Inform the Uninformed: Improving Online Informed Consent Reading with an AI-Powered Chatbot

Ziang Xiao
ziang.xiao@jhu.edu
Johns Hopkins University
Baltimore, United States
Microsoft Research
Montreal, Canada

Tiffany Wenting Li
wenting7@illinois.edu
University of Illinois Urbana-Champaign
Urbana, United States

Karrie Karahalios
kkarahal@illinois.edu
University of Illinois Urbana-Champaign
Urbana, United States

Hari Sundaram
hs1@illinois.edu
University of Illinois Urbana-Champaign
Urbana, United States

ABSTRACT

Informed consent is a core cornerstone of ethics in human subject research. Through the informed consent process, participants learn about the study procedure, benefits, risks, and more to make an informed decision. However, recent studies showed that current practices might lead to uninformed decisions and expose participants to unknown risks, especially in online studies. Without the researcher's presence and guidance, online participants must read a lengthy form on their own with no answers to their questions. In this paper, we examined the role of an AI-powered chatbot in improving informed consent online. By comparing the chatbot with form-based interaction, we found the chatbot improved consent form reading, promoted participants' feelings of agency, and closed the power gap between the participant and the researcher. Our exploratory analysis further revealed the altered power dynamic might eventually benefit study response quality. We discussed design implications for creating AI-powered chatbots to offer effective informed consent in broader settings.

CCS CONCEPTS

• **Human-centered computing** → **Human-Computer Interaction**; • **Computing methodologies** → **Intelligent agents**.

KEYWORDS

informed consent, conversational agents, AI-powered chatbot, human-AI interaction, power dynamic

1 INTRODUCTION

As a core cornerstone of ethics in human subject research [26], informed consent is the process that guards prospective participants' voluntary and informed participation decisions. Through the informed consent process, the participant understands the purpose of the study, procedures to be carried out, potential risks and benefits of participation, the extent of data collection and confidentiality, and their rights. Despite its importance, studies found people often sign the form without a thorough read [28, 52, 58, 63, 70, 76]. In this study, we examined the role of an AI-powered chatbot in improving the online informed consent process.

Consent form reading research, from high-stake studies with severe ramifications to low-stake studies with minimal risk, shows that participants do not thoroughly read consent forms [28, 52, 58, 63, 70, 76]. The participants become less informed when the consent process moves online [33, 63, 65]. For example, Ghandour et al. [34] found that 65% people spent less than 30 seconds reading an online consent form with over 800 words, which should have taken roughly seven minutes. Pedersen et al. [63] showed that compared to the in-person informed consent process, participants' ability to recognize and recall information from the consent form dipped further when the same consent form was presented online. As a result, the current informed consent process may be neither informed nor consensual.

Uninformed participation decisions put both the participant and the researcher at risk. The consequence could endanger the participant's health if they failed to notice the study procedure could induce stress on their pre-existing mental or physical conditions [21]. It could also create privacy risks to participants' personal data if the participant holds incorrect assumptions of the researcher's data-sharing practice [19]. Further, a lack of a good understanding of the consent form may inhibit the participant's autonomy in making the consent decision. Cassileth et al. [13] found more than 1/4 of participants in a medical study thought accepting the consent form was the only option to receive treatment even though the form indicated alternatives. Meanwhile, an ineffective informed consent

process could damage study validity and data quality [23, 41]. Failure to understand the study procedures and purposes may not only negatively impact performance on experimental tasks but also create confounding factors, especially for studies involving deception [23, 41]. Conversely, an effective informed consent process could improve participant engagement and promote trust and rapport between the participant and the researcher [23, 40, 67]. It could reduce the power asymmetry in participant-researcher relations by bridging the information gap of the study, informing participants of their rights, and guarding voluntary decisions [69]. Consequently, a successful informed consent process benefits both the participant and the researcher.

In recent years, especially during the COVID-19 pandemic, online informed consent has become more prevalent. More studies, both online and in-person, collected participants' consent remotely. And those studies may involve risky procedures or collect sensitive information, including people's mental health and personal genetic data. Therefore, improving online informed consent reading is becoming increasingly important.

However, improving consent-form reading is a challenging task, especially in an online environment [3, 29, 63]. Compared to an in-person setting where the researcher could directly interact with the participants, guide them through the consent form, and answer their questions, in an online environment, the absence of the researcher creates communication barriers and significantly demotivates and disincentivizes consent form reading [23, 65, 77]. In the past, researchers have experimented with different interventions to improve consent form reading, including simplifying the content, lowering reading grade level, and introducing interactive features [2, 64, 65]. However, two literature reviews of over 70 past studies suggest that the effect of those interventions was limited, and it is burdensome to design and develop compelling and effective interactive experiences [20, 29]. Therefore, exploring new techniques to improve online consent form reading is necessary.

We examined the role of an AI-powered chatbot [1] in the delivery of consent form content. We built the chatbot Rumi with a hybrid system that combines a rule-based system with AI components. Rumi can greet a participant, go through the consent form section by section, answer the participant's questions, and collect people's consent responses, similar to what an experienced researcher would do in a lab setting. Through a conversational interface, a chatbot could grab people's attention, deliver personalized experiences, and provide human-like interactions [83]. All those features could potentially benefit the online informed consent process. However, chatbots also bear several risks. First, a turn-by-turn chat requires extra time and effort to complete the informed consent process, which is a major challenge in consent form reading [29]. The risk is even higher for studies with paid participants, who would not be rewarded for taking longer to complete the study. Second, current chatbots are far from perfect. Their limited conversation capabilities may deliver incorrect answers or lead to user disappointment and frustration [39]. Therefore, it is yet unknown how a chatbot could affect the online informed consent process.

To explore the effectiveness of an AI-powered chatbot that guides a participant through an informed consent process, we asked three research questions (RQs),

RQ1: How would the participant's consent form reading differ in a study with the AI-powered chatbot-driven consent process vs. the form-based consent process? *(Consent Form Reading)*

RQ2: How would the participant's power relation with the researcher differ in a study with the AI-powered chatbot-driven consent process vs. the form-based consent process? *(Power Relationships)*

RQ3: How would response quality differ in a study with the AI-powered chatbot-driven consent process vs. the form-based consent process? (*Study Response Quality*)

To answer our research questions, we designed and conducted a between-subject study that compared the use of an AI-powered chatbot, Rumi, and a typical form-based informed consent process in an online survey study. Since no previous work has examined the use of chatbots that deliver informed consent, in this study, we focused on examining the holistic effect of a chatbot instead of the effect of individual chatbot features. With a detailed analysis of 238 study participants' informed consent experiences and their responses to the survey study, we found 1) Rumi improved consent form reading, in terms of both recall and comprehension, 2) participants who completed the consent form with Rumi perceived a more equal power-relation with the researcher, 3) by reducing the power gap, the improved informed consent experience ultimately benefited study response quality.

To the best of our knowledge, our work is the first that systematically compared the *holistic* effect of an AI-powered chatbot-driven informed consent process with that of a typical online informed consent process. Our work provides three unique contributions.

(1) An understanding of the holistic effect of AI-powered chatbots conducting online informed consent. Our findings extend prior work on creating an effective informed consent process and reveal the practical value of using an AI-powered chatbot for informed consent, especially in improving consent form recall and understanding, researcher-participant relation, and study response quality. We further provided empirical evidence that could attribute observed improvement in study response quality to the reduced power gap.

(2) Design implications of creating effective AI-powered chatbots for informed consent. We discussed design considerations, such as personalized reading assistance and power dynamics management, to further improve the online informed consent process.

(3) New opportunities for creating and operationalizing AI-enabled, consent experiences for a broader context. The demonstrated effectiveness of the chatbot-driven informed consent opens up opportunities for employing AI-powered chatbots for other types of consent in the age of AI.

2 RELATED WORK

2.1 Consent Form Reading

The concept of informed consent is embedded in the principles of many ethical guidelines including the Nuremberg Code, The Declaration of Helsinki, and The Belmont Report [26]. Four core elements

---

[1]In contrast to traditional rule-based chatbots, we define an AI-powered chatbot as a chatbot that leverages artificial intelligence (AI) technologies, including machine learning, natural language processing, and advanced analytics in the building process or the deployment environment of the chatbot.

ensemble the informed consent process, including disclosure, comprehension, voluntariness, and competency. Through the informed consent process, participants will learn about the study's purpose, procedure, risks, and benefits to make an informed decision.

Despite its importance, people often sign the consent form without a thorough read, regardless of whether it is about a clinical trial that may risk their physical and mental health or a study about their political opinions. For example, Lavelle-Jones et al. [52] found that 69% of patients preparing to undergo various surgical procedures signed the consent form without reading it carefully. Varnhagen et al. [76] showed that in a study regarding technology use, study participants could only recall less than 10% of the information contained in the form, and 35% of participants reported that they only skimmed the form or did not read it at all. Cummings et al. attributed the discrepancy between people's study participation decisions and their concerns regarding confidentiality, anonymity, data security, and study sensitivity in studies with open-data sharing practices to participants' inattentiveness to the consent form [19]. Consent form reading has become increasingly challenging as it moves online [33, 63, 65]. An online consent form is low-cost and easy to administrate, in addition to its broaden-reach across the internet. However, compared to the in-lab setting, no researcher will guide the participant through the consent form, explain the content, and clarify the participant's question [2, 20, 29]. Therefore, online informed consent often yields less informed participation decisions, especially when the study is more complicated and riskier.

Ineffective informed consent reading not only puts participants under unaware risks but also harms the study's validity and data quality [23, 40, 67]. Consent forms contain important information about study procedures and purposes, and comprehending such information may determine the success of later study manipulation, especially for deception studies with cover stories. Unaware confounding factors may further contribute to the replication crisis in social sciences.

Altogether, while the informed consent process plays a vital role for both participants and researchers in a research study, the current practice of conducting an informed consent process has many weaknesses, especially since more studies have started to collect participants' consent online. In this study, we aim to improve the online consent form reading by exploring novel interaction techniques, e.g., an AI-powered chatbot.

2.2 Improving Consent Form Reading

While many prior studies have focused on improving consent form reading, the study results were not always consistent [20, 29]. One group of researchers focused on the design of the consent form, including text readability [9, 37, 59], length [24, 64, 65], layout [18], and media [1, 31, 71]. Dresden and Levitt [24] found study participants could retain more information from a consent form with less unnecessary information and simpler vocabulary. However, other studies found a concise consent form may not yield a higher comprehension score [9, 64]. Although study participants advocate for a shorter form with simpler language, due to the regulation and the nature of a study, it is often difficult to achieve [33]. Researchers also explored converting text-based forms into multimedia. For example, Friedlander et al. found the utility of using video

to deliver a consent form in increasing people's engagement and comprehension [31]. However, there is no conclusive evidence of its effectiveness according to multiple meta-analyses [29, 62].

Another group of researchers brought interactivity into the informed consent process to create an engaging and personalized experience that facilitates consent form reading [2, 4, 7]. In an in-person setting, letting the researcher go through the consent form and answer the participant's questions is deemed the most effective and desirable [2]. As more studies move online, researchers have started to explore new interactive features to improve online consent [4, 7]. One effective intervention is to test the participant's knowledge about the consent form before signing [49]. Further, Balestra et al. used social annotations in online consent forms to facilitate consent form comprehension [4]. Bickmore et al. built an embodied agent that can explain a medical consent form to the reader and found people to be more satisfied with the agent that can tailor its explanation to the participant's existing knowledge [7]. Although not all attempts were successful, many studies emphasized the importance of interaction with the researcher to ensure participants' understanding of the consent information and to foster trust, especially for complex and risky studies [14].

