Tapping into Key Drivers
Self-Disclosure in Sensitive Health Conversations with ChatGPT

Kelly, Sage; White, Katherine M.; Kaye, Sherrie Anne; Oviedo-Trespalacios, Oscar

DOI
10.1080/10447318.2025.2499656

Publication date
2025

Document Version
Final published version

Published in
International Journal of Human-Computer Interaction

International Journal of Human–Computer Interaction

ISSN: 1044-7318 (Print) 1532-7590 (Online) Journal homepage: www.tandfonline.com/journals/hihc20

Tapping into Key Drivers: Self-Disclosure in Sensitive Health Conversations with ChatGPT

Sage Kelly, Katherine M. White, Sherrie-Anne Kaye & Oscar Oviedo-Trespalacios

To cite this article: Sage Kelly, Katherine M. White, Sherrie-Anne Kaye & Oscar Oviedo-Trespalacios (22 May 2025): Tapping into Key Drivers: Self-Disclosure in Sensitive Health Conversations with ChatGPT, International Journal of Human–Computer Interaction, DOI: 10.1080/10447318.2025.2499656

To link to this article: https://doi.org/10.1080/10447318.2025.2499656

INTERNATIONAL JOURNAL OF HUMAN–COMPUTER INTERACTION
https://doi.org/10.1080/10447318.2025.2499656

Taylor & Francis
Taylor & Francis Group

OPEN ACCESS Check for updates

Tapping into Key Drivers: Self-Disclosure in Sensitive Health Conversations with ChatGPT

Sage Kelly[a], Katherine M. White[a], Sherrie-Anne Kaye[b], and Oscar Oviedo-Trespalacios[c]

[a]Faculty of Health, School of Psychology & Counselling, Queensland University of Technology (QUT), Kelvin Grove, Queensland, Australia; [b]Faculty of Health, Centre for Accident Research and Road Safety – Queensland (CARRS-Q), School of Psychology & Counselling, Queensland University of Technology (QUT), Kelvin Grove, Queensland, Australia; [c]Faculty of Technology, Policy and Management, Delft University of Technology, Delft, The Netherlands

ABSTRACT

The rise of ChatGPT has prompted concerns over users' agency when revealing personal data to artificial intelligence. This study examined users' likelihood of disclosing their data to ChatGPT in physical and mental health scenarios. Participants ($N = 216$) completed a repeated measures survey where they viewed four vignettes of hypothetical scenarios and were asked to imagine disclosing health information (physical and mental health) at two sensitivity levels (low and high self-disclosure). A repeated measures ANOVA revealed participants were significantly more likely to provide their data when the information required low-disclosure than high-disclosure. Furthermore, participants were significantly more likely to report uploading their health information in the physical health scenario than in the mental health scenario. The findings suggest ChatGPT users exercise caution in disclosing data to the platform. Reluctance to upload information in sensitive scenarios reduces the training data for large language models, resulting in potential stagnation in technology development.

KEYWORDS

Artificial intelligence; ChatGPT; data; disclosure; healthcare

The development and deployment of artificial intelligence (AI) in healthcare is rapidly increasing, creating concerns for users' safety and privacy (Doraiswamy et al., 2019; Siau & Wang, 2020; Toma et al., 2023). Large language models (LLMs), such as ChatGPT, are AI tools that generate conversational output in response to a users' prompt (Farah et al., 2023). When using a LLM such as ChatGPT for healthcare advice, users are able to tailor their prompts to suit their concerns as well as the amount of information they wish to reveal. The more specified the information the user provides in their prompt, the better and more personalised the outcome (MIT Sloan Teaching & Learning Technologies, 2024). Prior studies have highlighted the beneficial potential of ChatGPT in healthcare, such as treatment recommendations and patient monitoring (Anwer, 2024; Javaid et al., 2023). However, health data are some of the most private information individuals can disclose (Johnson & Willey, 2011; Sunarti et al., 2021) and revealing sensitive information to AI tools can expose users to privacy breaches (Bang & Kim, 2023). For instance, ChatGPT can potentially identify and leak private individuals' names, payment information, phone numbers, addresses, and social media accounts in the training data (Carlini et al., 2021; Clark, 2023). It is, therefore, necessary to understand what scenarios, if any, users are likely to disclose their health information to ChatGPT. The findings from this study could potentially be used to tailor interfaces for health data collection to distinguish between low and high data sensitivity and privacy-first designs for AI health systems.

Self-disclosure of medical information

There is a trade-off between users and ChatGPT when seeking medical advice. In social frameworks, reciprocation of information is the foundation of self-disclosure benefits, while risk is the cost of self-disclosure (Posey et al., 2010). This framework can also be applied to technology use as a user will be rewarded with information (e.g., possible diagnosis) when seeking advice from a LLM but first has to risk disclosing personal information (e.g., symptomology). It follows that privacy concerns are the primary deterrent to users disclosing their medical information online in prior studies (Feng et al., 2019; Yuchao et al., 2020). Research shows that, as the sensitivity of information increases, consumer willingness to disclose such information decreases (Rifon et al., 2005). Social research categorises sensitive data as information for which the processing poses a higher risk for data subjects than other types of data (Hemphill et al., 2022; Quinn & Malgieri, 2021;

CONTACT Sage Kelly sage.kelly@hdr.qut.edu.au Queensland University of Technology (QUT), Faculty of Health, School of Psychology & Counselling, 130 Victoria Park Road, Kelvin Grove 4059 Queensland, Australia

Supplemental data for this article can be accessed online at https://doi.org/10.1080/10447318.2025.2499656.

Soni et al., 2021). Such risks include identification and privacy harms (Quinn & Malgieri, 2021). As such, self-disclosure refers to the intention to reveal sensitive, personally identifiable information (Liu et al., 2023; Posey et al., 2010). Despite calls for close collaborations between AI research and users who provide their health status information (Chancellor & De Choudhury, 2020; Yoo et al., 2024), the concept of self-disclosure is absent from existing research examining the use of ChatGPT in medicine. Consequently, this study aims to explore the consideration of sensitivity (i.e., low vs high) on the likelihood of using ChatGPT in both physical and mental healthcare scenarios. While the sensitivity of data often occurs when the data is considered private, the opposite effect has been shown in healthcare.

