Security and Privacy Perceptions of Mental Health Chatbots

Paulina Chametka
Carleton University
Ottawa, Canada
paulinachametka@gmail.com

Sana Maqsood
York University
Toronto, Canada
smaqsood@yorku.ca

Sonia Chiasson
Carleton University
Ottawa, Canada
chiasson@scs.carleton.ca

***Abstract*—Mental health chatbots are AI chatbots that aim to mimic human conversations about how a user feels, help a user work through issues they are facing, suggest wellness exercises to complete, and help track a user's mood over time. We compare the information disclosure practices and security and privacy concerns of adopters and non-adopters of mental health chatbots. We conducted a survey with 180 participants (30 adopters, 150 non-adopters), collecting data about what information they would hypothetically disclose to mental health chatbots, and concerns they had related to chatbots. We found that compared to non-adopters, adopters were more trusting of chatbots, were willing to reveal more information, perceived security and privacy risks to be less likely, and took fewer precautions.**

***Index Terms*—privacy, mental health chatbots, user study**

I. Introduction

According to research, 75-80% of patients benefit significantly from psychological treatment [1]. Mental health chatbots have received attention in recent years as a possible supplement or alternative to traditional therapy [2].

While the effectiveness of mental health chatbots remains under debate, these tools are readily available and their adoption requires users to divulge personal information and to trust chatbots to keep their communication confidential. This exposes users to security and privacy risks, such as the collection of sensitive personal information, use of the information in unexpected ways, and unauthorized parties gaining access to the information [3], [4]. Although privacy concerns have been noted for AI chatbots generally, little research exists on users' perspectives of the security and privacy implications for mental health chatbots where particularly sensitive information may be disclosed. Our main research questions are: (*RQ1:*) What are the differences between information disclosure practices of mental health chatbot adopters and non-adopters? and *(RQ2:)* What are the differences between the security and privacy concerns of mental health adopters and non-adopters?

We conducted an anonymous survey with 180 participants, including both adopters (30) and non-adopters (150). We refer to participants who indicated that they had used chatbots as *'adopters'*, and those who have never used chatbots as *'non-adopters'*. Adopters may have continued adoption, or may have tried chatbots and then discontinued their use. We found that adopters had more trust in chatbots compared to non-adopters. They were also willing to disclose more information to bots and took fewer precautions to protect their security and privacy. With regards to security and privacy risks, adopters perceived them to be less likely compared to non-adopters.

II. Background

A. Mental Health Chatbots

There are over 20 available mental health chatbots purportedly to help users manage different mental illnesses such as anxiety, postpartum depression, and substance abuse [5], [6], [7]. Two popular mental health chatbots are Woebot [8] and Wysa [9]. Developed in 2017, Woebot uses Cognitive Behavioural Therapy (CBT) to offer guided and empathetic prompts to users [8]. It gained popularity during the pandemic, when many users experienced increased distress. Developed in 2016, Wysa uses mindfulness techniques and CBT to offer mental health and emotional support [9]. It also offers the ability to talk to human psychologists through text messaging [9]. Both Woebot and Wysa provide self-help tools and exercises to improve mental health [9], [8].

Prior research provides evidence on the effectiveness of mental health chatbots in helping manage users' mental health [5], [6], [7], [10]. For example, Fitzpatrick et al. [5], found that Woebot could significantly reduce symptoms of anxiety and depression in young adults, while Inkster et al. [10] found Wysa to improve mood. However, these chatbots may also be harmful by enabling self-isolation, instead of building social connections or leading to co-dependence with the bot [11]. Due to their limited capability to handle complex or nuanced queries, chatbots can introduce risks to users with sensitive questions or information disclosures. For example, there have been reports of Woebot and Wysa responding inappropriately to queries around sexual abuse, suicidal ideation, eating disorders, and drug use [12].

Unlike licensed mental health professionals, chatbots are not required to comply with health regulations, such as privacy and confidentiality laws [13]. Thus, users' sensitive mental health data is at risk of misuse or exploitation. This is concerning because user disclosures to mental health chatbots may be similar to those made to therapists [13]. Given this, Stiefel [13] proposes a new legal framework and updated legislation protecting users' privacy and confidentiality.

B. Security and Privacy

The use of mental health chatbots raises issues relating to information disclosure and handling of private health data. For example, 65% of popular chatbots collect some form of user data, share the data with third-parties, and 70% do not de-identify user data [14]. The anthropomorphic qualities of chatbots also lead to more information disclosure from users, further risking their security and privacy [15].

