"I don’t trust it, but I have to trust it": The Paradox of Trust vs Use of Online Technology Across The Mental Health Spectrum

Gentry, E[1], Nichele, E. [2], Azim, T.[3], Naiseh, M.[4], Dowthwaite, L.[5], Armouch, M.S[1,7]. Greenway, F.[3,7], Weal, M.J.[3], Hughes, A.M.[6], and Palmer-Cooper, E.[7]

[1] School of Medicine, Institute of Mental Health, University of Nottingham, Nottingham, NG7 2TU

[2] Department of Management, College of Arts Social Sciences & Humanities, University of Lincoln, Lincoln, LN6 7TS

[3] School of Electronics and Computer Science, University of Southampton, Southampton, SO17 1BJ

[4] Department of Computing and Informatics, University of Bournemouth, Poole, BH12 5BB

[5] School of Computer Science, Faculty of Science, University of Nottingham, Nottingham NG8 1BB

[6] School of Health Sciences, University of Southampton, Southampton SO17 1BJ, UK

[7]School of Psychology, Centre for Innovation in Mental Health, University of Southampton, Southampton, SO17 1BJ

Corresponding Author: Emma Palmer-Cooper – epalmercooper@gmail.com

Keywords: eHealth, Privacy paradox, Self-awareness, Mental health, Trust

Highlights

- We demonstrate that the Privacy Paradox extends to attitudes and interactions with eHealth systems, whereby users are mistrustful of internet-based healthcare systems, and how personal data is used, but nonetheless use these systems for convenience or lack of alternatives.
- This effect appears regardless of mental health status.
- There are ethical considerations for development and prescription of eHealth interventions for mental health and wellbeing.

Abstract

Background: Interest in eHealth has grown since the Coronavirus (COVID-19) pandemic. Use of internet-based technologies (IBTs) and artificial intelligence (AI) has the potential to transform the delivery of mental healthcare services, however, trust remains a pivotal factor in public acceptance and adoption of these systems.

Aims: We investigated attitudes and behaviours towards eHealth services, with a focus on mental health and wellbeing provision, in the general population and individuals with experience of serious mental illness. Our investigation was underpinned by the extended technology acceptance model (TAM2), which included trust.

Methods: After trialling a cognitive training exercise involving autonomous feedback, we prompted participants’ views on and trust in IBTs, autonomous systems (AS) and AI for health care. We conducted 22 semi-structured interviews in total, including 8 individuals who declared having experience of a serious mental illness.

Results: We principally identified the privacy paradox extends to eHealth, whereby individuals engaged with IBTs despite distrusting them and/or having privacy concerns regarding them, and this was across all participants. Behaviours instead were driven by both convenience, ease of use, and lack of choice or alternatives.

Conclusions: Whilst trust is a factor in uptake and engagement with eHealth, there are other factors involved. It is concerning that individuals will utilise eHealth systems despite mistrusting them or their developers. There are clear ethical implications for both healthcare providers prescribing eHealth, and developers of these systems, with considerations relevant across the mental health and wellbeing spectrum. To foster trust in IBTs, particularly those using AI, a balance is needed between human and eHealth provision. This may lead to greater trust and acceptability of systems, yielding better outcomes for patients.

1. Introduction

eHealth is the delivery of healthcare services, using digital information and communication technologies, known as internet-based technologies (IBTs). eHealth often uses Artificial Intelligence (AI), the use of intelligent algorithms, to perform tasks or make decisions that would otherwise require human intelligence. Autonomous Systems (AS) are used in many sectors of the economy from social media to self-driving cars, or aspects of diagnosis and healthcare delivery. These systems can deploy a range of AI techniques, from expert system logic models through to deep-learning machine learning techniques such as Large Language Models (LLMs). Within eHealth service delivery, AI techniques have been used in the delivery of interventions[1] and the guiding of diagnostic processes[2]. The benefits of such techniques need to be matched by better understanding of issues around ethics, bias and trustworthiness[3]. eHealth for mental health can support a diverse range of needs, in a variety of health care settings, through providing AI support , from text-messaging, peer support and illness self-management[4], to adjunct or out-of-hours support and web-based psycho-education and therapeutic interventions, and be useful alongside well-established interventions[5], [6]. eHealth can improve medication adherence and clinical engagement with psychological services in serious mental illnesses (SMI)[7], [8]. eHealth can also improve symptoms of mental health when used as an adjunct to psychotherapy [9]. During the coronavirus pandemic, the rapid shift to online therapy further highlighted the importance and usefulness of wide-ranging applications of eHealth to support mental health[5].