We built `Rumi` with a state-of-art hybrid chatbot framework with AI-powered question-answering modules. Compared to prior work [7, 87] where the agent largely relies on a rule-based system without taking natural language input, recent language technologies enabled `Rumi` to answer a diverse set of questions in natural language and deliver engaging experiences with multiple conversational skills (for more details, see Sec. 3.3). Such a chatbot opens up a new opportunity to bring humanness back to the informed consent process and make the informed consent process more effective.

2.3 Power Relation in Human Subject Research

In a power relation, the person with lower social-formative power is often constrained by their superior counterpart [30]. In the context of human subject research, the researcher has often been viewed as having total authority and being able to decide the resource distribution whereas the participant is often sitting at the lower end [45]. The power asymmetry between the researcher and the participant results from the researcher's control over the participant's recruitment, treatment, data, and compensation [12]. Such a power gap inhibits the participant's autonomy, decreases study engagement, and deters authentic answers [12, 45].

Many researchers advocate for power redistribution to close the power gap in human subject research for both ethical and data quality considerations [25, 79]. By reducing the power gap, the participants could be more engaged during the study, more comfortable disclosing their true thoughts, and more cooperative with the study procedure, which may ultimately benefit the study quality. For example, Chen analyzed the interviewer's language use and found the reduced power gap encourages data richness [16]. However, some studies warn of the importance of maintaining the distance between the researcher and the participant for professional judgment [56, 75].

By sharing information about the study procedure, clarifying risks and benefits, and elaborating on the participant's rights, the informed consent process is designed to close the information gap

and ensure the participant's autonomy [69]. Additionally, as one of the earliest interactions happens between the researcher and the participants, the informed consent process provides an excellent opportunity to redistribute power and establish trust. Kustatscher [51] used visual aids to improve the informed consent for children and found such an engaging process altered the power relation and created a more comfortable environment for information disclosure.

We further extended prior knowledge on the power relation in the research setting by showing the improved informed consent process with an AI-powered chatbot could close the power gap. Through our path analysis, we found the observed effect on study quality can be attributed to the altered power relation.

2.4 Conversational AI as a Research Tool

Recent advances in natural language processing enabled more powerful chatbots for researchers. An emerging application is using chatbots as a tool for research success, including in-depth conversational surveys online [22, 83], or field studies in the real world [47, 72, 78]. Compared to traditional form-based interactions, a chatbot retains the advantage of scalability while providing more engaging and personalized experiences [83]. Specifically, the conversational interface provides interactivity through a turn-by-turn chat, which allows a chatbot to frame questions in a more personalized, conversational way, deliver human-like social interactions to encourage self-disclosure, and probe for in-depth information [81]. Tallyn et al. [72] demonstrated the effectiveness of using a chatbot to gather ethnographic data in the absence of a human ethnographer. Xiao et al. showed that compared to a form-based survey, an AI-powered chatbot could manage a conversational survey, collect higher quality data, and deliver a more engaging survey experience [83]. Chatbots have also been used for other research scenarios, including building psychological profiles [86], delivering interventions [27], and instructing study procedures [50]. For instance, psychologists used a chatbot to infer a participant's personality through conversation to avoid faking in questionnaire-based assessment [82, 86]. Lee et al. built chatbots to practice journaling in the context of mental health research [53]. In summary, past studies have shown a chatbot can potentially serve as a moderator like a research assistant or an interviewer to proactively manage a study process to collect high-quality data for research success.

Our study examined the utility of chatbots in another critical step in human subject research, the informed consent process. We demonstrated a new opportunity and provided design implications to streamline the use of conversational AI in research. We offered design implications to build better future AI for social science.

3 METHOD

To answer our RQs, we designed a between-subject study that compared the outcomes of two methods to deliver the consent form online, an AI-powered chatbot (*Chatbot Condition*) and a typical form (*Form Condition*), on consent form reading, participant-researcher power relation, and study response quality.

3.1 Dummy Survey Study Design

To understand how an AI-powered chatbot facilitates online informed consent reading and how it might influence study response

quality, we have to separate the study quality evaluation from the consent form reading evaluation. Therefore, we designed a dummy study dedicated to evaluating study response quality.

The dummy study is about problematic social media use. To complete the study, the participant first read a short article about problematic social media use. The goal is to familiarize participants with the issue. Then, the participant answered a survey with both close-ended questions and open-ended questions. The choice-based questions included six five-point Likert Scale questions adopted from Bergen Social Media Addiction Scale [55]. The open-ended questions are adopted from [6] with the goal of understanding people's attachment to social media and how it affects people's real life. Both question sets are widely used in social media research. We choose this topic for four reasons. First, it relates to most people online and is suitable to conduct online. Second, we could vary the level of psychological discomfort and data sensitivity to simulate a wider range of online studies. Third, the survey method is the most widely used research instrument, ensuring our finding's generalizability. Lastly, prior studies have provided us with established methods to robustly measure the response quality of a survey with both open-ended questions and choice-based questions [60, 83]. We also considered a genetic study used by [4] that deals with genetic data. Although mishandled genetic data may cause more tangible consequences, the level of risks is difficult to vary and the generalizability is limited.

We designed three versions of our dummy study for three common risk levels in online survey studies,

- *Low* - Non-sensitive data without personal identifiers
- *Medium* - Sensitive data without personal identifiers
- *High* - Sensitive data with personal identifiers

For the *Low* risk version, the survey will ask about people's *opinions* regarding others' problematic social media use. This version is designed to evoke minimal psychological discomfort by asking for opinions about other people instead of directly recalling their own experiences. For the *Medium* risk version, the participant will answer questions regarding *their own* problematic social media use. And in the *High* risk version, we will ask participants to additionally reveal their social media handles as a personal identifier for a follow-up study. The distinction was made clear in the study description and potential risks in the consent form.

3.2 Consent Form Design: Form Condition

The consent form was based on the Social Behavioral Research Consent Form template provided by the Institutional Review Board (IRB) at the University of Illinois Urbana-Champaign. We improved the design of the form-based consent form based on recommendations from prior work [18, 24, 64, 65]. Specifically, we broke the consent form into sections to reduce information overload and used a clear and simple format to ensure clarity and readability.

3.3 Consent Form Design: Chatbot Condition

We created a chatbot, Rumi, with the goal of simulating an in-person informed consent process experience where a researcher goes through the consent form section by section, asks if the participant has questions, and makes clarifications. In our study, Rumi first greeted the participant and informed the participant that it could

take questions at any time during the informed consent process. Then, Rumi went through the informed consent form section by section with the exact content in the Form Condition. Participants can click the "Next" button to proceed to the next section or type in the text box with their questions or other requests. During the process, Rumi proactively asked if the participant had any questions twice, one after the risk section and one at the very end. Participants can skip by pressing the "No Questions" button. Then, Rumi confirmed the participant's age and elicited for their consent to join the study. We included a video in the supplementary material to demonstrate how Rumi conducts the informed consent process.

We adopted a hybrid approach to build Rumi by combining a rule-based model with AI-powered modules [80]. We made this decision for the following reasons. First, rule-based models have limited capability to recognize participants' questions and deliver diverse and engaging responses, let alone handle non-linear conversations (e.g., side-talk, request to repeat information, etc.). Second, although an end-to-end generative system could lead to a more engaging experience, it may produce incorrect or non-factual answers. Those answers may cause severe issues in high-stake scenarios [8, 10], e.g., delivering a consent form. Third, the consent form content is typically specific to the situation or context in which the form is being used. Such a few-shot or zero-shot setting poses another challenge to building an effective generative model, even fine-tuning pre-trained models (e.g., DialoGPT [85]). As a result, we decide to prioritize the answer authenticity and build Rumi on a hybrid system with AI-powered modules to enable Rumi's ability to handle a broad set of questions and provide diverse and accurate answers.

Specifically, we built Rumi on the Juji platform, a hybrid chatbot building platform. Juji provides built-in AI-based functions for dialog management and effective Q&A [44]. Rumi will follow a rule-based conversation agenda to go through the consent form section by section and acquire the participant's consent. Given a question, Juji offers pre-trained Natural Language Understanding (NLU) models to identify relevant questions with known answers in a Q&A database and returns an answer or a follow-up question for clarification. When the chatbot is unsure about how to answer a question, it will recommend similar questions to give participants a chance to obtain desired answers and to learn more about the chatbot's capabilities [44]. Juji also offers a diverse set of conversation skills such as handling side talks and conversation repair to provide an engaging conversation experience [83].

We curated the Q&A database by creating a set of seed questions ourselves and piloting Rumi with 54 online participants. Since the goal is to gauge potential questions participants may ask, we asked our participants to ask as many questions as possible. Researchers on the team wrote answers for each question based on the consent form and added Q&A pairs into the database. Before deploying Rumi for this study, the Q&A database contains over 200 Q&A pairs. To further enhance Rumi ability to recognize questions that may be differently phrased, we leveraged a text-generation model GPT-3 (text-davinci-002) [61] to create question paraphrases. Similarly, we used the GPT-3 to create a candidate answer set for each question. We augmented the Q&A database with five question paraphrases and five candidate answers for each Q&A pair. We hand-checked all generated texts to ensure information authenticity.

3.4 Study Procedure

The complete study consists of two sections (Fig 1) and is approved by the IRB. In the first section, participants were randomly assigned to one of the conditions (Chatbot Condition vs. Form Condition) and one of the risk levels (Low vs. Medium vs. High). The study started with the informed consent process based on the assigned condition. Upon consent, the participant completed the Dummy Survey Study.

Our consent form evaluation is in the second section. The participants answered questions about their understanding of the consent form, their perceived relationship with the researcher, the informed consent process experience, and demographic information. In the end, we debriefed the participants with the real purpose of the study and asked them to complete an additional consent form about sharing their answers to the second section of the study. Before leaving the study, we also asked participants to give open-ended comments on their informed consent.

If the participant decided not to join the study in the first section, they were debriefed on the real purpose of the study and asked if they were willing to join our second section to evaluate their consent form reading. If they agreed to join, we would direct them directly to section two to complete the rest of the study.

3.5 Measures

3.5.1 ***Consent Form Reading.*** We measure participants' consent form reading in two dimensions, *recall* and *comprehension*.

Recall: The ability to recall information from the consent form suggests that people pay attention to the consent form. Similar to [23], to assess participants' recall of the form, we inserted two random statements "*watch a video with orange cat*" and "*read study materials in a blue background*" into the middle of the study procedure section and the risks section of the consent form respectively. Participants were tested on their recall of the color words, "*orange*" and "*blue*". We decided to use recall of two random statements as our measure of consent form reading because it could sensitively measure thoroughness and making a correct guess is difficult [23]. Studies showed participants were more likely to read the procedure and risks sections carefully [23, 76], which provides us an opportunity to measure the lower bound of consent form reading. The participants were asked to select one or two color words from five common color words ("green", "red", "white", "orange" and "blue"). The participants would receive a score of 2 if both color words were selected correctly, a score of 1 if the participants selected only one option and the answer was correct, and a score of 0.5 if the participant selected two options but was partially correct. Otherwise, they got a 0.

