through the consent process.

The HIPAA Privacy Rule gives patients rights over their health information and establishes boundaries for sharing health data in any form [35]. Certain "covered entities," including healthcare providers and health insurance companies, must comply with these rules. Before a covered entity discloses data in a way that is not explicitly allowed, the Privacy Rule requires that the entity get the patient's consent by presenting them with an "authorization" that meets specific requirements [39].

We present a case study of a consent flow for a chatbot being developed by a large U.S. health insurance company, referred to in this paper as "HealthCo." The chatbot's primary purpose is to answer customers' questions about their health insurance coverage. This chatbot uses a major cloud service provider in a manner that triggers a requirement for users to view and agree to a HIPAA authorization before using the chatbot [1]. In a three-phase interview and survey study, we tested four versions of the consent flow for this chatbot in order to explore whether changes rooted in our empirical observations and recommended best practices could meaningfully improve usability, user understanding, or user sentiment towards the chatbot.

We first conducted three sets of six remote interview sessions in which we observed each set of users interacting with a different prototype consent flow. Prototype 1 was the original version created by HealthCo, Prototype

***Corresponding Author: Sarah Pearman:** Carnegie Mellon University, E-mail: spearman@cmu.edu
Ellie Young: New College of Florida, E-mail: eleanor.young18@ncf.edu
Lorrie Faith Cranor: Carnegie Mellon University, E-mail: lorrie@cmu.edu

2 was a version that we simplified and modified to more clearly state data use details, and Prototype 3 was an even further simplified version. Qualitative results from the interviews suggested that Prototypes 2 and 3 could improve on Prototype 1 in usability and understanding.

In the second phase, we analyzed data from 761 participants who had been randomly assigned to use one of the three prototypes tested in the interview study. We asked questions about their experience, their understandings of what they had read, and whether they would be inclined to use such a chatbot if they encountered that consent information in real life. We then displayed the consent information again, asked participants to review it carefully, and asked them again about their understandings of and attitudes toward the chatbot. We found usability benefits to Prototype 3, and users could understand our redesigned prototypes better when asked to review documents carefully, but user understanding was poor across all prototypes on the first pass through the documents.

In the third phase, we surveyed an additional 456 participants to compare users' understandings and opinions of two consent flow prototypes, one of which omitted the word "HIPAA." We also asked questions to further explore what non-experts understood about HIPAA and Protected Health Information (PHI), finding that most respondents grossly overestimated the protections that HIPAA could provide.

In this paper we discuss multiple insights from this case study that may apply more broadly to creating consent flow interfaces and consent documents for tools that handle health data. First, we discuss ways in which user interface (UI) and wording simplifications can improve the user experience and help users understand what they are agreeing to. Second, we discuss disclosures that users are unlikely to understand after skimming a document while trying to get to their primary task quickly but which they can understand after reading more carefully. As some of these disclosures are likely to change users' decisions once they understand them, it is important to find ways to highlight main takeaways to users who are unlikely to read every word. A third major insight relates to users' inflated sense of confidence in the privacy and security of healthcare data and their lack of understanding of HIPAA, both of which suggest a need for extreme care in data use disclosures in healthcare contexts. Our findings help to inform the design of healthcare-related data sharing notices specifically, as well as provide insights on improving the effectiveness of privacy notices in a broader set of contexts.

2 Related work

A user who read all of the privacy notices for all of the products and services that they used could be expected to spend hundreds of hours per year on this task [27]. Unsurprisingly, less than a quarter of U.S. adults surveyed by Pew Research in 2019 reported that they always or often read privacy policies, and 36% reported *never* reading them [3]. Numerous researchers have explored methods of shortening and simplifying disclosures to help users efficiently grasp the most important details. Studies have suggested that approaches such as short-form policies and "privacy nutrition labels" can help users find and understand privacy information [17, 21]. In addition, a study on End User License Agreements (EULAs) found that users spent more time reviewing paraphrased EULAs than traditional long-form EULAs, and that paraphrased EULAs yielded more positive sentiments and higher comprehension [44]. However, in a study by Ben-Shahar and Chilton, policy simplifications had little to no effect on users' understanding or their consent choices. The authors of that study speculate that users may be entirely "numb" and "unmotivated to use privacy notices of any kind" [7].

While open questions remain, there is some consensus on best practices for notice and choice. Schaub et al. reviewed the literature on privacy notice design and [40] made a number of recommendations that could apply to a healthcare disclosure. We sought to apply several of Schaub et al.'s recommendations in our prototype redesigns in this study, including identifying and providing control over practices that might be unexpected, avoiding jargon, conducting user testing, and trying not to overwhelm users with multiple choices.

We have limited data on the effectiveness of these recommended practices for HIPAA authorizations, which may be an especially difficult space for users to make informed choices, both due to the subject matter knowledge required and because people may believe that not consenting will jeopardize their ability to receive healthcare.

In addition to length, technical concepts and vocabulary may also present challenges to informed consent for the average user. In a study of the adoption of secure communication tools, Abu-Salma et al. report on a number of incorrect mental models that might complicate informed consent in the context of healthcare data: for example, mental models that conflate "security," "privacy," and "safety," as well as models that confuse en-

cryption with the general concept of authentication [2]. Tang et al. also report that many users have incorrect understandings of vocabulary that is fundamental to the understanding of data use disclosures, such as "privacy policy," "anonymized," and "encryption." On the other hand, users may be more likely to understand terms such as "PII" and "personal information" [43].

Patients in the U.S. healthcare system may struggle to understand HIPAA's main tenets and the documents that providers are required to show them. Moore et al. argue that the HIPAA Privacy Rule "fails to give equal weight to individuals' reasonable expectations of privacy" and works to facilitate the flow of data among healthcare entities rather than to protect patient privacy. In examining HIPAA complaint records, they found that patient privacy complaints increased along with institutions' attempts to comply with HIPAA. They argue that this may be at least in part because HIPAA fails to capture patients' expectations of what healthcare privacy should be. They also argue that HIPAA's design was "reactive, not proactive," failing to anticipate the number of entities that might have access to health data in the modern digital age. Moore et al. made this argument in 2007, and the healthcare data landscape has expanded even further since then with the proliferation of digital health tools [30].

Pollio argues that the HIPAA Privacy Rule is problematic because it "relies on disclosure to individuals as the primary mechanism" for giving users control and simply requires that disclosures be written in "plain language," without explaining how to create documents that non-experts can understand. She also argues that HIPAA disclosures may overwhelm users while making too many assumptions about what users already understand of their rights under HIPAA [37].

An increasing amount of health data is being gathered in contexts that are not covered under HIPAA, with important healthcare and privacy implications. Internet companies and data brokers that are not under the purview of HIPAA control a great deal of health data from online sources such as health and fitness apps, search histories, and social media. Glenn and Monteith warn about privacy and security risks that result from the collection and storage of this data and argue that people still value medical privacy. They also describe research indicating that, as patients become increasingly worried about the privacy of their health data, they may make decisions that negatively affect healthcare outcomes such as not returning for needed followups or not seeking care at all [16]. Also, in a scoping review of literature on health-related conversational agents, May and Denecke wrote that the third-party technologies used to implement such chatbots can pose a number of privacy and security concerns for users, arguing that further research is needed in this space to ensure that these tools prioritize users' security and privacy [26].

3 Remote user study

We conducted a qualitative study in late 2020 and early 2021 to test the consent flow for a chatbot provided by a major health insurance company (referred to in our study as HealthCo). The study sessions, conducted over Zoom, consisted of user study tasks focused on usability assessment and semi-structured interviews exploring user understanding and sentiment. We iterated on the chatbot consent flow prototype over three rounds of study sessions, with six participants per round.

Our remote user study and subsequent surveys were approved by our Institutional Review Board. Participants were reminded to avoid disclosing private health information while responding to our questions.

3.1 Chatbot and HIPAA authorization

We provide here an overview of how HealthCo planned to implement the chatbot in real life, as well as highlights from the authorization text that appeared (with slight variations) in the three prototypes.

The chatbot would appear on HealthCo's website for customers who were signed into their accounts. It was intended to allow customers to get quick, personalized answers about their insurance coverage without calling a customer service phone line. HealthCo chose Google Cloud as the partner for this chatbot largely because it could provide conversational agent functionality via DialogFlow [9]. HealthCo would also send data about customers to Google Cloud storage to allow DialogFlow to personalize responses. Some of this data would be Personal Identifying Information (PII) or Personal Health Information (PHI). To ensure HIPAA compliance, HealthCo chose to implement an authorization step to notify customers about third party data sharing.

Once data was sent to Google Cloud, it was stored there long-term to allow for personalized responses to be retrieved quickly on the next website visit. Users could return to the website to revoke their consent for future data to be sent to Google Cloud but could not revoke, access, or amend data that had already been sent.

Google Cloud is not bound by HIPAA in this scenario. Prototype 1 alluded to this in its HIPAA authorization—e.g., "Healthco...must share this information with certain parties who are not subject to the same laws as HealthCo"—but did not make it explicit or prominent. Prototypes 2 and 3 featured this text at the start of the authorization: "I understand that Google Cloud is not subject to HIPAA or certain other healthcare information laws that HealthCo must follow."

While HIPAA did not apply, Google Cloud still had contractual obligations to HealthCo. It could not use data for advertising. It also took security precautions such as encrypting data in transmission and at rest. Prototype 1 did not make these points clear, but as discussed further in Section 3.5, we added additional details to Prototypes 2 and 3 to clarify these points.

3.2 Recruitment

Participants were recruited from the Pittsburgh, PA area through Craigslist. Recruitment posts invited potential participants to give feedback on designs for a healthcare website and online tools. Potential participants first answered a screening survey that confirmed their eligibility for the study: 18+ years of age; located in the U.S.; access to high-speed internet; and able to install and use Zoom, speak and read English, and view images and read text on a computer screen.

The screening survey also asked basic demographic questions and whether individuals had work experience or education in healthcare, healthcare administration, law, or technology. Using the screening survey responses, we invited individuals to participate based on a purposive sampling approach that sought to create a sample that was demographically diverse and that did not over-represent people with subject-matter expertise. See Table 1 for details about the sample demographics.

3.3 Study procedure

After individuals were selected via purposive sampling, each potential participant received an email invitation that included a link to an online consent form. Individuals who filled out the consent form were then directed to Calendly.com (an online scheduling platform) to choose an appointment time. Sessions lasted approximately 30 to 45 minutes. Each participant received a $15 Amazon gift code at the end of the study session.

We recorded audio and video via Zoom cloud recording. The audio recording was used for generating and correcting transcripts. The purpose of the video was to screen-record users' interactions with Adobe XD prototypes of our chatbot consent flows for further analysis. The study consent form notified each participant that they would be recorded before they signed up for an appointment, and we also obtained verbal consent before starting the recording. Participants were not required to leave their cameras on and were reminded that their faces would be recorded if they did turn them on.