Most research on the security and privacy of mental health chatbots focuses on their technical aspects, and there is a gap in looking at the user's perspective [16]. Recently, Prakash et al. [3] performed a thematic analysis of user reviews of Woebot and Wysa on the Google Playstore, and found that the chatbots sometimes increased users' safety risks, such as making them feel panic or by providing dangerous advice. Users also expressed privacy concerns, such as a lack of confidence in the privacy policy, discomfort in providing app permissions, and hesitation in sharing data with third-parties (e.g., Facebook messenger) [3]. Some limitations of this work include indirectly exploring users' perspective and focusing only on two chatbots. Belen et al. [4] surveyed 491 British users to explore concerns with how chatbot providers use sensitive data. Respondents were concerned about losing access to the data collected by the chatbot and how it might be used inappropriately.

In summary, existing research (i) focuses on certain aspects of security and privacy, such as trust, (ii) focuses on one homogeneous user group, (iii) only considers non-adopters, thus missing security and privacy insights from adopters. We address these gaps with a survey study about the information disclosure practices and security and privacy concerns of adopters and non-adopters, and we compare the two groups.

III. Methodology

We conducted an anonymous online survey using Qualtrics[1] so that participants felt comfortable disclosing their honest thoughts without fear of judgment, as some questions addressed sensitive topics such as mental health. The study was cleared by our University's Research Ethics Board (REB).

A. Participants

192 English-speaking adults completed the study. 150 had no experience with mental health chatbots (i.e., *non-adopters*) and 30 had used them (i.e., *adopters*); We discarded responses from 12 respondents who were unsure, leaving us with 180 responses. 60% of respondents were under 33 years of age. 52% were women, 42% were men, 9% were transgender or non-binary. The most common ethnicities were 62% European descent, 9% Chinese descent. For mental health, 83% of adopters and 77% of non-adopters reported at least one mental illness, with anxiety and depression most common.

Participants were recruited through Prolific[2], public social media communities about mental health (e.g., Reddit's "r/mental health"), snowball sampling, and the primary researcher's social media. As few individuals have experience with mental health chatbots (e.g., 4% of sample in [17]), we used multiple recruitment strategies to reach a substantial sample. One third of adopters were recruited from Prolific and the remainder from social media. The participants from Prolific were compensated 2.25 GBP; the other 94 were entered in a draw for $50 Amazon gift cards, maintaining odds of 1 in 50.

[1]Qualtrics: http://qualtrics.com

[2]Prolific: https://www.prolific.co/

B. Survey

The survey took on average 12 minutes to complete. Given the limited research on this topic, our survey[3]) was influenced by both prior literature [18], [19], [20] and the functionality of popular mental health chatbots, such as Woebot and Wysa.

- **Demographics (6 questions):** We collected participants' age, education level, gender, sexual orientation, ethnicity, and experiences with mental illness. Participants could select multiple options, no response, or self-describe.
- **Knowledge and experiences (6 questions):** We provided a definition of mental health chatbot before asking four closed-ended and two open-ended questions about participants' experience using mental health chatbots, and whether they would use one in the future. We compiled a list of 24 mental health chatbots from existing research [21], [22], [23], [24] and asked whether participants had heard about each and if they had used it.
- **Mental health disclosure practices (12 questions):** Seven closed-ended and four open-ended questions evaluated participants' comfort with disclosing specific types of mental health information (e.g., feelings, thoughts, relationship issues, addiction) to chatbots and to humans in their life. We also recorded participants' preference for using a mental health chatbot or a therapist.
- **Security and privacy (10 questions):** Three closed-ended and seven open-ended questions evaluated security and privacy perceptions. We addressed participants' perceived likelihood of specific risks occurring and their severity. We also asked about protective strategies and for suggested improvements to make participants feel more comfortable and safe using mental health chatbots.

Open-ended responses were analyzed using inductive thematic analysis. The themes most relevant to this paper included a lack of trust in the chatbot's abilities, and three themes related to security and privacy: concerns, precautions, and strategies for improvement.

IV. Results

A. Knowledge and experiences

Most (98%) non-adopters had never heard of any of the 24 chatbots presented. On average, adopters recognized or had used 16% of listed chatbots. The most recognized chatbots among adopters were Woebot [8] and Help4Mood [25] with 33% of adopters each, and Wysa [9] with 23% of adopters.