Privacy paradox

Although consumers report concern about their data privacy, little effort is made to protect their data, resulting in a privacy paradox (Barth & De Jong, 2017). Here, there is a discrepancy between users' attitudes and their actual behaviour due to risk-benefit calculations that aim to maximise the benefits and minimise the risk of data disclosure (Barth & De Jong, 2017). Zhu et al. (2021) state that healthcare is a prime example of the privacy paradox as they found that users' perceived benefits were a stronger predictor of disclosure intention to a mobile health device than privacy concerns. As there is a high perceived usefulness that positively informs users' behaviour (Kelly et al., 2023), it could be predicted that self-disclosure will occur when the benefits outweigh the risks. Li et al. (2024) conducted an online experiment where they compared medical providers (AI vs. human), emotional support (low vs. high), and information sensitivity (low vs. high). Their findings revealed that users have lower perceived privacy risks for AI devices than human doctors (Li et al., 2024). This finding disputes past studies that state that humans have heightened privacy concerns about AI systems and that there is a positive relationship between information sensitivity and privacy concerns (Bansal et al. 2010). The authors postulate that the privacy paradox applies to the health domain as users perceive the benefits to outweigh the risks (Li et al., 2024), supporting Zhu et al. (2021). However, in the context of disclosing health information, there may be further considerations and nuances to these risks and benefits, such as the type of information required and the health-seeking alternatives. To better understand what data healthcare systems will have to build models on, it is pertinent to research users' intentions to provide their data to AI systems.

Sensitive data

Lei and Liu (2025) state that an advantage of AI systems is their ability to make users disclose more information than they would to a real human. It has been found that people's increased disclosure to AI systems, compared to humans, is due to a perceived lack of judgement and social capability of AI systems (Kim et al., 2022). This finding is relevant to the disclosure of health information due to past reports that people are reluctant to expose health information to human experts due to a fear of judgement (Lee et al., 2024). Previous work has shown that people will turn to AI systems when the sensitivity of the information is high (e.g., discussion of sexually transmitted infections) due to a desire to conceal this information from humans and receive health advice from a non-judgemental device (e.g., an AI chatbot; Nadarzynski et al., 2020). Furthermore, Lee et al. (2024) recruited 89 South Korean participants and found that the ability to build rapport with an AI psychotherapy chatbot and the chatbot's empathic responses encouraged self-disclosure of mental health information. Therefore, it may be that individuals are driven to disclose their mental health information to AI systems due to their usefulness and perceived lack of judgement. Therefore, past research indicates that users will turn to AI devices, like ChatGPT, for health advice that they may be considered sensitive. However, less is known about whether individuals will remove identifying information from their health prompts.

As people perceive their mental health data as more stigmatised than physical health information (Kendell, 2001; Werner, 2015), it may be that people differ in their disclosure of mental and physical information to AI systems. However, a limitation of Lee et al. (2024) is that they did not also measure physical health disclosure. Further work is needed to determine what, if any, differences exist between mental and physical health information disclosure. While past work has identified the type of information (e.g., low sensitivity = weight, blood type; high sensitivity = genetic disorders, sexual orientation) as sensitive, the present study explored the concept of identification as the sensitive information that users may conceal or reveal to ChatGPT (Li et al., 2024). Further research is required to determine if users' likelihood of using ChatGPT for healthcare differs depending on the sensitivity of the information to understand what input and output these AI systems will be receiving and delivering. Our work aimed to explore the differences between identifiable information (high sensitivity) and non-identifiable information (low sensitivity). Furthermore, our work aimed to identify if there are cases in which users will reject AI for healthcare to better understand the impact of AI on traditional health services.

AI alternatives

Despite the power of AI systems and users' privacy paradox, it may be that some individuals continue to rely on alternative methods for their health needs. For instance, concerns about the security of personal data have been shown to deter individuals from using AI chatbots for healthcare. Nadarzynski et al. (2019) conducted semi-structured interviews and an online survey to understand users' behaviours regarding AI-led health services. The authors proposed that AI hesitancy would negatively influence the engagement around this technology and that this hesitancy would be due to concerns about confidentiality, cyber security, and lack of empathy (Nadarzynski et al., 2019). Furthermore, some may prefer a human doctor due to familiarity and pre-determined trust (Bansal et al., 2022; Riedl et al., 2024).

However, less is known about individuals' preferences around a generalised LLM, such as ChatGPT. Therefore, it is of interest to measure whether individuals will seek healthcare alternatives to ChatGPT. It is important to understand what data these LLMs will be collecting for resource allocation for our healthcare systems as well as to predict future research and development of these models.

Need for data

From a technological standpoint, large data sets are required for LLM analyses as they function by identifying patterns within the data (Kahneman et al., 2021). Consequently, the larger the data set, the higher the chances of anomalies being represented within the data, improving AI predictors of rare events and reducing the need for human supervision (Kahneman et al., 2021). LLMs learn from interactions with humans (Mitchell, 2019). Without this input and feedback from humans, the AI systems would have low-quality data to learn from, rendering them unreliable and invalid. However, the demand for data is often greater than the supply (Bansal et al., 2022) and obtaining user data is a primary priority of technology companies (Kim et al., 2022). While this disclosure of personal information is advantageous for technology companies in need of data, it poses risks to human safety to reveal identifying information to machines that may be used maliciously (Doherty, 2022; Greenberg, 2024). Therefore, there is an increasing need for human data production and submission to reduce data scarcity and improve predictions, thus supporting the need for research to help stakeholders understand under what circumstances users will disclose their personal information to AI systems.

Current study

This study aims to provide novel insights into adopting AI in healthcare by researching if users differentiate their likelihood to self-disclose personal information to ChatGPT for physical as opposed to mental healthcare. The medical scenarios had two levels of required disclosure: low and high. Health was chosen as the scenario as health data are inherently sensitive due to the ability to draw conclusions about an individual using this information (Quinn & Malgieri, 2021), potentially leading to harms such as exploitation by employers or insurance companies based on health status (Al-Saggaf, 2015). Furthermore, health information can be deanonymised, resulting in identity theft and threatening users' well-being (Johnson & Willey, 2011). In addition to these concerns, certain health conditions may carry a social stigma. This consideration is particularly relevant for mental health data, which deal with sensitive subjects such as domestic violence, depression, and suicidal thoughts. Prior research has demonstrated individuals are less open to using an AI mental healthcare application than an AI device that monitors the risk of a heart attack (Antes et al., 2021). It is, therefore, of interest to explore if users are discerning in the personal information they disclose to ChatGPT for physical as opposed to mental health-related services.

Due to the exchange of personal information required to use LLMs for health services, it is important to discover whether people self-censor when using ChatGPT for health advice. Prior work has suggested users should be circumspect during their interactions with ChatGPT to avoid revealing sensitive information (Fui-Hoon Nah et al., 2023). ChatGPT was chosen to represent AI technology due to its prevalence in the healthcare industry (Moulaei et al., 2024). While prior research has explored the ability to leak personal information in LLMs (Carlini et al., 2021; Lukas et al., 2023) and the consequential harms (Weidinger et al., 2021), this paper aimed to examine if users choose to self-censor when using ChatGPT for health advice based on the level of sensitivity of the information. Health was chosen as the target use behaviour due to the often sensitive nature of this information. Further, mental and physical health scenarios were used to compare the type of personal information due to users' perceived differences between the two types of healthcare (Antes et al., 2021). Participants were also asked about their alternative healthcare preferences. The following research questions guided the study:

Research Question 1: Are users more or less likely to provide their personal information in a low-disclosure scenario, compared to a high-disclosure scenario?