To avoid priming participants to pay extra attention to the chatbot consent process, we did not specifically mention concepts such as privacy or HIPAA until after participants had completed the task portion and had answered some general followup questions about usability and their overall impressions.

We first asked introductory questions about participants' internet use and past experiences with online healthcare tools. We then explained that we were "testing designs for a new tool that allows users to talk to a chatbot to get questions about their health insurance." We had them open Adobe XD prototypes on their own screens and share their screens via Zoom. (We provided technical support as needed to help participants enable screen sharing. In one interview where the participant could not share their screen, the researcher opened the prototype on their screen, shared it with the participant, and asked the participant to instruct the researcher on the actions that the participant wanted to take.)

In all three rounds we asked participants to imagine that they were experiencing an ankle injury that had not initially seemed like an emergency but had remained painful for some time, that they wanted to see a provider to get an X-ray, and that they wanted to use the HealthCo chatbot to find out what their insurance would cover before making an appointment. The chatbot required users to go through a consent flow in which HealthCo disclosed that they used Google Cloud, which was not subject to HIPAA, to provide interactive chat functionality and asked users to consent to the collection, use, and disclosure of their protected health information in connection with the chatbot. The interviewer directed the participant to vocalize their thought processes as they interacted with the prototype. During each step of the task, the interviewer prompted the participant to note anything useful or confusing about the page, explain what option they would select next if they wanted to use the chatbot, and explain what they thought would happen after taking that action. If the participant got stuck, the interviewer proceeded

through a series of hints, beginning with general nudges ("Is there anywhere else you would look / anything else you would try?") to increasingly specific hints ("Try looking in [location]") if more general hints did not help. Once participants completed the consent flow, we stopped them and proceeded to ask followup questions.

In the second and third rounds, we added an interview portion in which we asked participants to return to the text of the consent document(s), review it more carefully, and let us know if they noticed anything new. We also asked some additional followup questions at that time. Since we had observed that participants were unlikely to read the document carefully while focused on the primary task that they were asked to try to complete—to ask the chatbot questions about health insurance coverage—we wanted to better understand whether any misunderstandings of the document arose from simply not reading (as would be expected during a real-life consent interaction), or from portions of the text being fundamentally unclear even if read carefully.

We refer to the three consent flow versions used in the three rounds as **Prototype 1**, **Prototype 2**, and **Prototype 3**. Prototype 1 was the version given to us by HealthCo, which required users to view a HIPAA Authorization document that opened in a separate tab as a PDF. In Prototype 2, users saw first a summary and then the HIPAA Authorization document in a modal window. Prototype 3 showed users a single HIPAA Authorization document in a similar modal window.

3.4 Analysis

Transcripts were created via Zoom's automatic transcription mechanism. Using the original videos and notes from the interview sessions, researchers corrected the transcripts and annotated them to mark important user actions. We coded the interviews in a two-part process. Codes were recorded using Airtable.

First, we coded empirical observations of user actions during the study tasks. Code categories were primarily deductive, focused on whether tasks were completed successfully, whether any hints were required and what types of hints were given, and whether users' expressed expectations about the interface's behavior matched its actual behavior. An inductive code category was also created to identify common UI sticking points that were noted across multiple study sessions. See Appendix B for more details about codes.

For each of the three interview rounds, we used the following empirical coding process. The lead researcher coded one interview and created an initial codebook. Two assistant coders then reviewed this interview to understand how to apply codes. The two assistant coders both coded one of the remaining five interviews from that round. We confirmed that Krippendorff's alpha was greater than 0.80, which is a recommended threshold for reliability [24]: 0.88 for Round 1, 0.90 for Round 2, 1.00 for Round 3. The two coders worked together to reconcile disagreements in Rounds 1 and 2. Reconciled codes were used for analysis. (There were no disagreements in Round 3.) Each assistant coder then coded two more interviews. The lead researcher checked in with the coders regularly to answer questions, reviewed all coding, and identified any portions that had been left uncoded or any areas where the codebook required revision.

Second, we conducted a thematic analysis to identify other themes that were related to understandings of HIPAA, understandings of the consent documents, and privacy attitudes (detailed in Appendix B). The lead researcher created an initial codebook and collaboratively iterated on this codebook along with two assistant coders. Because this was a small dataset that was coded in a highly collaborative fashion, we did not calculate interrater reliability for this portion of the analysis [28].

3.5 Results

Below we describe qualitative results from the usability testing and interview questions. We do not provide the exact number of participants that fell into a particular category when reporting attitudes and understandings from the thematic analysis since we want to avoid misconceptions that the *frequencies* from this phase of the study are necessarily generalizable. We simply report whether "some" or "none" of the participants answered in particular ways to describe what types of answers were present or absent in our sample.

3.5.1 Prototype 1 results

The consent flow for all of the prototypes began with a screen showing what a HealthCo customer would see if they signed in to view information about their benefits. A chatbot prompt appeared in the lower right corner, and participants could then click "Let's Chat."

In Prototype 1, participants then had to choose between buttons that said "View Resources" and "View Digital HIPAA Authorization" (Figure 1). The "View Resources" choice was confusing for some. One partici-

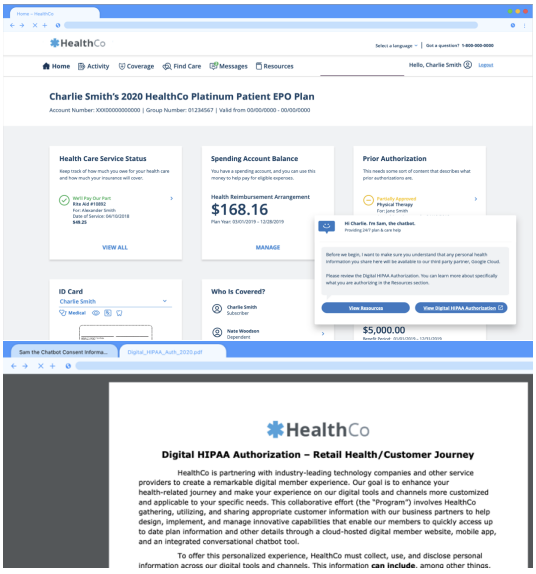

Fig. 1. Screenshots from Prototype 1 including the HIPAA authorization shown as a PDF.

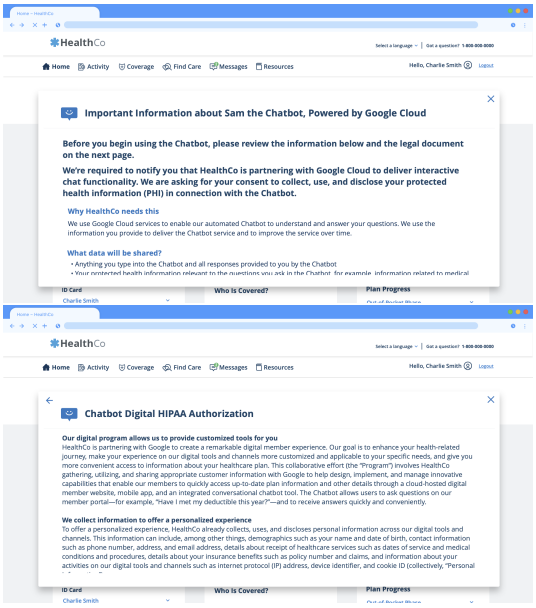

Fig. 2. Screenshots from Prototype 2: "Resources" summary and HIPAA Authorization documents, shown in modal window.

pant thought incorrectly that this might refer to "facts or FAQ," and another that it might contain information about "the programmers" or customer service associated with the tool. Only one participant actually selected "View Resources" and went through that optional portion of the flow that displayed a summary document.

After clicking "View Digital HIPAA Authorization," participants were shown a PDF of the document in a separate tab. Two participants required assistance from the interviewer to return to the main tab, and another found it only by accident and told us they thought they would have trouble finding it in real life.

After completing the consent flow, some participants correctly understood that Google Cloud would receive data. However, one expected that the data would be received only by healthcare providers and their staff. Another thought the data would mostly be received by "customer service agents" and "claims representatives."

When we asked participants about the purpose of the HIPAA authorization, none mentioned that data sent to Google Cloud was not subject to HIPAA, and some conveyed misconceptions. For example, one participant said the authorization was just intended to tell them how the insurance agency "conforms to HIPAA."

Participants might in some cases understand that Google Cloud would receive data but still not understand that Google Cloud was not subject to HIPAA, leading to a false sense of security. When asked how they felt about Google Cloud having access to their health information, one participant said, "I know HIPAA is pretty stringent, so I'm not too concerned with them having access." That person was also comforted that the third party was Google, noting that they were a well-known company with "really good security standards."

Not all participants had this false sense of security about how their data would be protected. While some participants had correct understandings of how the data was permitted to be used (e.g., personalization, aggregate metrics, and providing better chatbot functionality), others were concerned that data would likely be sold to additional parties and/or used for advertising.

There was also some confusion about *why* HealthCo had partnered with a third party: "Why is a third party involved, unless it's like the doctor or something like that?" Some participants understood that Google Cloud provided data storage but did not seem to understand other reasons that Google Cloud was involved.

3.5.2 Prototype 2: Formatting, clarity, and readability

We redesigned the "Resources" summary document and the HIPAA authorization based on observations from the first six interviews (Figure 2). We edited for clarity and brevity, focusing especially on using more common words to explain legal and technical concepts to a lay audience. In addition, we changed the flow so that clicking "Let's Chat" immediately displayed the "Resources" summary information in a modal window rather than

offering two buttons. We placed a "Continue" button at the bottom of that summary page, which, when clicked, loaded the full HIPAA authorization.

We made edits to emphasize two main points that some participants had missed: (1) that data was being shared with Google Cloud and (2) that the data shared with Google Cloud was not subject to HIPAA. We changed the title of the summary document from "Sam, the Chatbot" to "Important Information about Sam the Chatbot, Powered by Google Cloud." We added an overview at the top: "We're required to notify you that HealthCo is partnering with Google Cloud to deliver interactive chat functionality. We are asking for your consent to collect, use, and disclose your protected health information (PHI) in connection with the Chatbot." We also added large, bold subtitles: (1) "If you use the Chatbot, HealthCo shares data with Google Cloud," and (2) "We need your permission because data stored by Google Cloud is not covered under HIPAA."

The HIPAA authorization document had two parts, with explanatory text in the first section and a series of legal affirmations ("I understand / acknowledge / authorize. . . ") in the second. In the first portion, we added section subtitles in bold such as "HIPAA requires us to request your permission to disclose your data." In the second portion, we changed the initial sentences in the following ways to ensure that the important points were made in the first paragraph, as users may not read more:

> **Prototype 1:** I understand the nature of the digital member experience and the work HealthCo is conducting. I realize the value of HealthCo partnering with technology and service providers to create pioneering products and solutions, such as a chatbot tool, that will benefit consumers like me. I further understand that HealthCo must collect, use, and disclose my Personal Information and PHI in order to deliver this Program, and must also share this information with certain parties who are not subject to the same laws as HealthCo.

