

RESEARCH

Open Access



Generative AI in healthcare: effect of explanations and the role of AI on trust, perceived privacy, and intent to use

Ruijie Wang^{1*}, Alys Legg¹, Alyssia Jansz¹, Yuanchen Xu² and Liming Chen³

Abstract

Background The use of artificial intelligence (AI), such as ChatGPT, in healthcare has grown rapidly, yet public perceptions of trust, privacy, and intent to use AI for medical purposes remain underexplored. This study aimed to investigate how ChatGPT's role (diagnostic tool vs. explanation assistant) and the level of explanation provided (why, confidence, or none) influence these key factors.

Methods A within-subjects experimental design was employed using a Wizard-of-Oz methodology to systematically manipulate ChatGPT's role (diagnostic tool vs. explanation assistant) and level of explanation (none, confidence, why). Ninety-eight UK-based participants recruited from the general public evaluated hypothetical healthcare scenarios via an online survey. Dependent variables (trust, perceived privacy, and intent to use) were measured using validated scales with satisfactory internal consistency in the present study. Data were analysed using repeated-measures two-way ANOVA to test the main and interaction effects of AI role and explanation level.

Results Trust ($F(1, 97) = 12.40, p < 0.001$) and intent to use ($F(1, 97) = 14.61, p < 0.001$) were significantly higher when ChatGPT acted as an explanation assistant compared to a diagnostic tool. Perceived privacy was highest for "confidence" explanations overall ($F(1.84, 178.85) = 4.22, p = 0.019$) among all types of explanations. The interaction between AI role and explanation level was significant for all outcomes ($p < 0.001$). Among all conditions, a "why" explanation yielded the highest scores when ChatGPT was an explanation assistant for trust, perceived privacy, and intent to use. Interestingly, when ChatGPT was a diagnostic tool, trust, perceived privacy, and intent to use were higher when ChatGPT provided no explanation than when it provided "why" explanation.

Conclusions This study highlights the importance of context in AI design for healthcare. ChatGPT was trusted and preferred more as an explanation assistant, especially when providing "why" or "confidence" explanations. The findings suggest that AI integration into healthcare should emphasise transparency and context-sensitive roles to build public trust and optimise usability.

Keywords Generative AI, ChatGPT, Healthcare, Trust, Perceived privacy, Intent to use, Diagnostic tool, Explanation assistant, Public perceptions

*Correspondence:

Ruijie Wang
rwang3@bournemouth.ac.uk

¹Faculty of Media, Science and Technology, School of Psychology, Bournemouth University, Poole BH12 5BB, UK

²Faculty of Media, Science and Technology, School of Computing and Engineering, Bournemouth University, Poole, UK

³School of Computer Science and Technology, Dalian University of Technology, Dalian, China



Introduction

Advancements in Artificial Intelligence (AI) have significantly reshaped human-computer interaction, transforming various domains, including healthcare [1]. Among these advancements, generative AI, a category of models capable of creating new data by analysing patterns from large datasets, has gained prominence [2]. Large language models (LLMs) such as ChatGPT, trained on extensive text datasets, can generate human-like responses, answer complex questions, and support language-related tasks [3]. The potential applications of generative AI in healthcare are extensive, spanning diagnostic, clinical, surgical, predictive, and rehabilitative domains [4]. These systems can enhance diagnostic accuracy and speed, alleviate healthcare professional burnout, and reduce care costs by automating repetitive tasks [5, 6]. Moreover, AI has demonstrated the ability to tackle under-diagnosis of complex diseases by identifying subtle patterns and associations beyond human capabilities [7].

Generative AI also holds promise for increasing healthcare accessibility and delivering personalised care [8]. For healthcare professionals, AI provides opportunities to shift focus toward clinical creativity and critical thinking [6]. Nevertheless, significant barriers persist, including professional scepticism, limited understanding of AI's mechanisms, and concerns about biases and ethical implications [9]. Among patients, challenges such as biased responses, over-reliance, inappropriate content, and privacy threats have emerged as critical concerns, affecting trust and adoption [10]. One significant concern with generative AI is its inherent biases. For instance, ChatGPT has been found to disproportionately represent Caucasian individuals in medical advice, neglecting other ethnic groups [11]. Such biases often originate from the datasets used to train AI models, which reflect human cognitive biases [12]. This dual susceptibility to bias underscores the importance of addressing fairness and representativeness in training data to mitigate disparities in AI outputs. While scepticism persists, the potential for generative AI to update continuously with the latest medical advancements offers an opportunity for addressing such biases over time [13].

Patient trust plays a pivotal role in the adoption of generative AI in healthcare. Transparency is essential in emerging technologies to provide users with autonomy to facilitate informed decision making [14]. Studies have shown that trust increases when AI systems are transparent, and their advice is corroborated by human professionals. For instance, patients were more willing to trust AI diagnoses when human doctors confirmed the recommendations [15]. Similarly, trust was higher when patients could autonomously choose between AI and human diagnoses [16]. However, concerns about inadequate explanations for AI diagnoses and reduced

autonomy have been highlighted as barriers to trust [17]. These findings suggest that integrating ChatGPT as an explanation assistant, wherein a human doctor provides the initial diagnosis and the AI offers supplementary explanations, could enhance trust and willingness to use. Stereotypes surrounding AI also influence patient trust and adoption. Many users view AI systems as complex and opaque, leading to anxiety and avoidance, consistent with uncertainty reduction theory [18, 19]. Recent advancements in AI have amplified these concerns due to the public's limited understanding of its mechanisms [20]. Evidence suggests that providing detailed, example-based explanations for AI diagnoses can alleviate such apprehensions and foster trust [21]. However, the literature remains divided. While Bussone et al. [22] found that comprehensive explanations increased trust among healthcare professionals, Zhang et al. [23] reported no significant impact of explanations on patient trust when using AI-assisted radiology tools. These inconsistencies highlight the need for further research to explore how explanation formats influence trust in diverse healthcare contexts.