Research Question 2: Are users more or less likely to provide their personal information in a mental healthcare scenario, compared to a physical healthcare scenario?

Research Question 3: What are users' alternative physical and mental healthcare preferences?

Methodology

Participants and recruitment

A convenience sample of 216 participants (162 females) aged 18-77 years (M age $=$ 26.51 years, $SD = 11.24$) was recruited from the Australian population. The demographic information is presented in Appendix A. An a priori power analysis was conducted using G*Power (Faul et al., 2009) to evaluate the sample size for the repeated measures ANOVA. The observed statistical power was 0.95, $\alpha = 0.05$ (Beck, 2013; Cohen, 1988), for a sample of 105 participants, providing evidence for the robustness of the sample size ($N = 216$). Participants comprised of first-year psychology students at Queensland University of Technology recruited through an online university student research management system (SONA) and received 0.5-course credit for survey completion. Participants were also recruited from the general population and were invited to enter a prize draw with the chance to win one of six $50 (AUD) gift vouchers. Recruitment was conducted online and via word-of-mouth, with the online survey open from August 2023 to March 2024.

Design, measures and procedure

The study was approved by the Queensland University of Technology Ethics Committee (approval number: 6926).

This study used a one-way, within-groups experimental design. Participants were recruited via social media, including Facebook, Instagram, LinkedIn, and an online university student research management system (SONA). Participants' informed consent was first obtained, and they were directed to complete the online survey. The online survey platform, Qualtrics, was used for the study. The survey first asked for participants' demographic information (e.g., age and gender). The survey automatically ended for those who did not meet the eligibility criteria via self-report (i.e., 18 years or older, currently residing in Australia, and having used ChatGPT at least once before). Participants were then provided with a definition of AI, "an unnatural object or entity with the ability and capacity to meet or exceed the task's requirements when considering cultural and demographic circumstances," as well as a definition of ChatGPT, "an AI language model that generates human-like text based on the input provided by the user. The training data comes from a diverse range of texts on the internet. When you input your data into ChatGPT, it collects and stores the information to train future versions of this technology" (OpenAI, 2024). Participants were asked to imagine themselves in scenarios where they could use ChatGPT for (1) mental healthcare and (2) physical healthcare services. They were asked to read a corresponding vignette (high-disclosure or low-disclosure). ChatGPT-3.5 was used to generate these vignettes. The four vignettes were randomised via Qualtrics to control for order and fatigue effects. Participants were assigned to all categories (high-disclosure mental healthcare, low-disclosure mental healthcare, high-disclosure physical healthcare, low-disclosure physical healthcare). Participants were asked how likely they were to upload their mental or physical health information to ChatGPT as per the examples displayed in Table 1.

Likelihood

After viewing the physical health vignette,[1] participants were asked, "In this scenario, how likely are you to upload your physical health report to ChatGPT to help you identify your options for recovery, as per the example above?" After viewing the mental health vignette, participants were asked, "In this scenario, how likely are you to upload your mental health information to ChatGPT to help you understand what you are experiencing, as per the example above?" Responses were measured on a 7-point Likert scale from *very unlikely* to *very likely*.

Alternatives

Following the vignettes, participants were asked, "Do you have someone or something else you would ask for physical health advice in this situation rather than ChatGPT?" They could respond "yes" or "no." If the participants responded yes, they were asked, "What or who would you ask for physical health advice in this situation instead of ChatGPT?" Responses were, "Google," "general practitioner," "other healthcare service (e.g., WebMD or Lifeline)," "A friend, partner, or family member," "I would not seek further advice," and "other." Multiple responses were allowed.

Data analyses

All assumptions were met and significance was assessed at $p < 0.05$, unless otherwise stated. The Statistical Package for the Social Sciences (SPSS) Version 28 was used to conduct all analyses. Bivariate correlations are reported first, followed by repeated measures ANOVA and the descriptive statistics.

Results

Bivariate relationships

The bivariate correlations between the independent and dependent variables can be found in Appendix B. Categorical demographic information (e.g., gender, race, sexual orientation, income, and education) were converted to binary items (e.g., female and other, white Australian and other, straight and other, above and below the national median income, tertiary educated and other). All four scenarios were significantly and positively related (physical high/physical low $r = 0.54$, $p < 0.001$; physical high/mental high $r = 0.75$, $p < 0.001$; physical high/mental low $r = 0.48$, $p < 0.001$; mental high/mental low $r = 0.56$, $p < 0.001$; mental high/physical low $r = 0.45$, $p < 0.001$; physical low/mental low $r = 0.58$, $p < 0.001$). All of the demographic variables were either not correlated significantly or only weakly correlated with disclosure likelihood across all four scenarios. No corrections were applied.

Repeated measures ANOVA

Visual assessment of the residual histograms indicated that data were normally distributed. The residual and pairwise scatterplots confirmed linearity. Inspection of the residual scatterplots indicated that the data were normal and homoscedastic. Skewness and kurtosis values were between the recommended value of ±2. Collinearity tests indicated that the assumption of multicollinearity was met (i.e., VIF > 10, Tolerance < 0.1).

A repeated measures ANOVA was conducted to examine the effects of healthcare (physical and mental) on disclosure (low and high) across the scenarios. The ANOVA revealed significant main effects of healthcare, $F(1, 216) = 17.51$, $p < 0.001$, and disclosure, $F\ (1, 216) = 71.35$, $p < 0.001$. The interaction between healthcare and disclosure was not significant, $F(1, 216) = 3.14$, $p = 0.08$.[2] These results reveal that healthcare and disclosure both independently influence participants' likelihood of using ChatGPT, but do not interact with each other. Estimated marginal means demonstrate that participants were more likely to use ChatGPT for physical healthcare than mental healthcare (see Table 2). Meanwhile, participants were more likely to use ChatGPT when the information they are uploading requires low-disclosure than high-disclosure. The mean scores were relatively low overall, demonstrating an overall reluctance by participants to upload their health information to ChatGPT.

Table 1. Vignettes.