Across diagnoses, dropout rates from psychotherapy, such as Cognitive Behavioural Therapy (CBT), vary from 15.9%, at pre-treatment, and 26.2%, during treatment [10], where practical factors, such as setting (in-patient vs. outpatient), format (in-person vs. online), and duration influence the rate of drop-outs. Therapeutic eHealth technologies can allow individuals to complete interventions with greater convenience and may overcome practical and interpersonal barriers[11]. eHealth also has positive implications for destigmatisation and patient empowerment through managing treatment in a variety of formats [12]. Nonetheless, clear individual differences drive a preference for in-person versus eHealth support and must be carefully considered, in addition to personal experiences, abilities and beliefs [13], [14], [15]. Yet, a significant barrier to engaging with eHealth is trust.

The future role that eHealth should play in the delivery of therapeutic interventions, particularly considering the use of AI such as chatbots[16], remains an open debate. Even though automation may support improved access and adherence to healthcare, the views and experiences of users across the mental health spectrum (from good mental health to severe, negative mental health experiences) must be considered in their design, development and implantation.

1.1. Trust: Attitudes and Behaviours Towards IBTs

Trust involves both a trustor and a trustee that may be a person, computer, or agent[17]. Trust may be defined as "the willingness of a party to be vulnerable to the actions of another party based on the expectation that the other will perform a particular action important to the trustor, irrespective of the ability to monitor or control the other parties" [p.712]. Its conceptualisation may equally involve the condition of risk [18] and a reliance on others as a condition of that risk [19]. A distinction has been made between trust in online and offline environments[20] where the object of offline trust is typically a human or an entity (organisation), whilst in online trust the object is ’the internet’. This raises questions on whether it is truly possible for humans to experience trust in a nonhuman entity [21]. Trust in an agent can be influenced by an expectation of capability, or discrepancy between individuals’ expectations and the agent's real capabilities[22], posing a challenge in eHealth where repeated engagement is vital for patient treatment, especially among populations that may experience a lack of trust in both online and offline settings. Individuals’ propensity to trust and source trustworthiness influence this so-called *intention* to trust[23], where propensity to trust is related to clarity of source trustworthiness. Individual level factors, the health system and its reputation, and interaction-related factors also play a part in influencing trust in eHealth[24].

Understanding patients’ perspectives on internet-based technologies (IBTs) and eHealth, can help us understand how IBTs are perceived and experienced, and what aspects are helpful or barriers to engagement. This can, in turn, help to develop more acceptable interventions. For instance, work on respiratory, cardiovascular and chronic pain conditions has emphasized the need for user-centred flexible design and integration with existing healthcare systems [25], [26] alongside the pivotal role of trust in healthcare providers and technology, underscoring the intricate interplay between technological features and the human element in healthcare[27]. Investigations into digital platforms for monitoring mental health conditions revealed that ease of

use, accessibility, and tailored content were facilitators, while challenges like technical issues, a lack of human interaction, and privacy concerns hindered engagement [28]. As the 'Human-AI' interaction continues to advance, the role of trust in the practical success of AI tools for eHealth cannot be overstated [29]. Research indicates public scepticism about reliance on AI, preferring human expert involvement in critical decision-making scenarios, even when these experts may occasionally err [30], [31]. Consequently, operationalising AI systems in real-world applications necessitates an understanding of human trust in this technology. Intersecting the consideration of trust in eHealth use is the observation that behaviour does not always reflect attitudes. For example, the privacy paradox describes the tendency for users to be lax around privacy-protecting measures, despite holding attitudes of concern [32].

Several factors contribute to the need to understand and address trust in AI systems. Firstly, research has consistently demonstrated that trust serves as a prerequisite for the acceptance, adoption, and effective use of technology, in general, [33], [34] and AI, in particular[35]. Secondly, the adoption of AI-based support systems is not a one-time event; rather, trust is an ongoing, dynamic process characterized by fluctuations between under-trust and over-trust[36]. Therefore, it has become increasingly essential in the literature to quantify when, how, why, and under what conditions individuals tend to exhibit under-trust or over-trust in these systems[37].

Integration of AI and autonomous systems into widespread eHealth applications is not a completely new concept. One example is the NHS COVID-19 tracing app which incorporated autonomous decision-making. While many may be familiar with the application itself, a recent study found that only one-fifth of participants were aware of the autonomous element [33] and that distrust impeded participants' willingness to accept and subsequently adopt the app on a long-term basis. The authors additionally found a lower sense of trust in the NHS among BAME participants, prompting concerns around accessibility and inclusivity with online healthcare systems. We endeavour to build upon this work by looking at the role of individual differences in trust, as a way to develop more inclusive systems by design.

1.2. Intersectional Impact of Trust and Paranoia

The experience of suspicion and paranoia is most common in individuals living with serious mental health problems. Whilst 0.7% of the adult population is diagnosed with psychosis-related

disorders[38] , trait paranoia exists on a continuum that is measurable in the general population. Further, cyber-paranoia and cyber-fear [39] related to paranoia and fear about ’cyber’ or IBTs, is present in the general population, and distinct from paranoid, psychosis-related delusions. These phenomena are understood to be more frequent for those with less understanding of how the IBTs work. Internet-related paranoia and suspicion have been reported since the introduction of the internet, [40] but have been more recently identified as a barrier to clinical engagement via telehealth systems, during the Coronavirus pandemic [41]. The extent to which these beliefs impact the use of IBTs and eHealth, in clinical and general populations, remains unknown and warrants further investigation. Similarly, additional research is needed to understand how internet-related paranoid beliefs may interact with trust in IBTs.