> **Prototype 2:** I understand that HealthCo must collect, use, and disclose my Personal Information and Personal Health Information (PHI) to Google Cloud in order to deliver this Program and provide the Chatbot. I understand that Google Cloud is not subject to HIPAA or certain other healthcare information laws that HealthCo must follow.

Our edits resulted in the Flesch-Kincaid grade level (a readability measurement [23, 32]) dropping from 17.4 to 15 for the HIPAA authorization and from 15 to 12.8 for the summary text. While an improvement, our revised version required college-level reading skills. Guidelines for the readability of healthcare information and computer interfaces commonly recommend an eighth-grade reading level or lower [4, 29, 31].

3.5.3 Prototype 2: Data sharing and legal details

We edited the authorization to include more elements from the 2018 Model Privacy Notice published by the U.S. Department of Health and Human Services [36]. We consulted with HealthCo's lawyers and developers to ensure that our revisions accurately reflected the backend behavior of the chatbot and HealthCo's internal practices and policies. This resulted in adding information about data encryption, access, and deletion. We also added text stating that Google Cloud employees could not access data without HealthCo's permission.

Some participants were concerned that data would be sold or used for advertising, and the initial authorization document did not dispel these concerns, since it contained sentences like the following: "I further acknowledge that these third parties may be able to use my information for their own commercial purposes." We confirmed with HealthCo that there were actually contractual restrictions on how Google Cloud could use the data, and for Prototypes 2 and 3, we removed the aforementioned sentence and added the following: "Google Cloud will not sell your PHI or use it for advertising."

Prototype 2 and 3 did, however, still need to contain a piece of legal boilerplate about the legal definition of a "sale," which may have caused confusion for any participants who read this portion: "I acknowledge that under certain laws or regulations, HealthCo's disclosure of my information could be deemed a 'sale' of information. I authorize HealthCo's disclosure of my information in connection with Program activities even if such disclosure constitutes a 'sale' under applicable law."

3.5.4 Prototype 2 results

All participants were able to complete all of the consent steps without hints in Round 2. However, having two documents was a negative factor for multiple participants. One said that the process was so long and "cumbersome" that they would probably give up and call on the phone. Some expected to be able to use the chatbot after clicking Continue on the summary page and were surprised to see another document. On the other hand, one participant thought the summary was "helpful," a good length, and that people would read more of it than of the actual HIPAA authorization.

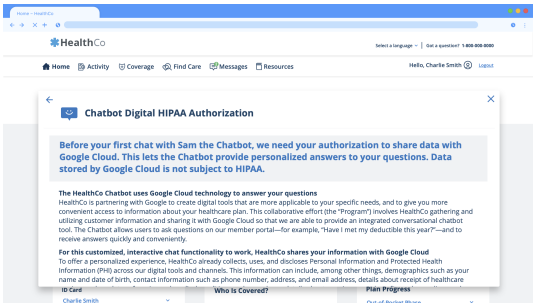

Fig. 3. Screenshot from Prototype 3: HIPAA Authorization document in modal window.

Several participants understood without being primed that the documents were telling them that data would go to a third party. All but one noticed before being primed that data was not subject to HIPAA. Some indicated confusion about why this would be the case. The sixth noticed this during the text review step.

As with Prototype 1, multiple participants expressed concerns about advertising and tracking. While referring to the part of the document that said this would not happen, one participant said: "They're saying [Google Cloud is] not permitted to access your data without your permission or sell your PHI or use it for advertising. With these big tech companies, they do that all the time... My level of trust is not high with that."

3.5.5 Creating Prototype 3

The primary change made between Prototype 2 and Prototype 3 was to remove the summary document that came before the HIPAA authorization: participants just saw the single document, scrolled to the bottom, and clicked "I agree." (See Figure 3.)

Because we removed the summary document, we incorporated some of its text into the HIPAA authorization, including borrowing simpler language and section titles from the summary and incorporating them into the introduction of the authorization. We took the "Powered by Google Cloud" language from the title of the summary and added it to the initial chatbot window that appeared before the authorization.

We added the following summary sentences to the top of the HIPAA authorization, in bold, colored font in a shaded box to draw attention: "Before your first chat with Sam the Chatbot, we need your authorization to share data with Google Cloud. This lets the Chatbot provide personalized answers to your questions. Data stored by Google Cloud is not subject to HIPAA."

3.5.6 Prototype 3 results

All participants were able to complete all of the consent steps without hints in Round 3. There were also relatively few negative usability comments in Round 3. One participant complained about the amount of text: "I would think something nefarious is going on because of how much fine print there is."

Some participants still did not understand that data would not be covered under HIPAA when in the possession of Google Cloud. This could cause a false sense of security, since, as one participant put it, a common perception is that "HIPAA is privacy." One participant noted that they received similar documents at the doctor's office and signed them without reading.

Some participants understood that Google Cloud would receive data, but some did not. Some expected only healthcare or insurance providers to receive it.

3.5.7 Additional themes

Below we describe patterns of responses that were not specific to individual prototypes and that have important implications for informed consent.

Surprised about consent step. Throughout the interviews, some participants were surprised to encounter a consent step at all, likely because most consumer chatbot tools that they had encountered (e.g., for retail stores) were not subject to HIPAA.

Purpose of consent process/document. While some participants correctly reported one or more of the main purposes of the consent document, including informing users that data was being collected and notifying users that chatbot data would be shared with a third party, some demonstrated concerning misunderstandings. For example, several participants stated that the consent document was notifying them that the chatbot data could *not* be shared with third parties. One participant reported, "It was a HIPAA authorization," and when asked what that meant, said, "That private health information can't be shared." Another summarized the document by saying, "It is trying to tell me I'm directly speaking to a representative of the insurance company and anything I say is confidential." This came up in all three prototype conditions.

Purpose of data collected during use of the chatbot. Many participants identified at least one of the main purposes of the data collected during use of the chatbot, e.g., creating a personalized experience and gathering metrics to inform improvements to the

chatbot and the website. However, several participants thought data from the chatbot would be used for advertising or tracking, even though the HIPAA authorization stated this would not happen. Some also thought that data would be shared with other companies beyond Google Cloud: "I would think that it's stored and sold to other parties to see if they can collect information on . . . people in their 40s and 50s with these type of health issues, and this is who you market your product to."

Purpose of Google Cloud partnership. When asked why HealthCo had partnered with Google Cloud, some participants understood that Google Cloud would provide data storage. Most participants did not mention Google Cloud's artificial intelligence or natural language processing capabilities, although a small number did mention concepts such as Google's "AI expertise."

Risk perceptions related to Google Cloud. Some participants were concerned about risks to data stored with Google Cloud, either from outside attackers or Google Cloud employees. The authorization stated that Google Cloud employees were not permitted to access their data directly without HealthCo's explicit permission, but some participants were still concerned: "I question what happens when, say, a rogue Google employee decides to sell 200,000 social security numbers to so-and-so credit card company." Lack of familiarity with Google Cloud could also make risk assessment difficult:

> "It says HealthCo is sharing my information with Google Cloud. . . what is Google Cloud?. . . Google Cloud supposedly doesn't follow the HIPAA laws, so that may be something that makes me a bit uncomfortable. Though. . . Google is. . . a big search engine and. . . they collect data on everything. So maybe HealthCo is using Google Cloud just to store their data. So it should be all right."

Understanding of the concept of *chatbot.* Most participants understood that a chatbot uses automation to answer questions. However, a small number thought that talking to a chatbot meant talking directly to a person: "I just assume that there's a person on the other side of these." Operating from this premise could make understanding the other details challenging.

4 Surveys

We conducted two crowdsourced surveys to collect data from a larger sample of participants. The first survey (**Survey 1**) compared the three consent flow prototypes developed in the interview study. The second survey (**Survey 2**) further investigated users' understandings of HIPAA by comparing Prototype 3 to a new **Prototype 4** which was identical to Prototype 3 except that it avoided using the term "HIPAA."

4.1 Recruitment

In early 2021, we recruited participants from Prolific for a 20-minute task whose purpose was described as "evaluating the usability of an online healthcare information tool." Participants needed to be 18 years of age or older, located in the U.S., able to read and write English, and able to view prototype images on a tablet, laptop, or desktop computer. We used Prolific's prescreening tools to show the study task only to adults who reported being located in the U.S. when they signed up for Prolific.

We recruited 1,199 participants for our first survey, 1,050 of whom completed the entire task and were compensated. This survey phase was originally intended to contain four study conditions, but a technical error was detected in one of the conditions after data collection and participant compensation, so responses from 293 participants who had been assigned to that condition were dropped. In the rest of this section we report only on the three remaining conditions. Thirty additional responses were dropped due to technical errors that prevented participants from viewing the prototype properly or markers of bot and low-quality data (determined using manual review of free response answers as well as Qualtrics reCAPTCHA bot detection, duplicate scores, and fraud detection scores). For Survey 1, we ultimately analyzed data from 761 participants.

In fall 2021, we recruited for Survey 2 on Prolific using similar criteria to Survey 1, but we also prescreened to exclude those who had participated in Survey 1. Additionally, due to a disruption to the demographics of the Prolific participant pool that occurred shortly before this study launched [25], we needed to balance demographics manually using Prolific prescreening. We ran batches to gather data from 12 groups determined by gender—48% men, 48% women, and 4% those who selected other gender options in Prolific prescreening (*Genderqueer/Gender Non Conforming, Different Identity, Rather not say*)—and approximately equal percentages from four age buckets (18-32, 33-47, 48-62, 63+). We had to reject portions of some batches due to data quality and also had trouble filling the quota for the group of the oldest age category and third gender category, so after ensuring we would have at least 48% women and 48% men and at least 110 responses in each

age bucket, we re-posted the small number of remaining slots with fewer demographic restrictions.

We did not filter for past task approval rate when selecting participants in either of these surveys: prior to this study we had not found it necessary on Prolific, and we did not want to exclude research-naïve participants. Unfortunately, we found that bot and inauthentic responses were more prevalent in these surveys than in our past use of Prolific. Approval rate filtering may now be beneficial on that platform.

We recruited 582 total participants for Survey 2, 496 of whom completed the task and were compensated. After excluding duplicates, bot and low-quality responses, and instances of technical errors with the prototypes, we included 456 responses in the dataset.

We paid participants $5 for Survey 1 and $3.75 for Survey 2 via Prolific's direct payment platform. For analyzed responses, the median completion time was 19 minutes for Survey 1 and 12.5 minutes for Survey 2. See Table 1 for demographic details from both surveys.

4.2 Survey task

We implemented the surveys using Qualtrics. In the beginning of each survey, we introduced an ankle injury scenario similar to the one in the interview (as described in Section 3.3) and asked participants to open the prototype link and try to ask the chatbot a question about their healthcare coverage as related to the diagnosis or treatment of the ankle injury. In attempting that, participants encountered one of the consent flows, based on their assigned condition. After completing the consent flow, they were asked a set of questions about their experiences, understandings, and attitudes.