Physical healthcare	
High-disclosure	SA Casey Watson (DOB: 19/05/1993) tore their right ACL while running around New Farm Park on Saturday, the 8th of June. Their mother, Patricia Watson, brought them to the emergency room at the Mater Hospital. Casey cannot load weight onto the right knee, reports hearing a pop at the time of injury, and the knee has been swelling since their arrival at the hospital. They have no history of previous knee or joint injury. They have been referred to Dr D'Arcy at Brisbane Knee Clinic.
Low-disclosure	SA I have torn my ACL running and cannot bear weight on the knee. What are my treatment options?
Mental healthcare	
High-disclosure	SA Casey Watson (DOB: 19/05/1993) presented to the practice on Saturday, the 8th of June. Their mother, Patricia Watson, brought them to New Farm Family Care General Practice. For the past three weeks, Casey has been experiencing symptoms that align with Major Depressive Disorder, such as trouble sleeping, feelings of worthlessness, and low concentration. They have no history of any psychological disorders. They have been referred to Dr Mee at River City Psychology Clinic. What are the treatment options for Casey?
Low-disclosure	SA I am having trouble sleeping and feel worthless. What are my treatment options?

Table 2. Estimated marginal means for likeliness to use ChatGPT.

Measure	Condition	Mean *(SE)*
Disclosure	Low	3.45 *(0.12)*
	High	2.59 *(0.11)*
Health	Physical	3.20 *(0.12)*
	Mental	2.84 *(0.11)*

Note: 7-point range scale (1 = *very unlikely*, 7 = *very likely*). *SE* = standard error.

Frequencies

The majority of participants (mental healthcare $N = 177$, 81.9%; physical healthcare $N = 181$, 83.8%) reported they would use alternatives to ask for health advice rather than ChatGPT in both scenarios (Figure 1). The majority (mental healthcare $N = 131$, 60.5%; physical healthcare $N = 114$, 52.8%) stated they would seek help from a general practitioner or a friend, partner, or family member (Figure 2). Ten respondents who selected "other" for mental health said they would speak to a psychologist. Three respondents who selected "other" for physical healthcare said they would speak to a physiotherapist, and two participants said they would speak to a medical student friend.

Discussion

The current study examined users' likelihood of uploading their health information to ChatGPT in four scenarios. Overall, the findings demonstrated that participants differ in their likelihood to use ChatGPT depending on the health service and required disclosure. Specifically, participants were significantly more likely to use ChatGPT for physical healthcare than mental healthcare. Furthermore, individuals were more likely to upload their information when it required low-disclosure than high-disclosure. Thus, ChatGPT users are circumspect about the data they directly provide to ChatGPT. This study is one of the first to compare users' likelihood to upload personal information to a LLM in different health scenarios.

Disclosure

Participants were significantly more likely to report providing their personal information in a low-disclosure scenario than in a high-disclosure scenario. Thus, people were more likely to submit their personal information to ChatGPT when personal identifying factors were removed from the prompt. The finding that participants were less likely to disclose identifying information (than anonymised data) demonstrates that people proactively protect their personal data when using AI, supporting Nadarzynski et al. (2019) and contesting the privacy paradox. It may be that in the case of ChatGPT, the benefits of disclosing identifiable information for health services do not outweigh the risks (Kelly et al., 2023; Posey et al., 2010). Consequently, individuals will exercise varying degrees of control over their sensitive data, choosing to disclose or conceal it based on the circumstance. Our finding challenges prior work (Li et al., 2024; Zhu et al., 2021), and alludes to a disconnect in the privacy paradox (Barth & De Jong, 2017). It could be suggested that by manipulating the sensitivity of identifying information (high sensitivity = identifying, low sensitivity = non-identifying), as opposed to the type of information (e.g., low sensitivity = weight, blood type; high sensitivity = genetic disorders, sexual orientation; Li et al., 2024), this study found an

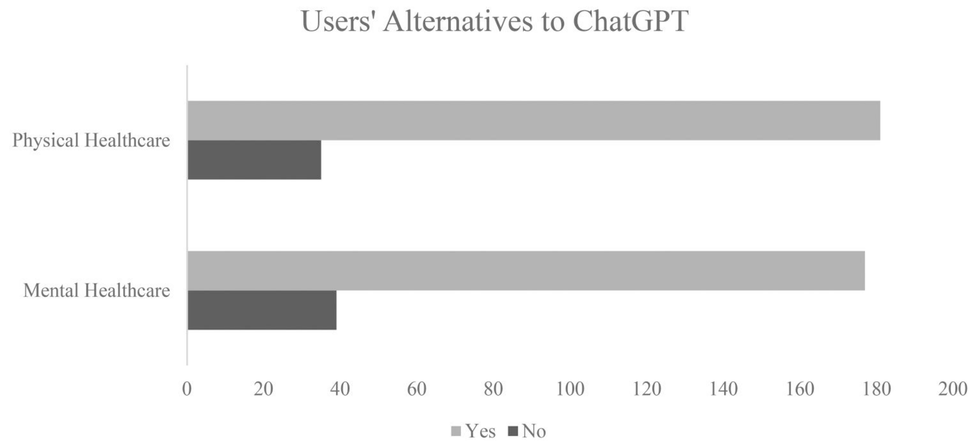

Figure 1. Users' alternatives to ChatGPT.

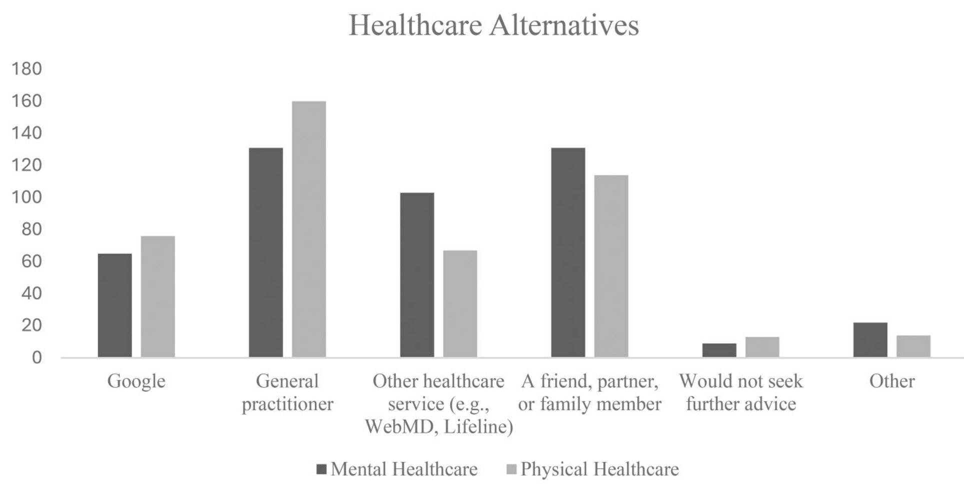

Figure 2. Healthcare alternatives.
Note: Participants could select multiple answers.

exception to the privacy paradox in regards to the likelihood of disclosing personal health information to AI.

