Check for updates

Trust by Interface: How Different User Interfaces Shape Human Trust in Health Information from Large Language Models

Xin Sun
University of Amsterdam
the Netherlands
Centrum Wiskunde &
Informatica (CWI)
the Netherlands
x.sun2@uva.nl

Yunjie Liu
Eindhoven University of Technology
the Netherlands
y.liu12@tue.nl

Jan de Wit
Tilburg University
the Netherlands
j.m.s.dewit@tilburguniversity.edu

Jos A. Bosch
University of Amsterdam
the Netherlands
j.a.bosch@uva.nl

Zhuying Li
Southeast University
China
zhuying9405@gmail.com

ABSTRACT

The integration of Large Language Models (LLMs) with Conversational User Interfaces (CUIs) has significantly transformed health information seeking, offering interactive access to health resources. Despite the importance of trust in adopting health advice, the impact of user interfaces on trust perception in LLM-provided information remains unclear. Our mixed-methods study investigated how different CUIs (text-based, speech-based, and embodied) influence trust when using an identical LLM source. Key findings include (a) higher trust levels in information delivered via text-based interface compared to others; (b) a significant correlation between trust in the interface and the information provided; (c) participant's prior experience, processing approach for information with different modalities and presentation styles, and usability level were key determinants of trust in health-related information. Our study sheds light on trust perceptions in health information from LLMs and its dissemination, underscoring the importance of user interface in trustworthy and effective health information seeking with LLM-powered CUIs.

CCS CONCEPTS

• **Human-centered computing → Empirical studies in HCI**.

KEYWORDS

Conversational user interface, Large language model, Healthcare, Human trust perception

ACM Reference Format:
Xin Sun, Yunjie Liu, Jan de Wit, Jos A. Bosch, and Zhuying Li. 2024. Trust by Interface: How Different User Interfaces Shape Human Trust in Health Information from Large Language Models. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (CHI EA '24), May 11–16, 2024, Honolulu, HI, USA.* ACM, New York, NY, USA, 7 pages. https://doi.org/10.1145/3613905.3650837

CHI EA '24, May 11–16, 2024, Honolulu, HI, USA

ACM ISBN 979-8-4007-0331-7/24/05
https://doi.org/10.1145/3613905.3650837

1 INTRODUCTION

As online sources offer convenient and quick responses, there has been a growing demand for such sources to deliver reliable health-related information. As a result, various conversational user interfaces (CUIs) [9, 22, 31] have emerged, becoming important ways for people to obtain health-related information on a daily basis. CUIs, including chatbots and virtual health assistants [12, 16, 19], provide essential information on diseases and treatments[9, 41], enhancing patient engagement and access to health information. These CUIs aid in various health-related tasks, from scheduling appointments to symptom checking [1, 2, 47]. Moreover, the generative AI and large language models (LLMs) [46] are now key in healthcare dialogue and question-answering tasks[5]. ChatGPT, a prominent LLM-powered CUI, is increasingly used in health contexts [4, 18, 26, 37, 44]. These LLMs, trained on vast datasets, excel in natural language processing [8]. They support various healthcare applications, from patient education to mental health support, by transforming information searches into conversational interactions [4, 18, 19, 26]. The impact of this change is significant in personal health, where the value of information can greatly affect individual health and well-being. The increasing research into LLM-powered CUIs highlights their growing importance in accessing health information [4, 18, 19, 26, 37, 44].

While LLMs and CUIs offer great benefits, understanding how much people trust these systems is complex and not well-studied [41]. Trust in online health information significantly affects how people perceive and use this information, especially in making health decisions. This study focuses on trust in LLM-powered CUIs, examining three main aspects: first, how people's trust in health information varies across different interfaces powered by an identical LLM; second, the relationship between trust in the information and the interfaces, including how usability affects this trust; and third, identifying interface factors that influence trust in the information and how these factors shape overall trust perception [15, 38]. Given the importance of trust in health information from LLM-powered CUIs, our research aims to answer these research questions: (RQ1)

(a) text-based interface **(b) speech-based interface** **(c) embodied interface**

Figure 1: Three user interfaces used in the lab study for participants to complete the health-related search tasks.

How do user-perceived trust levels in health information from a large language model vary when it is delivered through different interfaces (text-based, speech-based, and embodied)? (RQ2) How does the user interface impact people's trust perception of health information from an identical LLM?

To address our research questions, we conducted a mixed-methods study combining a lab session for quantitative findings and interviews for qualitative insights. Each participant interacted with three interfaces: text-based, speech-based, and embodied, for health information searches. The lab study provided insights into how participants trust differently with interfaces. The follow-up interviews gave us a deeper look into individual experiences and perceptions of trust. We identified several factors affecting trust in health information from these interfaces, such as participants' prior experience with the interfaces, how they processed information differently with each interface, and the interfaces' usability. We also discussed the importance of physical presence in health information delivery and gained insights into different aspects of LLM-powered CUIs in health information seeking.