After the initial task and first set of questions, Survey 1 participants were asked to return to the consent document portion of the prototype to review the text more carefully. They were then presented with some of the same questions again and could change their answers if their understandings or opinions had changed. They were also asked additional questions about their understanding of the consent document(s). These questions were placed at the end to avoid priming.

The main goal of Survey 2 was to compare the third prototype of the consent flow to Prototype 4, which was identical to Prototype 3 except for having the "Digital HIPAA Authorization" title replaced with "Chatbot Data Sharing Authorization" and "HIPAA" in the body of the document replaced with "federal healthcare privacy laws." (See Figure 4.) We evaluated Prototype 4 to further examine some of the ideas that arose in the previous phases regarding the word "HIPAA" potentially creating a false sense of security or making it difficult for participants to understand that data was *not* covered by HIPAA when held by a cloud service provider. This survey task included very similar questions to Survey 1, but it was shorter and did not ask participants to return to the consent document.

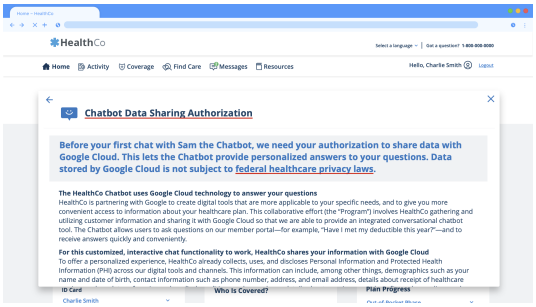

Fig. 4. Annotated screenshot from Prototype 4. Identical to Prototype 3 as shown in Figure 3, except for new title and replacement of instances of "HIPAA" with "federal healthcare privacy laws." (Some of these appear in portions of the text that are not pictured here that appear when the user scrolls down.)

4.3 Data analysis

We performed similar quantitative and qualitative analyses for both surveys, as detailed below.

4.3.1 Statistical analysis

We performed quantitative analysis tasks using R v4.0.2 and RStudio v1.2.5019. We used $\alpha = 0.05$ for statistical significance. We applied Holm-Bonferroni corrections for multiple hypothesis testing to the set of omnibus tests and when running pairwise tests. The p-values reported below are the corrected values.

We binned Likert data before analysis as follows: "agree" and "strongly agree" were combined into an "agree" bin, and "neutral," "disagree," and "strongly disagree" into a "did not agree" bin. In some cases (described further below) when using tests that assumed binary outcome variables, we binned "yes," "no," and "not sure" responses into binary bins.

Analyses with the prototype condition as the independent variable and a categorical outcome were performed using Chi-squared tests of independence. Effect sizes for χ^2 tests were calculated using Cramér's V, for which a value of 0 indicates no association and a value of

1 indicates perfect association. Regardless of statistical significance, we only consider results with Cramér's V > 0.10 to be meaningful, as values below that indicate only negligible association [38]. For χ^2 tests on tables larger than 2x2 where results were significant after omnibus testing and HC correction was complete, we also ran posthoc pairwise χ^2 tests using `rstatix` [19, 20].

Analyses of categorical outcomes before and after the text review step were performed using the McNemar test for paired nominal data. Effect sizes for McNemar are determined using odds ratios (OR), for which OR=1 reflects the null hypothesis, OR<1 indicates lower odds of outcome after "treatment" (the text review step), and OR>1 indicates higher odds of outcome [42]. Effect sizes for this test were categorized by converting to Cohen's d [8, 10] using the `effectsize` package in R [5, 6]. In this paper we do not consider results with effect sizes considered "very small" (according to Chen et. al [8]) to be evidence of meaningful associations. For "small" effect sizes, we report the results but note the size.

4.3.2 Qualitative analysis

We performed a qualitative coding process with a lead researcher and a team of five assistant coders. This coding process had both inductive and deductive elements: we were interested in coding for certain concepts based on what we observed in the quantitative survey results as well as the interview results—e.g., "knows that Google Cloud gets data"—but we also allowed for inductive category creation since we expected to find additional attitudes and understandings in this dataset that we had not yet observed. Code categories are described further in Appendix C. We used Airtable to tag survey responses. Codebooks were created for each free response question either by the lead researcher alone or by an assistant coder who worked with the lead researcher to iterate on a codebook and then applied it to all responses from that question. The coder and lead researcher discussed iterative additions or revisions to each set of codes as needed. We did not calculate an interrater reliability statistic for the survey coding: this coding was straightforward, often based on the presence or absence of a word or phrase (e.g., "HIPAA") [28].

4.4 Survey results

We present the results of both surveys, focusing on document reading behavior, usability, user attitudes, understanding of chatbot data use, and understanding of HIPAA and other legal concepts. We observed some usability improvements in Prototypes 2 and 3, but limited impact on participant understanding as they accessed the chatbot. Participants made many incorrect assumptions about HIPAA protections as well as Google Cloud's data practices. We saw a number of large changes in responses in Part 2 of Survey 1 after participants revisited the authorization documents, suggesting that some misconceptions would be corrected if people actually read the documents. We saw no significant differences between Prototypes 3 and 4 in Survey 2.

4.4.1 Document reading behavior

In both surveys, we asked two questions to get an understanding of participants' self-reported reading behavior when presented with healthcare consent documents.

We asked how much people had read of the consent document presented to them during the initial study task. Across both surveys, 41.2% responded "I skimmed it," 10.4% said they did not read any of the document, and 16.8% reported reading only titles or headlines. In contrast, 17.9% said they read most of it, and only 13.6% said they read all of the document.

We also asked, "When you see legal documents like this related to healthcare, how much of them do you usually read?" 44.5% responded "I skim them," 7.8% reported not reading these documents at all, and 14.5% only titles or headlines. 22.3% said they read most of these documents, and 9.3% said they read all of them.

4.4.2 Usability

In Part 1 of Survey 1, we asked participants two questions related to usability of accessing the chatbot.

First we asked participants to indicate their level of agreement with the statement "I found it easy to access this chatbot." 94.7% of participants who saw Prototype 3 agreed, while only 83.6% of participants who saw Prototype 1 agreed, a small but statistically significant difference ($p < 0.01, V = 0.15$). Differences between Prototypes 1 and 2 or 2 and 3 were not statistically significant. (See Tables 2, 3, and 5 in the appendices for more details about this and other Likert items.)

We also asked participants to indicate their level of agreement with the statement "The process to get to the chatbot took too long." Across all conditions, 12.8% of participants agreed or strongly agreed. Differences be-

tween prototypes were not statistically significant after Holm-Bonferroni correction ($p = 0.07$).

4.4.3 User attitudes about chatbot

Understanding of *chatbot*. Because there had been some evidence in the interviews that not all participants understood that a chatbot offered computer-generated answers, we asked participants in Survey 1, "How do you think the answers shown to you by the chatbot would be created?" Most participants (649, 85.2%) answered correctly, "Automatically generated by a computer program." A small number, however, answered that the answers would be typed by a human (65, 8.5%).

Willingness to use chatbot. In both parts of Survey 1, as well as in Survey 2, we asked, "Based on the information you saw, would you use this chatbot?" We did not find a significant difference between prototypes in either part of Survey 1 or in Survey 2. However, across all conditions, there was a large decrease in willingness to use the chatbot from Part 1 to Part 2 of Survey 1. In Part 1, 481 (63.1%) said they would use the chatbot, and in Part 2, only 332 (43.6%) said they would use it. We used a McNemar test on paired data that was bucketed into "yes" and non-yes ("no" or "not sure") categories. ($p < 0.01$, OR=0.10). In Survey 2 (which had only one part), 298 participants (65.4%) said they would use the chatbot, 82 (18.0%) were not sure, and 76 (16.7%) would not use it. (See Tables 2, 3, and 5 in the appendices for more details about this question.)

In both surveys, we also asked a followup free-response question to explore why participants said they would or would not use the chatbot. Among those who said they would not use the chatbot, most mentioned privacy concerns. Most of those privacy concerns related to the lack of HIPAA protections or to concerns about the data being shared with Google. Other categories of privacy concerns included feeling like there was too much disclosure text, objection to data sharing in general, or non-specific concerns about privacy or security.

We observed 103 participants (13.5%) change their answers from "Yes" to "No" in Part 2 of Survey 1, and 62 (8.1%) from "Yes" to "Not sure." Of these participants who changed their answers in a negative direction, 46% indicated that their reasons included objections to the data not being subject to HIPAA, discomfort with Google storing health data, or other privacy concerns.

Chatbot versus other methods of finding information. We asked participants in both parts of Survey 1 to rate their agreement with the statement "I would prefer to find health insurance information in another way instead of using the chatbot." There was no significant difference between prototypes in either part. However, there was a large increase from Part 1 to Part 2 in the percentage who agreed they would rather *not* use the chatbot (42.3% vs. 61.3%, $p < 0.01$, OR=17.1).

Privacy confidence. In both parts of Survey 1, we asked participants to rate their level of agreement with the statement "I am confident that the privacy of my data will be protected if I use the chatbot." There was no statistically significant difference between prototype conditions for either time that this question was asked. However, there was a large and statistically significant decrease in privacy confidence after the text review step, with 44.6% of participants agreeing with this statement in Part 1, and only 27.4% agreeing in Part 2 ($p < 0.01$, OR=0.08, 95% CI for OR [0.03, 0.14]). Only 35 participants (4.6%) changed their answer to this Likert item in a positive direction between Part 1 and Part 2.

We asked a free-response question to explore participants' reasons for their answers to this Likert item. For the small number of participants whose sentiment changed in a positive direction, some noted reasons for changing their opinion that were correctly gleaned from the disclosure: that data would not be sold or used for ads, that there were limitations on who data could be shared with, and that data would be encrypted. Some also noted incorrect takeaways from the disclosure, e.g., "HIPAA compliance." Among participants who changed their answers in a negative direction, the most common reasons associated with that change were that data would be shared with Google Cloud and that data would not be subject to HIPAA or other laws protecting health data, both correct points that the disclosure was intended to clarify. Some were also concerned about data being sold or shared outside of HealthCo in general, or concerned that data would be available to multiple third parties, not just Google Cloud. Some mentioned security concerns such as "hacking" or "leaks." A few participants mentioned reasons that explicitly contradicted the disclosure, believing that Google Cloud was allowed to use data for ad targeting, or that Google Cloud employees were free to view and potentially misuse data.

4.4.4 Understanding of chatbot data use

Who gets data? In both parts of Survey 1 and in Survey 2, participants were asked a "check all that apply" question regarding what people or companies would receive copies of data if they used the chatbot. We were

primarily concerned with the number of participants who checked (1) only "No people or companies would have these records," implying that they did not even understand that HealthCo or healthcare providers might have their records or (2) "Google Cloud." We did not find differences between prototype conditions in either survey. Most people understood that *someone* would receive data. However, many did not realize that Google Cloud would receive data until they revisited the authorization in Part 2.