Privacy concerns remain a critical challenge for the integration of generative AI in healthcare. Theoretical models such as the Privacy Calculus Model provide a framework for understanding how individuals balance perceived risks and benefits before disclosing personal information [24]. For example, users weigh factors such as system trust, perceived informativeness, and privacy risks when deciding whether to share health data with AI systems [25]. Privacy concerns are particularly pronounced in healthcare due to the sensitive nature of health information, including mental health, substance abuse, and sexual health data [26]. AI-based healthcare tools exacerbate these concerns by requiring large amounts of personal data for training and operation, increasing the risk of data breaches [1]. Cheng and Jiang [27] found that users hesitated to share information with AI chatbots due to fears of misuse or unpredictable outcomes, which significantly affected user satisfaction. Strategies to address these concerns include enhancing transparency, implementing robust data protection policies, and offering users greater control over their information [28]. For instance, Luo et al. [29] demonstrated that high perceived information control reduced privacy concerns and increased satisfaction in AI-based medical consultations. Despite these concerns, the Privacy Calculus Model suggests that individuals may overlook privacy risks when the perceived benefits of AI systems outweigh potential harms [30]. This "privacy paradox" has been observed in various contexts, including personal health record systems, where users with chronic conditions were more willing to share data due to the perceived benefits of improved care [31, 32]. These findings underscore

the importance of designing AI systems that prioritise transparency, accuracy, and user empowerment to mitigate privacy concerns.

The rise of generative AI, particularly ChatGPT, has necessitated a re-evaluation of its role in healthcare from the user's perspective. ChatGPT's widespread adoption as a tool for medical advice, including self-diagnosis, raises critical questions about its reliability, usability, and ethical implications. For instance, AI symptom checkers often missed crucial medical history details and struggled to interpret nuanced symptom descriptions, posing risks for self-diagnosis [33]. To address these challenges, the current study seeks to contribute to the exploration for optimal integration of generative AI in healthcare by examining its use as a self-diagnosis tool versus an explanation assistant. Specifically, this study aims to investigate how varying levels of explanation provided by ChatGPT influence user trust, perceived privacy, and intent to use. The study aims to answer the following research questions:

- How do different levels of explanation and ChatGPT's role (diagnostic tool vs. explanation assistant) influence user trust?
- How do different levels of explanation and ChatGPT's role influence perceived privacy?
- How do different levels of explanation and ChatGPT's role influence users' intent to use ChatGPT for healthcare purposes?

Methods

Design and materials

This study employed a 3 (level of explanation: "why", "confidence", and no explanation) × 2 (AI role: diagnostic tool versus explanation assistant) within-subject design. The independent variables were the type of explanation provided by the AI and the operational role of the AI, while the dependent variables included trust, perceived privacy, and intent to use. Participants were exposed to all experimental conditions in a randomised order, ensuring control for order effects and minimising potential biases in an online anonymous survey. A Wizard of Oz (WoZ) approach, as used by Bussoni et al. [22], was employed in the present study to ensure controlled and consistent experimental manipulation. In WoZ studies, participants "interact" with a system, while the system behaviour is predetermined or human-controlled. The method is widely used in human-computer interaction and user-experience research before full system implementation [34]. The approach allowed us to predefine ChatGPT prompts and responses to ensure that variations in explanation levels and system roles were fully controlled rather than generated dynamically by the model. This approach simulated ChatGPT's diagnostic or explanatory

responses while maintaining experimental control. In the diagnostic tool condition, ChatGPT was treated as a tool to provide a diagnosis as the output based on patient symptoms as the input. In the explanation assistant condition, ChatGPT was treated as an assistant to provide explanation as the output to an existing diagnosis and symptoms as the input. Each output was accompanied by one of the explanation levels.

Participants were presented with twelve hypothetical clinical scenarios, divided equally between the two AI roles: diagnostic tool (Table 1) and explanation assistant (Table 2), with each role featuring six scenarios (two per explanation level). The scenarios were adapted from Bussoni et al. [22] to include relevant medical contexts while maintaining diversity to reduce potential demand characteristics and fatigue effects:

- In the diagnostic tool condition, each scenario included patient background information, symptom input, and a diagnosis with one of the three types of explanations as ChatGPT's output.
- In the explanation assistant condition, the scenarios included the same elements as the diagnostic tool condition but with the addition of patient-provided examination results. The output in this condition focused on explaining an existing diagnosis rather than providing a new one.

Explanations include three levels:

- "Why" explanation: A detailed rationale behind the diagnosis, simulating a ChatGPT output.
- "Confidence" explanation: A diagnostic confidence score (81%), derived from Haze et al. [35] that established ChatGPT's accuracy on a medical examination at this level.
- No explanation: A simple diagnostic output with no accompanying explanation.

To assess the study's dependent variables, validated scales were used:

- Perceived Privacy: Measured using an 11-item, 7-point Likert scale taken from Dinev et al. [36]¹,

¹Items were adapted from the original scale by replacing "Web sites" with "ChatGPT" for the present research context. Original publication of the scale: Dinev, T., Xu, H., Smith, J. H., & Hart, P. (2013). Information privacy and correlates: an empirical attempt to bridge and distinguish privacy-related concepts. *European Journal of Information Systems*, 22(3), 295–316. <https://doi.org/10.1057/ejis.2012.23>.

Copyright © 2013 The Operational Research Society, reprinted by permission of Informa UK Limited, trading Taylor & Francis Group, <https://www.tandfonline.com> on behalf of The Operational Research Society. Conditional permission to use and adapt the scale items in Appendix A of the original paper has been granted by the publisher and the original authors.

Table 1 Clinical scenarios: AI as a diagnostic tool

Age and Gender	Symptoms input	Diagnosis	Explanation type	ChatGPT output
50, male	'I have been feeling very tired and feeling thirsty, I have noticed blurred vision when doing work in my office. When I injured myself last month, it took a long time to heal. What's my diagnosis?'	Diabetes-type 2	Why	'I'm your personal doctor, and the symptoms you have described are indicative of diabetes. Diabetes is a condition characterised by high blood sugar levels which can lead to these symptoms.'
19, female	'I've been very nauseous and tired, and I have a metallic taste in my mouth. I used to like smoking but now I can't stand the thought of a cigarette as it makes me want to throw up. What could this be?'	Pregnant	Why	'I'm your personal doctor, there are several possible explanations for the symptoms you are experiencing, such as nicotine withdrawal. But the most likely outcome is that you're pregnant. Nausea and changes in taste are symptoms of pregnancy. It's advisable to take a pregnancy test.'
23, male	'I've had a dry cough recently and I haven't been able to smell my wife's cooking in so long. Just now I've walked up the stairs and felt like I couldn't breathe. I think I know what I have but what do you think it could be?'	COVID19	Confidence	'I'm your personal doctor, given the current global situation, I am 81% sure you have COVID 19.'
14, male	'A week ago, my cat scratched my arm and now the cut is really swollen and looks red and yellow. It hurts way more than it did last week, I am scared. What's wrong with it?'	Cut infection	Confidence	'I'm your personal doctor, I am 81% sure your scratch is infected.'
31, female	'My 2-year-old daughter won't stop scratching her stomach. There is a rash of red spots there that seem to be spreading and she's burning up. Some of her peers have been off sick too. What's going on?'	Chicken pox	No	'I'm your personal doctor, your child has chickenpox.'
20, female	'I've been having itchy flare-ups on my skin, especially after I've been swimming or when I'm stressed. It makes my skin really dry, sore and cracked. It's really annoying me, what is it?'	Eczema	No	'I'm your personal doctor, you have eczema.'

covering the three constructs of the information privacy model—perceived privacy (PRIV), perceived information control (PCTL), and perceived privacy risk (RISK), where questions for PCTL were originally developed by Xu [37]. These constructs (PRIV, PCTL, RISK) were adopted for measuring Perceived Privacy in the present study as Dinev et al. [36] found that perceived information control and perceived privacy risk are salient determinants of perceived information privacy.