This work is motivated by the need to better understand how people trust LLM-provided health information from various CUIs, based on prior work [7, 30], as we move towards advanced multimodality and humanoid technologies like Pepper [36] and Optimus [45] Our findings are expected to inform the design of trustworthy CUIs and deepen the understanding of trust in LLMs in health information seeking. This research fills a current gap and sets the stage for future work on the reliability and trustworthiness of LLM-driven health communication.

2 RELATED WORK

2.1 Conversational user interfaces for health information seeking and dissemination

Conversational User Interfaces (CUIs), characterized by their capacity for natural language interaction, have significantly transformed the landscape of health information seeking [11, 32]. CUIs like chatbots [24, 32] and voice assistants [12] offer intuitive access to health information through dialogue exchanges, driven by advancements in AI [16, 23]. Notable examples include Babylon Health's chatbot [2] and Ada Health's symptom checker [1]. Despite the promise of CUIs, challenges in ensuring information credibility and managing user expectations regarding the limitations of these systems are ongoing areas of research [29, 35]. Current research and development focus on refining CUIs for clarity and effectiveness, balancing automated interaction with human oversight in delivering health information [17, 38, 40].

2.2 Trust in conversational user interfaces

Trust is vital for user interaction with CUIs in healthcare [25]. The trust model by Mayer, Davis, and Schoorman, focusing on competence, benevolence, and integrity, is foundational for understanding trust in automation [42]. Trust in CUIs involves the user's belief in the system's reliability and credibility, and ability to perform tasks effectively. Luger highlights the need for intelligent and contextually aware behavior in CUIs to build trust [29]. Accurate and credible information is key, as mistakes can quickly erode trust. Torning and Oinas-Kukkonen point out that design elements like social presence and expertise enhance trust in CUIs [43]. Health-related CUIs require rigorous testing and validation due to the sensitive nature of health information. Overall, fostering trust in health CUIs demands careful design of the interface, capable dialogue systems, and a consistently positive user experience.

2.3 Trust of large language models in health

Conversational information seeking [11] steps into the field of health, with LLMs like ChatGPT demonstrating capability in handling health-related queries [4, 18, 26, 37]. A systematic review [18] outlines ChatGPT's potential in healthcare, including enhancing health literacy and supporting medical research. The LLM is especially promising for transforming front-line medical services by providing automated patient consultations, initial diagnoses, and health advice [4, 26].

However, the understanding of human trust in these LLMs, especially their impact on health outcomes, is limited. Zhu et al.[48] found ChatGPT's responses to prostate cancer queries are more accurate than other models. Another study [44] reported that 40% of medical experts preferred ChatGPT's answers over Google's. Yet, these studies primarily involve medical professionals, leaving the trust of the general public in LLMs unclear. In addition, most previous work focused on the quality of information from LLMs

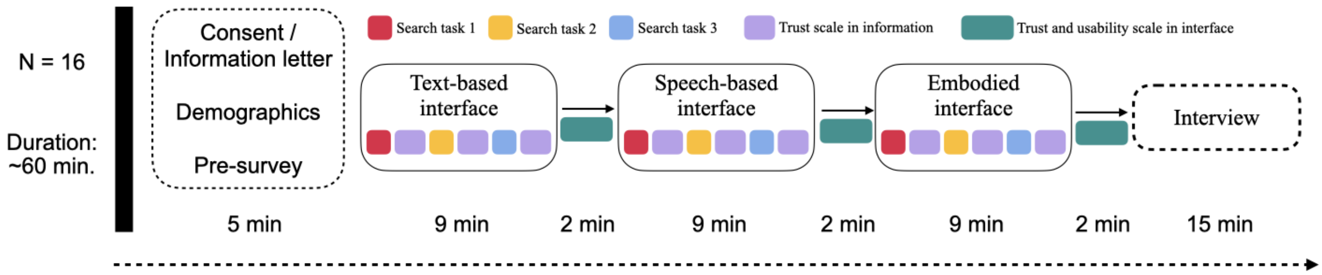

Figure 2: The procedure of the lab study and semi-structured interview.

rather than how users perceived trust in information and its dissemination. While there is extensive research on trust in online health information [14], the factors influencing trust in LLMs remain underexplored. Studies on chatbots suggest that trust in chatbots is affected by their communication style, medical knowledge, and interaction approach [38]. The advent of LLM-powered CUIs, agents, and humanoids [36, 45] highlights the need to study how different interfaces affect trust in information from the identical LLM.

3 METHODS

3.1 Participants

We conducted a mixed-methods study employing the within-subjects design to observe how each participant interacts with three interfaces for health information seeking. We calculated the sample size by G*Power [13], which indicated a need for 15 participants to detect a medium effect with 80% power. We recruited participants (N=16) from our institute, ensuring voluntary participation. Each participant received 12.5 EUR for their involvement. English fluency was a prerequisite for participation. The study was approved by our institute's ethics and data protection committee.