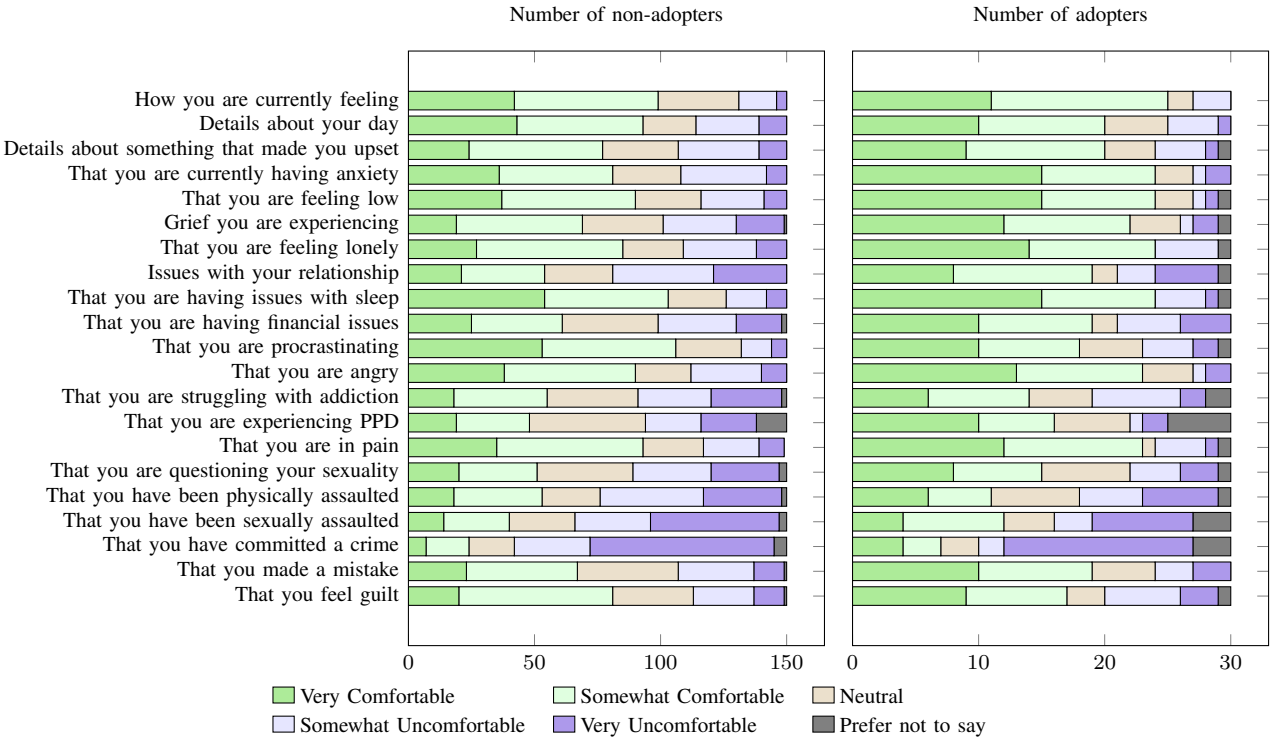

Fig. 1. Participant comfort level with hypothetically disclosing types of information to a chatbot.

B. Mental Health Disclosure Practices

Participants shared their willingness to disclose certain information to chatbots, and the factors which may encourage or prevent information disclosure.

1) Willingness to disclose: We asked twenty-one 5-point Likert scale questions about the types of information participants would be willing to disclose to a mental health chatbot. As shown in Figure 1, participants were most comfortable sharing information about how they were feeling, details about their day, or something that made them upset. In contrast, they were least comfortable sharing deeply personal information, which could have negative consequences if disclosed to unauthorized parties (e.g., committing a crime, experiencing physical or sexual assault, and details about their sexuality). We computed an *information disclosure* score out of 105 by adding participants' responses to all 21 questions. A one-sided Mann Whitney test found a significant difference ($p < .01061, Z = 2.2958, r = 0.17$) between the scores of the adopters ($M = 77, SD = 16.15$) and non-adopters ($M = 67.83, SD = 20.24$), with adopters more comfortable disclosing than non-adopters. Adopters' prior experience with chatbots may lead to increased trust and comfort, or it may be that individuals who are less privacy-conscious are more likely to use chatbots.

Open-ended responses revealed that participants were particularly concerned about the information being used against them in the future. As one non-adopter explained: *"I would censor myself a lot with a chatbot. They might be malicious and collecting my data to give to Mark Zuckerberg to make money, or it could leak and I'll get blackmailed" (p174, non-adopter)*. Another worried about how the information being used to institutionalize them: *"Data leaks, would be worried about being put in involuntary hold" (p190, non-adopter)*.

2) Factors encouraging disclosure to chatbots: We asked eight 5-point Likert-scale questions presenting reasons for disclosure to chatbots. For each reason, participants selected how much it applied to them, where 1 = *Does not apply to me at all* and 5 = *Definitely applies to me*. We aggregated responses to give each participant a score out of 40. Scores for the adopters ($M = 23.73, SD = 7.41$) and non-adopters ($M = 21.2, SD = 8.22$) were compared using a one-sided Mann-Whitney test, which found no significant difference ($p = 0.06, Z = 1.55, r = 0.15$) between the two groups. The most common reasons selected by both groups were that they would be embarrassed to tell a person, would feel judged by a person, would not trust a person, or would not feel ready to talk to a person about their problems.