Comprehension: Participants' comprehension of the consent form reflects the effectiveness of reading. To comprehend the consent form, the participants need to process the text, understand its meaning, and integrate it with their prior knowledge [36]. Inspired by [36], we measured comprehension with six questions that required participants to process the information presented in the consent form beyond simple recollection. These questions asked about the study procedure, potential risks, and actions to take if a certain scenario happens, e.g., how the participants could protect their privacy if a data breach happens or what a participant should do if they
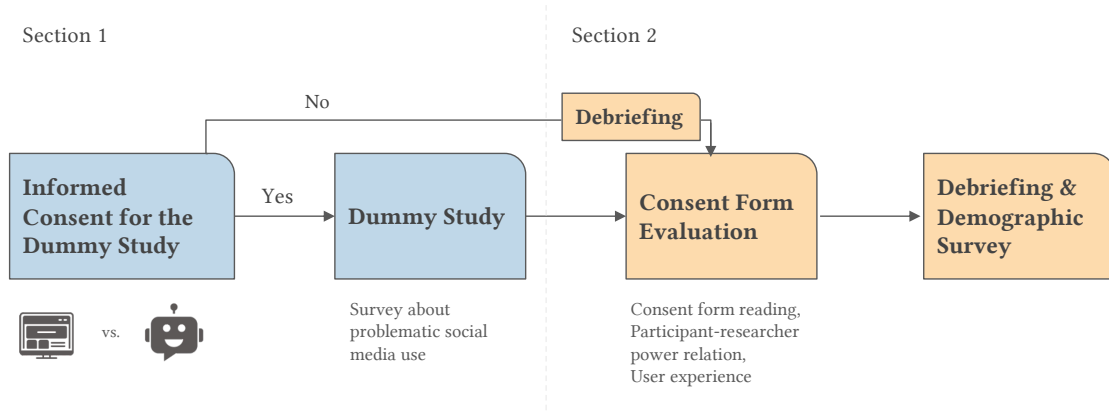
Section 1

Section 2

Figure 1: The figure shows the overall study procedure. In Section 1, based on the assigned condition, the participant interacted with Rumi **or the Form for the informed consent process and then completed a dummy study about social media use. In Section 2, participants answered questions about their consent form reading experience. We debriefed our participants and collected demographic information at the end. If the participants decided not to join the study in Section 1, they would be invited to start Section 2.**

decide to withdraw from the study. A total of four multiple-choice questions were presented. The comprehension score measured how accurately a participant answered those questions. The final score is the percentage of correct answers, ranging from 0% to 100%.

3.5.2 *Participant-Researcher Power Relation*.

We measure the participant-researcher power relation by participants' perceived relationship with the researcher in the study and their feeling of agency and control.

We measure two aspects of the perceived relationship, Partnership and Trust. Based on [42], we asked if the participants perceived their relationship with the researcher who ran the study as partners. Adopted from [4], we measured the trust by asking to what degree the participants trusted the researcher would handle their data properly.

According to [12, 45], the power asymmetry between research parties inhibits participants' autonomy. By power redistribution, we would expect participants to regain agency and control, two constructs of autonomy. We measured the perceived agency and control from both sense of positive agency and the sense of negative agency. We adapted scales from [74] based on the context of the informed consent process. All items adopted a 7-point Likert scale from Strongly Disagree to Strongly Agree.

3.5.3 *Study Response Quality*.

We measured the study response quality by examining participants' responses to both choice-based questions and open-ended questions in the problematic social media use survey.

Non-differentiation: Non-differentiation is a survey satisficing behavior where the respondents fail to differentiate between the items by giving nearly identical responses to all items using the same response scale [48]. Non-differentiation deteriorates both the reliability and validity of question responses. It further inflates intercorrelation among items within the questionnaire and suppresses differences between the items [84]. We used the mean root

of pairs method [60] to measure the non-differentiation in choice-based questions. We calculated the mean of the root of the absolute differences between all pairs of items in a questionnaire. The metric ranged from 0 (The least non-differentiation) to 1 (The most non-differentiation).

Response Quality Index (RQI): To measure the response quality of open-ended questions, we created a *Response Quality Index* based on [83]. It measures the overall response quality of N responses given by a participant on three dimensions, relevance, specificity, and clarity, derived from Gricean Maxim [38]:

$$\text{RQI} = \sum_{n=1}^{N} \text{relevance}[i] \times \text{clarity}[i] \times \text{specificity}[i]$$
(N is the number of responses in a completed survey) $\qquad$ (1)

Relevance. A good response should be relevant to the context. For an open-ended question, a quality response should be relevant to the survey question. Irrelevant responses not only provide no new information but also burden the analysis process. We manually assessed the relevance of each open-text response on three levels: 0 – Irrelevant, 1 – Somewhat Relevant, and 2 – Relevant.

Specificity. Quality communication is often rich in details. Specific responses provide sufficient details, which help information collectors better understand and utilize the responses and enable them to acquire more valuable, in-depth insights. We manually assessed the specificity of each open-text response on three levels: 0 – Generic description only, 1 – Specific concepts, and 2 – Specific concepts with detailed examples.

Clarity. Clarity is another important axis. Each text response should be easily understood by humans without ambiguity, regardless of its topical focus. We manually scored each free-text response on three levels: 0 – Illegible text, 1 – Incomplete sentences, and 2 – Clearly articulated response.

Coding Process: We coded a total of 1428 open-ended responses. Two researchers with a background in human-computer interaction and expertise in open-ended survey analysis conducted the coding process. They first randomly selected 10% of the responses

	Low	Medium	High	Total
Chatbot Condition	41	38	40	119
Form Condition	39	40	40	119
Total	80	78	80	238

Table 1: The table shows the participant distribution across conditions. Participants were randomly assigned to one condition based on the consent form conditions and the risk level. A total of 238 participants were included in the final analysis.

and created a codebook on the above three dimensions with definitions and examples. Then, two researchers coded the rest of the data independently and were blind to the condition. Krippendorff's alpha ranged from 0.83 to 0.98 for each set of coding. The final disagreement was resolved by discussion.

*3.5.4 **Participant Experience and Demographics**. Time and Effort:* Injecting interactivity often means the participants need to spend extra time and effort to interact with the system, which is a major trade-off [20, 29]. Therefore, to measure the perceived time and effort of the informed consent process, we adapted the ASQ scale with two items [54] on how satisfied people were with the time and effort spent on the informed consent process, which we later averaged into a single score.

Future Use: People's willingness to use the same system in the future is a strong indicator of good user experience and satisfaction. We used a single-item 7-point Likert scale to ask if the participant would use the chatbot or the form to complete an informed consent process in the future.

Demographics: Prior studies on chatbots suggests individual differences in people's experience [83] may moderate their chatbot experience. We collected basic demographic information, including age, gender, education level, and annual household income.

3.6 Participant Overview

We recruited fluent English speakers from the United States on Prolific [2]. Of the 278 participants who opened our link and started the informed consent process, 252 completed the informed form. Two participants in the Chatbot Condition explicitly declined to join the study and left the study immediately.

Out of the 250 participants who started the study, 238 (Denoted as P#) completed the study and passed our attention and duplication check. Our following analysis is based on those 238 valid responses (Table 1). Among those 238 participants, 97 identified as women, 136 identified as men, and 5 identified as non-binary or third gender. The median education level was a Bachelor's degree. The median household income was between $50,000 - $ 100,000. And the median age of participants was between 25 - 34 years old. On average, our participants spent 1.24 mins (SD = 3.03) completing the informed consent process in the Form Condition and 7.75 mins (SD = 7.06) with Rumi. We compensated our participants at the rate of $12/hr.

3.7 Data Analysis

We used Bayesian analysis to compare the distributions of effects on consent form reading (RQ1), the participant's power relation with

the researcher (RQ2), and study response quality (RQ3) between two consent methods. We were motivated to use Bayesian analysis for the following reasons. First, Bayesian models allow us to foreground all aspects of the model; No modeling assumptions need checking that are not already foregrounded in the model description. Second, compared to the null hypothesis significance testing (NHST), the Bayesian analysis focuses on "how strong the effect size is" instead of "if there is an effect". It better fits the exploratory nature of our study. Third, Bayesian models facilitate the accumulation of knowledge within the research community as study outcomes can be used as informative priors later. Kay et al. provide a detailed review of the Bayesian method's advantages in HCI research [46].

We formulated a hierarchical Bayesian model for each outcome measure. We build two types of hierarchical Bayesian models, linear regression models for continuous measures and ordinal logistic regression for ordinal measures. For Recall, Comprehension, Agency and Control [3], Non-differentiation, and RQI, we modeled the data as a Normal distribution and used linear regression models to estimate the Normal distribution means for both the Chatbot Condition and the Form Condition. By contrasting the posterior distributions of the means for the two conditions, we would know how the consent method affects outcome variables. Furthermore, we estimated the effect size of the difference of the posterior distribution with Cohen's d for Bayesian linear regression models.

For Partnership, Trust, Time and Effort, and Future Use, we used ordinal logistic regression models to estimate the posterior distributions of the cumulative odds for a given value on the ordinal scale. Our Bayesian analysis goal was to compare whether the rating distribution was significantly different between conditions with respect to the neutral midpoint of the scale (*Neither agree nor disagree*). This would tell us whether participants were more likely to disagree with the statement in one condition over the other. We constructed the distribution of the difference between the cumulative odds of a rating of 4 or below (the midpoint of a 7-point Likert scale) between the Chatbot Condition and the Form Condition. Negative values indicate that participants in one condition had less odds of providing a neutral or negative response to the other condition. We used this for all the above measures except Recall where we estimated the posterior distributions of the cumulative odds of getting the recall question 50% correct or lower. Based on the posterior distributions, we calculated the Odds Ratio (OR) as the effect size [4].

In all models, we controlled for the following covariates: study risk levels, participants' age, gender, education level, and annual household income. We controlled for these demographics as prior studies on conversational agents suggest individual differences may affect their interaction with a conversational agent [83]. Full mathematical descriptions of each type of model are provided in the Supplementary Material. We performed the Bayesian analysis using NumPyro [5], a popular Bayesian inference framework. We used

---

[2]www.prolific.co

[3]We modeled the perceived Agency and Control as a continuous variable as it is a composite score from an 11-item scale.
[4]We interpreted the magnitudes of odds ratios based on Chen et al. [15] where OR = 1.68, 3.47, and 6.71 are equivalent to Cohen's d = 0.2 (small), 0.5 (medium), and 0.8 (large), respectively
[5]https://num.pyro.ai/

Markov Chain Monte Carlo (MCMC), a stochastic sampling technique to sample the posterior distribution P(θ|D), the distribution functions of the parameters in the likelihood function given the data observations D. In particular, we used the No-U Turn Sampler (NUTS) for sampling.

We supplemented our quantitative analysis with qualitative evidence by analyzing participants' chat transcripts in the Chatbot Condition. We performed the thematic analysis [35] on the questions people asked. A member of the research team first performed open coding on the data and then refined these codes in an iterative and reflexive process. The same person then used axial coding to group these codes into larger categories to extract common themes.

4 RESULTS

Overall, Rumi improved participants' consent form reading. Our participants who interacted with Rumi could recall more information from the consent form and take more correct actions based on the consent form compared to those in the traditional form-based informed consent process. We also found that in the Chatbot condition, participants perceived themselves as having a more equal power relation with the researcher and offered higher-quality responses in the Dummy Survey Study. Our exploratory path analysis revealed a potential mechanism where the chatbot-based consent method improves response quality by reducing the power gap.