3.2 Materials

3.2.1 Search tasks. In this study, search tasks are personal health-related questions that we ask participants to find answers to. Each participant needs to complete nine search tasks. These questions are sourced from an open-sourced dataset [3], focusing on three pre-selected categories. We selected 25 questions from each category to ensure diversity (the list of questions used as search tasks is included as supporting material). Three categories of health-related questions are: Informational health questions: These seek general facts about personal healthcare topics. Example questions are: "Do you have information about Weight Control"; "Do you have information about Vitamin D?" Symptom and Cause-Related Health Questions: These are about symptoms or causes of health conditions. Example questions are: "What are the symptoms of eating disorder?"; "What causes Memory loss?" Treatment-Related Health Questions: Focusing on treatments or advice for specific health issues. Example questions are: "What are the treatments for dry eye syndrome?"; "How can I lower my heart rate?"

3.2.2 User interfaces. Three user interfaces were used in this study (as shown in Fig 1): text-based, speech-based, and embodied interfaces. We first created a text-based interface with a chatting web page, using the GPT-4 model [34] for backend query processing. This text-only setup allows us to assess how people trust information in pure text form. The second interface was an Amazon Echo Dot, offering voice interactions. Connected to the same GPT-4 model, it was able to respond to verbal queries. This provided insights into trust and perceptions in voice-based conversational interactions. Lastly, we used a self-made embodied 'robot' with a internal speaker for conversational interactions vocally, linked to the same GPT-4 model. This setup, with its physical presence, allows us to explore how a tangible interface impacts user trust in the delivered information, especially compared to interfaces without a physical presence.

3.3 Measures

Before the lab study, we collected demographic information and details on participants' online health information-seeking experience and familiarity with the CUIs in the study. To assess their general trust in technology, participants completed a pre-study questionnaire using a 6-item scale from [20], rated on a 5-point Likert scale. We also evaluated participants' AI and eHealth literacy using two 5-point Likert scales. The AI literacy scale [6] (11 items) and the eHealth literacy scale [33] (8 items) measured familiarity with AI and proficiency in using digital health resources respectively.

After each search task, participants rated their trust in the health information using an 11-item scale from the 'Trust of online health information' questionnaire [21, 39]. This scale assessed the information's credibility, usefulness, and readability. Upon completing tasks with each user interface, participants evaluated their trust in the interface using a 15-item scale adapted from the questionnaire [10]. For usability, we used two items from the UMUX-Lite survey [28] to assess each interface's usability level. Participants also indicated their intention to use these interfaces for future health information seeking on a 5-point Likert scale with a self-constructed item.

3.4 Procedure

Before the lab session, we obtained informed consent from all participants, as per our institute's guidelines. They interacted with three types of user interfaces: a text-based, a speech-based, and an embodied interface in a counterbalanced order to avoid bias. Participants undertook three search tasks per user interface, selected from a set list (see section 3.2.1). They had the freedom to ask follow-up questions, mimicking real-life information searches. After each task, they rated their trust in the provided health information through a brief survey. Following all tasks with an interface, they assessed

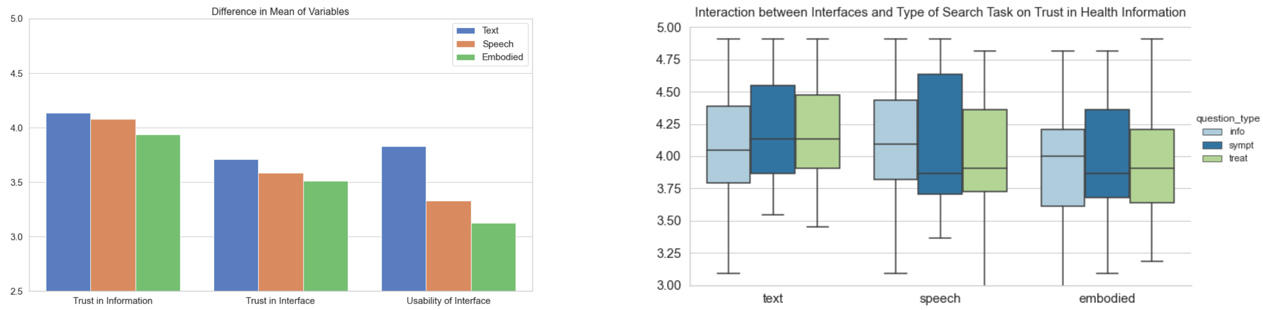

Figure 3: Left: Mean of trust in health information, trust in the user interface, and the usability level of each user interface for health information search. Right: Mean of trust in health information between three user interfaces across three categories of search tasks.

their overall perception of that interface, focusing on trust and usability. The lab study ended with a 20-minute semi-structured interview per participant, exploring their experiences and trust perceptions with each interface. Discussions involved factors influencing trust perceptions, a comparative analysis of interfaces, and suggestions for improving the trustworthiness of information from these interfaces in health contexts.

The overview of the study procedure is detailed in Fig 2.