Participants' open-ended responses expanded on how their experiences with feeling judged by a person and need to remain anonymous impacted their decision. One explained: *"I lean towards a mental health chatbot because I'm more comfortable remaining anonymous. I do have concerns about the bot collecting personal data, but this is not enough to stop me from choosing this option." (p111, adopter).*

3) Factors preventing disclosure to chatbots: We asked nine 5-point Likert-scale questions presenting different reasons

[3]https://chorus.scs.carleton.ca/wp-content/2023PST_survey.pdf

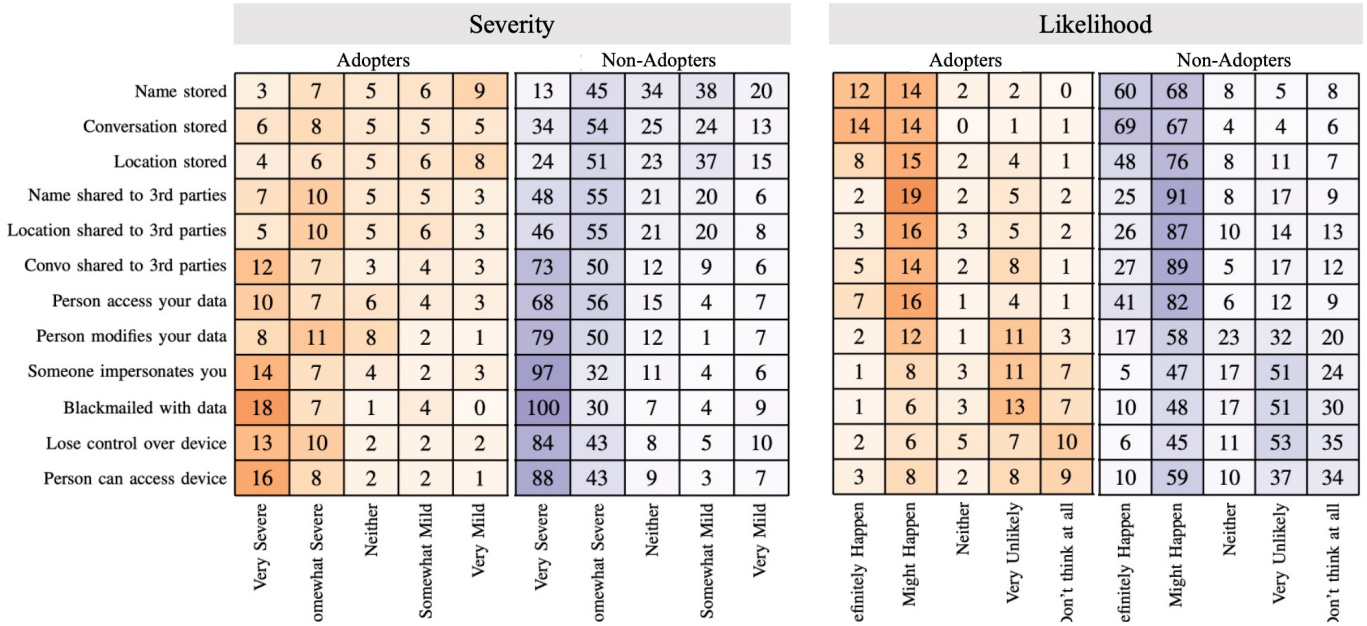

Fig. 2. Participants' perceived severity of security and privacy risks (left) and the likelihood of them occurring while using mental health chatbots (right).

which may prevent someone from disclosing information to mental health chatbots. Neither group feared judgement or embarrassment with regards to disclosing to chatbots, but worried about their effectiveness in terms of understanding concerns, and providing an adequate response. We computed an aggregate score out of 45 per participant. A one-sided Mann Whitney test found a significant difference ($p < 0.02, Z = -2.1172$) between the scores of the adopters ($M = 25.37, SD = 6.08$) and non-adopters ($M = 27.88, SD = 5.87$), with non-adopters feeling more strongly that the provided reasons could prevent them from disclosing information. For example, non-adopters were less trusting of chatbots and felt that they did not need help from chatbots.