Chi-squared tests did not show differences between prototypes for the "Google Cloud" answer option in either part of Survey 1 or in Survey 2. However, the number of participants who checked "Google Cloud" nearly doubled between Part 1 and Part 2 of Survey 1 (from 39.2% to 79.1%, $p < 0.01$, OR=51.7).

"No people or companies would have these records" was an uncommon response overall: only 31 (4.1%) people checked that answer in Part 1 and only 16 (2.1%) in Part 2. There were too few discordant pairs to conduct a McNemar test comparing Parts 1 and 2.

Whose employees can look at data? We then asked participants who would be able to look at their data without additional permission. Participants could check any options that applied: "My doctor or people at my doctor's office," "HealthCo employees," "Amazon Web Services employees," "Facebook employees," "Google Cloud employees," "Other," or "No one would look at the data" (which would uncheck other options). We tested for differences in whether participants checked: (1)"no one," i.e., did not even understand that their healthcare entities would have access to data and (2) "Google Cloud employees," implying that they did not understand or notice the statement that Google Cloud employees could not look at data without permission. In Part 1, most people understood correctly that someone would be able to look at their data, and the number who believed incorrectly that Google Cloud employees could look at their data was relatively small. However, after revisiting the authorization in Part 2, many more believed incorrectly that Google Cloud employees could look at their data, regardless of condition.

In Survey 1, the only significant difference between conditions occurred for the "no one" answer in Part 2, where those who saw Prototypes 2 or 3 were slightly more likely than those who saw Prototype 1 to say that no one would look at their data ($p = 0.03$, $V = 0.15$): 2.8%, 10.4%, and 10.3% respectively for Prototypes 1, 2, and 3. The difference between P2 and P3 for the "no one" answer was not statistically significant. There were not significant differences between conditions in either part of Survey 1 for the "Google Cloud" answer.

In Survey 2, there were no significant differences between prototypes for the "no one" or "Google Cloud" answers. Across the two conditions, 103 participants (22.6%) answered that they thought Google Cloud employees could look at their data without permission.

We also ran McNemar tests to compare Part 1 to Part 2 responses for Survey 1. The number of participants who incorrectly thought that Google Cloud employees could look at their data increased significantly—from 21.3% to 53.9%—from Part 1 to Part 2 ($p < 0.01$, OR 8.52). On the other hand, the number of participants who incorrectly thought that no one would see their data decreased slightly ($p < 0.01$, OR=0.36) between Part 1 (13.5%) and Part 2 (7.9%).

What will data be used for? Participants were asked to check all that applied for a question asking, "If you used the chatbot, what do you think your chatbot-related data would be used for?" This list included 10 options, but for analysis we focused on three that the document explicitly stated would not occur: "Selling it to other companies," "Sharing it with other companies (without selling it for money)," and "Advertising."

In Survey 1 Part 1, there was no significant difference between conditions in the number of participants who had the misconception that data from their use of the chatbot could be sold (28.1%), shared (19.3%), or used for advertising (26.5%). In Survey 2, there was also no difference between conditions. Overall 16.7% thought data could be sold, 13.2% thought it could be shared, and 22.6% thought it could be used for advertising.

We did see some significant differences in beliefs about data use between conditions in Survey 1 Part 2: Prototypes 2 and 3 led to lower rates of misconceptions than Prototype 1. A McNemar test also showed an increase in misconceptions from Part 1 to Part 2.

See Tables 6, 7, 8, and 9 in the appendices for more details about these results.

Why is HealthCo partnering with Google Cloud? We asked participants a free-response question about why they thought HealthCo was partnering with Google Cloud. Across both surveys, about 9% assumed profit-oriented motivations on the part of HealthCo or Google Cloud: e.g., "Google is interested in marketing the info for profit" or "HealthCo can make money by giving Google access to people's data."

4.4.5 Understanding of HIPAA and legal concepts

We explored participants' levels of understanding of HIPAA in the context of the consent flow they saw.

What does "federal healthcare privacy laws" refer to in Prototype 4? Before any questions mentioning HIPAA, Survey 2 participants who saw Prototype 4 were asked what they thought "federal healthcare privacy laws" was referring to. About half gave free responses that mentioned HIPAA. A few mentioned other relevant concepts (e.g., PHI), or legal or health concepts or entities that were not directly relevant (e.g., hippocratic oath, patient-doctor confidentiality, OSHA). The rest said they didn't know or gave vague or irrelevant answers (e.g., "health law," "federal privacy laws," "this kind of law comes from government").

What does "not subject to HIPAA" mean? In both parts of Survey 1, we asked all participants what they thought the sentence "Data stored by Google Cloud is not subject to HIPAA" meant. In Survey 2, we asked that question to the participants in the Prototype 3 condition, and we asked Prototype 4 participants what they thought "not subject to federal healthcare privacy laws" meant. (This question appeared before questions where we asked directly about HIPAA to avoid priming Prototype 4 participants.) About 40% of participants across both surveys interpreted these phrases to mean there were essentially no protections or rules applying to the data at all or that the data would be in grave danger, e.g., "they can sell the data to whomever." A few in each group (about 3% in Survey 1, and about 5% in Survey 2) misinterpreted these phrases in the other direction, giving answers indicating that the data would have *more* protections: "It doesn't violate hipaa and won't share personal information about health or a diagnosis." In every condition across both surveys, we saw responses indicating that participants' expectations of HIPAA were violated, for example: "it means that hipaa doesnt apply to data on the cloud which is CRAZYYY" and "This is very wrong, It should be covered by HIPAA."

When is data subject to HIPAA? In both parts of Survey 1 and in Survey 2 we asked, "Is data about your use of this chatbot always subject to HIPAA?" The differences between conditions were not significant in Part 1 of Survey 1 or Survey 2. However, the percentage of Survey 1 participants who answered this question correctly showed a large increase from Part 1 to Part 2 (from 20.9% to 59.1%, $p < 0.01$, OR=58.29). Also, a Chi-squared test showed significant differences between conditions in Part 2 ($p < 0.01$, $V = 0.30$). Of participants who saw Prototypes 2 or 3, the respective percentages who answered correctly were 73.1% and 65.3%, while less than half (38.8%) of those who saw Prototype 1 who answered correctly. The difference between Prototypes 2 and 3 was not statistically significant. (See Tables 6, 7, and 8 for more details about these results.)

Participants who correctly answered that their data was not always subject to HIPAA were asked a free response question about the circumstances under which they believed their data was not subject to HIPAA. Most of these participants gave accurate explanations, indicating that the data shared with Google Cloud was not subject to HIPAA, that the data they consented to share via this authorization was not subject to HIPAA, or a similar response. A few participants in this group, however (11% in Survey 1, and 20% in Survey 2) provided reasons that indicated misconceptions. These people indicated limited circumstances when data would not be subject to HIPAA, such as only if the data was not personal or medical, only in the case of emergency, or only in the case of legal demands such as subpoenas.

What is HIPAA? When we asked Survey 2 participants, "What is HIPAA?" 65.6% of participants selected "A U.S. federal law that prevents anyone or any company from sharing health data without the patient's permission," grossly overstating what HIPAA mandates and which entities must follow it. Only 26.8% of participants selected the answer we considered to be correct, "A U.S. federal law that regulates how healthcare data can be shared" [35]. (See Table 10 in the appendices.)

Similarly, when we asked Survey 2 participants, "If HIPAA defines a piece of information as Protected Health Information, what does that mean?" 58.3% selected "that no person or company can share it without the patient's permission." Only 24.3% selected the correct answer, "That a healthcare provider can only share it with another person or organization under certain conditions" [13]. (See Table 11 in the appendices.)

5 Limitations

This study design did not allow us to isolate the effects of individual differences between prototypes in our quantitative results, since there were multiple differences between each pair of prototype versions. However, our results still offer useful insights as a case study.

The Adobe XD prototypes were not accessible to people with low vision or blindness. More research is necessary to better understand how to implement accessible consent flows.

Participants were required to use devices with larger screens, such as laptops or tablets, to participate in the study. Some of the design observations may not be as applicable to smartphones or other small devices. However, we would expect that many of the aspects that users found challenging would still be equally challenging or perhaps *more* challenging on smaller devices.

Especially in the first study phase, observer effects may have caused users to evaluate the prototypes more positively than they otherwise would have. We mitigated this by emphasizing that we wanted honest feedback and would not take personal offense at critiques.

While we made efforts to maximize the diversity of our study samples, they are not representative of the U.S. population. They likely exclude most participants with low technical literacy, since participants needed computers or tablets as well as technical knowledge to participate in a study via Zoom or Prolific.

6 Discussion

We evaluated four prototypes of a healthcare data consent flow, one provided to us by a healthcare company and three variants we developed and tested iteratively. We redesigned the original prototype to follow usability guidelines and recommended best practices for healthcare disclosures, and we clarified language in the disclosure document in consultation with lawyers and technical experts at the company. While our user studies showed evidence that our design changes clearly improved usability and resulted in disclosures that were better understood by some participants after they were directed to re-read them, our design changes did not appear to significantly improve understanding by participants who encountered the disclosures in the course of making a consent decision. In this section we discuss potential reasons that our design changes failed to provide a more informed consent experience and suggest other approaches that may be worth exploring.

6.1 Reasons for misunderstandings

One factor that likely contributed to participants' misunderstandings is that they did not read the document carefully. We did not expect that participants would read the entire document, and doing so would arguably be irrational (both in the study and in real life). However, we were surprised that most participants did not seem to read even the summary box at the top of the document in Prototype 3. We saw in the first survey that many participants could understand the main points if they did take time to read the documents.

Another possible reason for misunderstandings is that these authorization documents seem to violate users' expectations that HIPAA protects privacy unconditionally. Qualitative evidence from all three phases of the study indicated that participants have high privacy expectations for healthcare data and often believe that laws forbid data sharing broadly. Some participants seemed to find it unnecessary to read any HIPAA-related documents, assuming they all state that their data is incontrovertibly private. This complicates the entire concept of a HIPAA authorization requesting consent to share data with a third party. In the second survey, although we replaced the word "HIPAA," many participants still inferred that we were referring to HIPAA and thus had the same high expectations of privacy, and we did not see significant effects on users' understanding or attitudes.

People's lack of prerequisite technical knowledge or their incorrect mental models can also affect the ability to understand this type of disclosure. For example, it's impossible for someone to understand the nature and purpose of this data sharing if they don't understand that the chatbot is serving automated answers. Some people also don't understand why a third party would be involved, or have only a very high-level concept of what the "cloud" is, whose responsibility it is, and what might be good or bad or risky about storing data in it. People may also have trouble separating the consumer-facing offerings of third parties like Google or Amazon from their commercial services, and they may not understand that commercial tools can be subject to different terms or special agreements with business customers. For example, some people assume (even when the document explicitly says the opposite) that data is going to be sold for ads. In this case this would at least cause them to be overly cautious rather than making an unwanted disclosure, but this misunderstanding still hinders making an informed decision, and it might keep someone from using a tool that could be useful to them.