4.1 People engaged with Rumi and were satisfied with the experience

Although our participants spent more time chatting with Rumi to complete the informed consent process, they were, in general, satisfied with the Time and Effort (Chatbot: M = 4.72, SD = 1.54; Form: M = 4.48, SD = 1.57) used in completing the informed consent process. They also indicated that they are willing to use such a chatbot for future informed consent experiences (Chatbot: M = 4.21, SD = 1.89; Form: M = 4.03, SD = 1.84). We modeled both measures as ordinal variables and contrasted the posterior distributions of the cumulative likelihoods of a rating of 4 for both conditions. We found that participants' perceived time and effort and their indicated future use did not differ significantly between the two conditions. The High-Posterior Density Interval (HPDI) [6] for the cumulative odds difference overlapped with the ROPE (Region of Practical Equivalence) [7] of 0 ± 0.05 in all cases (Time and Effort: M = 0.05, 94% HPDI: [0.01, 0.08]; Future Use: M = -0.22, 94% HPDI: [-0.49, 0.01]), indicating that participants were not significantly more likely to disagree or agree in one condition over the other for both Time and Effort and Future Use.

---

[6]The HPDI is the location of 94% of the posterior density. It is similar to, but different from, the idea of the confidence interval used in non-Bayesian Statistics. In non-Bayesian Statistics, a 94% confidence interval is informally interpreted as "with 94% probability the parameter of interest lies in a specific interval; the tails are of equal width (i.e., 3%)"; the HPDI is the *densest* interval covering 94% of the posterior. The HPDI is guaranteed to include the most likely value, but this is not always true for confidence intervals; see McElreath [57]. For a more careful definition of the confidence interval, see Hoekstra et al. [43].

[7]Unlike non-Bayesian Statistics, where one can ask, if the two means for two conditions are different $P(\mu_1 \neq \mu_2)$, in Bayesian statistics, one asks if the HPDI of the distribution $P(\mu_1 - \mu_2)$, the distribution of the difference of the means of the two conditions, excludes an interval where we can consider the two treatments equivalent. This equivalence interval is domain-dependent. A posterior distribution HPDI that lies outside the ROPE is considered a significant result in Bayesian data analysis.

We dug into the participants' transcripts to further understand participants' interaction patterns. We found our participants engaged actively with Rumi. Our participants raised a total of 449 questions (M = 3.77, SD = 2.56), and Rumi answered 389 (85.97%) of them. We identified four major categories of questions, Rumi's capability(12.69%), research team information (11.58%), study information (56.15%), and side-talking (19.59%). Our participants ask about what Rumi can do and what questions could Rumi answer (e.g., *"What do you know?"*[P107]). Another type of question is about the research team (e.g., *"Who is [Researcher Name]?"*[P57]). Through those questions, the participants could learn more about the research team to start rapport building. Unsurprisingly, people asked the most questions regarding the study itself. Specifically, our participants asked questions about the study procedure (45.23%; e.g., *"What do I do after the survey?"*[P41]), risks (28.17%; e.g., *"Will my information be safe?"*[P83]), compensation details (12.30%; e.g., *"What will I get after this?"*[P94]), and general information about the study (14.30%; e.g., study purpose, survey topic, etc.). Interestingly, some participants started side-talking with Rumi, such as *"How's your day?"*[P43] or *"I didn't sleep well yesterday. Do you sleep?"*[P67], which suggests an even higher engagement. Those questions indicate our participants were willing to spend the effort interacting with Rumi. And Rumi helped our participant to clarify important information regarding the consent form.

Prior work suggests that introducing interactivity often creates user burdens which may deter user experience, especially for consent form reading where the required time and effort is one major roadblock [20, 29]. Our results indicate that people are willing to actively engage with Rumi to complete the informed consent process, even though it took a longer time than normal online informed consent. We believe scale and speed should not be the only value in the informed consent process. Sacrificing speed and scale for a more engaging experience and grabbing people's attention, especially in this high-stake scenario, is important to consider.

4.2 Rumi lead to better consent form reading

Overall, similar to prior work [23, 63, 64], the consent form reading is poor. In terms of recall, only 26 participants in both conditions in total recalled both "orange" and "blue" correctly from the color phrase list. On average, people scored 0.63 out of 2 (SD = 0.64) on the recall task and scored 53.22% (SD = 24%) on the comprehension task.

Compared to the static form, going through the consent form with Rumi leads to significant improvement in consent form reading. Participants who interacted with Rumi recalled more correct color words from the consent form (Chatbot: M = 0.76, SD = 0.71; Form: M = 0.51, SD = 0.54). We modeled Recall as an ordinal variable and estimated the posterior distributions of the cumulative likelihoods of getting 50% of the total score (1 out of 2). The HPDI for the cumulative odds difference excluded the ROPE of 0 ± 0.05 (M = -3.84, 94% HPDI: [-6.26, -1.55]). The estimated effect size (Odds Ratio) is small (M = 1.75, 94% HPDI: [1.47, 2.02], excluding the ROPE of 0 ± 0.05; See Fig 2). The result indicates our participants could recall better with Rumi.

Participants in the Chatbot Condition also better comprehended the consent form (Chatbot: M = 61%, SD = 22%; Form: M = 46%,
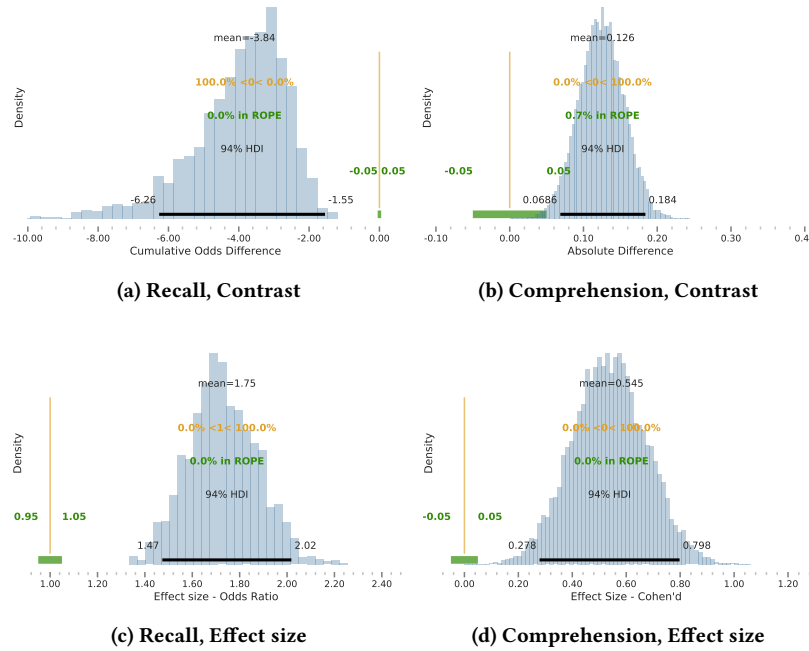
Figure 2: The first row represents the posterior distribution contrasts between the Chatbot Condition and the Form Condition for the cumulative odds of achieving less than 50% accuracy in the Recall task and for the means of Comprehension score. The second row shows the effect size distribution, Odds Ratio for Recall and Cohen's d for Comprehension. Plot (a), (b), and (d) show an orange vertical line located at 0 with green bars indicating ROPE. Similarly, plot (c) has an orange vertical line located at 1 and a green ROPE interval. Effect distribution falling into these ROPE regions suggests no difference between conditions or no effect. Note that the x-axis is not the same scale for all plots. Main finding: Compared to the participants in the Form condition, participants in the Chatbot Condition provided more correct answers in the recall task and achieved a higher comprehension score. The differences are both statistically significant.

SD = 23%). We estimated the posterior distributions of the mean for the Chatbot Condition and the Form Condition. We found the difference between the Chatbot Condition (M = 51%, 94% HPDI: [39%, 62%]) and the Form Condition (M = 51%, 94% HPDI: [27%, 48%]) is statistically significant (excluding ROPE 0 ± 0.05), with a medium to large effect size (Cohen's d) (M = 0.55, 94% HPDI: [0.28, 0.80], excluding a ROPE of 0 ± 0.05; See Fig 2). This suggests that the participants understood the content and could take better actions according to the consent form to protect their rights.

The results answer RQ1 clearly. The chatbot-driven informed consent process improves consent form reading in terms of both recalling information from the consent form and comprehending its content. Two factors may play a role in the observed improvement. First, the participant may be better engaged. The interactive features of Rumi were designed to simulate an in-person experience in which the research assistant actively engaged with the participant. Many of our study participants enjoyed this human-like experience and commented *"the bot is very friendly."*[P61] and *"i liked how the bot talks to me."*[P58]. And participants appreciated that Rumi went through the consent form with them, *"It was easier and nicer to read the consent form with the bot using texts other than a wall of text. Thank you :D"*[P15]. As we showed in Sec. 4.1, in the Chatbot Condition, the participant spent significantly more time during the informed consent process. Although time spent may not always lead to better reading [20], it could suggest higher

engagement which plays a key role in reading comprehension [36]. Second, Rumi is designed to answer people's questions. A total of 449 questions were raised and Rumi answered 85.97% of them. Our participants appreciated Rumi's ability to answer their questions in real-time, *"It's pretty cool that the chatbot can answer my questions right away"[P32]* but some participants mentioned that Rumi cannot fully understand their questions. Although the chatbot has limited capability, the ability to answer people's questions on the fly might contribute to the improved consent form reading.