4 RESULTS

4.1 Quantitative findings

4.1.1 Trust in health information differs by user interfaces. After checking the statistical reliability, we used a two-way repeated-measures ANOVA to compare the differences in trust levels among the three interfaces. The main effect of the user interface on trust in health-related information was found to be significant between text-based and embodied interface, with $F(1, 16) = 6.32, p = .024$, and $\eta^2 = .046$. However, for the other two interfaces, no significant difference in trust levels in health information was found. The trust score of retrieved information from the text-based interface is marginally higher compared to the other two interfaces: $F(1, 16) = 0.245, p = .628$, and $\eta^2 = .004$ for the text-based interface and speech-based interface; $F(1, 16) = 2.69, p = .122$, and $\eta^2 = .020$ for speech-based and embodied interface. This indicates a user preference for text-based information in health contexts. The difference was not statistically significant among the search tasks with $F(2, 30) = 0.95, p = .393$, and $\eta^2 = .002$. This finding implies that the trust in the retrieved information remained consistent regardless of the category of health-related questions, as shown in Fig 3.

4.1.2 Trust in the interface and trust in the information provided. Through Pearson correlation analysis, a significant relationship was identified between participants' trust in information and trust in each specific type of interface: text-based (r(20)=0.56, p<.05), speech-based (r(20)=0.52, p<.05), and embodied interfaces (r(20)=0.71, p<.01). This suggests that trust in the interface itself plays a crucial role in the perceived trust in the information provided.

4.1.3 Usability level of the interface and trust in information. Usability level of each interface has a significant impact on the trust levels in the interface as well, as shown in Fig 3. Utilizing Pearson correlation analysis, it showed that the usability level of text-based (r(16)=0.53, p<.05), speech-based (r(16)=0.64, p<.01), and embodied (r(16)=0.52, p<.05) interfaces significantly correlated to the perceived trust in the interfaces. In the case of the embodied interface, the usability level showed a significant correlation with the trust in the information provided as well (r(16)=0.66, p<.01).

4.2 Qualitative findings

The majority of participants from the interviews expressed the greatest trust in the text-based interface, followed closely by the speech-based interface. Trust in information from the embodied interface was significantly lower. We thereby further explored the factors that user interfaces influence the trust perceptions of information provided.

4.2.1 Prior experience with the interface affects the trust in information the most. A primary determinant in trust levels appears to be people's prior experiences and familiarity level with the interface. Our interviews revealed that participants' familiarity with different interfaces and their personal habits in seeking online health information predominantly shaped their trust (P2-3,P5,P8-10,P11-12,P14). For instance, many participants were accustomed to using Google or professional websites for online health information and some were familiar with text-based health apps, thus finding text-based interfaces more familiar and trustworthy. Moreover, the similarity of text-based interfaces to popular communication tools like WhatsApp, and the resemblance to the process of seeking advice from human experts or professional literature online (P3,P6,P7), significantly enhanced their trust in text-based interfaces.

4.2.2 Information modality determines the way information is processed, thereby affecting trust. The inherent characteristics of different information modalities and how human process these modalities play a crucial role in trust. Participants found that text information easier to process than speech or visual information, as they felt reading was simpler than listening (P5,P10,P12,P20). This ease of processing allows them to focus on the content, thereby enhancing trust perception: "*I prefer to read book like paper rather than listening, because if I'm listening to it, then I can do like a million other things in the meantime (distract). - P12*" Additionally, text information circumvents the limitations of context retrieval and timeliness inherent in speech, allowing users to compare information within the same interface. This comparative process not only deepens understanding but also incrementally builds trust (P1,P4,P5,P7,P9,P11,P14): "*I'm only picking up certain things [...], whereas the text one, I can see everything and I know exactly what I want to ask another question about. Because I can see and go back to and refer to the information if I forget anything. - P7*" Moreover, the text-format information can also be shared with the human professionals and friends (P11).

4.2.3 Information presentation in various interfaces influences the trust in health information provided. The presentation style of information in various interfaces affects trust. The information source remained consistent across interfaces in our study avoiding the bias of information presentation style. Participants suggested that while a numbered format works well in text interfaces, it might be less effective in speech or embodied interfaces, where continuous listening can lead to information overload and forgetfulness (P5-6,P10,P12). Adapting presentation styles, such as storytelling or summarizing key points at the beginning of vocal information, could potentially increase their appeal and trust perception (P5,P10,P13). Moreover, adding the source information would significantly increase the trust perception (P1,P5,P10,P12). "*It is better to present the input quite amount of information more like a storytelling, [[...] Just like an essay, if you don't put it in bullet points but you say like you use connection words. - P10*"

4.2.4 The usability of interfaces and contexts affect trust in information. The usability of interfaces emerges as another important factor influencing trust. Participants generally preferred text-based interfaces for health information search due to ease of use and convenience (P1,P4-P7,P9-11,P18). With the rapid development of LLMs, users have already experienced this trust-building process with text-based interfaces. However, for speech and embodied interfaces, where familiarity is still developing, usability becomes a key determinant in whether users are willing to foster trust through usage (P1,P13,P18,P20). "*I think this is more likely to use it. [...], because more likely to use like add supplementary questions, you can sort of build trust on the information. - P8*" Moreover, the using context affects trust perception. For example, speech interfaces are preferred for convenience in everyday use, while text-based interfaces are favored for serious contexts: "*[...] for speech or embodied [...] for like everyday usage I would more likely use the vocal ones. - P2*"