- Trust and Intent to Use: Developed by Choudhury and Shamszare [38]², this included a 7-item scale measuring Trust (e.g., “ChatGPT is providing the information I need”) and a 3-item scale measuring Intent to Use (e.g., “I am willing to use ChatGPT in the future”). We adopted a 5-point Likert scale to capture participants' responses to all questions,

ranging from 1 = *strongly disagree* to 5 = *strongly agree*, allowing participants to provide a neutral response, i.e., 3 = *neither agree nor disagree*.

Participants

A total of 133 participants provided informed consent and attempted to complete the study. Data were collected between January 2024 and March 2024. Participants were recruited from the United Kingdom through Bournemouth University's SONA research participation system and social media channels. The inclusion criteria were being at least 18 years old and not being a medical student or working in a medical profession, to reduce potential bias in evaluating AI-related healthcare interactions from prior medical knowledge or personal health experience, allowing the study to focus on general user perceptions while maintaining experimental control. The final sample included 98 participants for analysis after accounting for attrition and missing data (80 females, 16 males, and 2 nonbinary individuals; age range = 18–70 years, $M = 25.19$, $SD = 13.40$). All participants were based in the UK, where language and healthcare terminology are largely standardised, minimising cultural or linguistic variation. The sample size ensured adequate statistical power (0.95) to detect a medium effect size ($f = 0.25$),

²Original publication of the scales: Choudhury, A., & Shamszare, H. (2023). Trust and Intent to Use Scale. Journal of Medical Internet Research, 25, e47184. Available from: <https://www.jmir.org/2023/1/e47184>.

© Avishek Choudhury, Hamid Shamszare. Originally published in the Journal of Medical Internet Research (<https://www.jmir.org>), 14.06.2023. Licensed under the Creative Commons Attribution License (<https://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. Disclaimer of Warranties applies (<https://creativecommons.org/licenses/by/4.0/legalcode.en>, Sect. 5).

Table 2 Clinical scenarios: AI as an explanation assistant

Age and Gender	Symptoms input and Examination	Diagnosis	Explanation type	ChatGPT output
47, female	'I've had stiff and painful knees for a long time now. I went to my local GP, and they gave me a C-reactive protein blood test, which showed high inflammation levels. My doctor told me I have Rheumatoid Arthritis but I'm still not sure, would you be able to explain this further?'	Rheumatoid arthritis	Why	'I'm your healthcare assistant and can provide information on Rheumatoid arthritis. Your symptoms of symmetrical joint pain in both knees are common in RA. Elevated C-reactive protein levels can indicate inflammation but can be caused by other various conditions, not just RA. I believe your doctor is right, however it is important to get an X-ray and consult further with your doctor if you are unsure.'
39, male	'My partner noticed that my breathing stops and starts in my sleep, and I have been absolutely knackered all day. I went to the doctor, and they gave me some sort of device to monitor my heart and breathing. They later diagnosed me with sleep apnoea. Is this right? What will this mean for me?'	Sleep apnoea	Why	'I'm your healthcare assistant. If you doctor diagnosed you with sleep apnoea based on a monitoring device, they likely observed abnormal patterns in your breathing and heart rate during sleep. This, combined with your symptoms, highly suggests that your doctor is correct in their diagnosis. It is recommended that you make lifestyle changes such as, avoiding alcohol and sedatives before bed. Your doctor may also suggest continuous positive airway pressure (CPAP), which will help keep your airway open during sleep.'
23, male	'My knees have been really dry and flaky, it looks like they're covered in scales. I went to the doctor and they said I have a skin condition called psoriasis. They gave this cream this cream to test out. I'm still confused about the condition, is my doctor correct?'	Psoriasis	Confidence	'I'm your healthcare assistant. I am 81% sure that your doctor is right and that you have psoriasis.'
45, male	'My son just started playing football but has been wheezing and coughing after matches. The GP said he has asthma and that it can be worse after exercise. He had to do a FeNO test which showed inflammation in his lungs. I'm worried it will impact his football performance. Are you sure the GP is correct?'	Asthma	Confidence	'I'm your healthcare assistant. If diagnosed after completing the FeNO test, I am 81% sure that your GP is right and your son has asthma.'
28, female	'Recently I went to my doctor because I was experiencing muscle stiffness, tingling feelings and vision problems. I had an MRI scan which confirmed a diagnosis of multiple sclerosis. Is this right?'	Multiple sclerosis	No	'I'm your healthcare assistant. Yes, this is right you have multiple sclerosis.'
72, male	'I had problem with my face last week, where it started to droop on one side, my wife called 999 and I had to go to hospital. They gave me a blood test and told me my high cholesterol gave me a stroke, I find this is all very dramatic. Is this even true? Did I actually have a stroke?'	Stroke	No	'I'm your healthcare assistant. The hospital is right, you had a stroke.'

as recommended by Cohen [39] and consistent with research on similar AI topics (e.g [40]).

Procedure

The study was conducted via an online survey platform (Qualtrics), guiding participants step-by-step through the hypothetical ChatGPT interactions and the subsequent dependent variable measures. Completion time ranged from approximately 15–30 minutes. Participants recruited via the SONA system received 0.5 SONA credit as compensation, while participants recruited through social media participated voluntarily without compensation. Participants first reviewed an information sheet outlining the study's information (e.g., purpose, eligibility criteria, time taken, data collection and handling, contact

details) and provided informed consent. Eligibility was verified based on age and exclusion criteria. Participants were instructed to imagine themselves as "patients" interacting with ChatGPT in a series of text-based clinical scenarios. Each scenario featured a well-known health issue. Across the six diagnostic tool scenarios, ChatGPT's output varied by explanation level ("why", "confidence", no explanation). In the diagnostic tool condition, ChatGPT always began with "I'm your personal doctor" whereas in the diagnostic explanation assistant condition, ChatGPT always began with "I'm your healthcare assistant". For the diagnostic tool condition, the "patient" gave ChatGPT a list of health symptoms and asked for a diagnosis. For the explanation assistant condition, the "patient" listed their symptoms and examinations and asked for an

explanation of their existing health diagnosis. An example scenario is illustrated in Fig. 1, where ChatGPT, as a diagnostic tool, provides a “confidence” explanation.