4.3 Rumi aided participant-researcher relationship

The participants who interacted with Rumi perceived themselves as having a more equal power relation with the researcher in charge of the study. Our results indicated that people in the Chatbot Condition trust the researcher more (Chatbot: M = 5.60, SD = 1.40; Form: M = 4.92, SD = 1.93) and believed their relationship with the researcher is more like a partnership (Partnership) compared to the Form Condition, which indicates a smaller power gap (Chatbot: M = 4.26, SD = 1.70; Form: M = 3.63, SD = 1.74). Similar to Sec 4.1, we treated both variables as ordinal and estimated the posterior distributions of the cumulative likelihoods. Again, we constructed the distribution of the difference between the cumulative odds of a rating of 4 (the midpoint of the 7-point Likert scale) in the Chatbot condition and the cumulative odds of the rating in the Form condition. Based on
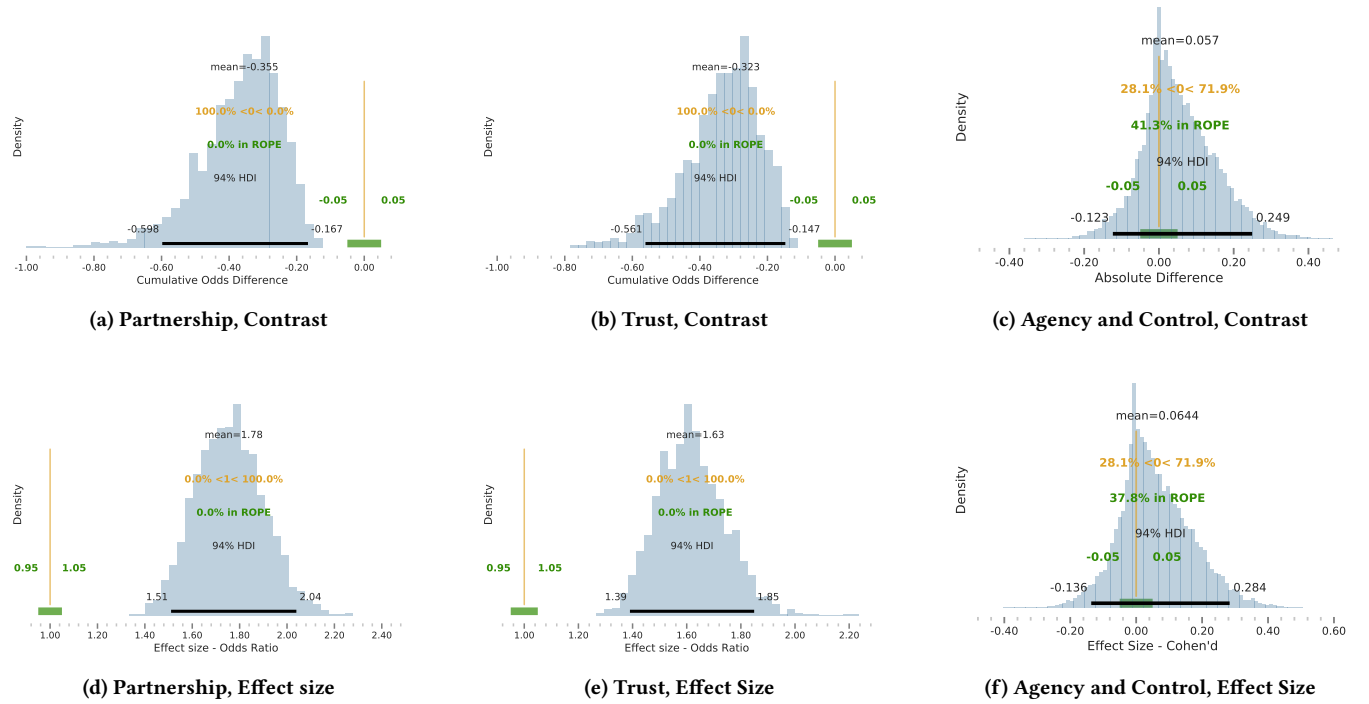
Figure 3: The first row represents the posterior distribution contrasts between the Chatbot Condition and the Form Condition for the cumulative odds of observing a neutral or lower rating in Partnership and Trust and for the means of Agency and Control. The second row shows the effect size distribution, Odds Ratio for Partnership and Trust and Cohen's d for Agency and Control. Plots (a), (b), (c), (f) show an orange vertical line located at 0 with green bars indicating ROPE. Similarly, plots (d) and (e) have an orange vertical line located at 1 and a green ROPE interval. Effect distribution falling into these ROPE regions suggests no difference between conditions or no effect. Note that the x-axis is not the same scale for all plots. Main finding: Compared to the participants in the Form Condition, participants who interacted with Rumi perceived their relationship with the researcher more like a partnership and trusted the researcher more. The difference is significant. Although participants in the Chatbot Condition reported higher perceived agency and control, the difference is not statistically significant.

the estimated posterior distributions, we found the differences of the cumulative likelihoods in both measures are statistically significant (Partnership: M = -0.36, 94% HPDI: [-0.60, -0.17], excluding the ROPE of 0 ± 0.05; Trust: M = -0.67, 94% HPDI: [-1.14, -0.28], excluding the ROPE of 0 ± 0.05). The odd ratio suggests a small effect size for both measures (Partnership: M = 1.78, 94% HPDI: [1.51, 2.04], excluding the ROPE of 0 ± 0.05; Trust: M = 1.63, 94% HPDI: [1.39, 1.85], excluding the ROPE of 0 ± 0.05).

However, we did not observe a significant difference in participants' feelings of agency and control after the informed consent process. The participant who interacted with Rumi (M = 5.15; SD = 0.78) reports a higher feeling of agency and control, measured by a composite score of positive and negative agency, compared to the Form Condition (M = 5.04; SD = 0.91). Since we measured participants' feelings of agency and control with a composite score of an 11-item scale, we treated the score as a continuous variable and contrasted the mean between the Chatbot Condition and the Form Condition. We estimated the posterior distribution of the mean difference and found the difference is not statistically significant (M = 0.06, 94% HPDI: [-0.12, 0.25], overlapping the ROPE of 0 ± 0.05).

Going through the online informed consent process with Rumi increases the trust between the participant and researcher and closes

the power gap. Two potential factors may explain the observed effect. First, as mentioned in [69], a more effective consent form reading could reduce the power gap by bridging the information gap and assuring a voluntary decision. We did observe a significant correlation between participants' consent form reading and their power relation with the researcher (Recall: r(236) = 0.14, p = 0.03; Comprehension: r(236) = 0.21, p < 0.01). The observed difference in the power relation could potentially be attributed to more effective communication. Secondly, the humanness of Rumi's design may help with rapport building between the researcher and the participants, which potentially reduces the power gap [45]. In our study, Rumi is framed as a virtual research assistant and represents the research team. As the first interaction between the researcher and the participant, the informed consent process may also serve the role of rapport building beyond informing the study participant.

4.4 Rumi lead to better survey response quality

Participants who interacted with the chatbot provided higher-quality responses to the dummy survey. For choice-based questions, participants in the Chatbot Condition exhibit less survey satisficing behavior (Chatbot: M = 0.42, SD = 0.22; Form: M = 0.47, SD = 0.25).

(a) Non-differentiation, Contrast

(b) RQI, Contrast

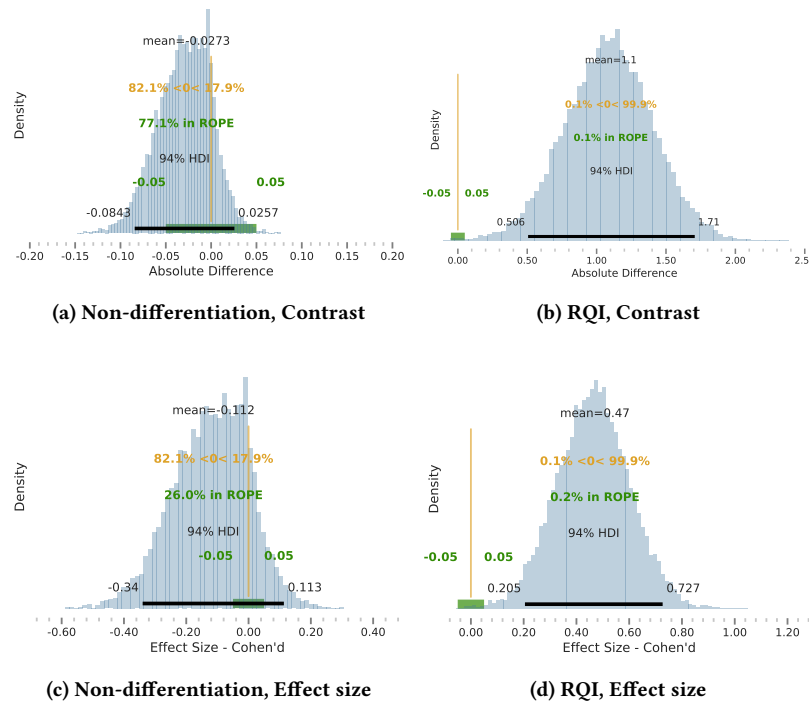**(c) Non-differentiation, Effect size**

(d) RQI, Effect size

Figure 4: The first row represents the posterior distribution contrast of two means between the Chatbot Condition and the Form Condition for the Non-differentiation and Response Quality Index (RQI). The second row shows the effect size distribution, Cohen's d, for both measures based on the posterior distribution. Each plot shows an orange vertical line located at 0 with green bars indicating ROPE. This represents that there was no difference between conditions or no effect. Note that the x-axis is not the same scale for all plots. Main finding: Participants who interacted with Rumi displayed less non-differentiation in close-ended questions but the difference is not statistically significant. We also observed participants in the Chatbot Condition contribute higher quality answers to the open-ended questions. The posterior distributions indicate the difference is statistically significant with a medium effect size.

The posterior distributions show the observed difference is not statistically significant (M = - 0.03, 94% HPDI: [-0.08, 0.03], overlapping a ROPE of 0 ± 0.05).

For open-ended questions, participants in the Chatbot Condition provided higher-quality responses (Chatbot: M = 5.38, SD = 2.29; Form: M = 4.17, SD = 2.43). The posterior distribution on the difference of RQI between two conditions shows a mean contrast of 1.1 with HPDI of [0.50, 1.71]. Since the 94% HPDI lies outside a significant ROPE of 0 ± 0.05, the result implies a significant effect with a medium effect (Cohen's d: M = 0.47, 94% HPDI: [0.21, 0.73], excluding the ROPE of 0 ± 0.05).

The results answered our RQ3. The participants who interacted with the chatbot during the informed consent process contributed significantly higher-quality responses to the open-ended questions. Participants who are in the High-risk condition were also willing to elaborate on their answers. Based on existing work, the observed effect could be due to a more effective consent form reading [21, 23] and a reduced power gap in the researcher-participant relationship [12, 45]. We further examined two potential pathways in Sec. 4.5.

4.5 Reduced power gap may explain the study response quality improvement

Existing studies suggest that the power gap between the researcher and the participant plays a key role in the participant's engagement and level of self-disclosure [12, 45]. A more equal power dynamic promotes trust and enhances the participant's feeling of agency and control, potentially benefiting the study response quality. On the other hand, better consent form reading may also improve study response quality as the participant understands the study procedure better [21, 23]. Therefore, we explore two potential pathways that may mediate the effect of chatbot-driven informed consent on study response quality: through reduced power gap and through improved consent form reading.

We built a Structural Equation Model (SEM) model to answer our question. As shown in Fig 5, the Consent Method (Chatbot vs. Form) was set to predict the Researcher-Participant Power Relation and Consent Form Reading. The Researcher-Participant Power Relation is a composite score of the Researcher-participant relationship and the participant's perceived agency and control. The Consent Form Reading combines recall and comprehension into one single measure with normalization. The Study Response Quality combines Non-differentiation for choice-based questions and Study
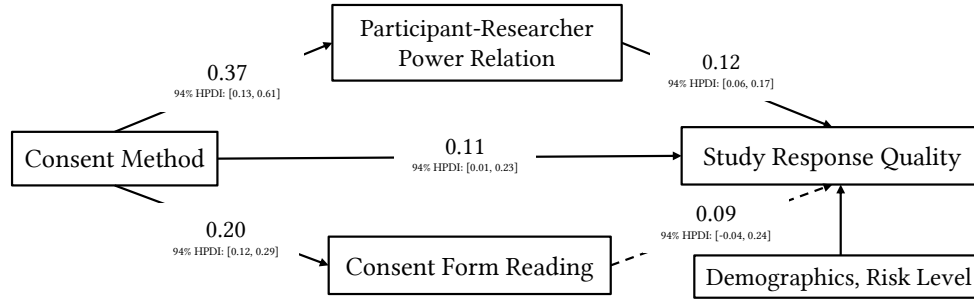
Figure 5: The figure shows a Bayesian Structural Equation Model (SEM) model to understand why chatbot-driven informed consent leads to improved study quality. As suggested by previous literature, we examined two potential pathways, 1) reduced power gap in the researcher-participant relationship, and 2) improved consent form reading. The results indicate a significant partial mediation effect from the reduced power gap.