5 DISCUSSION

5.1 Limitations

Some limitations need to be acknowledged for our study. Firstly, the embodied interfaces lacked advanced features like body movements, possibly affecting the interactive naturalness. Future research could explore more sophisticated embodied interfaces with dynamic movements for deeper insights into physical presence and trust. Additionally, we used a consistent native female voice across both speech-based and embodied interfaces to avoid voice bias. However, this might not have fully matched the embodied interface's appearance, potentially influencing user perception of trust. Future studies should aim for a more cohesive vocal and visual design. Lastly, our study did not assess the accuracy of responses from LLMs, despite participant concerns about potential AI hallucination phenomena [27]. Future steps should include an evaluation component for the accuracy of LLM responses and their impact on the trust perception. These limitations should be considered. Addressing them in future steps will enrich our understanding of trust in health information from LLM-powered CUIs and aid in the development of more effective and user-centered CUIs.

5.2 Interface shapes user interaction and influences trust perception

Understanding the reason and how LLM and its powered CUIs can elicit higher trust levels is crucial not only for academic research but also for the practical development of LLM-driven health information systems and smart conversational interfaces. Identifying trust-building factors can guide future design and development of the LLM-powered CUIs. The alignment of our qualitative and quantitative findings provides a comprehensive insight of trust perceptions in health information delivery via different interfaces.

The qualitative analysis reveals a strong preference for text-based interfaces, rooted in familiarity with traditional search engines and medical websites for personal health information. This habitual reliance on text leads to a higher trust level, as one participant noted: "*I think we are used a lot to this text thing. - P14*", this habitual reliance on text contributes to a higher trust level. However, emerging modalities like speech-based and embodied interfaces struggle to gain similar trust, particularly in health-related contexts. Participants' varying experiences with each interface significantly shaped their trust. While open to using speech interfaces for general queries, there's hesitancy for health-related use due to a lower tolerance for errors and time-sensitive, as highlighted by a participant: "*My stereotype is to use text bot, but if it's about the other questions like uh, do you have any suggestions for the vacation or like sometimes it's more flexible - P2*". The need for accuracy and reliability in health information makes the errors in these interfaces more detrimental to trust.

In addition, the presentation format of information is another critical factor in trust-building. Users prefer interfaces that facilitate efficient processing, retention, and referencing of information, a crucial aspect in healthcare settings. Notably, trust levels did not vary significantly across different types of health questions, suggesting that trust is more dependent on the perceived credibility of the interfaces rather than the nature of the health questions themselves. A significant correlation existing between trust in the information and trust in the interface delivering it further evidence the argument. As one participant noted, "*I kinda have the almost the same trust levels for the information to similar trust level in different interface. - P3*". This finding underscores the critical role of interface design in trust perception and emphasizes the importance of enhancing overall interface quality and credibility to foster and maintain trust.

These findings underscore the necessity of focusing on prior experience, effective information presentation style, dissemination and processing for LLM-powered CUIs. Enhancing these aspects is key to improving the LLM-powered CUIs for health information dissemination, thereby building people's perceived trust and acceptance levels.

5.3 Implications and next steps

Our study highlights several future directions as the next steps. A critical area is the role of physical presence in information dissemination. Future work could explore how LLM information is perceived through tangible mediums versus virtual ones, focusing on more advanced physical embodiment that incorporate interactive visual expression, body movements, voice matching, and human-like features to understand their impact on trust. Besides, the impact of multimodal information delivery on trust in health information is an interesting area for further exploration. Understanding how different characteristics of voice and visual modalities as well as their coherence, affect trust can help in creating more effective multimodal CUIs. This includes examining vocal and visual features such as voice pitch, age, gender and matching of these modalities to influence trust. As humanoid evolve [36, 45], another direction could be the influence of humanization in advanced physical embodiment. Understanding the need for human-like features and their relationship to trust in the information dissemination process becomes vital. This research could reveal how the balance between humanization and trust perception can be optimized in LLM-powered CUIs. Exploring the long-term impacts of interactions with different interfaces on human trust perception will be also beneficial. Understanding these long-term effects is key to designing interfaces that maintain and enhance trust consistently, contributing to the development of reliable, user-centered LLM-powered CUIs for health information seeking.

Ethical and privacy considerations are also paramount, especially as physical embodiment and multimodal information become more sophisticated. As CUIs can access to increasing amounts of personal information, addressing how these AI access, manage, and distribute personal data is crucial to protect user privacy and adhere to ethical standards.