It must also be acknowledged that the mean likelihood scores were relatively low overall (i.e., 2.59 to 3.45 on a 7-point scale), demonstrating an overall reluctance by participants to upload their information. While privacy self-management protects users from privacy breaches, it would also result in their receiving less personalised care from their conversation with ChatGPT. Data privacy is not compatible with the growth of AI (King & Meinhardt, 2024). As LLMs are trained on user data (Carlini et al., 2021), it may be that future models have less specified information to learn from. AI companies may, therefore, need to implement features that positively increase users' behavioural intentions if they want to receive users' health information. One such method would be via privacy preferences that allow individuals to exercise control over the use of their data (Bonney et al., 2009).

Healthcare

Participants were more likely to provide their personal information in a physical healthcare scenario and less likely in a mental healthcare scenario. This result demonstrates users are discerning in what information they will provide AI. Health conditions can carry social stigmas and biases (Noor, 2020; Reihl et al., 2015). This point is particularly relevant for mental health data, which may deal with sensitive subjects such as domestic violence, depression, and suicidal thoughts (Habicht et al., 2024). Therefore, any breaches can have severe consequences, potentially leading users to be more cautious in sharing mental health information than physical health information. Supporting Rifon et al. (2005), disclosure willingness decreases as data sensitivity increases. As such, and fitting with prior studies, it may be that users are more likely to use ChatGPT for physical healthcare than mental healthcare due to the perceived sensitivity of their mental health data (Lee et al., 2024). This finding additionally contests the privacy paradox (Barth & De Jong, 2017) as well as disputing prior work that states that users will use ChatGPT for assistance with sensitive topics (Lee et al., 2024; Nadarzynski et al., 2020). Instead, it suggests that users feel less protective of their physical health data than their mental health data. Alternatively, it may be that participants had alternate avenues of service for their mental health concerns that they would prefer to sharing their information online. This finding demonstrates that users may opt out of interacting with potentially beneficial technologies in certain scenarios. However, approximately 20% of participants stated that they would not seek an alternative to ChatGPT, indicating that some users would use ChatGPT for both physical and mental health services.

There are various repercussions from the disuse or underutilisation of AI systems to seek mental healthcare advice. Firstly, low user uptake of AI for mental health services may result in missed opportunities for early intervention and self-management of mental health issues. Though the mean difference is small between likelihood in the physical and mental health conditions, the bias towards physical health data would also result in an incomplete understanding of a user's overall well-being, resulting in ineffective treatment strategies. Furthermore, reluctance to adopt technology may result in the continuation of mental health disparities between communities that cannot access traditional mental healthcare. Thus, if users do not intend to use ChatGPT to seek mental health services, the weight of care remains on traditional service providers, who are already understaffed and overworked (Habicht et al., 2024). From a technological perspective, AI systems cannot exist without training data. Therefore, LLMs will face significant limitations if they cannot access certain information (Marcus, 2018), restricting their ability to address complex clinical challenges (Flanagan & Warren, 2022). As such, effort must be placed into the factors that predict technology acceptance of AI.

Alternative to ChatGPT

The low (i.e., below average) mean scores demonstrate that ChatGPT users feel that it was "unlikely" or "somewhat unlikely" that they would upload their health information to ChatGPT, even when identifying information was removed from the prompt. The majority of participants demonstrated that they would seek care from alternative sources. The most common responses were that participants would seek help from a general practitioner or a friend, partner, or family member. As such, it seems that in both mental and physical healthcare, the participants had an existing person in their life to turn to for advice, which may have reduced their need to use ChatGPT. Another consideration is that at the time of writing, ChatGPT is a generalised LLM that offers a large range of services but does not "remember" past interactions. Alternatively, other LLMs such as Replika or Wysa develop ongoing relationships with their users and market themselves as "empathetic" and "caring" chatbots that remember prior conversations. As such, users may be more likely to reveal personal information to chatbots with whom they have created relationships. Further research may wish to differentiate disclosure between chatbots that build relationships with their users and those that do not.

Implications

The findings from the present study carry practical implications for stakeholders that should be addressed. The findings demonstrate that there will be disparate amounts of data collected via the use of ChatGPT. The preference for sharing non-identifiable rather than personalised health data with ChatGPT may result in reduced personalisation of responses, limited accuracy, and slower advancement of AI-driven health solutions. While this approach enhances privacy, it may hinder the development of nuanced, equitable AI systems, particularly for underrepresented or complex health conditions. The decline in efficient outcomes may reduce use behaviour of AI systems for healthcare. Furthermore, as users are less likely to share their mental health data than their physical health data, there will be less available examples for LLMs to learn from. If mental health data are underrepresented, the resulting models may lack the nuanced understanding required to address the complex and varied nature of mental health issues. This lack of representation can lead to biases, misdiagnoses, or ineffective interventions when these AI systems are deployed as mental and physical health are so closely entwined. For instance, models trained on predominantly physical health data might fail to capture the linguistic, emotional, and cultural subtleties that are critical in mental healthcare. This could exacerbate existing inequalities in access to quality care and diminish trust in AI-based solutions for mental health, further limiting their adoption and efficacy. This predicted disparity underscores the need for targeted efforts to safeguard user privacy to encourage data sharing when appropriate in sensitive contexts, such as mental healthcare. Furthermore, stakeholders in mental healthcare must actively participate in shaping AI applications to ensure these technologies are relevant and effective for mental health use cases, driving users to experience the usefulness of these systems.

Limitations and future recommendations

Limitations of the present study warrant mention. Firstly, the study's likelihood measurement using one item may not fully capture the complexity of participants' perceptions towards uploading information on ChatGPT. Future studies should employ more comprehensive and nuanced measures to assess participants' predicted behaviours. For instance, factors like users' trust in technology may impact their likelihood of using ChatGPT. However, the robustness of the findings is supported by the consistency of responses across various demographic groups and experimental conditions. Secondly, the participants were not asked if they had any pre-existing health conditions. As such, there is no way of knowing if this community faced ongoing health concerns that would change the urgency of care. Future studies should explore whether the likelihood of using ChatGPT for health advice is influenced by need. Third, this sample was recruited from the Australian population, which offers universal healthcare. Future research may explore if access to healthcare and trust in healthcare institutions influence the likelihood of seeking alternative Westernised healthcare services, like ChatGPT. Fourth, this research relied on convenience sampling to recruit participants who were primarily female undergraduate psychology students and members of the online population. These participants may not fully represent the population from which the sample has been drawn. Future research should conduct different sampling techniques and data collection to study a more diverse range of participants.