Response Quality Index for open-ended questions. Both Researcher-Participant Power Relation and Consent Form Reading were set to predict the Study Response Quality, which is also predicted by the Consent Method. We believe that the Researcher-Participant Power Relation predicts Study Response Quality because the power dynamics influence trust and autonomy, two key factors in self-disclosure [12, 45]. We believe that the Consent Form Reading predicts Study Response Quality by how much the participant understands and follows the study instruction. Whether the Consent Method predicts Study Response Quality is our main question; the other two pathways are paths that mediate the effect of the Consent Method on Response Quality. Like all other models, we controlled participants' demographics and risk levels. All variables were treated as manifest variables and modeled as Normal Distributions. We used Bayesian inference to fit the proposed SEM model with the No-U-Turn Sampler [8]. The model fits perfectly, with posterior predictive p-value = 0.48; BRMSEA = 0.01, 94% HPDI: [0.00, 0.06]; BMc = 0.99, 94% HPDI: [0.98, 1.00]; adjB$\hat{\Gamma}$ = 0.99, 94% HPDI: [0.96, 1.00].

Results indicated a significant total effect of the Consent Method on the Study Response Quality (β = 0.18, 94% HPDI: [0.08, 0.29], excluding the ROPE of 0 ± 0.05). Notably, as shown in Fig 5, there was an indirect path from the Consent Method to the Study Response Quality via Participant-Researcher Power Relation (β = 0.12, 94% HPDI: [0.06,0.17], excluding the ROPE of 0 ± 0.05). This path reduces the total effect by 39%. However, the direct effect remained significant, reflecting only partial mediation. Meanwhile, we did not observe a significant path from the Consent Method to the Study Response Quality via the Consent Form Reading (β = 0.09, 94% HPDI: [-0.04, 0.24], overlapping the ROPE of 0 ± 0.05).

The Bayesian SEM model shows that the potential reason that chatbot-driven informed consent improved later study response quality may be the reduced power asymmetry between the researcher and the participant. This finding aligns with prior studies that as one of the early interactions between the participant and the researcher, the informed consent process could bridge the information gap and shape the researcher-participant power relation, which may ultimately benefit study response quality and data richness

[12, 45, 69]. However, the model does not indicate a potential pathway from the improved consent form reading. We believe there are two potential explanations. First, our study procedure is straightforward. An improved consent form reading may not have a strong effect in preparing the participants for the later study. Second, our study is low-stake compared to medical trials. Existing studies' findings on clinical trials may not generalize [21].

5 DISCUSSION

In this study, we found that a chatbot-driven informed consent process could effectively improve consent form reading, reduce the power gap, and ultimately benefit study response quality. In this section, we discuss design implications for a more effective informed consent experience with conversational AI.

5.1 Personalized Chatbot-driven Informed Consent Experience

In a turn-by-turn conversation, a chatbot can ask questions about participants' experiences and preferences to initiate personalization [86]. In our study, we found our participants enjoyed the feeling of a personalized experience. For example, one participant commented, "*Never seen a bot like this. I like the feeling* Rumi *is talking to ME.*" *[P11]*. Given this opportunity, we should consider incorporating personalization into the chatbot-driven informed consent experience. However, we need to be extremely cautious about unwanted persuasive effects [73] that violate the principle of voluntary participation when introducing personalization into online informed consent. The participation decision should be fully voluntary. The goal of personalization should be to facilitate the understanding of the consent form, not to *nudge* the participant to participate.

An AI-powered chatbot could highlight important content based on the participant's previous experiences with consent forms. On the one hand, existing studies suggest that more experienced study participants spent less time reading the consent form and often missed important information [66] because they tended to assume all consent forms were similar. Such incorrect assumptions exposed participants to unwanted risks. In this case, their pre-existing experience hindered consent form reading. On the other hand, regulations, such as General Data Protection Regulation (GDPR), often require a

---

consent form to contain information that a participant may already be familiar with. In this case, a chatbot could summarize those materials to help participants better allocate their attention to new content. Therefore, it will be useful for the chatbot to learn about participants' prior informed consent experiences, analyze the new consent form, summarize contents a participant may be familiar with, and highlight new content to ensure a thorough read. An interesting idea to explore is a centralized chatbot that helps participants to manage all consent forms while preserving anonymity from prior study experiences. Although the IRB requires the researcher to draft the consent form in layman's language, some study procedures are indeed complicated. And sometimes certain terminology is necessary for clear communication, especially for high-stake clinical trials. Participants in our study appreciated Rumi's ability to offer clarifications and answer their questions. Therefore, the chatbot should include interactive features to help participants understand the consent form. We could also borrow a thread of chatbot research focused on education where the chatbot helps people study new content and review materials [17, 68]. For example, a chatbot could first assess the participant's existing knowledge about the study topic to determine the necessary explanations and ask questions at the end to ensure learning outcomes.

However, as time and effort are among the biggest hurdles in consent form reading, the trade-off between the benefits and risks of the interactive features needs proper calibration and consideration. Specifically, we need to consider the time cost a participant will spend on consent form reading. A participant's compensation is often associated with time spent in a study procedure. If the total compensation is fixed, the interactive informed consent procedure will reduce the pay rate. In our study, we compensated participants in the Chatbot condition with an extra bonus at the end to ensure the promised pay rate. Although the research team may need to budget more for each participant, we argue, as the study shows, the improved response quality will ultimately benefit the research results. It saves time for researchers to clean up low-quality data from participants who did not read the study procedure carefully, especially in cases where unattended participants create confounding factors that endanger the study quality [23].

5.2 Managing Power Dynamics

Our results echo prior studies on the role of an effective informed consent process for closing the power gap in the researcher-participant power relation and its benefit in study response quality. We believe that the chatbot could influence researcher-participant relations by adjusting its own power relation with the participant. We could further extend the utility by designing an informed consent chatbot that actively manages the power dynamic.

From a power relation point of view, one's identity has a strong influence on their relative power over their counterpart [30]. Cassell mentioned that a researcher's chosen identity, e.g., Interviewer, Facilitator, Initiator, Researcher, could change the power dynamic with the study participant [12]. We could carefully design an informed consent chatbot's identity to suit various contexts. For example, to reduce the power gap, we could design the chatbot as a research partner rather than as a researcher. However, designing a virtual agent's identity is complicated. Many design dimensions, including

appearance, language style, etc., need to be carefully considered. Any incoherence may mar the entire experience.

The informed consent process could be considered as a negotiation about information disclosure between the researcher and the participant [45]. The researcher holds the information about the study, and the participants gain the knowledge and experience needed to perform the study. Karnieli-Miller et al. pointed out that such negotiation has the potential to change power relations by giving participants more information [45]. Thus, we should prepare an informed consent chatbot with conversation skills for such negotiation, so that the chatbot could understand participants' requests, clarify their information needs, and actively manage information disclosure about the study.

We could further empower the participant by considering the ownership of the informed consent chatbot. In this study, the chatbot acted on the researcher's behalf and as a part of the research team. Although in most chatbot use cases the chatbot is owned by the creator instead of the user, a participant-owned informed consent chatbot may provide several benefits. First, the participant will have total authority over the conversation history. In this case, the participant could have a safe space to ask questions without feeling judged. Second, such an informed consent chatbot could become the central hub for all informed consent needs. It could act on behalf of the participant, analyze consent forms based on the participant's preferences, and proactively ask the researcher questions to satisfy the participant's information needs.

5.3 Combining Human Expertise with LLMs

Many of our participants liked Rumi's ability to respond to their questions in real-time with answers grounded in the consent form content. However, creating a chatbot that can accurately answer people's questions, especially in high-stake contexts, is challenging. Due to limited natural language understanding ability, the current Q&A functionality for most commercial chatbot building platforms relies on a database of handcrafted Q&A pairs. It is especially time-consuming in the informed consent context as participants' questions are specific to the consent form, further limiting the reusability of a Q&A database. Although some questions could be reused, for example, an institution may share the same template, and some study procedures could be similar, future studies are needed to design tools to support such a sharing practice.

Large language models (LLMs), like GPT-3, show promise in a new way to build conversational agents to answer people's natural language questions. Although one could use off-the-shelf LLMs with in-context learning to build a chatbot to answer a wide range of domain-specific questions, LLMs sometimes generate non-factual information and have limited capability to memorize a long document [11]. Both shortcomings should be avoided in high-stake contexts, e.g., delivering consent forms, as non-factual information could mislead a participant to make an uninformed decision. It is not only an ethical concern but also could lead to severe consequences. For example, a participant who agrees to join a study without full knowledge of the specific study procedures may experience unexpected extreme physical or mental stress.

Therefore, we should consider leveraging LLMs carefully with the above shortcomings in mind. One framework to consider is

combining LLMs with human expertise. In our study, we used GPT-3 to augment Q&A pair generation to empower Rumi. To ensure correctness, we acted as a validation layer to check if the GPT-3 generated paraphrased questions and answers were correct and appropriate. The augmented Q&A database enables Rumi to capture more participant questions and delivered more diverse answers. We believe LLMs could facilitate more chatbot development tasks by teaming with human experts. For example, one could use an LLM as a testing tool by generating question sets to identify issues and develop fix. Such method could enable a faster iteration that traditionally relies on bootstrapping conversations on the fly [80]. Besides using human expertise as a gatekeeper, we should study better LLMs control mechanisms for factual Q&A. For example, we could leverage a knowledge-driven approach [32] by parsing a plain text consent form into a structured knowledge graph and using the graph to steer LLMs to generate factual answers that are grounded in the consent form content. Again, given the shortcomings of generative models, we believe a human-in-the-loop framework is preferred to safely take the advantage of generative models for more capable informed consent chatbots.

In summary, future work should study effective human-in-the-loop frameworks that can support research teams, especially teams without AI expertise and resources, to build and test an informed consent chatbot that consistently delivers factual answers.

5.4 Future Directions

5.4.1 Chatbot as a Virtual Research Assistant. In addition to prior work that studies the uses of chatbots for research, including for conversational surveys [83] and for ethnographic studies [72], we did see a future opportunity to build virtual research assistants that could help researchers manage human-subject studies from beginning to end. A virtual research assistant could help researchers reach a worldwide population (if necessary for the study), engage with the participant and build rapport, deliver the intervention, collect high-quality data, and debrief the participants. Beyond moderating the study process, in our study, we found some participants wanted to communicate with the researcher through the virtual agent, one participant said to Rumi, *"let your owner know i like this study."*[5]. Such a virtual research assistant could be especially helpful for longitudinal field studies where keeping participants engaged and collecting high-quality data over a long period of time is particularly difficult and expensive.

However, creating such a virtual research assistant is challenging. First, repeated interactions with a virtual research assistant at different stages of a study pose new challenges to interaction design. The agent needs to adapt and react to unique individual experiences over time. Second, personalization is a double-edged sword in the context of human subject research. Although it could increase engagement, a highly personalized chatbot could induce unwanted confounding factors due to the inconsistency across participants. Third, mediating the communication between the participants and the researchers requires a new interaction paradigm. The agent needs to mediate the communication and actively engage both parties for study success. Though challenging, such a virtual research assistant could help researchers conduct scalable, robust, and engaging human-subject studies.

5.4.2 Online Informed Consent Beyond the Research Context. Personal data, from personal health information to web browsing histories, has become increasingly valuable. It powers millions of intelligent applications. As the world becomes more connected, personal data becomes a new source of power. As a result, while system builders are eager to collect data, policymakers and users are more cautious about personal data sharing. For example, GDPR [9] regulations explicitly require data collectors to ask for users' consent before collecting and storing any personal data. However, the current practice of data consent online is largely flawed [5]. For example, users often need to go through a lengthy document without any guidance. Therefore, how to empower user giving meaningful and informed consent about their data becomes an emerging challenge. In this study, we found an AI-powered chatbot could deliver effective informed consent in the context of human subject research. In the future, we should explore the potential of such an agent for data-sharing consent in broader contexts.