To minimise order effects, the 12 scenarios were presented in a randomised order. After completing the scenarios in each condition, participants answered questionnaires assessing their trust, perceived privacy, and intent to use ChatGPT for healthcare in the future. Participants were then debriefed with information about the purpose and design of the study including the WoZ approach used, support services for physical and mental health, emphasising the importance of seeking professional healthcare advice for any medical concerns and the fact that ChatGPT is not a diagnostic tool and should not replace medical consultation.

Data preparation and analysis

All collected data were processed and analysed using SPSS Statistics for Windows, version 28 (IBM Corp, Armonk, NY, USA). Descriptive statistics were calculated for the dependent variables: trust, perceived privacy, and intent to use. The Shapiro-Wilk test was used to check normality, and Bonferroni correction was applied to adjust for multiple comparisons where variables were not normally distributed ($p < 0.05$). For sphericity assumptions, Mauchly’s test was employed, and Huynh-Feldt corrections were applied when sphericity was violated ($p < 0.001$). Repeated-measures two-way ANOVA was conducted separately for each dependent variable, examining the main effects and interaction effects of AI role and level of explanation. Internal consistency was assessed using Cronbach’s alpha. Overall, all the scales used in our study demonstrated satisfactory internal consistency (Trust $\alpha = 0.934$; Intent to Use $\alpha = 0.934$; Perceived Privacy $\alpha = 0.762$), exceeding the commonly accepted threshold (Cronbach’s $\alpha > 0.70$; [41, 42]), and indicating

satisfactory reliability of the measures in the present sample.

Results

The effect of role of AI and level of explanation on trust

Figure 2 shows the mean scores with standard deviation (SD) as error bars for trust in the six conditions. Repeated-measures two-way ANOVA found a significant main effect of AI role on trust ($F(1, 97) = 12.40, p < 0.001, \eta_p^2 = 0.113$). Participants trusted AI more when it acted as an explanation assistant ($M = 2.93$) compared to when it served as a diagnostic tool ($M = 2.79$). The main effect of explanation level on trust was not significant ($F(1.56, 150.94) = 3.15, p = 0.06, \eta_p^2 = 0.031$).

A significant interaction was observed between AI role and explanation level on trust ($F(1.68, 162.71) = 42.42, p < 0.001, \eta_p^2 = 0.304$). Post-hoc paired t-tests with Bonferroni correction revealed that when AI acted as an explanation assistant, trust was highest for the “why” explanation ($M = 3.14$) compared to the “confidence” explanation ($M = 2.89, p < 0.001$) and no explanation ($M = 2.75, p < 0.001$). Conversely, when AI acted as a diagnostic tool, trust was lower for the “why” explanation ($M = 2.66$) compared to “confidence” explanation ($M = 2.85, p < 0.001$) and no explanation ($M = 2.85, p = 0.002$). For the “no explanation” condition, trust was higher for the diagnostic tool ($M = 2.85$) than for the explanation assistant ($M = 2.75, p = 0.045$). For the “why” explanation, trust was higher for the explanation assistant ($M = 3.14$) than for the diagnostic tool ($M = 2.66, p < 0.001$).

The effect of role of AI and level of explanation on perceived privacy

Figure 3 shows the mean scores with SD as error bars for perceived privacy in the six conditions.

```
Input from the Patient(14, male) : "A week ago, my cat scratched
my arm and now the cut is really swollen and looks red and
yellow. It hurts way more than it did last week, I am scared.
What's wrong with it?"
```

Click the right arrow to see ChatGPT output.

Page Break

Q50

```
ChatGPT Output: "I'm your personal doctor, I am 81% sure your
scratch is infected."
```

Fig. 1 An example of patient input and ChatGPT output presented to participants

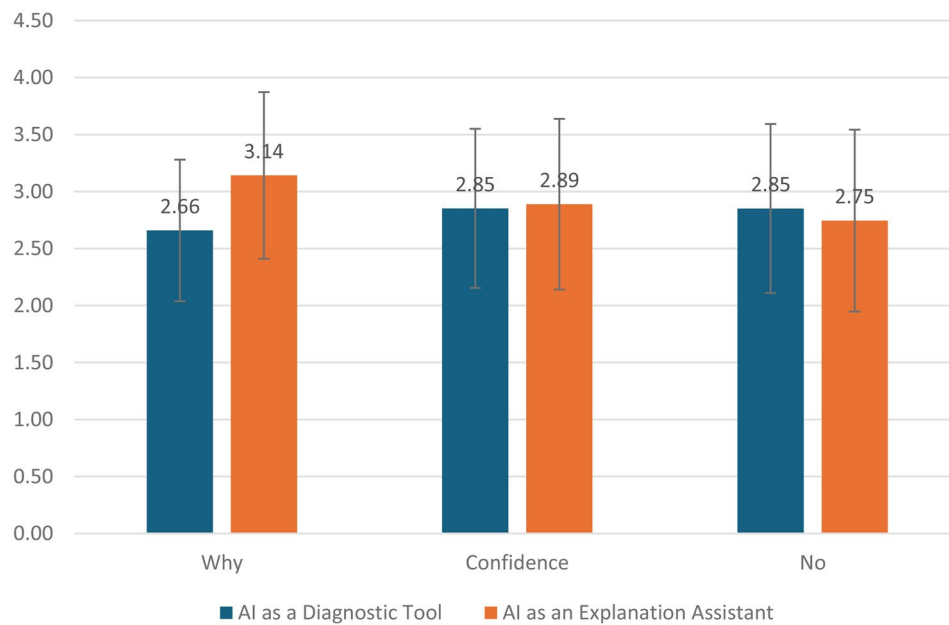


Fig. 2 Trust: the mean score in each condition with an error bar (SD)