Conclusion

The current study aimed to provide novel insights into AI adoption in healthcare by researching if users' likelihood to use ChatGPT differentiates between the required self-disclosure (high vs low) of personal information to ChatGPT and in differing medical scenarios (physical health vs mental health). An online survey demonstrated ChatGPT users differentiate their likelihood of using the device depending on the required disclosure and healthcare scenario. Specifically, users were more likely to disclose their health data to ChatGPT when that information was non-identifiable, with individuals reporting significantly lower mean scores in their likelihood of revealing identifying information, such as their name and age, to ChatGPT than more generic descriptions. The survey also revealed that users were more likely to provide their personal information in a physical healthcare scenario than in a mental healthcare scenario. This demonstrates that LLMs like ChatGPT may have disparate data across healthcare fields. Overall, participants currently prefer to seek help from a general practitioner or a friend, partner, or family member rather than ChatGPT. Further work is required to understand what factors contribute to increased willingness to provide personal information to LLMs to help the adoption and subsequent advancement of AI while preventing the misuse of this technology.

Notes

1. The viewing order of the four vignettes was randomised.
2. A figure of the results is displayed in the online supplement.

Disclosure statement

No potential conflict of interest was reported by the author(s).

ORCID

Sage Kelly http://orcid.org/0000-0002-1022-0467
Katherine M. White http://orcid.org/0000-0002-0345-4724
Sherrie-Anne Kaye http://orcid.org/0000-0002-2709-3345
Oscar Oviedo-Trespalacios http://orcid.org/0000-0001-5916-3996

References

Al-Saggaf, Y. (2015). The use of data mining by private health insurance companies and customers' privacy: An ethical analysis. *Cambridge Quarterly of Healthcare Ethics: CQ: The International Journal of Healthcare Ethics Committees*, *24*(3), 281–292. https://doi.org/10.1017/S0963180114000607

Antes, A. L., Burrous, S., Sisk, B. A., Schuelke, M. J., Keune, J. D., & Dubois, J. M. (2021). Exploring perceptions of healthcare technologies enabled by artificial intelligence: An online, scenario-based survey. *BMC Medical Informatics and Decision Making*, *21*(1), 221. https://doi.org/10.1186/s12911-021-01586-8

Anwer, M. S. (2024). Opportunities & challenges of artificial intelligent-powered technology in healthcare. *Medical Research Archives*, *12*(3), 1–12. https://doi.org/10.18103/mra.v12i3.5141

Bang, J., & Kim, J.-Y. (2023). Metaverse ethics for healthcare using ai technology: Challenges and risks [Paper presentation]. In M. Rauterberg (Ed.), *Culture and Computing. International Conference on Human-Computer Interaction*, Switzerland. https://doi.org/10.1007/978-3-031-34732-0_28

Bansal, G., Zahedi, F., & Gefen, D. (2010). The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decision Support Systems*, *49*(2), 138–150. https://doi.org/10.1016/j.dss.2010.01.010

Bansal, M. A., Sharma, D. R., & Kathuria, D. M. (2022). A systematic review on data scarcity problem in deep learning: Solution and applications. *ACM Computing Surveys*, *54*(10s), 1–29. https://doi.org/10.1145/3502287

Barth, S., & De Jong, M. D. (2017). The privacy paradox–investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review. *Telematics and Informatics*, *34*(7), 1038–1058. https://doi.org/10.1016/j.tele.2017.04.013

Beck, T. W. (2013). The importance of a priori sample size estimation in strength and conditioning research. *Journal of Strength and Conditioning Research*, *27*(8), 2323–2337. https://doi.org/10.1519/JSC.0b013e318278eea0

Bonney, R., Ballard, H., Jordan, R., McCallie, E., Phillips, T., Shirk, J., Wilderman, C. C. (2009). *Public participation in scientific research: Defining the field and assessing its potential for informal science education. A caise inquiry group report*. Center for Advancement of Informal Science Education. https://eric.ed.gov/?id=ED519688

Carlini, N., Tramer, F., Wallace, E., Jagielski, M., Herbert-Voss, A., Lee, K., Roberts, A., Brown, T. B., Song, D., & Erlingsson, U. (2021). *Extracting training data from large language models*. USENIX Security Symposium, Vancouver, BC, Canada. https://www.usenix.org/system/files/sec21-carlini-extracting.pdf

Chancellor, S., & De Choudhury, M. (2020). Methods in predictive techniques for mental health status on social media: A critical review. *NPJ Digital Medicine*, *3*(1), 43. https://doi.org/10.1038/s41746-020-0233-7

Clark, M. (2023). *Chatgpt's history bug may have also exposed payment info, says openai*. The Verge. Retrieved March 28 from https://www.theverge.com/2023/3/24/23655622/chatgpt-outage-payment-info-exposed-mondays

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge. https://doi.org/10.4324/9780203771587

Doherty, B. (2022). Medical data hacked from 10m australians begins to appear on dark web. *The Guardian*. https://www.theguardian.com/world/2022/nov/11/medical-data-hacked-from-10m-australians-begins-to-appear-on-dark-web

Doraiswamy, P. M., London, E., Varnum, P., Harvey, B., Saxena, S., Tottman, S., Campbell, P., Ibáñez, A. F., Manji, H., & Al Olama, M. A. A. S. (2019). Empowering 8 billion minds: Enabling better mental health for all via the ethical adoption of technologies. *NAM perspectives*. https://doi.org/10.31478/201910b

Farah, L., Davaze-Schneider, J., Martin, T., Nguyen, P., Borget, I., & Martelli, N. (2023). Are current clinical studies on artificial intelligence-based medical devices comprehensive enough to support a full health technology assessment? A systematic review. *Artificial Intelligence in Medicine*, *140*, 102547. https://doi.org/10.1016/j.artmed.2023.102547

Faul, F., Erdfelder, E., Buchner, A., & Lang, A.-G. (2009). Statistical power analyses using g* power 3.1: Tests for correlation and regression analyses. *Behavior Research Methods*, *41*(4), 1149–1160. https://doi.org/10.3758/BRM.41.4.1149

Feng, C.-L., Cheng, Z.-C., & Huang, L.-J. (2019). An investigation into patient privacy disclosure in online medical platforms. *IEEE Access*, *7*, 29085–29095. https://doi.org/10.1109/ACCESS.2019.2899343

Flanagan, A. J., Warren, S. (2022). *Advancing digital agency: The power of data intermediaries*. World Economic Forum. https://www.weforum.org/publications/advancing-digital-agency-the-power-of-data-intermediaries/#:~:text=Levers%20of%20action%20for%20both,the%20technology%20that%20serves%20them