6.2 Alternative approaches

Our results suggest that disclosures can be made more usable by applying standard best practices in this space such as shortening disclosures, simplifying language, and highlighting important points. However, these prac-

tices do not seem to offer much benefit towards informed consent when users are more focused on another primary task. To achieve informed consent for HIPAA authorizations, we may need alternative approaches.

Standardization. Standardized disclosures, sometimes referred to as "nutrition labels" after standard food labels [15], aim to provide disclosures in succinct, consistent (often tabular) formats that allow for easy comparison [12, 14, 21, 22]. If HIPAA authorizations could be distilled into a standard set of options, users might be able to glance at them and quickly determine whether any concerning or surprising boxes are checked. In the case of the HealthCo authorization, users might see that data will be disclosed to a third party (identified as Google Cloud), may not be shared further, and may be used for only limited purposes which specifically exclude advertising. Research is needed to determine what information is most important to include in a simple, standardized disclosure and whether users would actually review these disclosures before providing consent.

Multimedia, quizzes, and conversational approaches. Efforts towards improved informed consent for clinical research may offer ideas that could be explored to improve online HIPAA authorizations and other online consent flows. Methods that have been explored for improving informed consent in clinical research include multimedia formats, quizzes with feedback, discussion with study staff, and "enhanced" consent forms or leaflets (often written at a middle-school reading level), with the latter two showing the most promising evidence [34]. In the context of a chatbot or conversational agent, we recommend research into ways that quizzes or discussion could be implemented within the conversational functionality. Harkous et al. explored how chatbots could be implemented to offer users privacy notice and choice in a more conversational fashion [18]. Perhaps this approach could have helped to highlight important points in the HIPAA authorization and answer user questions. Games may also offer inspiration for conversational approaches to privacy choices, with some such as Animal Crossing: New Horizons incorporating important privacy disclosures and choices directly into in-game interactions [33].

Privacy assistants. Expecting users to read disclosures and make choices about them in the moment may not be the ideal solution for some contexts, no matter how much effort is taken to improve the user experience. Automated user agents such as Personalized Privacy Assistants [11] may someday offer a way to better align consent decisions with users' actual preferences because they can be designed to automatically consent, withhold consent, or prompt users based on previously configured preferences, reducing the need for users to spend time reviewing authorization details for every service. In some cases these agents may use machine learning to infer user preferences or may take recommendations from a user's trusted friends or organizations.

Avoid unnecessary data sharing. We recommend that companies—whether subject to HIPAA or not—avoid unnecessary data sharing in healthcare contexts whenever possible. The goal should be to avoid violating people's expectations that healthcare data will be protected and shared only when truly necessary, and to avoid the need for disclosure. An alternative solution here would have been for the company to negotiate a contract with their cloud service provider that *did* fall completely within HIPAA business associate standards. This might have been a better solution for both the patients and the company. In this scenario completely deidentifying data might not be possible due to the need to look up personalized insurance information, but data deidentification and other privacy enhancing approaches should also be pursued when possible.

7 Conclusion

We described a three-phase study in which we compared four versions of a consent flow for a healthcare tool. While we were able to improve the usability of the consent flow in our redesigns by following best practices and clarifying the disclosure text, users' ability to understand essential points of the disclosures while encountering them in a task context did not seem to be meaningfully improved by these redesigns. Although users did understand many of the points when instructed to read the documents carefully, in some cases they also picked up new misunderstandings when re-reading. Reasons for misunderstandings seemed to include not reading the document, the document violating users' privacy expectations in the context of healthcare and HIPAA, and other factors such as a lack of necessary technical knowledge. Given that following best practices and clarifying language was not sufficient to ensure informed consent, we propose that future research on HIPAA authorizations and other health data disclosures should explore alternate approaches such as standardized disclosures; methods borrowed from clinical research contexts such as multimedia formats, quizzes, and conversational approaches; and automated privacy assistants.

Acknowledgements

This research was supported in part by the National Science Foundation under grant CCF-1852260, Carnegie Corporation of New York, Innovators Network Foundation, and Highmark Health. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation or other funders. We also thank Brigette Bernagozzi, Sanjnah Ananda Kumar, Jiamin Wang, James Ray, and Bobby Zhang for their assistance with qualitative coding, and Julia Petrich and Tarannum for their consultation throughout this project.

References

[1] 45 CFR § 164.508 - Uses and disclosures for which an authorization is required. https://www.law.cornell.edu/cfr/text/45/164.508.

[2] Ruba Abu-Salma, M. Angela Sasse, Joseph Bonneau, Anastasia Danilova, Alena Naiakshina, and Matthew Smith. Obstacles to the adoption of secure communication tools. In *2017 IEEE Symposium on Security and Privacy (SP)*, pages 137–153, May 2017.

[3] Brooke Auxier, Lee Rainie, Monica Anderson, Andrew Perrin, Madhu Kumar, and Erica Turner. 4. Americans' attitudes and experiences with privacy policies and laws. Technical report, Pew Research Center, November 2019.

[4] Sameer Badarudeen and Sanjeev Sabharwal. Assessing Readability of Patient Education Materials: Current Role in Orthopaedics. *Clinical Orthopaedics and Related Research*, 468(10):2572–2580, October 2010.

[5] Mattan S. Ben-Shachar, Dominique Makowski, and Daniel Lüdecke. Effectsize: Indices of effect size and standardized parameters. https://easystats.github.io/effectsize/.

[6] Mattan S. Ben-Shachar, Dominique Makowski, and Daniel Lüdecke. Effectsize::interpret_oddsratio. https://easystats.github.io/effectsize/reference/interpret_oddsratio.html.

[7] Omri Ben-Shahar and Adam Chilton. Simplification of privacy disclosures: An experimental test. *Journal of Legal Studies*, 45:27, June 2016. https://www.ftc.gov/system/files/documents/public_comments/2017/11/00022-141740.pdf.

[8] Henian Chen, Patricia Cohen, and Sophie Chen. How big is a big odds ratio? interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics - Simulation and Computation*, 39(4):860–864, March 2010.

[9] Google Cloud. Dialogflow Documentation, 2022.

[10] Jacob Cohen. *Statistical Power Analysis for the Behavioral Sciences*. L. Erlbaum Associates, Hillsdale, N.J, 2nd ed edition, 1988.

[11] 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, New York, NY, USA, April 2020. Association for Computing Machinery.

[12] Lorrie Faith Cranor, Pedro Giovanni Leon, and Blase Ur. A large-scale evaluation of U.S. financial institutions' standardized privacy notices. *ACM Trans. Web*, 10(3), August 2016.

[13] Digital Communications Division (DCD). What is PHI?, February 2013.

[14] 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*, pages 1–12, Glasgow, Scotland, UK, May 2019. ACM.

[15] FDA. Nutrition facts label better informs your food choices, 2016.

[16] Tasha Glenn and Scott Monteith. Privacy in the digital world: Medical and health data outside of HIPAA protections. *Current Psychiatry Reports*, 16(11):494, November 2014.

[17] 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 *Proceedings of the Twelfth Symposium on Usable Privacy and Security (SOUPS 2016)*, page 21, Denver, CO, USA, June 2016. https://www.usenix.org/conference/soups2016/technical-sessions/presentation/gluck.

[18] Hamza Harkous and Kassem Fawaz. PriBots: Conversational Privacy with Chatbots. In *Workshop on the Future of Privacy Indicators, at the Twelfth Symposium on Usable Privacy and Security (SOUPS) 2016*, 2016.

[19] Alboukadel Kassambara. Proportion test (prop_test). https://rpkgs.datanovia.com/rstatix/reference/prop_test.html.

[20] Alboukadel Kassambara. Pipe-friendly framework for basic statistical tests [R package Rstatix version 0.7.0], February 2021. https://CRAN.R-project.org/package=rstatix.

[21] Patrick Gage Kelley, Joanna Bresee, Lorrie Faith Cranor, and Robert W. Reeder. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security (SOUPS 2009)*, pages 1–12, New York, NY, USA, July 2009. Association for Computing Machinery.

[22] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3393–3402, 2013.

[23] J. Kincaid, Robert Fishburne, Richard Rogers, and Brad Chissom. Derivation of new readability formulas (Automated Readability Index, Fog Count And Flesch Reading Ease Formula) for Navy enlisted personnel. *Institute for Simulation and Training*, January 1975.

[24] Klaus Krippendorff. *Content Analysis: An Introduction to Its Methodology*. SAGE Publications, second edition, 2004.

[25] Rafi Letzter. A teenager on TikTok disrupted thousands of scientific studies with a single video. *The Verge*, September 2021.

[26] Richard May and Kerstin Denecke. Security, privacy, and healthcare-related conversational agents: A scoping review. *Informatics for Health and Social Care*, pages 1–17, October 2021.

[27] 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*, page 22, 2008. http://www.is-journal.org.

[28] Nora Mcdonald, Sarita Schoenebeck, and Andrea Forte. Reliability and inter-rater reliability in qualitative research: Norms and guidelines for CSCW and HCI practice. *Proc. ACM Hum.-Comput. Interact.*, 3(CSCW):24, November 2019. https://doi.org/10.1145/3359174.

[29] George R. Milne, Mary J. Culnan, and Henry Greene. A longitudinal assessment of online privacy notice readability. *Journal of Public Policy & Marketing*, 25(2):238–249, September 2006.

[30] Ilene N. Moore, Samuel Leason Snyder, Cynthia Miller, and Angel Qui An. Confidentiality and privacy in health care from the patient's perspective: Does HIPAA help? *Health Matrix: Journal of Law-Medicine*, 17(2):215–272, 2007.

[31] Jakob Nielsen. Lower-literacy users: Writing for a broad consumer audience, 2005.

[32] Jakob Nielsen. Legibility, readability, and comprehension: Making users read your words, 2015.

[33] Nintendo. Animal Crossing™: New Horizons for the Nintendo Switch™ system – Official Site, 2021.

[34] Adam Nishimura, Jantey Carey, Patricia J. Erwin, Jon C. Tilburt, M. Hassan Murad, and Jennifer B. McCormick. Improving understanding in the research informed consent process: A systematic review of 54 interventions tested in randomized control trials. *BMC Medical Ethics*, 14(1):28, July 2013.

[35] Office for Civil Rights (OCR). 187-What does the HIPAA Privacy Rule do, October 2015. https://www.hhs.gov/hipaa/for-individuals/faq/187/what-does-the-hipaa-privacy-rule-do/index.html?language=en.

[36] Office of the National Coordinator for Health Information Technology. 2018 Model Privacy Notice, 2018. https://www.healthit.gov/sites/default/files/2018modelprivacynotice.pdf.

[37] Marie C Pollio. The inadequacy of HIPAA's Privacy Rule: The plain language notice of privacy practices and patient understanding. *New York University Annual Survey of American Law*, 60(3):42, 2005.