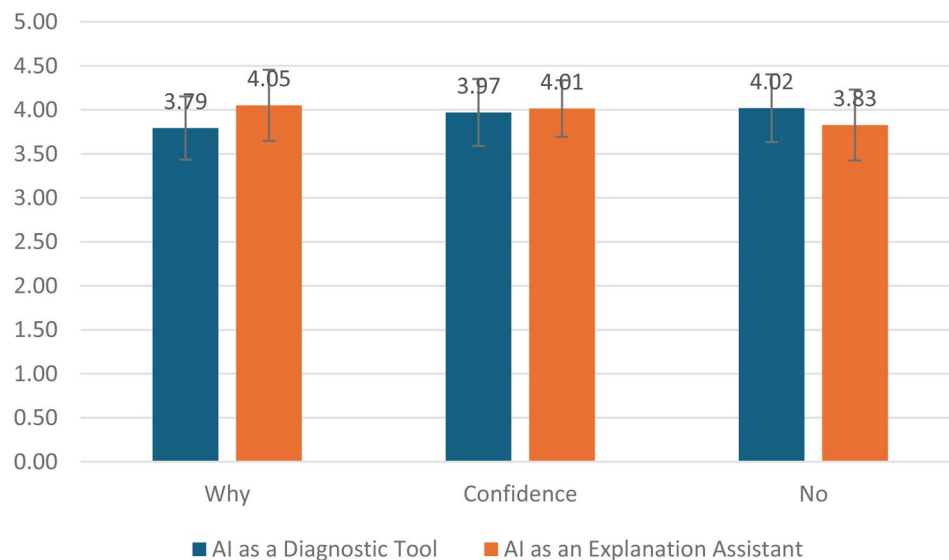


Fig. 3 Perceived privacy: the mean score in each condition with an error bar (SD)

Repeated-measures two-way ANOVA found that the main effect of AI role on perceived privacy was not significant ($F(1, 97) = 1.22, p = 0.272, \eta_p^2 = 0.012$). A significant main effect of explanation level was found ($F(1.84, 178.85) = 4.22, p = 0.019, \eta_p^2 = 0.042$). Perceived privacy was highest for the “confidence” explanation ($M = 3.99$) compared to the “why” explanation ($M = 3.92$) and no explanation ($M = 3.93$).

A significant interaction was observed between AI role and explanation level on perceived privacy ($F(2, 194) = 33.87, p < 0.001, \eta_p^2 = 0.259$). Post-hoc paired t-tests with Bonferroni correction revealed that when AI

acted as an explanation assistant, perceived privacy was higher for the “why” explanation ($M = 4.05, p < 0.001$) or “confidence” explanation ($M = 4.01, p < 0.001$) compared to no explanation ($M = 3.83$). Conversely, when AI acted as a diagnostic tool, perceived privacy was lower for the “why” explanation ($M = 3.79$) compared to “confidence” explanation ($M = 3.97, p < 0.001$) and “no” explanation ($M = 4.02, p < 0.001$). For the “no explanation” condition, perceived privacy was higher when the AI acted as a diagnostic tool ($M = 4.02$) compared to the explanation assistant ($M = 3.83, p < 0.001$). In contrast, for the “why” explanation, perceived privacy was higher for the

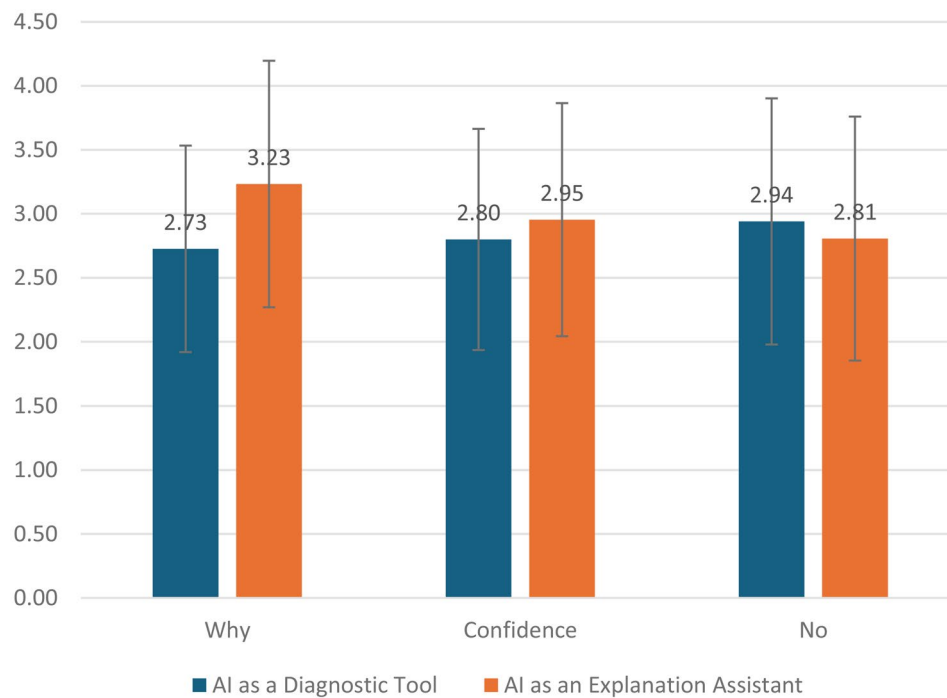


Fig. 4 Intent to use: the mean score in each condition with an error bar (SD)

explanation assistant ($M = 4.05$) compared to the diagnostic tool ($M = 3.79$, $p < 0.001$).

The effect of role of AI and level of explanation on intent to use

Figure 4 shows the mean scores with SD as error bars for intent to use in the six conditions. Repeated-measures two-way ANOVA found a significant main effect of AI role on intent to use ($F(1, 97) = 14.61$, $p < 0.001$, $\eta_p^2 = 0.131$). Participants had a higher intent to use AI when it acted as an explanation assistant ($M = 3.00$) compared to a diagnostic tool ($M = 2.82$). The main effect of explanation level is not significant ($F(1.47, 143.23) = 2.46$, $p = 0.105$, $\eta_p^2 = 0.025$).

The interaction effect between explanation level and AI role on intent to use was also significant ($F(1.74, 168.61) = 28.60$, $p < 0.001$, $\eta_p^2 = 0.228$). Post-hoc paired t-tests with Bonferroni correction revealed that when AI acted as an explanation assistant, intent to use was highest for the “why” explanation ($M = 3.23$) compared to the “confidence” explanation ($M = 2.95$, $p < 0.001$) and no explanation ($M = 2.81$, $p = 0.004$). In contrast, when AI acted as a diagnostic tool, intent to use was lower for the “why” explanation ($M = 2.73$, $p = 0.009$) or “confidence” explanation ($M = 2.80$, $p = 0.023$) compared to no explanation ($M = 2.94$). For the “why” explanation, intent to use was higher for the explanation assistant ($M = 3.23$) than for the diagnostic tool ($M = 2.73$, $p < 0.001$). Likewise, for the “confidence” explanation, intent to use was higher for the explanation assistant ($M = 2.95$) compared

to the diagnostic tool ($M = 2.80$, $p = 0.027$). However, for the “no explanation” condition, intent to use was higher for the diagnostic tool ($M = 2.94$) than for the explanation assistant ($M = 2.81$, $p = 0.03$).