Fui-Hoon Nah, F., Zheng, R., Cai, J., Siau, K., & Chen, L. (2023). Generative ai and chatgpt: Applications, challenges, and ai-human collaboration. *Journal of Information Technology Case and Application Research*, *25*(3), 277–304. https://doi.org/10.1080/15228053.2023.2233814

Greenberg, A. (2024). *Microsoft's recall feature is even more hackable than you thought*. Wired. Retrieved June 11, 2024, from https://www.wired.com/story/microsoft-windows-recall-privilege-escalation/

Habicht, J., Viswanathan, S., Carrington, B., Hauser, T. U., Harper, R., & Rollwage, M. (2024). Closing the accessibility gap to mental health treatment with a personalized self-referral chatbot. *Nature Medicine*, *30*(2), 595–602. https://doi.org/10.1038/s41591-023-02766-x

Hemphill, L., Schöpke-Gonzalez, A., & Panda, A. (2022). Comparative sensitivity of social media data and their acceptable use in research. *Scientific Data*, *9*(1), 643. https://doi.org/10.1038/s41597-022-01773-w

Javaid, M., Haleem, A., & Singh, R. P. (2023). Chatgpt for healthcare services: An emerging stage for an innovative perspective. *BenchCouncil Transactions on Benchmarks, Standards and Evaluations*, *3*(1), 100105. https://doi.org/10.1016/j.tbench.2023.100105

Johnson, M. E., & Willey, N. (2011). Usability failures and healthcare data hemorrhages. *IEEE Security & Privacy Magazine*, *9*(2), 35–42. https://doi.org/10.1109/MSP.2010.196

Kahneman, D., Sibony, O., & Sunstein, C. R. (2021). *Noise*. HarperCollins.

Kelly, S., Kaye, S.-A., & Oviedo-Trespalacios, O. (2023). What factors contribute to the acceptance of artificial intelligence? A systematic review. *Telematics and Informatics*, *77*, 101925. https://doi.org/10.1016/j.tele.2022.101925

Kendell, R. E. (2001). The distinction between mental and physical illness. *The British Journal of Psychiatry: The Journal of Mental Science*, *178*(6), 490–493. https://doi.org/10.1192/bjp.178.6.490

Kim, T. W., Jiang, L., Duhachek, A., Lee, H., & Garvey, A. (2022). Do you mind if i ask you a personal question? How ai service agents alter consumer self-disclosure. *Journal of Service Research*, *25*(4), 649–666. https://doi.org/10.1177/10946705221120232

King, J., Meinhardt, C. (2024). *White paper: Rethinking privacy in the ai era*. Stanford University Human-Centered Artificial Intelligence. https://hai.stanford.edu/sites/default/files/2024-02/White-Paper-Rethinking-Privacy-AI-Era.pdf

Lee, J., Lee, D., & Lee, J-g (2024). Influence of rapport and social presence with an ai psychotherapy chatbot on users' self-disclosure. *International Journal of Human–Computer Interaction*, *40*(7), 1620–1631. https://doi.org/10.1080/10447318.2022.2146227

Lei, X., & Liu, F. (2025). How service robots facilitate user self-disclosure: The roles of personality, animacy, and automated social presence. *International Journal of Human–Computer Interaction*, *41*(4), 2135–2148. https://doi.org/10.1080/10447318.2024.2316368

Li, S., Mou, Y., & Xu, J. (2024). Disclosing personal health information to emotional human doctors or unemotional ai doctors? Experimental evidence based on privacy calculus theory. *International Journal of Human–Computer Interaction*, 1–13. https://doi.org/10.1080/10447318.2024.2411619

Liu, Z., Zhang, L., Wu, Z., Yu, X., Cao, C., Dai, H., Liu, N., Liu, J., Liu, W., Li, Q., Shen, D., Li, X., Zhu, D., & Liu, T. (2023). Surviving Chatgpt in healthcare. *Frontiers in Radiology*, *3*, 1224682. https://doi.org/10.3389/fradi.2023.1224682

Lukas, N., Salem, A., Sim, R., Tople, S., Wutschitz, L., & Zanella-Béguelin, S. (2023). Analyzing leakage of personally identifiable information in language models. *IEEE Symposium on Security and Privacy (SP)*. https://doi.org/10.48550/arXiv.2302.00539

Marcus, G. (2018). Deep learning: A critical appraisal. arXiv preprint arXiv:1801.00631. https://doi.org/10.48550/arXiv.1801.00631

MIT Sloan Teaching & Learning Technologies (2024). *Effective prompts for ai: The essentials*. MIT. Retrieved November 30, 2024, from https://mitsloanedtech.mit.edu/ai/basics/effective-prompts/

Mitchell, M. (2019). *Artificial intelligence: A guide for thinking humans*. Penguin Random House.

Moulaei, K., Yadegari, A., Baharestani, M., Farzanbakhsh, S., Sabet, B., & Afrash, M. R. (2024). Generative artificial intelligence in healthcare: A scoping review on benefits, challenges and applications. *International Journal of Medical Informatics*, *188*, 105474. https://doi.org/10.1016/j.ijmedinf.2024.105474

Nadarzynski, T., Bayley, J., Llewellyn, C., Kidsley, S., Graham, C. A. (2020). Acceptability of artificial intelligence (ai)-enabled chatbots, video consultations and live webchats as online platforms for sexual health advice. *BMJ Sexual & Reproductive Health*. http://search.proquest.com/docview/2351038616/

Nadarzynski, T., Miles, O., Cowie, A., & Ridge, D. (2019). Acceptability of artificial intelligence (ai)-led chatbot services in healthcare: A mixed-methods study. *Digital Health*, *5*, 2055207619871808. https://doi.org/10.1177/2055207619871808

Noor, P. (2020). Can we trust ai not to further embed racial bias and prejudice? *BMJ (Clinical Research ed.)*, *368*, m363. https://doi.org/10.1136/bmj.m363

Posey, C., Lowry, P. B., Roberts, T. L., & Ellis, T. S. (2010). Proposing the online community self-disclosure model: The case of working professionals in france and the uk who use online communities. *European Journal of Information Systems*, *19*(2), 181–195. https://doi.org/10.1057/ejis.2010.15

Quinn, P., & Malgieri, G. (2021). The difficulty of defining sensitive data—The concept of sensitive data in the eu data protection framework. *German Law Journal*, *22*(8), 1583–1612. https://doi.org/10.1017/glj.2021.79

Reihl, K. M., Hurley, R. A., & Taber, K. H. (2015). Neurobiology of implicit and explicit bias: Implications for clinicians. *The Journal of Neuropsychiatry and Clinical Neurosciences*, *27*(4), A6–253. https://doi.org/10.1176/appi.neuropsych.15080212