[38] Louis M. Rea and Richard A. Parker. *Designing and Conducting Survey Research: A Comprehensive Guide*. Jossey-Bass, fourth edition, 2014.

[39] Office for Civil Rights (OCR). 264-What is the difference between consent and authorization under the HIPAA Privacy Rule, October 2015. https://www.hhs.gov/hipaa/for-professionals/faq/264/what-is-the-difference-between-consent-and-authorization/index.html.

[40] Florian Schaub, Rebecca Balebako, Adam L. Durity, and Lorrie Faith Cranor. A design space for effective privacy notices. In *Proceedings of the Eleventh USENIX Conference on Usable Privacy and Security*, SOUPS '15, pages 1–17, USA, July 2015. USENIX Association.

[41] Katta Spiel, Oliver L. Haimson, and Danielle Lottridge. How to do better with gender on surveys: A guide for HCI researchers. *Interactions*, 26(4):62–65, June 2019.

[42] Magdalena Szumilas. Explaining odds ratios. *Journal of the Canadian Academy of Child and Adolescent Psychiatry*, 19(3):227–229, August 2010. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2938757/.

[43] Jenny Tang, Hannah Shoemaker, Ada Lerner, and Eleanor Birrell. Defining privacy: How users interpret technical terms in privacy policies. *Proceedings on Privacy Enhancing Technologies*, 2021(3):70–94, July 2021.

[44] T. Franklin Waddell, Joshua R. Auriemma, and S. Shyam Sundar. Make it simple, or force users to read?: Paraphrased design improves comprehension of end user license agreements. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, pages 5252–5256, San Jose, CA, USA, May 2016. ACM.

A Artifacts

Artifacts including our interview scripts, surveys, full HIPAA authorization text, and prototype flows are available at https://github.com/chatbot-study/chatbot.

B Interview codebooks

The interview data was coded in two separate phases. The first portion was an empirical and predominantly deductive approach to analyzing participants' actions during the user study tasks. The second was a thematic analysis of the semi-structured interview discussion.

In most cases, a segment could receive multiple codes if it contained multiple pieces of information, except in the case of exclusionary (e.g., *yes/no*) codes.

We also explored a Revoke Consent option as part of these interviews, and there were additional codes applied to that section of the interview, but that topic is outside the scope of this paper, so the codes for that portion are not detailed here.

B.1 Empirical coding of UI interactions

For the empirical portion, a lead coder first broke each transcript down into relevant snippets that corresponded to predetermined sets of task steps. These steps differed slightly between interview phases due to the three distinct versions of the prototype. (The lead coder also used researcher notes and the original video

during this step to clarify user actions.) Each snippet was then assigned codes for the following categories:

- **Type:** *action*, *expectation*, *other user comment*
- **Did without hints**: *yes* or *no*
- **Hints given**: *general "try again" or "keep exploring," suggest location on screen*, *suggest specific element*, *other*, *hint related to Adobe XD prototype interface rather than actual chatbot interface*
- **UI sticking point**: This codebook category was created inductively as we noticed two specific sticking points that appeared repeatedly: *finding the previous tab in the Prototype 1 HIPAA PDF screen*, and *finding a consent-revoking feature in a task outside of the scope of this paper*
- **Comments about usability opinions:** *easy*, *hard/confusing*, *too many steps or too long*, *would rather use phone*, *other*
- **UI expectations met?:** *yes* or *no* (This refers to whether the actual behavior of the UI corresponded to what participants said they expected would happen when they performed their next action.)

B.2 Thematic analysis

Semi-structured interview responses were first chunked by the sections of the interview scripts that they corresponded to, then coded by **topic** and **type**:

- **Topic:** *general info*, *View Resources button* (Prototype 1 only), *Resources text* (Prototype 1 only), *HIPAA authorization button*, *authorization text*, *third party sharing*, *data uses*, *Revoke Consent* (outside of the scope of this paper), *other*
- **Type:** *general info*, *understandings*, *feelings*, *reading comment*, *revoke consent understandings* (outside of the scope of this paper), *other*

Each chunk was then assigned a code from at least one of the following categories, depending on what type of information it contained. Some of these codes were focused on categories of answers to specific questions, some of which were just intended to provide background information—e.g., *frequent internet user*, *most frequent device*, *uses healthcare tools*—or to specific discussion topics related to factual understandings, e.g., *who gets data*. We also included higher-level thematic categories that were not tied to particular questions: for example, *privacy feelings valence* and *privacy feelings categories*, both of which were codes that could be applied anytime a participant discussed privacy-related sentiments.

B.2.0.1 General information codes

- **Frequent internet user**: *yes* or *no* (All interviews ultimately received a "yes" here)
- **Device used most frequently for accessing the internet:** *computer*, *tablet*, *smartphone*, *other/unknown*
- **Uses healthcare tools regularly?**: *yes*, *no*, *other/unknown*

B.2.0.2 Understanding codes

- **Understanding of HIPAA authorization purpose:** *it's routine/standard*, *inform that data is being collected*, *inform that data may go to third party*, *inform that data can't be shared*, *liability release*, *other*
- **Understanding of data use purpose:** *aggregate metrics / research*, *improving chatbot or "Program"*, *personalization*, *tracking*, *advertising*, *other*
- **Understanding of purpose of partnering with third party:** *doesn't understand purpose of partnership at all*, *AI/ML*, *data storage*, *other*
- **Understanding of View Resources button (Prototype 1 only):** *mostly accurate understanding*, *little or no understanding*, *unclear/other*
- **Who gets data:** *Google Cloud*, *doctors/providers*, *HealthCo*, *other healthcare entities*, *customer service*, *other*
- **Knows data is not covered under HIPAA:** *yes*, *no*, *other/unknown*
- **Knows bot is bot (versus thinking a human is typing answers):** *yes*, *no*, *unclear*

B.2.0.3 Feelings codes

- **Would agree in real life:** *yes*, *no*, *unsure*, *other/unknown*
- **Privacy feelings valence:** *positive*, *negative*, *neutral/ambivalent*, *other*
- **Privacy feelings categories:**
  - **Negative:** *resigned to data collection or privacy invasion*, *concerned about cross-site tracking*, *compromising to meet medical needs*, *concerned about data collection in general (even if only collected by HealthCo)*, *concerned about data storage in general*, *other negative*
  - **Positive:** *data is secure*, *trusts the chatbot*, *other positive*

B.2.0.4 Round 3 only

In Round 3 only, we asked some additional exploratory questions, leading to the addition of the following code categories:

- **First steps participants would take when seeking health insurance information in real life**: *phone*, *search online*, *contact doctor's office*, *health insurance company website*, *other*
- **Types of questions they would feel comfortable asking the chatbot:** *services covered by insurance*, *other*
- **Types of questions they would not ask:** *specific medical conditions*, *location-based questions*, *other*
- **Seems like something they would use in real life:** *yes*, *would rather use phone*, *would rather use some other method*
- **Would they trust this chatbot?:** *yes*, *no*, *other*
- **Reason for trusting or not trusting:**
  - **Yes reasons:** *no reason not to*, *because it's a machine*
  - **No reasons:** *because Google isn't bound by HIPAA*, *because Google may share my info*, *I don't trust Google*
- **Would they expect to have to go through the authorization process every time they used the chatbot?:** *yes*, *no*, *hope not / not sure*, *other*
- **Main points remembered from document seen on previous page:** *something about HIPAA*, *means health info can't be shared*, *authorizes HealthCo to access my data*, *legal disclaimer*, *data being shared*, *data being stored by Google Cloud*, *I can opt out anytime*, *other*

In the third round of interviews, we also introduced an additional, final section of the interview in which we asked participants to return to the authorization text, review it, note anything surprising that they did not notice the first time, and answer a few additional questions. This served as exploratory data to help us build the Part 1 / Part 2 structure of Survey 1. The following codes were applied to responses from that section:

- **Would that affect what types of questions they would be comfortable asking the chatbot?:** *yes (changed type of questions)*, *yes (wouldn't use chatbot at all)*, *no*, *other*
- **Most important things noticed:** *data is safe or other reassuring comment*, *data will be used for other purposes*, *other*
- **Meaning of summary sentences at top:** *data shared with Google Cloud*, *data not covered under HIPAA*, *my permission is needed*, *info can be used for personalization*, *other*
- **Effect of summary sentences on feelings:** *less likely to use chatbot*
- **Meaning of "not subject to HIPAA":** *Google is not subject to same laws as HealthCo*, *Google has more freedom to share data (but not infinite)*, *Google can do whatever they want with no restrictions*, *other*
- **Reason that data is not subject to HIPAA:** *because you willingly shared it*, *because Google Cloud isn't "compatible" with HIPAA*, *they are a separate entity or third party*, *other*
- **Effects of fact that data is not subject to HIPAA on sentiments about using the chatbot:** *less likely to use*, *more likely to use*, *no effect*, *other*

C Survey codebooks

We coded responses to free-response survey questions with codebooks specific to each question. A survey response could generally receive multiple codes from that question's codebook, except in the case of code sets that contain mutually exclusive codes such as *yes* versus *no*.

Surveys 1 and 2 shared codebooks for the following questions, some of which are outside the scope of this paper:

- **Why they would or would not use the chatbot** (Followup to question that asked "Would you use the chatbot" with multiple choice answer options): *useful, easy to use, positive privacy sentiments, not useful, difficult to use, privacy concerns related to Google, privacy concerns related to HIPAA, other privacy concerns, other personal preferences, other*
- **What is your understanding of why a legal agreement is necessary to use this chatbot?:** We initially coded this in a more granular way that is outside of the scope of this paper, but we ultimately created a higher-level set of codes focused on whether participants indicated that the data *was not subject to HIPAA* or *was subject to HIPAA*, whether they made some *other HIPAA-related comment*, whether they *did not mention HIPAA* at all, or whether the question was left *blank*
- **What is your understanding of why HealthCo has partnered with Google Cloud for this chatbot?:** coded broadly to indicate

whether responses indicated a *mostly correct* understanding of this, an *incorrect understanding that assumed nefarious motives* or *some other incorrect understanding*, or whether they were *too vague to categorize*

- **Under what circumstances would data about your use of the chatbot not be subject to HIPAA?:** *if aggregate or demographic, if PHI, when shared with Google Cloud, if it's an emergency, legal reasons broadly, in case of subpoena, if not medical, when shared with third parties, if I consented, if not a healthcare provider, if about insurance plan, if deceased, if not personal or PII, if it's in the cloud, the data IS subject to HIPAA, the data is never subject to HIPAA, other, I don't know, none/NA/empty*
- **What does the following sentence mean to you?: "Data stored by Google Cloud is not subject to HIPAA.":** coded broadly to indicate whether response suggested that this meant *more restrictions on data use*, *less/no restrictions on data use*, or *did not indicate either of those*
- **If you have any other thoughts or feedback about this process or the information you viewed, please let us know here** (Final question of survey): topics observed were *requests for visual changes*, *privacy concerns*, *reading*, *optimistic sentiment*, *phone comment*, *technical comment*, *HIPAA*, *Google*, *participant plans to change future behavior*, *other*, *none/NA/empty*