Discussion

The study aimed to explore participants’ trust in ChatGPT for healthcare queries, their perceived privacy regarding medical data, and intent to use ChatGPT for future healthcare purposes. It examined how these factors were influenced by the AI’s role (explanation assistant or diagnostic tool) and the level of explanation (why, confidence, or none). The findings provide meaningful insights into public perceptions of ChatGPT in healthcare settings and contribute to the understanding of how explanation levels and AI roles shape trust, privacy, and usage intentions.

The results revealed that trust and intent to use were significantly higher when ChatGPT functioned as an explanation assistant compared to a diagnostic tool. This aligns with Yang et al. [15], who observed that trust and intent to use increased when AI was positioned as a supportive tool for human-led diagnoses rather than a standalone diagnostic system. These findings suggest that participants value ChatGPT more as a complementary tool, particularly when it supports a human professional’s judgment. This also reflects prior research [43] showing that the context of AI use can shape user trust and adoption. Moreover, ChatGPT’s role as an explanation assistant likely amplified perceived benefits of the technology,

which is a known driver of intent to use [31]. By providing comprehensive and understandable explanations, ChatGPT may have been perceived as more helpful and engaging. This explanation role capitalises on ChatGPT's ability to simplify complex medical information, as noted by Kasneci et al. [3]. It also highlights that participants likely viewed ChatGPT as more reliable and transparent in its supportive role, consistent with Shahsavar and Choudhury's finding [44] about higher user satisfaction with ChatGPT in an explanation capacity. However, the lower trust and intent to use ChatGPT as a diagnostic tool could be attributed to its lack of recognition as a standard diagnostic system [44] and potential participant scepticism about its standalone diagnostic reliability. This echoes Seoni et al. [45] who noted that public trust in AI is limited due to stereotypes and concerns about its capabilities.

The study identified a significant effect of explanation level on perceived privacy, with the "confidence" explanation resulting in the highest perceived privacy. This somewhat contrasts with prior research [22] suggesting that more detailed explanations, such as the "why" explanation, might enhance privacy perceptions by fostering trust. The difference may be explained through the Privacy Calculus Model [24], which posits that perceived privacy reflects a balance between risks and benefits. "Why" explanations, while detailed, may increase privacy concerns as they provide abundant information, potentially raising fears about how this data could be used [27], and "no explanations" provide minimal information, which may cause uncertainty and diminish trust. Therefore "confidence" explanations strike a balance by offering enough information to establish trust while avoiding overexposure of sensitive details. Moreover, uncertainty reduction theory [18, 19] suggests that users prefer information that reduces ambiguity without introducing new sources of cognitive strain or perceived vulnerability. "Confidence" explanations may strike this optimal balance: they may provide enough information to reduce uncertainty about the system's reliability, while not exposing the internal reasoning in a way that could trigger concerns about sensitive data use or model transparency risks. From a risk-benefit trade-off perspective [24, 46], "confidence" explanations may present low perceived risk (minimal detail shared, reducing feelings of exposure) while still offering a meaningful benefit (clarity regarding how certain or uncertain the AI is). This balance may explain why they ranked highest on perceived privacy across conditions. Interestingly, perceived privacy was higher for no explanation when ChatGPT served as a diagnostic tool. This unexpected finding may reflect context sensitivity: in diagnostic scenarios, detailed explanations may be seen as particularly intrusive and implying deeper access to personal information. Thus, the absence

of explanation may make users feel less exposed, particularly in sensitive contexts like healthcare diagnoses. This result further supports the idea that interaction context is a critical factor influencing privacy perceptions [43]. These findings highlight that explanation design in healthcare AI is not simply a linear relationship; instead, perceived privacy may depend on the type of explanation and the context in which it is delivered, shaped by users' risk assessments, uncertainty management needs, and expectations of appropriate information flow.

The interaction between AI role and explanation level significantly influenced all three variables (trust, perceived privacy, and intent to use). When ChatGPT served as an explanation assistant, "why" and "confidence" explanations consistently led to higher trust, privacy, and intent to use, likely due to the detailed and supportive nature of these explanations. Conversely, when ChatGPT acted as a diagnostic tool, "no explanation" was associated with higher perceived privacy and intent to use, highlighting the contextual relevance of explanation types. These results suggest that users' preferences for information vary depending on the perceived role of AI in the interaction. These findings align with the Privacy Calculus Model [24], but also suggest a need to update this model to incorporate the context of AI interaction as a factor influencing privacy perceptions. While perceived privacy is often cited as a barrier to adopting AI in healthcare [8], this study suggests that strategically designed AI interactions can mitigate these concerns.

The findings have several practical implications for the design and deployment of AI in healthcare. First, the results indicate the importance of positioning generative AI tools like ChatGPT as explanation assistants rather than standalone diagnostic tools, suggesting that initial implementations should focus on supportive, explanatory roles. This positioning can enhance public trust, alleviate privacy concerns, and increase intent to use. Second, the significant interaction between AI role and explanation level suggests that explanations should be context-sensitive and tailored to the AI's intended role, avoiding overly detailed justifications when AI is providing diagnostic outputs while providing transparent explanations when used as an explanation assistant. Moreover, clear information about AI's confidence in its output is valued by users as it can enhance their perceptions of privacy. Finally, the findings highlight the potential of generative AI to reduce healthcare professional burnout by taking on explanatory tasks, as noted by Alabi et al. [5]. This role could increase patient satisfaction by improving the accessibility and comprehensibility of medical information. Our findings provide actionable insights for developers, healthcare organisations, and policymakers to consider for AI usability, trust, and acceptance in real-world healthcare settings.