5.5 Limitations

We recognize several limitations in our approach. First, as the first study of this kind, our main goal was to explore the potential benefits and limitations of the informed consent process driven by an AI-powered chatbot. Through an SEM model, we further explored the potential path that may explain our observed effect; namely, the chatbot-driven informed consent process improved the study response quality by altering the power relationship. However, due to the exploratory nature of this study, our study could not infer strong causal relationships. Future confirmatory studies are needed to confirm the observed effects and explain the mechanism.

Second, the scope of our study design was limited to online studies with surveys. Although we designed three risk levels to simulate studies that collect different types of data, compared to high-stake clinical trials that may involve severe ramifications, the risk of an uninformed decision is lower in our case. Two factors may limit the generalizability of such a design. First, people tend to pay less attention to consent forms for lower stake studies [63]. We may observe a smaller difference if people were more attentive to the consent form in both conditions. Second, the study procedure was simple and straightforward, e.g., complete a survey. Although it represents the majority of online studies, some studies may include more complicated study procedures, for example, playing a game, where a good understanding of the procedural details may play a stronger effect in the later study. Thus, we need to study chatbot-driven informed consent under various contexts and studies with different levels of complexity.

Third, participants who declined to join the study are missing from our analysis. Even though, in our study design, participants who decided not to join the study were redirected to Section 2 and offered the opportunity to complete the consent form evaluation. However, some participants may have closed the consent form without answering it. Although other factors may play a role, e.g., usability issues, those participants could read the consent form carefully and make an informed not-to-participate decision. The current study design did not include those participants in the analysis. In our study, 26 out of 278 participants opened our consent

---

[9]https://gdpr-info.eu/

form (Chatbot Condition: N = 18; Form Condition: N = 8) without completing it. We believe the effect of this potential confounding factor on our results is limited, but a future study is necessary.

Fourth, our study was designed to investigate the *holistic* effect of using an AI-powered chatbot to lead the informed consent process. However, the design of a chatbot (e.g., language style, name, and appearance) and its capability (e.g., natural language interpretation, question answering, dialogue management) are important design dimensions that may have an effect on the final outcome. As the first step, we aimed to build Rumi to deliver the best possible experience. The data collected in this study were inadequate to tease apart and quantify the contribution of each individual design factor. Since each of the interaction features have both benefits and risks [83], it is valuable to rigorously quantify the contribution of different features. This, however, requires additional, fully controlled experiments that are beyond the scope of the current study.

Lastly, although chatbots are increasingly adopted in our daily lives, from customer service to conversational surveys [39], it is still uncommon to use chatbots to conduct an informed consent process. Used in the first study of its kind, Rumi was novel to most participants. Since we could not control for the novelty effect in our current study design, we did not know the impact of novelty factors. While we are planning longitudinal studies to examine the influence of the novelty effect, the novelty effect may wear off, like any novel technology, as chatbots become a norm. In our case, as chatbot-driven informed consent becomes more common, the effect is most likely to wear off, similar to how more studies are using e-consent forms today [20].

6 CONCLUSION

In this paper, we examine the role of an AI-powered chatbot in improving informed consent online. We built, Rumi, an AI-powered chatbot that can greet a participant, go through the consent form section by section, answer the participant's questions, and collect people's consent responses, to simulate an in-person informed consent experience. We designed and conducted a between-subject study that compared Rumi with a typical form-based informed consent process in the context of an online survey study about people's social media use to examine the holistic effect of a chatbot in leading an online informed consent process. We found Rumi improved consent form reading, promoted a more equal power relationship between the participant and the researcher, and improved the study response quality. Our exploratory path model indicated the improved study response quality may be attributed to the reduced power gap by the chatbot-driven informed consent process. Given our study results and the simplicity of creating such a chatbot, our work suggests a new and promising method for conducting effective online informed consent. As chatbots become more popular, our results also present important design implications for creating more effective informed consent chatbots.

ACKNOWLEDGMENTS

REFERENCES

[1] Patricia Agre, Frances A Campbell, Barbara D Goldman, Maria L Boccia, Nancy Kass, Laurence B McCullough, Jon F Merz, Suzanne M Miller, Jim Mintz, Bruce Rapkin, et al. 2003. Improving informed consent: the medium is not the message. *IRB: Ethics & Human Research* 25, 5 (2003), S11–S19.

[2] Emily E Anderson, Susan B Newman, and Alicia K Matthews. 2017. Improving informed consent: Stakeholder views. *AJOB empirical bioethics* 8, 3 (2017), 178–188.

[3] Michela Assale, Erica Barbero, and Federico Cabitza. 2019. Digitizing the Informed Consent: the challenges to design for practices. In *2019 IEEE 32nd International Symposium on Computer-Based Medical Systems (CBMS)*. IEEE, 609–615.

[4] Martina Balestra, Orit Shaer, Johanna Okerlund, Madeleine Ball, and Oded Nov. 2016. The effect of exposure to social annotation on online informed consent beliefs and behavior. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*. 900–912.

[5] Solon Barocas and Helen Nissenbaum. 2014. Big data's end run around anonymity and consent. *Privacy, big data, and the public good: Frameworks for engagement* 1 (2014), 44–75.

[6] Baylor University. 2016. Are you addicted to social media? Six questions. *Science Daily* (Oct. 2016).

[7] Timothy Bickmore, Dina Utami, Shuo Zhou, Candace Sidner, Lisa Quintiliani, and Michael K Paasche-Orlow. 2015. Automated explanation of research informed consent by virtual agents. In *International Conference on Intelligent Virtual Agents*. Springer, 260–269.

[8] Timothy W Bickmore, Ha Trinh, Stefan Olafsson, Teresa K O'Leary, Reza Asadi, Nathaniel M Rickles, and Ricardo Cruz. 2018. Patient and consumer safety risks when using conversational assistants for medical information: an observational study of Siri, Alexa, and Google Assistant. *Journal of medical Internet research* 20, 9 (2018), e11510.

[9] Else Bjørn, Peter Rossel, and Soren Holm. 1999. Can the written information to research subjects be improved?–an empirical study. *Journal of Medical Ethics* 25, 3 (1999), 263–267.

[10] Su Lin Blodgett, Solon Barocas, Hal Daumé III, and Hanna Wallach. 2020. Language (Technology) is Power: A Critical Survey of "Bias" in NLP. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 5454–5476.

[11] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.

[12] Joan Cassell. 1980. Ethical principles for conducting fieldwork. *American anthropologist* 82, 1 (1980), 28–41.

[13] Barrie R Cassileth, Robert V Zupkis, Katherine Sutton-Smith, and Vicki March. 1980. Informed consent—why are its goals imperfectly realized? *New England journal of medicine* 302, 16 (1980), 896–900.

[14] Cindy Chen, Pou-I Lee, Kevin J Pain, Diana Delgado, Curtis L Cole, and Thomas R Campion Jr. 2020. Replacing paper informed consent with electronic informed consent for research in academic medical centers: a scoping review. *AMIA Summits on Translational Science Proceedings* 2020 (2020), 80.

[15] Henian Chen, Patricia Cohen, and Sophie Chen. 2010. How big is a big odds ratio? Interpreting the magnitudes of odds ratios in epidemiological studies. *Communications in Statistics—simulation and Computation®* 39, 4 (2010), 860–864.

[16] Shu-Hsin Chen. 2011. Power Relations Between the Researcher and the Researched: An Analysis of Native and Nonnative Ethnographic Interviews. *Field methods* 23, 2 (May 2011), 119–135.

[17] Francesco Colace, Massimo De Santo, Marco Lombardi, Francesco Pascale, Antonio Pietrosanto, and Saverio Lemma. 2018. Chatbot for e-learning: A case of study. *International Journal of Mechanical Engineering and Robotics Research* 7, 5 (2018), 528–533.

[18] Cathy A Coyne, Ronghui Xu, Peter Raich, Kathy Plomer, Mark Dignan, Lari B Wenzel, Diane Fairclough, Thomas Habermann, Linda Schnell, Susan Quella, et al. 2003. Randomized, controlled trial of an easy-to-read informed consent statement for clinical trial participation: a study of the Eastern Cooperative Oncology Group. *Journal of Clinical Oncology* 21, 5 (2003), 836–842.

[19] Jorden A Cummings, Jessica M Zagrodney, and T Eugene Day. 2015. Impact of open data policies on consent to participate in human subjects research: Discrepancies between participant action and reported concerns. *PLoS One* 10, 5 (2015), e0125208.

[20] Evelien De Sutter, Drieda Zaçe, Stefania Boccia, Maria Luisa Di Pietro, David Geerts, Pascal Borry, Isabelle Huys, et al. 2020. Implementation of electronic informed consent in biomedical research and stakeholders' perspectives: systematic review. *Journal of medical Internet research* 22, 10 (2020), e19129.

[21] Marcela G Del Carmen and Steven Joffe. 2005. Informed consent for medical treatment and research: a review. *The oncologist* 10, 8 (2005), 636–641.

[22] David DeVault, Ron Artstein, Grace Benn, Teresa Dey, Ed Fast, Alesia Gainer, Kallirroi Georgila, Jon Gratch, Arno Hartholt, Margaux Lhommet, et al. 2014.

SimSensei Kiosk: A virtual human interviewer for healthcare decision support. In *Proceedings of the 2014 international conference on Autonomous agents and multi-agent systems*. 1061–1068.

[23] Benjamin D Douglas, Emma L McGorray, and Patrick J Ewell. 2021. Some researchers wear yellow pants, but even fewer participants read consent forms: Exploring and improving consent form reading in human subjects research. *Psychological Methods* 26, 1 (2021), 61.

[24] Graham M Dresden and M Andrew Levitt. 2001. Modifying a standard industry clinical trial consent form improves patient information retention as part of the informed consent process. *Academic Emergency Medicine* 8, 3 (2001), 246–252.

[25] Catherine A Ebbs. 1996. Qualitative research inquiry: Issues of power and ethics. *Education* 117, 2 (1996), 217–223.

[26] Ruth R Faden and Tom L Beauchamp. 1986. *A history and theory of informed consent*. Oxford University Press.

[27] Ahmed Fadhil and Silvia Gabrielli. 2017. Addressing challenges in promoting healthy lifestyles: the al-chatbot approach. In *Proceedings of the 11th EAI international conference on pervasive computing technologies for healthcare*. 261–265.

[28] Carlos Miguel Ferreira and Sandro Serpa. 2018. Informed consent in social sciences research: Ethical challenges. *Int'l J. Soc. Sci. Stud.* 6 (2018), 13.

[29] James Flory and Ezekiel Emanuel. 2004. Interventions to improve research participants' understanding in informed consent for research: a systematic review. *Jama* 292, 13 (2004), 1593–1601.

[30] Michel Foucault. 1982. The subject and power. *Critical inquiry* 8, 4 (1982), 777–795.

[31] Joel A Friedlander, Greg S Loeben, Patricia K Finnegan, Anita E Puma, Xuemei Zhang, Edwin F De Zoeten, David A Piccoli, and Petar Mamula. 2011. A novel method to enhance informed consent: a prospective and randomised trial of form-based versus electronic assisted informed consent in paediatric endoscopy. *Journal of medical ethics* 37, 4 (2011), 194–200.

[32] Yubin Ge, Ziang Xiao, Jana Diesner, Heng Ji, Karrie Karahalios, and Hari Sundaram. 2022. What should I Ask: A Knowledge-driven Approach for Follow-up Questions Generation in Conversational Surveys. *arXiv preprint arXiv:2205.10977* (2022).