Riedl, R., Hogeterp, S. A., & Reuter, M. (2024). Do patients prefer a human doctor, artificial intelligence, or a blend, and is this preference dependent on medical discipline? Empirical evidence and implications for medical practice. *Frontiers in Psychology*, *15*, 1422177. https://doi.org/10.3389/fpsyg.2024.1422177

Rifon, N. J., LaRose, R., & Choi, S. M. (2005). Your privacy is sealed: Effects of web privacy seals on trust and personal disclosures. *Journal of Consumer Affairs*, *39*(2), 339–362. https://doi.org/10.1111/j.1745-6606.2005.00018.x

Siau, K., & Wang, W. (2020). Artificial intelligence (AI) ethics: Ethics of AI and ethical AI. *Journal of Database Management*,), *31*(2), 74–87. https://doi.org/10.4018/JDM.2020040105

Soni, H., Ivanova, J., Grando, A., Murcko, A., Chern, D., Dye, C., & Whitfield, M. J. (2021). A pilot comparison of medical records sensitivity perspectives of patients with behavioral health conditions and healthcare providers. *Health Informatics Journal*, *27*(2), 14604582211009925. https://doi.org/10.1177/14604582211009925

Sunarti, S., Rahman, F. F., Naufal, M., Risky, M., Febriyanto, K., & Masnina, R. (2021). Artificial intelligence in healthcare: Opportunities and risk for future. *Gaceta Sanitaria*, *35 Suppl 1*, S67–S70. https://doi.org/10.1016/j.gaceta.2020.12.019

Toma, A., Senkaiahliyan, S., Lawler, P. R., Rubin, B., & Wang, B. (2023). Generative ai could revolutionize health care—But not if control is ceded to big tech. *Nature*, *624*(7990), 36–38. https://doi.org/10.1038/d41586-023-03803-y

Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P.-S., Cheng, M., Glaese, M., Balle, B., & Kasirzadeh, A. (2021). Ethical and social risks of harm from language models. arXiv preprint arXiv:2112.04359. https://doi.org/10.48550/arXiv.2112.04359

Werner, S. (2015). Public stigma and the perception of rights: Differences between intellectual and physical disabilities. *Research in Developmental Disabilities*, *38*, 262–271. https://doi.org/10.1016/j.ridd.2014.12.030

Yoo, D. W., Woo, H., Pendse, S. R., Lu, N. Y., Birnbaum, M. L., Abowd, G. D., & De Choudhury, M. (2024). Missed opportunities for human-centered ai research: Understanding stakeholder collaboration in mental health ai research. *Proceedings of the ACM on Human-Computer Interaction*, *8*(CSCW1), 1–24., Article 95. https://doi.org/10.1145/3637372

Yuchao, W., Ying, Z., & Liao, Z. (2020). Health privacy information self-disclosure in online health community. *Frontiers in Public Health*, *8*, 602792. https://doi.org/10.3389/fpubh.2020.602792

Zhu, M., Wu, C., Huang, S., Zheng, K., Young, S. D., Yan, X., & Yuan, Q. (2021). Privacy paradox in mhealth applications: An integrated elaboration likelihood model incorporating privacy calculus and privacy fatigue. *Telematics and Informatics*, *61*, 101601. https://doi.org/10.1016/j.tele.2021.101601

About the authors

Sage Kelly is a researcher in human factors. Her PhD research centred on user acceptance and decision-making processes concerning AI technologies.

Katherine M. White is a Professor of Psychology in the School of Psychology and Counselling at Queensland University of Technology. Her research focuses on the application of decision-making models across broad settings including social, health, and technology contexts.

Sherrie-Anne Kaye is a Senior Research Fellow at the MAIC-QUT Road Safety Research Collaboration. Her research has focused on user acceptance of advanced technologies, risky driving behaviours, and pro-social driving behaviours.

Oscar Oviedo-Trespalacios is an Asst/Professor in Human Factors at TU Delft. His research focuses on the interaction between humans and technology, with a particular interest in preventing technology misuse and improving safety in complex sociotechnical systems.

Appendix A

Demographics

		N	%
Gender	Females	162	75.0%
	Males	47	21.8%
	Non-Binary	4	1.9%
	Transgender	1	0.5%
	Prefer not to say	1	0.5%
Race	White Australian	140	65.8%
	Asian	32	14.8%
	Multiple races	12	5.6%
	Australian Aboriginal, South Sea Islander, Torres Strait Islander or Norfolk Islander	8	3.7%
	Asian Indian	8	3.7%
	Hispanic	5	2.3%
	Prefer not to say	6	2.8%
	Other	5	2.3%
Sexual orientation	Straight	152	70.4%
	Bisexual	25	11.6%
	Homosexual	14	6.5%
	Pansexual	12	6.0%
	Asexual	4	1.9%
	Other	1	0.5%
	Prefer not to say	7	3.2%
Monthly income	Nil income	11	5.1%
	$1-$149	1	0.5%
	$150-$299	4	1.9%
	$300-$399	6	2.8%
	$400-$499	6	2.8%
	$500-$649	8	3.7%
	$650-$799	7	3.2%
	$800-$999	6	2.8%
	$1000-$1249	15	6.9%
	$1250-$1499	13	6.0%
	$1500-$1749	10	4.6%
	$1750-$1999	12	5.6%
	$2000-$2999	29	13.4%
	$3000 or more	67	31.0%
	Prefer not to say	21	9.7%
Education	Year 11 or below	6	2.7
	Year 12	109	48.7
	Bachelor degree	54	24.1
	Advanced diploma or diploma	21	9.4%
	Postgraduate degree	34	15.2%

Appendix B

Bivariate correlations

	1	2	3	4	5	6	7	8	9	10
1. Physical Health High- Disclosure Likelihood	1	0.544**	0.747**	0.481**	−0.010	−0.086	0.047	−0.137*	−0.029	0.204**
2. Physical Health Low- Disclosure Likelihood			0.452**	0.576**	0.042	−0.024	0.033	−0.020	−0.019	−0.022
3. Mental Health High-Disclosure Likelihood				0.585**	0.074	−0.129	0.040	−0.028	−0.034	0.079
4. Mental Health Low-Disclosure Likelihood					0.025	−0.041	0.062	−0.079	0.030	0.004
5. Gender						0.295**	0.009	0.091	−0.036	−0.070
6. Sexual orientation							−0.074	0.148*	−0.036	−0.100
7. Race								−0.010	0.190**	−0.089
8. Education									0.198**	−0.552**
9. Income										−0.194**
10. Age										

**. Correlation is significant at the 0.001 level (2-tailed).
*Correlation is significant at the 0.05 level (2-tailed).