The present study focuses on public perspectives, providing valuable insights into the factors shaping trust, privacy, and adoption of generative AI in healthcare. The precise manipulation of independent variables enabled causal inferences and added to the experimental rigor. Nonetheless, the study has its limitations. The lack of detailed demographic data (e.g., health status) limits the generalisability of findings. Prior research [47] has shown that health status can influence perceived privacy, suggesting that results might differ for healthy versus unhealthy participants. Moreover, the participants were predominantly young, female, and members of the general public rather than medically trained individuals or real patients. Although appropriate for our focus on general user perceptions in the present study, this constrained demographic profile may not represent the broader range of healthcare users who may interact with AI in real settings. Furthermore, although it enabled a well-controlled experimental design, the use of pre-defined scenarios and a WoZ design may not fully capture real-world interactions, limiting the applicability of findings to actual patient-AI interactions and reducing ecological validity. While strong for identifying trends, quantitative methods may overlook the nuanced reasons behind participants' preferences [38]. Another limitation of the present study is that we did not collect data on participants' race/ethnicity or education level. While the sample was recruited from the United Kingdom, where language and healthcare terminology are largely standardised, it is possible that cultural, educational, or demographic factors could influence how participants perceive and interact with AI-generated healthcare information. Future research should consider including these variables to examine potential moderating effects and to enhance the generalisability of findings across diverse populations. Additionally, the outcomes measured were self-reported attitudes and behavioural intentions rather than actual behaviours. While these provide valuable indicators of user perceptions, they do not necessarily translate into real-world adoption or usage patterns, and the use of different Likert-scale formats across dependent variables may constrain direct cross-variable comparison and may introduce minor differences in response tendencies. To address these limitations, future studies should involve real patients interacting with ChatGPT or similar generative AI systems to provide more ecologically valid insights in more authentic settings. Future studies can also incorporate behavioural measures or longitudinal designs to assess whether stated intentions correspond to real usage. Combining quantitative and qualitative methods can further provide deeper insights into users' preferences, concerns, and motivations. Additionally, further research should explore how generative AI can be

tailored to specific healthcare contexts to maximise trust, privacy, and usability.

Conclusions

This study contributes to the wider adoption and effective integration of generative AI in healthcare by examining the effects of ChatGPT's role (explanation assistant vs. diagnostic tool) and explanation level (why, confidence, or none) on trust, perceived privacy, and intent to use in healthcare contexts. The findings highlight that trust and intent to use were significantly higher when ChatGPT acted as an explanation assistant, reflecting user preference for AI systems that complement rather than replace human professionals. These results align with previous research suggesting that context and perceived benefits strongly influence trust and adoption of AI in healthcare [15, 39]. Perceived privacy was highest with "confidence" explanations, suggesting users prefer a balance of sufficient but not excessive information. However, when ChatGPT acted as a diagnostic tool, providing no explanation unexpectedly resulted in higher perceived privacy and intent to use. This highlights the importance of tailoring AI design to specific contexts, as users' preferences shift depending on the AI's role and the sensitivity of the task. Future studies should test these effects with real patients and more diverse populations in authentic clinical settings.

In conclusion, positioning ChatGPT as an explanation assistant with carefully designed explanations appears to enhance trust, perceived privacy, and intent to use. These findings have important implications for optimising generative AI in healthcare, fostering its ethical adoption, alleviating burdens on healthcare professionals, helping to balance patient satisfaction, data privacy, and healthcare efficiency.

Acknowledgements

Not applicable.

Author contributions

Conceptualization of the study: R. W. and Y. X.; Design of the study and methodology: R. W., A. L., and A. J.; Investigation and data collection: A. L., and A. J.; Formal analysis and interpretation: R. W., A. L., and A. J.; Writing of the main manuscript – original draft: R. W., A. L., and A. J.; Review and editing: Y. X. and L. C.; Visualization: R. W. All authors reviewed the manuscript.

Funding

No funding.

Data availability

The dataset utilized in the analysis of this study could be requested and made available in accordance with the ethical guidelines. Subject to reasonable requests and ethical approval, the author will provide the anonymous dataset associated with this study. Please contact the corresponding author at rwang3@bournemouth.ac.uk to request access to the anonymous data.

Declarations

Ethics approval and consent to participate

Ethics approval was obtained from the School of Psychology at Bournemouth University in accordance with the standards and guidelines set out by the British Psychological Society (BPS) Code of Ethics and Conduct. The present study complied with the Declaration of Helsinki principles. Informed consent to participate was obtained from all of the participants in the study.

Consent for publication

Not applicable.

Competing interests

The authors declare no competing interests.