In both groups, open-ended responses revealed security and privacy concerns preventing disclosure to chatbots. Specifically, participants feared that information collected by the chatbot could be accidentally leaked to the public, which can potentially result in negative personal or professional consequences. Surprisingly, even adopters were very concerned about non-AI entities having access to the information they shared with the chatbot. Both groups were concerned with how their data is shared with entities outside of the chatbot AI. For example, one explained: *"[I need] clear rules on what info goes where and who has access to my info. Rules on what happens if the bot thinks you may hurt yourself, someone else, commit a petty or serious crime, etc." (p192, adopter).*

C. Security and Privacy

Participants rated the likelihood of 12 risks related to mental health chatbots and the severity of the consequences associated with these risks. They also rated protective strategies.

1) Likelihood of risks: Figure 2 shows the list of risks presented and participants' rating of each risk. Both groups identified certain risks as more likely to occur. They found having their name, conversation, and location stored by the chatbot the most likely privacy risks. Conversely, impersonation, blackmail, losing control of their device, and having their device accessed by unauthorized users were all perceived to be relatively unlikely. Each participant's likelihood responses for all the risks were aggregated for a total score out of 60. The scores of adopters ($M = 39.43, SD = 8.53$) and non-adopters ($M = 42.05, SD = 9.5$) were compared using a one-sided Mann Whitney test, which found a significant difference ($p < 0.046, Z = -1.68$, r=0.13), with non-adopters believing security and privacy risks more likely.

2) Severity of consequences: As shown in Figure 2, both groups generally perceived the consequences of the security and privacy risks to be severe. Responses to all questions were aggregated to assign each participant a severity score out of 60. A one-sided Mann-Whitney test found a significant difference ($p < 0.012, Z = 2.24, r = 0.17$) between the severity scores of adopters ($M = 29.17, SD = 10.62$) and non-adopters ($M = 24.84, SD = 9.85$), with non-adopters perceiving the consequences to be more severe.

In general, the severity ratings were inversely proportional to the likelihood of the risk occurring while using mental health chatbots. Specifically, risks which were perceived to be less likely were rated as having the most severe consequences. These included impersonating the user, using data for blackmail, losing control of their device, and having their device accessed by unauthorized users. Risks that were perceived as most likely (i.e., storage of name, conversation, or location)

were perceived to have the least severe consequences.

3) Protective strategies : Shown in Figure 3, we provided a list of nine protective strategies and asked participants whether they would use each strategy. We tabulated the number of "Yes" responses for a total score out of 9 per participant. We compared adopters ($M = 5.5, SD = 2.27$) and non-adopters ($M = 6.5, SD = 2.15$) scores, and found a statistically significant difference ($p < 0.02, Z = -2.30, r = 0.17$). Adopters took fewer precautions compared to the non-adopters, which could make them more vulnerable to security threats.

D. Improving Security and Privacy

Through two open-ended questions, we explore what would make participants more comfortable using chatbots and how their security and privacy could be improved. Analysis revealed three main themes: improving transparency, providing assurance, and providing the ability to manage data.

1) Improving transparency: A privacy concern for both groups was not knowing the protection policies applicable to the data they had shared with the chatbot. Participants were concerned about how their data may be used by the chatbot company for profit or be shared with third-parties without consent. A non-adopter explained that they wanted: *"Some guarantee that data wouldn't leak, some confirmation that the bot isn't actually a person posing as a bot to get my info and blackmail me" (p181, non-adopter)*. Participants' fears may be justifiable considering recent incidents of data collected by an AI accessed by individuals within the organization [26].

Participants wanted greater transparency in terms of data collection, retention, storage, and sharing. Participants considered their mental health data to be extremely private and sensitive. Given this, they wanted more transparency in terms of what happens when certain types of information is shared with the chatbot. As one participant explained: *"[Need] info about what happens with the data. Will you be put in involuntary hold for typing suicidal thoughts? Can you be persecuted for anything typed? " (p190, non-adopter)*.

They also wanted more information about the technical implementation of the chatbot, such as the datasets used to train the AI and the security mechanisms used to protect users' information (e.g., encryption, SSL). One chatbot user wanted to know *"how the chatbot was trained and what type of data was deliberately censored from its training by the operators"* (p191, adopter). Some participants with technical expertise explained that making the code of the chatbot open-source would improve their trust and understanding of the system.

2) Providing assurance: Participants wanted assurance from chatbot companies that their data will be kept secure, and they wanted legal consequences for the company in case of data leaks or privacy violations. Some participants expected chatbot companies to provide the same type of confidentiality provided by in-person therapists. For example, a non-adopter was leery because: *"Human therapists are required to follow various ethical guidelines concerning a person's private data. A chatbot will likely be aggregating such data for the company that owns/operates it" (p78, non-adopter)*. This lack of assurance was a reason some non-adopters preferred in-person therapists over mental health chatbots.