6 CONCLUSION

We explored how different user interfaces (text-based, speech-based, and embodied) affect human perceived trust in health information from an identical LLM in this study. We found that text-based interfaces are highly trusted due to their familiarity and ease of information processing. Speech-based interfaces, while convenient and natural for some users, still face trust barriers. Embodied interfaces introduce a novel interactive dimension but also bring additional trust factors, like privacy concerns and questions about authenticity. This study highlights the importance of interface design in more common LLM-driven health tools, emphasizing that the dissemination interfaces can significantly affect user trust in LLM-generated health information. Our findings provide a basis for future research, emphasizing the need for developing more trusted and effective LLM-powered user interfaces in the realm of digital health information.

REFERENCES

[1] adaTakeCare. 2024. Take care of yourself with Ada. More than a symptom checker. — ada.com. https://ada.com/app/.

[2] Babylon. 2024. Symptom Checker | eMed UK — babylonhealth.com. https://www.babylonhealth.com/en-us/what-we-offer/chatbot.

[3] Asma Ben Abacha and Dina Demner-Fushman. 2019. A Question-Entailment Approach to Question Answering. *BMC Bioinform.* 20, 1 (2019), 511:1–511:23. https://bmcbioinformatics.biomedcentral.com/articles/10.1186/s12859-019-3119-4

[4] Som S. Biswas. 2023. Role of ChatGPT in Public Health. *Annals of Biomedical Engineering* 51 (2023), 868–869. https://doi.org/10.1007/s10439-023-03172-7

[5] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, Sandhini Agarwal, Ariel Herbert-Voss, Gretchen Krueger, Tom Henighan, Rewon Child, Aditya Ramesh, Daniel Ziegler, Jeffrey Wu, Clemens Winter, Chris Hesse, Mark Chen, Eric Sigler, Mateusz Litwin, Scott Gray, Benjamin Chess, Jack Clark, Christopher Berner, Sam McCandlish, Alec Radford, Ilya Sutskever, and Dario Amodei. 2020. Language Models are Few-Shot Learners. In *Advances in Neural Information Processing Systems*, H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin (Eds.), Vol. 33. Curran Associates, Inc., 1877–1901. https://proceedings.neurips.cc/paper_files/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[6] Astrid Carolus, Martin J. Koch, Samantha Straka, Marc Erich Latoschik, and Carolin Wienrich. 2023. MAILS - Meta AI literacy scale: Development and testing of an AI literacy questionnaire based on well-founded competency models and psychological change- and meta-competencies. *Computers in Human Behavior: Artificial Humans* 1, 2 (2023), 100014. https://doi.org/10.1016/j.chbah.2023.100014

[7] Astrid Carolus and Carolin Wienrich. 2022. "Imagine this smart speaker to have a body": An analysis of the external appearances and the characteristics that people associate with voice assistants. *Frontiers in Computer Science* 4 (2022). https://doi.org/10.3389/fcomp.2022.981435

[8] K. R. Chowdhary. 2020. *Natural Language Processing*. Springer India, New Delhi, 603–649. https://doi.org/10.1007/978-81-322-3972-7_19

[9] R. J. W. Cline and K. M. Haynes. 2001. Consumer health information seeking on the Internet: the state of the art. *Health Education Research* 16, 6 (12 2001), 671–692. https://doi.org/10.1093/her/16.6.671 arXiv:https://academic.oup.com/her/article-pdf/16/6/671/9809432/160671.pdf

[10] Cynthia Corritore, Susan Wiedenbeck, Beverly Kracher, and Robert Marble. 2007. Online Trust and Health Information Websites. *International Journal of Technology and Human Interaction* 8 (01 2007). https://doi.org/10.4018/jthi.2012100106

[11] Jeffrey Dalton, Sophie Fischer, Paul Owoicho, Filip Radlinski, Federico Rossetto, Johanne R. Trippas, and Hamed Zamani. 2022. Conversational Information Seeking: Theory and Application. In *Proceedings of the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval* (Madrid, Spain) *(SIGIR '22)*. Association for Computing Machinery, New York, NY, USA, 3455–3458. https://doi.org/10.1145/3477495.3532678

[12] Caroline de Cock, Madison Milne-Ives, Michelle Helena van Velthoven, Abrar Alturkistani, Ching Lam, and Edward Meinert. 2020. Effectiveness of conversational agents (virtual assistants) in health care: Protocol for a systematic review. *JMIR Res. Protoc.* 9, 3 (March 2020), e16934.

[13] Franz Faul, Edgar Erdfelder, Albert-Georg Lang, and Axel Buchner. 2007. G*Power 3: a flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behav. Res. Methods* 39, 2 (May 2007), 175–191.

[14] Asbjørn Følstad, Cecilie Bertinussen Nordheim, and Cato Alexander Bjørkli. 2018. What Makes Users Trust a Chatbot for Customer Service? An Exploratory Interview Study. In *Internet Science*, Svetlana S. Bodrunova (Ed.). Springer International Publishing, Cham, 194–208.

[15] Ann L Fruhling and Sang M Lee. 2006. The influence of user interface usability on rural consumers' trust of e-health services. *International journal of electronic healthcare* 2, 4 (2006), 305–321.