Survey 1 had codes for these questions that did not appear in Survey 2:

- **What was easy or hard about using the chatbot?:** *generally easy to use, easy to find, generally hard to use, hard because of reading, hard because of consent process, hard because of HIPAA concerns, hard to see, other*
- **Please describe anything you would have changed about the process to get to the chatbot:** visual change, reading, consent/agreement, Google, other, nothing
- **What additional information would have been helpful?:** *UI, real versus bot, how to use or how bot can be used, insurance details, alternatives to bot, data storage, less text, limitations of bot, opt-out option, disclosure of disclosure, privacy, other, nothing/blank/I don't know*
- **What makes you feel that way? (Followup to a Likert item: "I am confident that the privacy of my data will be protected if I use the chatbot."):** We went through an iterative process of assigning codes broadly to all responses for this question and then refining the codebook further and focusing on coding responses according to the nature of the changes that participants made to their answers between Part 1 and Part 2 (i.e., after reviewing the text again more carefully).
  - **Codes for responses that corresponded to no change in Likert response or to a negative change:** *resignation, not subject to HIPAA, not protected by laws (without naming HIPAA), can sell data, data shared with Google Cloud, data shared to multiple third parties, data shared generally, Google Cloud employees can access data, security concerns, can use data for ads, other*
  - **Codes for responses corresponding to positive change in Likert response:** *protected by HIPAA, Google employees can't access, can't use for ads, authorization's explanations or thoroughness, Google Cloud can't share data, can't sell, encryption, other data protections*

Survey 2 had codes for these two questions that were asked only to participants in the Prototype 4 condition that omitted the word "HIPAA":

- **When the information provided mentioned "federal healthcare privacy laws," what laws do you think it was referring to?**: *HIPAA*, *other named concept* (e.g., "PHI," "OSHA," "doctor-patient confidentiality"), *other*, *I don't know*
- **What does the following sentence mean to you? "Data stored by Google Cloud is not subject to federal healthcare privacy laws."**: coded broadly to indicate whether response suggested that this meant *more restrictions on data use*, *less/no restrictions on data use*, or *did not indicate either of those*

In Survey 1, some of these questions were included only in Part 1 or 2, and some were repeated in both parts so that participants had the option to change their answers after reviewing the authorization text again. When the questions were asked in both parts, we wrote scripts to identify whether answers had changed, and if they had, we coded the Part 2 answer separately from the Part 1 answer so that we could assess the nature of the changes. (This is not applicable to Survey 2, which did not contain a text-review step or re-ask any questions.) To see the full survey questionnaires, please visit the online artifacts (Appendix A).

D Tables

	Age			Gender*			Education	Specialized Knowledge		
	Min	Median	Max	Women	Men	NB/GNC	$\geq$ 4-yr degree	Tech	Health	Law
Interviews (n=18)	18	37	76	8 (44%)	9 (50%)	1 (6%)	14 (78%)	5 (28%)	2 (11%)	0 (0%)
Survey 1 (n=761)	18	31	79	384 (50%)	345 (45%)	29 (4%)	414 (54%)	179 (24%)	117 (15%)	32 (4%)
Survey 2 (n=456)	18	44	82	221 (48%)	221 (48%)	13 (3%)	276 (61%)	123 (27%)	89 (20%)	18 (4%)

Table 1. Demographics for each study phase.
*We implemented an inclusive gender question according to Spiel et al.'s recommended practices [41]. To simplify this table, "Woman" and "Man" in this table include those who checked only "Woman" or "Man." "NB/GNC" includes those who selected "Non-binary," "Agender," "Genderqueer," or "Genderfluid," who self-described, or who selected multiple gender options. Percentages may not add to 100% due to rounding and omission of "Prefer not to respond" answers.

	Yes / Agree	Significant diff. between conditions?	p	V	Prototype 1	Prototype 2	Prototype 3
Based on the information you saw, would you use this chatbot?	63.1%	No	0.90	–	–	–	–
I found it easy to access this chatbot.	89.6%	Yes	<0.01 (P1-P3)	0.15	83.6%	90.4%	94.7%
The process to get to the chatbot took too long.	12.8%	No	0.07	–	–	–	–
I would prefer to find health insurance information in another way instead of using the chatbot.	42.3%	No	1.00	–	–	–	–
I am confident that the privacy of my data will be protected if I use the chatbot.	44.6%	No	1.00	–	–	–	–

Table 2. Willingness to use question and Likert items from Survey 1, Part 1. Summary stats and Chi-squared comparisons between prototype conditions.

	Yes / Agree	Significant diff. between conditions?	p	V	Prototype 1	Prototype 2	Prototype 3
Based on the information you saw, would you use this chatbot?	43.5%	No	0.30	–	–	–	–
I would prefer to find health insurance information in another way instead of using the chatbot.	61.3%	No	1.00	–	–	–	–
I am confident that the privacy of my data will be protected if I use the chatbot.	27.4%	No	1.00	–	–	–	–

Table 3. Likert items from Survey 1, Part 2. Summary stats and Chi-squared comparisons between prototype conditions.

	% Yes or Agree, Part 1	% Yes or Agree, Part 2	Sig. change?	p	OR
Based on the information you saw, would you use this chatbot?	63.1%	43.5%	Yes	<0.01	0.10 (large)
I would prefer to find health insurance information in another way instead of using the chatbot.	42.3%	61.3%	Yes	<0.01	17.1 (large)
I am confident that the privacy of my data will be protected if I use the chatbot.	44.6%	27.4%	Yes	<0.01	0.08 (large)

Table 4. Understanding of data use and terms: Survey 1: Comparing Part 1 (initial task context) to Part 2 (after reviewing text carefully). Results were obtained using McNemar tests.
* This item did not have enough discordant pairs to run a valid McNemar test as the rate of this response was very low in both parts.

	Yes	Significant diff. between conditions?	p
Based on the information you saw, would you use this chatbot?	65.4%	No	1.00

Table 5. Willingness to use question from Survey 2: Summary stats and Chi-squared comparisons between Prototypes 3 and 4.

	Overall	Significant diff. between conditions?	p	V	Prototype 1	Prototype 2	Prototype 3
Correct: Know Google Cloud gets data	39.2%	No	0.25	–	–	–	–
Believe no one gets data	4.1%	No	1.00	–	–	–	–
Believe GC employees can look at data	21.3%	No	1.00	–	–	–	–
Believe no one can look at data	13.5%	No	1.00	–	–	–	–
Believe data can be shared with other companies	19.3%	No	1.00	–	–	–	–
Believe data can be sold	28.1%	No	1.00	–	–	–	–
Believe data can be used for ads	26.5%	No	1.00	–	–	–	–
Correct: Know data is not always subject to HIPAA	20.9%	No	0.06	–	–	–	–

Table 6. Understanding of data use and terms: Survey 1, Part 1: Summary stats and comparisons between prototype conditions. Prototypes were compared using Chi-squared tests. In Survey 1 Part 1, none of these comparisons between conditions showed significant results.

	Overall	Significant diff. between conditions?	p	V	Prototype 1	Prototype 2	Prototype 3
Correct: Know Google Cloud gets data	79.1%	No	0.98	–	–	–	–
Believe no one gets data	2.1%	No	1.00	–	–	–	–
Believe GC employees can look at data	53.9%	No	0.51	–	–	–	–
Believe no one can look at data	7.9%	Yes (P1-P2, P1-P3)	0.03	0.15	2.8%	10.4%	10.3%
Believe data can be shared with other companies	38.1%	Yes (P1-P2, P1-P3)	<0.01	0.27	56.0%	33.3%	25.6%
Believe data can be sold	36.0%	Yes (P1-P2, P1-P3)	0.02	0.14	45.2%	29.7%	33.2%
Believe data can be used for ads	33.8%	Yes (all pairs)	<0.01	0.18	44.0%	23.7%	33.2%
Correct: Know data is not always subject to HIPAA	59.1%	Yes (P1-P2, P1-P3)	<0.01	0.30	38.8%	73.1%	65.3%

Table 7. Understanding of data use and terms from Survey 1, Part 2. Summary stats and Chi-squared comparisons between prototype conditions.

	Part 1 Overall	Part 2 Overall	Sig. change?	p	OR
Correct: Know Google Cloud gets data	39.2%	79.1%	Yes	<0.01	51.67 (large)
Believe no one gets data	4.1%	2.1%	*	*	*
Believe GC employees can look at data	21.3%	53.9%	Yes	<0.01	8.52 (large)
Believe no one can look at data	13.5%	7.9%	Yes	<0.01	0.36 (small)
Believe data can be shared with other companies	19.3%	38.1%	Yes	<0.01	9.94 (large)
Believe data can be sold	28.1%	36.0%	Yes	<0.01	2.62 (small)
Believe data can be used for ads	26.5%	33.8%	Yes	<0.01	2.90 (small)
Correct: Know data is not always subject to HIPAA	20.9%	59.1%	Yes	<0.01	58.29 (large)

Table 8. Understanding of data use and terms from Survey 1: Comparing Part 1 (initial task context) to Part 2 (after reviewing text carefully). Results were obtained using McNemar tests.
* This item did not have enough discordant pairs to run a valid McNemar test as the rate of this response was very low in both parts.

	Overall frequency of response	Significant diff. between conditions?	p
Correct: Know Google Cloud gets data	36.8%	No	1.00
Believe no one gets data	4.4%	No	1.00
Believe GC employees can look at data	22.6%	No	1.00
Believe no one can look at data	9.6%	No	1.00
Believe data can be shared with other companies	13.1%	No	1.00
Believe data can be sold	16.7%	No	1.00
Believe data can be used for ads	22.6%	No	0.86
Correct: Know data is not always subject to HIPAA	51.5%	No	0.27

Table 9. Understanding of data use and terms from Survey 2: Summary stats and Chi-squared comparisons between Prototypes 3 and 4. In Survey 2, none of these comparisons between conditions showed significant results.

answer	count	percent
A U.S. federal law that prevents anyone or any company from sharing health data without the patient's permission	299	65.6%
Correct: A U.S. federal law that regulates how healthcare data can be shared	122	26.8%
A U.S. federal law that requires that emergency rooms treat all patients regardless of whether they can pay	6	1.3%
A U.S. federal law that established the American Health Benefits Exchange	4	0.9%
I don't know	25	5.5%

Table 10. Results of Survey 2 multiple choice question: "What is HIPAA?"

answer	count	percent
That no person or company can share it without the patient's permission	266	58.3%
Correct: That a healthcare provider can only share it with another person or organization under certain conditions	111	24.3%
That a healthcare provider can't share it with a third-party company even with the patient's permission	48	10.5%
That I can't share that data with anyone without my doctor's permission	3	0.7%
I don't know	28	6.1%

Table 11. Results of Survey 2 multiple choice question: "If HIPAA defines a piece of information as Protected Health Information, what does that mean?"