Participants explained that they would feel more comfortable using a mental health chatbot if it was developed by or had oversight from a trusted party, such as a hospital, mental health professional, or their country's healthcare system. For chatbots developed by for-profit companies, participants explained that they would feel more comfortable if the chatbot was endorsed by a trusted party. One non-adopter wanted *"Assurance of privacy and security, and perhaps some endorsement from some mental health or psychological organisation" (p84, non-adopter)*. Another non-adopter wanted *"Some type of recognized designation that the tool has been approved by the health ministry or some type of health organization that is trusted" (p151, non-adopter)*.

3) Providing ability to manage data: Participants expressed that having privacy controls to manage their data on a mental health chatbot would make them feel more comfortable, as they would be able to define rules around how data could be shared with third parties. Some requested controls included the ability to store data from therapy sessions locally on their own computer rather than on a company server, the ability to specify that data will not be shared with anyone, and the ability to delete their data. In chatbots that do not maintain user profiles, participants wanted data automatically destroyed: *"The information is destroyed as soon as the chat is terminated" (p48, adopter)*.

Given the sensitive nature of the information provided to a mental health chatbot, many adopters and non-adopters would prefer being able to specify that none of their data to be stored by the chatbot or shared with anyone: *"Being assured that my data is safe and the conversation will not be stored. Even though it may help to go back and build on the last conversation, it is more secure this way." (p60, adopter)*.

V. Discussion

In our work, we look at the information disclosure practices, and security and privacy concerns of chatbot adopters and non-adopters, to examine similarities and differences between the two groups. We use our findings to propose design recommendations for mental health chatbots.

A. Addressing RQ1: Security and privacy concerns

Adopters and non-adopters shared similar security and privacy concerns, though to different extents. Both showed a similar understanding of the risks presented, found most risks to be at least somewhat likely with mental health chatbots, and found the consequences of some risks to be severe. However, non-adopters were more concerned about these risks compared to adopters, though the effect was small.

Participants expressed a need for anonymization, transparent policies, and data protection. They were particularly concerned of "data leaks" and having unauthorized parties use their mental health data in ways which may have severe negative consequences on their personal and professional lives, such as harassment, blackmail, and stalking. While using any form

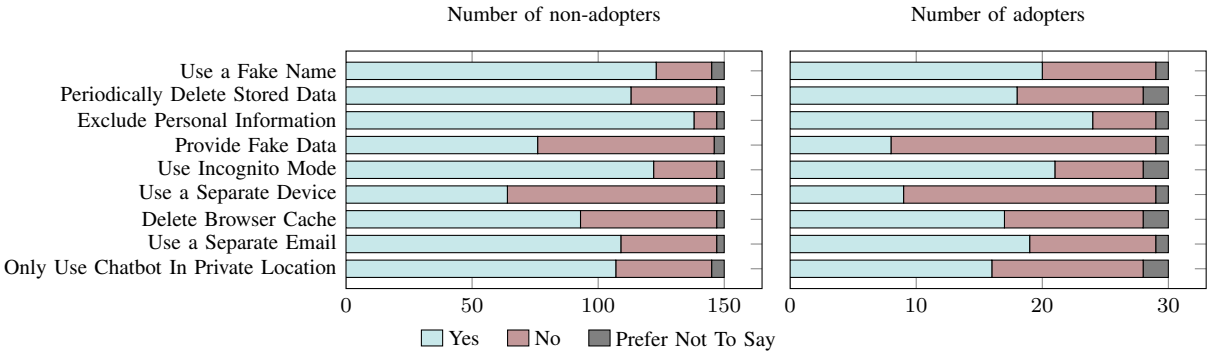

Fig. 3. Precautions participants would hypothetically take when using chatbots.

of technology presents some security and privacy risks to users, we found that the sensitive nature of the data disclosed to mental health chatbots made these risks more severe for our participants. Mental health is still very stigmatized, so users need to feel safe and trust the chatbot before disclosing personal and private information to them. Our participants expressed that they found the security and privacy mechanisms currently offered by chatbots to be inadequate, and provided suggestions for improvement.

B. Addressing RQ2: Information disclosure practices

Compared to non-adopters, adopters were more comfortable in disclosing personal information to chatbots, though the effect was small. This difference could be because adopters develop emotional bonds with their chatbots over time [27], and thus feel more comfortable disclosing sensitive information. These bonds might be further deepened if the chatbot exhibits anthropomorphic qualities. This, combined with the finding that adopters do not take some relatively simple precautions (e.g., using fake name), increases their security and privacy risks. As some users with mental illness may be more susceptible to cybersecurity crimes, technologies designed to improve mental health should provide appropriate measures to protect users during a vulnerable time. However, our participants did not feel protected by current technologies. Instead, they felt that their deeply personal data, such as their emotional pain, was being used by chatbot companies for monetary gain. Despite these privacy concerns, many adopters expressed that they would continue to use chatbots due to the benefits on their mental health and a lack of alternative supports.