[16] Luciano Gamberini, Patrik Pluchino, Luigi Porto, Filippo Zordan, Alessandro Monterosso, Kumara Palanivel, and Adriano Fontanari. 2021. *PatchAi: An e-Health Application Powered by an AI Virtual Assistant to Support Patients in Their Clinical Trials.* 315–322. https://doi.org/10.1007/978-3-030-78645-8_39

[17] Oscar A. Garcia Valencia, Supawadee Suppadungsuk, Charat Thongprayoon, Jing Miao, Supawit Tangpanithandee, Iasmina M. Craici, and Wisit Cheungpasitporn. 2023. Ethical Implications of Chatbot Utilization in Nephrology. *Journal of Personalized Medicine* 13, 9 (2023). https://doi.org/10.3390/jpm13091363

[18] Ravindra Kumar Garg, Vijeth L Urs, Akshya Anand Agrawal, Sarvesh Kumar Chaudhary, Vimal Paliwal, and Sujita Kumar Kar. 2023. Exploring the Role of ChatGPT in patient care (diagnosis and Treatment) and medical research: A Systematic Review. *medRxiv* (2023). https://doi.org/10.1101/2023.06.13.23291311

[19] Zhongxuan He. 2023. From ELIZA to ChatGPT: The Evolution of Chatbots in Public Health. *XRDS* 29, 3 (apr 2023), 59. https://doi.org/10.1145/3589658

[20] Sarah Jessup, Tamera Schneider, Gene Alarcon, Tyler Ryan, and August Capiola. 2019. *The Measurement of the Propensity to Trust Automation.* 476–489. https://doi.org/10.1007/978-3-030-21565-1_32

[21] Frances C. Johnson, Jennifer E. Rowley, and Laura Sbaffi. 2015. Modelling trust formation in health information contexts. *Journal of Information Science* 41 (2015), 415 – 429. https://api.semanticscholar.org/CorpusID:206454953

[22] Haruna Kibirige and Lila Ghemri. 2014. Trust Ranking of Medical Websites. In *Proceedings of the 4th ACM Conference on Data and Application Security and Privacy* (San Antonio, Texas, USA) *(CODASPY '14)*. Association for Computing Machinery, New York, NY, USA, 151–154. https://doi.org/10.1145/2557547.2557584

[23] Michael R King. 2023. The Future of AI in Medicine: A Perspective from a Chatbot. *Annals of Biomedical Engineering* 51, 2 (Feb. 2023), 291–295.

[24] Liliana Laranjo, Adam G Dunn, Huong Ly Tong, Ahmet Baki Kocaballi, Jessica Chen, Rabia Bashir, Didi Surian, Blanca Gallego, Farah Magrabi, Annie Y S Lau, and Enrico Coiera. 2018. Conversational agents in healthcare: a systematic review. *Journal of the American Medical Informatics Association* 25, 9 (07 2018), 1248–1258. https://doi.org/10.1093/jamia/ocy072 arXiv:https://academic.oup.com/jamia/article-pdf/25/9/1248/34150600/ocy072.pdf

[25] John D Lee and Katrina A See. 2004. Trust in automation: designing for appropriate reliance. *Hum. Factors* 46, 1 (2004), 50–80.

[26] Peter Lee, Sebastien Bubeck, and Joseph Petro. 2023. Benefits, Limits, and Risks of GPT-4 as an AI Chatbot for Medicine. *New England Journal of Medicine* 388, 13 (2023), 1233–1239. https://doi.org/10.1056/NEJMsr2214184 arXiv:https://doi.org/10.1056/NEJMsr2214184 PMID: 36988602.

[27] Florian Leiser, Sven Eckhardt, Merlin Knaeble, Alexander Maedche, Gerhard Schwabe, and Ali Sunyaev. 2023. From ChatGPT to FactGPT: A Participatory Design Study to Mitigate the Effects of Large Language Model Hallucinations on Users. In *Proceedings of Mensch Und Computer 2023* (Rapperswil, Switzerland) *(MuC '23)*. Association for Computing Machinery, New York, NY, USA, 81–90. https://doi.org/10.1145/3603555.3603565

[28] James R. Lewis, Brian S. Utesch, and Deborah E. Maher. 2013. UMUX-LITE: when there's no time for the SUS. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Paris, France) *(CHI '13)*. Association for Computing Machinery, New York, NY, USA, 2099–2102. https://doi.org/10.1145/2470654.2481287

[29] Ewa Luger and Abigail Sellen. 2016. "Like Having a Really Bad PA": The Gulf between User Expectation and Experience of Conversational Agents. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 5286–5297. https://doi.org/10.1145/2858036.2858288

[30] Michal Luria, Guy Hoffman, and Oren Zuckerman. 2017. Comparing Social Robot, Screen and Voice Interfaces for Smart-Home Control. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 580–628. https://doi.org/10.1145/3025453.3025786

[31] MAYO CLINIC. 2023. *MAYO CLINIC*. https://www.mayoclinic.org/

[32] Joao Luis Zeni Montenegro, Cristiano André da Costa, and Rodrigo da Rosa Righi. 2019. Survey of conversational agents in health. *Expert Systems with Applications* 129 (2019), 56–67. https://doi.org/10.1016/j.eswa.2019.03.054

[33] Cameron D Norman and Harvey A Skinner. 2006. EHEALS: The eHealth Literacy Scale. *J. Med. Internet Res.* 8, 4 (Nov. 2006), e27.