[33] Caitlin Geier, Robyn B Adams, Katharine M Mitchell, and Bree E Holtz. 2021. Informed Consent for Online Research—Is Anybody Reading?: Assessing Comprehension and Individual Differences in Readings of Digital Consent Forms. *Journal of Empirical Research on Human Research Ethics* 16, 3 (2021), 154–164.

[34] Lilian Ghandour, Rola Yasmine, and Faysal El-Kak. 2013. Giving consent without getting informed: a cross-cultural issue in research ethics. *Journal of Empirical Research on Human Research Ethics* 8, 3 (2013), 12–21.

[35] Graham R Gibbs. 2007. Thematic coding and categorizing. *Analyzing qualitative data* 703 (2007), 38–56.

[36] William Grabe. 2008. *Reading in a second language: Moving from theory to practice*. Cambridge University Press.

[37] Bradford H Gray, Robert A Cooke, and Arnold S Tannenbaum. 1978. Research Involving Human Subjects: The performance of institutional review boards is assessed in this empirical study. *Science* 201, 4361 (1978), 1094–1101.

[38] Herbert P Grice. 1975. Logic and conversation. In *Speech acts*. Brill, 41–58.

[39] Jonathan Grudin and Richard Jacques. 2019. Chatbots, humbots, and the quest for artificial general intelligence. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–11.

[40] Scott D Halpern, Jason HT Karlawish, and Jesse A Berlin. 2002. The continuing unethical conduct of underpowered clinical trials. *Jama* 288, 3 (2002), 358–362.

[41] Jeremy D Heider, Jessica L Hartnett, Emmanuel J Perez, and John E Edlund. 2020. Perceptions and understanding of research situations as a function of consent form characteristics and experimenter instructions. *Methods in Psychology* 2 (2020), 100015.

[42] Saras Henderson. 2003. Power imbalance between nurses and patients: a potential inhibitor of partnership in care. *Journal of clinical nursing* 12, 4 (2003), 501–508.

[43] Rink Hoekstra, Richard D Morey, Jeffrey N Rouder, and Eric-Jan Wagenmakers. 2014. Robust misinterpretation of confidence intervals. *Psychonomic bulletin & review* 21, 5 (2014), 1157–1164.

[44] Juji. 2020. Juji document for chatbot designers. https://docs.juji.io/. [Online; accessed 14-June-2020].

[45] Orit Karnieli-Miller, Roni Strier, and Liat Pessach. 2009. Power relations in qualitative research. *Qualitative health research* 19, 2 (2009), 279–289.

[46] Matthew Kay, Gregory L Nelson, and Eric B Hekler. 2016. Researcher-centered design of statistics: Why Bayesian statistics better fit the culture and incentives of HCI. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. 4521–4532.

[47] Soomin Kim, Joonhwan Lee, and Gahgene Gweon. 2019. Comparing data from chatbot and web surveys: Effects of platform and conversational style on survey response quality. In *Proceedings of the 2019 CHI conference on human factors in computing systems*. 1–12.

[48] Yujin Kim, Jennifer Dykema, John Stevenson, Penny Black, and D Paul Moberg. 2019. Straightlining: overview of measurement, comparison of indicators, and effects in mail–web mixed-mode surveys. *Social Science Computer Review* 37, 2 (2019), 214–233.

[49] Michael M Knepp. 2018. Using questions to improve informed consent form reading behavior in students. *Ethics & Behavior* 28, 7 (2018), 560–577.

[50] V Manoj Kumar, A Keerthana, M Madhumitha, S Valliammai, and V Vinithasri. 2016. Sanative chatbot for health seekers. *International Journal Of Engineering And Computer Science* 5, 03 (2016), 16022–16025.

[51] Marlies Kustatscher. 2014. Informed consent in school-based ethnography: Using visual magnets to explore participation, power and research relationships. *International Journal of Child, Youth and Family Studies* 5, 4.1 (2014), 686–701.

[52] Christine Lavelle-Jones, Derek J Byrne, Peter Rice, and Alfred Cuschieri. 1993. Factors affecting quality of informed consent. *British Medical Journal* 306, 6882 (1993), 885–890.

[53] Yi-Chieh Lee, Naomi Yamashita, Yun Huang, and Wai Fu. 2020. "I Hear You, I Feel You": encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. 1–12.

[54] James R Lewis. 1991. Psychometric evaluation of an after-scenario questionnaire for computer usability studies: the ASQ. *ACM Sigchi Bulletin* 23, 1 (1991), 78–81.

[55] Chung-Ying Lin, Anders Broström, Per Nilsen, Mark D Griffiths, and Amir H Pakpour. 2017. Psychometric validation of the Persian Bergen Social Media Addiction Scale using classic test theory and Rasch models. *Journal of behavioral addictions* 6, 4 (2017), 620–629.

[56] Yvonna S Lincoln and Egon G Guba. 1985. *Naturalistic inquiry*. sage.

[57] Richard McElreath. 2015. *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

[58] Narjara Conduru Fernandes Moreira, Camila Pacheco-Pereira, Louanne Keenan, Greta Cummings, and Carlos Flores-Mir. 2016. Informed consent comprehension and recollection in adult dental patients: a systematic review. *The Journal of the American Dental Association* 147, 8 (2016), 605–619.

[59] Gary R Morrow. 1980. How readable are subject consent forms? *Jama* 244, 1 (1980), 56–58.

[60] Kenneth Mulligan, Jon A Krosnick, Wendy Smith, Melanie Green, and George Bizer. 2001. Nondifferentiation on attitude rating scales: A test of survey satisficing theory. *Unpublished manuscript* (2001).

[61] Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *arXiv preprint arXiv:2203.02155* (2022).

[62] Barton W Palmer, Nicole M Lanouette, and Dilip V Jeste. 2012. Effectiveness of multimedia aids to enhance comprehension during research consent: A systematic review. *Irb* 34, 6 (2012), 1.

[63] Eric R Pedersen, Clayton Neighbors, Judy Tidwell, and Ty W Lostutter. 2011. Do undergraduate student research participants read psychological research consent forms? Examining memory effects, condition effects, and individual differences. *Ethics & Behavior* 21, 4 (2011), 332–350.

[64] Evan K Perrault and Seth P McCullock. 2019. Concise consent forms appreciated—still not comprehended: Applying revised common rule guidelines in online studies. *Journal of Empirical Research on Human Research Ethics* 14, 4 (2019), 299–306.

[65] Evan K Perrault and Samantha A Nazione. 2016. Informed consent—uninformed participants: shortcomings of online social science consent forms and recommendations for improvement. *Journal of Empirical Research on Human Research Ethics* 11, 3 (2016), 274–280.

[66] Kyle R Ripley, Margaret A Hance, Stacey A Kerr, Lauren E Brewer, and Kyle E Conlon. 2018. Uninformed consent? The effect of participant characteristics and delivery format on informed consent. *Ethics & Behavior* 28, 7 (2018), 517–543.

[67] Michael C Rowbotham, John Astin, Kaitlin Greene, and Steven R Cummings. 2013. Interactive informed consent: randomized comparison with paper consents. *PloS one* 8, 3 (2013), e58603.

[68] Sherry Ruan, Liwei Jiang, Justin Xu, Bryce Joe-Kun Tham, Zhengneng Qiu, Yeshuang Zhu, Elizabeth L Murnane, Emma Brunskill, and James A Landay. 2019. Quizbot: A dialogue-based adaptive learning system for factual knowledge. In *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–13.

[69] Peter H Schuck. 1994. Rethinking informed consent. *Yale Law Journal* (1994), 899–959.

[70] Anne Sherlock and Sonya Brownie. 2014. Patients' recollection and understanding of informed consent: a literature review. *ANZ journal of surgery* 84, 4 (2014), 207–210.

[71] Alan R Tait, Terri Voepel-Lewis, Stanley J Chetcuti, Colleen Brennan-Martinez, and Robert Levine. 2014. Enhancing patient understanding of medical procedures: evaluation of an interactive multimedia program with in-line exercises. *International Journal of Medical Informatics* 83, 5 (2014), 376–384.

[72] Ella Tallyn, Hector Fried, Rory Gianni, Amy Isard, and Chris Speed. 2018. The ethnobot: Gathering ethnographies in the age of IoT. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.

[73] Kar Yan Tam and Shuk Ying Ho. 2005. Web personalization as a persuasion strategy: An elaboration likelihood model perspective. *Information systems research* 16, 3 (2005), 271–291.

[74] Adam Tapal, Ela Oren, Reuven Dar, and Baruch Eitam. 2017. The sense of agency scale: A measure of consciously perceived control over one's mind, body, and the immediate environment. *Frontiers in psychology* 8 (2017), 1552.

[75] Vasti Torres and Marcia B Baxter Magolda. 2002. The evolving role of the researcher in constructivist longitudinal studies. *Journal of College Student Development* (2002).

[76] Connie K Varnhagen, Matthew Gushta, Jason Daniels, Tara C Peters, Neil Parmar, Danielle Law, Rachel Hirsch, Bonnie Sadler Takach, and Tom Johnson. 2005. How informed is online informed consent? *Ethics & Behavior* 15, 1 (2005), 37–48.

[77] James Walkup and Elinor Bock. 2009. What do prospective research participants want to know? What do they assume they know already? *Journal of Empirical Research on Human Research Ethics* 4, 2 (2009), 59–63.

[78] Alex C Williams, Harmanpreet Kaur, Gloria Mark, Anne Loomis Thompson, Shamsi T Iqbal, and Jaime Teevan. 2018. Supporting workplace detachment and reattachment with conversational intelligence. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[79] Diane L Wolf. 2018. Situating feminist dilemmas in fieldwork. In *Feminist dilemmas in fieldwork*. Routledge, 1–55.

[80] Ziang Xiao, Q Vera Liao, Michelle X Zhou, Tyrone Grandison, and Yunyao Li. 2023. Powering an AI Chatbot with Expert Sourcing to Support Credible Health Information Access. *arXiv preprint arXiv:2301.10710* (2023).

[81] Ziang Xiao, Michelle X Zhou, Wenxi Chen, Huahai Yang, and Changyan Chi. 2020. If i hear you correctly: Building and evaluating interview chatbots with active listening skills. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–14.

[82] Ziang Xiao, Michelle X Zhou, and Wai-Tat Fu. 2019. Who should be my teammates: Using a conversational agent to understand individuals and help teaming. In *Proceedings of the 24th International Conference on Intelligent User Interfaces*. 437–447.

[83] Ziang Xiao, Michelle X Zhou, Q Vera Liao, Gloria Mark, Changyan Chi, Wenxi Chen, and Huahai Yang. 2020. Tell me about yourself: Using an AI-powered chatbot to conduct conversational surveys with open-ended questions. *ACM Transactions on Computer-Human Interaction (TOCHI)* 27, 3 (2020), 1–37.

[84] T Yan. 2008. Nondifferentiation. *Encyclopedia of survey research methods* 2 (2008), 520–521.

[85] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and Bill Dolan. 2019. Dialogpt: Large-scale generative pre-training for conversational response generation. *arXiv preprint arXiv:1911.00536* (2019).

[86] Michelle X Zhou, Gloria Mark, Jingyi Li, and Huahai Yang. 2019. Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems (TiiS)* 9, 2-3 (2019), 1–36.

[87] Shuo Zhou, Timothy Bickmore, Michael Paasche-Orlow, and Brian Jack. 2014. Agent-user concordance and satisfaction with a virtual hospital discharge nurse. In *International conference on intelligent virtual agents*. Springer, 528–541.