C. Design recommendations

Recommendations regarding the appropriateness and effectiveness of mental health chatbots are outside the scope of this paper. However, if this technology is to be adopted, users should not have to compromise their security and privacy. Instead, these technologies should be designed to support users' security and privacy from the beginning. Recognizing that companies developing chatbots have corporate interests which may be counter to those of individual users, we nonetheless provide recommendations that prioritize incorporating security and privacy by design in mental health chatbots.

User control and freedom for data management: The primary concern among participants was not knowing how their data was being used by the chatbot and feeling a lack of control over their personal information. We recommend that mental health chatbots provide greater control to users for managing their data, and support different security and privacy data management needs. For example, some participants wanted complete anonymity due to fears of judgement, while others wanted services which may require some tracking (e.g., wanting to connect to a human therapist from the chatbot). Chatbots should provide an *anonymous* mode allowing users to interact with the chatbot in complete anonymity (i.e., no user data collected). A mechanism should be provided to allow users to switch between the *anonymous* and regular mode should they wish to access services which require tracking, similar to *Incognito mode* on web browsers. If the chatbot needs to track session data in *anonymous* mode, it should be destroyed as soon as the session ends.

Furthermore, when users choose to use the *regular* (i.e., tracking mode), they should be able to control which types of data gets collected, decide with whom the data is shared, and define their personal data retention preferences. For example, a user may decide that the chatbot is allowed to share their data with a mental health professional or a suicide prevention hotline if they disclose suicide ideation, while others may want to remain anonymous regardless of what information is disclosed to the chatbot. Even if users choose to store and share their data, they should be able to change these settings at any time in the future. For example, users may share a great deal of personal data when seeking help from a chatbot, but decide later that they do not want their emotionally sensitive data to be saved forever, thus choose to delete it.

Trust building mechanisms: Chatbots should incorporate mechanisms to improve trustworthiness. One mechanism is to provide greater transparency for data collection, data sharing, and technical implementation to ensure users' security and privacy. For example, our participants expected chatbots to use end-to-end encryption, authentication, and self destructing messages, but were uncertain whether these were actually be-

ing used by chatbots. While the closed-source design of mental health chatbots makes it difficult to determine whether they use adequate security and privacy protocols, a literature review of technology-based solutions for health care chatbots found that only a small percentage discussed patient safety issues, such as security and privacy [16]. In terms of transparency, chatbots should also provide information to users about whether their data is being used to improve the AI of the chatbot, and users should need to opt-in to this feature if they wish to participate.

Chatbots could also be improved by having their privacy and security assessed by a third-party in the security and privacy or mental health domains. It is unclear, however, which organization would be deemed trustworthy. For example, a participant from Germany explained that they would trust a chatbot more if it was reviewed by the *Chaos Computer Club*, which is a European association of hackers. Assessments by government organizations responsible for protecting consumer privacy, such as the Federal Trade Commission in the US, could also help improve user trust for some. Finally, compliance with government legislation (e.g., HIPPA in the US, and GDPR in Europe) for protecting the security and privacy of consumer healthcare data could further improve users' trust in mental health chatbots.

D. Limitations

As in other studies (e.g., [3]), we use a broad definition of 'adopter', so adopters may have varying experience with mental health chatbots. Despite attempts to recruit from communities with higher likelihood of having used mental health chatbots, our sample is limited in the number of adopters who responded to our survey. Our sample was also more educated than the general population and was biased towards individuals of white, European decent.

VI. Conclusion

In this study, we used an anonymous survey to compare adopters' and non-adopters' perceptions of mental health chatbots and to identify differences between the two groups regarding their willingness to disclose information and privacy and security concerns. Perhaps unsurprisingly, adopters were more trusting of chatbots in most regards than non-adopters. Adopters were significantly more willing to reveal information to the bots, they perceived security and privacy risks to be less likely, and they indicated that they would take fewer precautionary measures than non-adopters. Through our qualitative analysis, we found that both adopters and non-adopters voiced similar concerns about the security and efficacy of the bots, though adopters attributed less weight to these concerns.

References

[1] H. Chodos, "Options for improving access to counselling, psychotherapy and psychological services for mental health problems and illnesses," *Mental Health Commission of Canada*, 2017.