[34] OpenAI. 2023. GPT-4 Technical Report. arXiv:2303.08774 [cs.CL]

[35] Adam Palanica, Peter Flaschner, Anirudh Thommandram, Michael Li, and Yan Fossat. 2018. Physicians' Perceptions of Chatbots in Health Care: Cross-Sectional Web-Based Survey. *Journal of Medical Internet Research* 21 (11 2018). https://doi.org/10.2196/12887

[36] Amit Kumar Pandey and Rodolphe Gelin. 2018. A Mass-Produced Sociable Humanoid Robot: Pepper: The First Machine of Its Kind. *IEEE Robotics & Automation Magazine* PP (07 2018), 1–1. https://doi.org/10.1109/MRA.2018.2833157

[37] Abhishek Vaish Raju Vaishya, Anoop Misra. 2023. ChatGPT: Is this version good for healthcare and research? *Diabetes and Metabolic Syndrome: Clinical Research and Reviews* 17 (2023), 1871–4021. https://doi.org/10.1016/j.dsx.2023.102744

[38] Minjin Rheu, Ji Youn Shin, Wei Peng, and Jina Huh-Yoo. 2020. Systematic Review: Trust-Building Factors and Implications for Conversational Agent Design. *International Journal of Human-Computer Interaction* 37 (09 2020), 1–16. https://doi.org/10.1080/10447318.2020.1807710

[39] Jennifer E. Rowley, Frances C. Johnson, and Laura Sbaffi. 2015. Students' trust judgements in online health information seeking. *Health Informatics Journal* 21 (2015), 316 – 327. https://api.semanticscholar.org/CorpusID:21888204

[40] James Schaffer, John O'Donovan, James Michaelis, Adrienne Raglin, and Tobias Höllerer. 2019. I Can Do Better than Your AI: Expertise and Explanations. In *Proceedings of the 24th International Conference on Intelligent User Interfaces* (Marina del Ray, California) *(IUI '19)*. Association for Computing Machinery, New York, NY, USA, 240–251. https://doi.org/10.1145/3301275.3302308

[41] Elizabeth Sillence, Pam Briggs, Peter Richard Harris, and Lesley Fishwick. 2007. How do patients evaluate and make use of online health information? *Social Science and Medicine* 64, 9 (2007), 1853–1862. https://doi.org/10.1016/j.socscimed.2007.01.012

[42] Matthias Söllner, Axel Hoffmann, Maximilian Eike, Hirdes, Liudmila Rudakova, Stefanie Leimeister, and Jan Marco Leimeister. 2012. Towards a Formative Measurement Model for Trust. (10 2012).

[43] Kristian Torning and Harri Oinas-Kukkonen. 2009. Persuasive system design : State of the art and future directions, Vol. 350. 30. https://doi.org/10.1145/1541948.1541989

[44] Liesbet Van Bulck and Philip Moons. 2023. What if your patient switches from Dr. Google to Dr. ChatGPT? A vignette-based survey of the trustworthiness, value, and danger of ChatGPT-generated responses to health questions. *European Journal of Cardiovascular Nursing* (04 2023), zvad038. https://doi.org/10.1093/eurjcn/zvad038

[45] Wikipedia. 2024. Optimus (robot) — Wikipedia, The Free Encyclopedia. http://en.wikipedia.org/w/index.php?title=Optimus%20(robot)&oldid=1195180995.

[46] Tianyu Wu, Shizhu He, Jingping Liu, Siqi Sun, Kang Liu, Qing-Long Han, and Yang Tang. 2023. A Brief Overview of ChatGPT: The History, Status Quo and Potential Future Development. *IEEE/CAA Journal of Automatica Sinica* 10, 5 (2023), 1122–1136. https://doi.org/10.1109/JAS.2023.123618

[47] Yue You, Chun-Hua Tsai, Yao Li, Fenglong Ma, Christopher Heron, and Xinning Gui. 2023. Beyond Self-Diagnosis: How a Chatbot-Based Symptom Checker Should Respond. *ACM Trans. Comput.-Hum. Interact.* 30, 4, Article 64 (sep 2023), 44 pages. https://doi.org/10.1145/3589959

[48] Lingxuan Zhu, Weiming Mou, and Rui Chen. 2023. Can the ChatGPT and other large language models with internet-connected database solve the questions and concerns of patient with prostate cancer and help democratize medical knowledge? *Journal of Translational Medicine* 21, 1 (19 Apr 2023), 269. https://doi.org/10.1186/s12967-023-04123-5