Received: 23 December 2024 / Accepted: 3 June 2026

Published online: 11 June 2026

References

- Chen Y, Esmailzadeh P. Generative AI in Medical Practice: In-Depth Exploration of Privacy and Security Challenges. *J Med Internet Res*. 2024;26:e53008.
- Feuerriegel S, Hartmann J, Janiesch C, Zschech P. Generative AI. *Bus Inf Syst Eng*. 2024;66:111–26.
- Kasneji E, Sessler K, Küchemann S, Bannert M, Dementieva D, Fischer F, et al. ChatGPT for good? on opportunities and challenges of large language models for education. *Learn Individ Differ*. 2023 Apr;103:102274.
- Secinaro S, Calandra D, Secinaro A, Muthurangu V, Biancone P. The role of artificial intelligence in healthcare: a structured literature review. *BMC Med Inform Decis Mak*. 2021;21:125.
- Alabi RO, Almangush A, Elmusrati M, Mäkitie AA. Deep Machine learning for oral cancer: from precise diagnosis to precision medicine. *Front Oral Heal*. 2022 Jan;2:794248.
- Meskó B, Drobni Z, Bényei É, Gergely B, Gyórfy Z. Digital health is a cultural transformation of traditional healthcare. *mHealth*. 2017;3:38.
- Doyle OM, Leavitt N, Rigg JA. Finding undiagnosed patients with hepatitis C infection: an application of artificial intelligence to patient claims data. *Sci Rep*. 2020;10:10521.
- Singh JP. Quantifying Healthcare Consumers' Perspectives: An Empirical Study of the Drivers and Barriers to Adopting Generative AI in Personalized Healthcare. *Res Rev Sci Technol*. 2022;2:171–93.
- Singh RP, Hom GL, Abramoff MD, Campbell JP, Chiang MF. Current challenges and barriers to real-world artificial intelligence adoption for the healthcare system, provider, and the patient. *Transl Vis Sci Technol*. 2020;9:1–6.
- Nah FF-H, Zheng R, Cai J, Siau KL, Chen L. Generative AI and ChatGPT: Applications, challenges, and AI-human collaboration. *J Inf Technol Case Appl Res*. 2023;25:277–304.
- Liu Z, Zhang L, Wu Z, Yu X, Cao C, Dai H, et al. Surviving ChatGPT in healthcare. *Front Radiol*. 2024;3:1–7.
- Kliegr T, Bahník Š, Fůrnkranz J. A review of possible effects of cognitive biases on interpretation of rule-based machine learning models. *Artif Intell*. 2021;295:103458.
- Davenport T, Kalakota R. The potential for artificial intelligence in healthcare. *Futur Healthc J*. 2019;6:94–8.
- Wang R, Bush-Evans R, Arden-Close E, Bolat E, McAlaney J, Hodge S, et al. Transparency in persuasive technology, immersive technology, and online marketing: Facilitating users' informed decision making and practical implications. *Comput Human Behav*. 2023 Jan;139:107514.
- Yang K, Zeng Z, Peng H, Jiang Y. Attitudes Of Chinese Cancer Patients Toward The Clinical Use Of Artificial Intelligence. *Patient Prefer Adherence*. 2019;13:1867–75.
- Juravle G, Boudouraki A, Terziyska M, Rezlescu C. Trust in artificial intelligence for medical diagnoses. *Prog Brain Res*. 2020;253:263–82.
- Nelson CA, Pérez-Chada LM, Creadore A, Li SJ, Lo K, Manjaly P, Pournamdari AB, Tkachenko E, et al. Patient Perspectives on the Use of Artificial Intelligence for Skin Cancer Screening: A Qualitative Study. *JAMA dermatology*. 2020;156:501–12.
- Berger CR, Calabrese RJ. Some explorations in initial interaction and beyond: Toward a developmental theory of interpersonal communication. *Hum Commun Res*. 1975;1:99–112.
- Griffin MA, Grote G. When is more uncertainty better? A model of uncertainty regulation and effectiveness. *Acad Manag Rev*. 2020;45:745–65.
- Richardson JP, Smith C, Curtis S, Watson S, Zhu X, Barry B, et al. Patient apprehensions about the use of artificial intelligence in healthcare. *NPJ Digit Med*. 2021;4:140.
- Alam L, Mueller ST. Examining the effect of explanation on satisfaction and trust in AI diagnostic systems. *BMC Med Inform Decis Mak*. 2021 Jun 3;21(1):178.
- Bussone A, Stumpf S, O'Sullivan D. The role of explanations on trust and reliance in clinical decision support systems. In: 2015 international conference on healthcare informatics (ICHI); 2015 Oct 21–23; Dallas, TX. Piscataway (NJ): IEEE. 2015;160–9.
- Zhang Z, Genc Y, Wang D, Ahsen ME, Fan X. Effect of AI explanations on human perceptions of patient-facing ai-powered healthcare systems. *J Med Syst*. 2021 Jun 2;45(7):73.
- Dinev T, Hart P. An Extended Privacy Calculus Model for E-Commerce Transactions. *Inf Syst Res*. 2006;17:61–80.
- Xu H, Teo HH, Tan BCY, Agarwal R. The Role of Push-Pull Technology in Privacy Calculus: The Case of Location-Based Services. *J Manag Inf Syst*. 2009;26:135–74.
- Esmailzadeh P. The Effects of Public Concern for Information Privacy on the Adoption of Health Information Exchanges (HIEs) by Healthcare Entities. *Health Commun*. 2018;34:1202–11.
- Cheng Y, Jiang H. How Do AI-driven Chatbots Impact User Experience? Examining Gratifications, Perceived Privacy Risk, Satisfaction, Loyalty, and Continued Use. *J Broadcast Electron Media*. 2020;64:592–614.
- Chan CKY, Hu W. Students' voices on generative AI: perceptions, benefits, and challenges in higher education. *Int J Educ Technol High Educ*. 2023 Jul 13;20(1):43.
- Luo Y, Li X, Ye Q. The Impact of Privacy Calculus and Trust on User Information Participation Behavior in AI-based Medical Consultation - The Moderating Role of Gender. *J Electron Commer Res*. 2023;24:48–67.
- Hew JJ, Lee VH, Ooi KB, Lin B. Mobile social commerce: The booster for brand loyalty? *Comput Human Behav*. 2016 Jun;59:142–154.
- Shen N, Bernier T, Sequeira L, Strauss J, Silver MP, Carter-Langford A, et al. Understanding the patient privacy perspective on health information exchange: A systematic review. *Int J Med Inf*. 2019;125:1–12.
- Zhang Y, Lang P, Zheng D, Yang M, Guo R. A Secure and Privacy-Aware Smart Health System with Secret Key Leakage Resilience. *Secur Commun Networks*. 2018. <https://doi.org/10.1155/2018/7202598>.
- You Y, Gui X. Self-Diagnosis through AI-enabled chatbot-based symptom checkers: user experiences and design considerations. *AMIA Annu Symp Proc*. 2021 Jan 25;2020:1354–63.
- Schlögl S, Doherty G, Luz S. Wizard of Oz experimentation for language technology applications: Challenges and tools. *Interact Comput*. 2015;27:592–615.
- Haze T, Kawano R, Takase H, Suzuki S, Hirawa N, Tamura K. Influence on the accuracy in ChatGPT: Differences in the amount of information per medical field. *Int J Med Inf*. 2023;180:105283.
- Dinev T, Xu H, Smith J, Hart P. Information privacy and correlates: an empirical attempt to bridge and distinguish privacy-related concepts. *Eur J Inf Syst*. 2013;22:295–316.
- Xu H. The effects of self-construal and perceived control on privacy concerns. *ICIS 2007 Proc - Twenty Eighth Int Conf Inf Syst*. 2007:1–14.
- Choudhury A, Shamszare H. Investigating the Impact of User Trust on the Adoption and Use of ChatGPT: Survey Analysis. *J Med Internet Res*. 2023;25:e47184.
- Cohen J. *Statistical power analysis for the behavioral sciences*. 2nd ed. New York (NY): Lawrence Erlbaum Associates; 1988.
- Hamm P, Klesel M, Coberger P, Wittmann HF. Explanation matters: An experimental study on explainable AI. *Electron Mark*. 2023;33:17.
- Nunnally JC, Bernstein IC. *Psychometric theory*. 3rd ed. New York (NY): McGraw-Hill; 1994.
- Taber KS. The Use of Cronbach's Alpha When Developing and Reporting Research Instruments in Science Education. *Res Sci Educ*. 2018;48:1273–96.
- Schwesig R, Brich I, Buder J, Huff M, Said N. Using artificial intelligence (AI)? Risk and opportunity perception of AI predict people's willingness to use AI. *J Risk Res*. 2023;26:1053–84.
- Shahsavari Y, Choudhury A. User Intentions to Use ChatGPT for Self-Diagnosis and Health-Related Purposes: Cross-sectional Survey Study. *JMIR Hum factors*. 2023;10:e47564.

45. Seoni S, Jahmunah V, Salvi M, Barua PD, Molinari F, Acharya UR. Application of uncertainty quantification to artificial intelligence in healthcare: A review of last decade (2013–2023). *Comput Biol Med.* 2023;165:107441.
46. Acquisti A, Brandimarte L, Loewenstein G. Privacy and human behavior in the age of information. *Science.* 2015;347:509–14.
47. Bansal G, Zahedi F, Mariam, Gefen D. The impact of personal dispositions on information sensitivity, privacy concern and trust in disclosing health information online. *Decis Support Syst.* 2010;49:138–50.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.