[2] E. M. Boucher, N. R. Harake, H. E. Ward, S. E. Stoeckl, J. Vargas, J. Minkel, A. C. Parks, and R. Zilca, "Artificially intelligent chatbots in digital mental health interventions: a review," *Expert Review of Medical Devices*, vol. 18, pp. 37–49, 2021.

[3] A. V. Prakash and S. Das, "Intelligent conversational agents in mental healthcare services: A thematic analysis of user perceptions," *PAJAIS*, vol. 12, p. 1, 2020.

[4] R. Belen Saglam, J. R. Nurse, and D. Hodges, "Privacy concerns in chatbot interactions: When to trust and when to worry," in *HCI International*, 2021, pp. 391–399.

[5] K. K. Fitzpatrick, A. Darcy, and M. Vierhile, "Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial," *JMIR Ment Health*, vol. 4, no. 2, 2017.

[6] K. Daley, I. Hungerbuehler, K. Cavanagh, H. G. Claro, P. A. Swinton, and M. Kapps, "Preliminary evaluation of the engagement and effectiveness of a mental health chatbot," *Frontiers in Digital Health*, 2020.

[7] J. J. Prochaska, E. A. Vogel, A. Chieng, M. Kendra, M. Baiocchi, S. Pajarito, and A. Robinson, "A therapeutic relational agent for reducing problematic substance use (woebot): Development and usability study," *J. Med. Internet Res*, vol. 23, no. 3, 2021.

[8] "Woebot health," https://woebothealth.com/.

[9] "Wysa," https://www.wysa.com/.

[10] B. Inkster, S. Sarda, and V. Subramanian, "An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study," *JMIR Mhealth Uhealth*, vol. 6, no. 11, Nov 2018.

[11] J. E. Brown and J. Halpern, "AI chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare," *SSM - Mental Health*, vol. 1, 2021.

[12] "Child advice chatbots fail to spot sexual abuse," https://www.bbc.com/news/technology-46507900.

[13] S. Stiefel, "The chatbot will see you now: protecting mental health confidentiality in software applications," *Colum. Sci. & Tech. L. Rev.*, vol. 20, p. 333, 2018.

[14] S. Saini, D. Panjwani, and N. Saxena, "Mobile mental health apps: Alternative intervention or intrusion?" in *PST*, 2022.

[15] C. Ischen, T. Araujo, H. Voorveld, G. van Noort, and E. Smit, "Privacy concerns in chatbot interactions," in *Chatbot Research and Design*. Springer, 2020.

[16] S. D. Aishwarya Surani, "Understanding privacy and security postures of healthcare chatbots," in *CHI*. ACM, 2022.

[17] A. Schick, J. Feine, S. Morana, A. Maedche, and U. Reininghaus, "Validity of chatbot use for mental health assessment: Experimental study," *JMIR mHealth and uHealth*, vol. 10, no. 10, 2022.

[18] T. Eagle, ""like talking to a person": User-perceived benefits of mental health and wellness mobile apps," Master's thesis, University of California, 2021.

[19] K. Kretzschmar, H. Tyroll, G. Pavarini, A. Manzini, I. Singh, and NeurOx Young People's Advisory Group, "Can your phone be your therapist? young people's ethical perspectives on the use of fully automated conversational agents (chatbots) in mental health support," *Biomedical informatics insights*, vol. 11, 2019.

[20] A. Palanica, P. Flaschner, A. Thommandram, M. Li, and Y. Fossat, "Physicians' perceptions of chatbots in health care: cross-sectional web-based survey," *JMIR*, vol. 21, no. 4, 2019.

[21] A. A. Abd-Alrazaq, M. Alajlani, N. Ali, K. Denecke, B. M. Bewick, and M. Househ, "Perceptions and opinions of patients about mental health chatbots: Scoping review," *J Med Internet Res*, vol. 23, no. 1, 2021.

[22] E. I. Konstantinidis, M. Hitoglou-Antoniadou, A. Luneski, P. D. Bamidis, and M. M. Nikolaidou, "Using affective avatars and rich multimedia content for education of children with autism," in *PETRA*. ACM, 2009.

[23] S. Mujeeb, M. Javed, and T. Arshad, "Aquabot: A diagnostic chatbot for achluophobia and autism," *IJACSA*, vol. 8, 2017.

[24] E. Hudlicka, "Virtual training and coaching of health behavior: Example from mindfulness meditation training," *PEC*, vol. 92, no. 2, 2013.

[25] "Help4mood," https://joinup.ec.europa.eu/collection/ehealth/document/help4mood-help4mood/.

[26] J. Morse, "Amazon employees listening to your alexa chats can see where you live," https://mashable.com/article/amazon-alexa-device-location-data.

[27] Y.-C. Lee, N. Yamashita, and Y. Huang, "Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional," *HCI*, May 2020.