In this context, our qualitative investigation focuses on trust in AI, AS and IBT tools, employed in eHealth applications, exploring how users’ trust in these systems is constructed and evolves. Particularly, this research aims to shed light on the nuanced dynamics of trust, in the realm of mental health and eHealth technologies, contributing valuable insights for the effective integration and acceptance of such technologies in practical settings. This was achieved by asking participants to engage with a simple AI cognitive training programme and providing their thoughts on the system, as well as the broader concept of IBTs, AI and AS.

We drew on the following TAM2 extended framework [33] which includes the added variable of trust. However, in order to make these systems more inclusive by design, we must understand whether experiences of trust differ across the mental health spectrum.

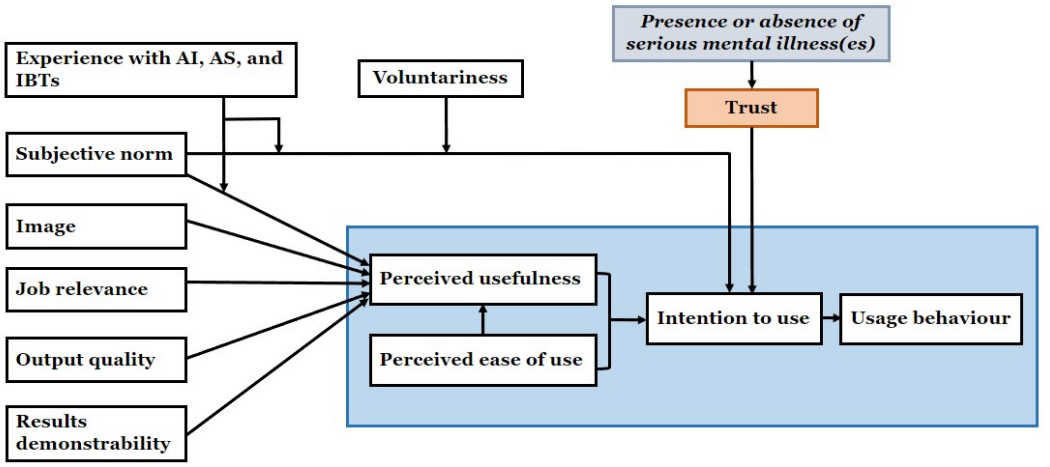

Figure 1: A proposed model for the influence of serious mental illness on trust, experience, and intention to use eHealth applications, based on TAM2 with the added component of trust [33]

1.3. Research Questions

We set out to understand the attitudes towards eHealth services among people with experience of serious mental illness (SMI), as well as the general population.

RQ1: What attitudes do individuals have towards trusting eHealth services?

RQ2: Do attitudes towards eHealth services differ between people with experience of serious mental illness (SMI) and the general population?

2. Methods

2.1. Design

This online interview study invited participants who completed a quantitative questionnaire to participate in an interview. Participants were asked to spend 20 minutes using an eHealth cognitive training application before the interview.

2.1.1. Public Involvement Activities

The study was conceived after discussions with mental health practitioners and people with experience of serious mental illness (SMI) on the role of trust in eHealth for SMI. Feedback on the study concept was sought from members of the general public with experience of SMI, via public engagement events, to understand their perspectives and co-develop research questions, and understand what interview topics would be of interest. Feedback included themes to cover, ways to ask the questions and clarity of study description and instructions. Public collaborators also supported the interpretation of results. These activities informed the final version of our study, as outlined below.

2.2. Participants

Participants were recruited via an online recruitment website and social media. Participants were over 18 years of age. Twenty-two participants took part in a semi-structured interview- eight had a diagnosis of SMI (two had a self-reported diagnosis of psychosis, six had a self-reported diagnosis of depression), and 14 reported no previous history of mental health problems.

57% of the participants identified as women, and in the control group this figure was 73%. The majority of the sample identified as being from a White ethic background, (PSD = 91.4%., controls = 92.3%). Participants had a mean age of 46 years (Range = 23-75), ~71% identified as female, and the majority of the sample identified as being from a White ethic background (76.2%), 14.3% identified as being from a mixed background, 4.7% identified as being from an Asian ethnic background, and 4.7% identified as being from a Black ethnic background.

Participant Number	Clinical Condition	Total Participants
Participants 1-2	Psychosis	2
Participants 3-8	Depression	6
Participants 9-22	None(Control)	14

2.3. Materials

Participants engaged with a 20-minute eHealth cognitive training application ahead of the interview (See supplementary material S1). This was done to provide insights into eHealth system factors, including levels of autonomy and automation. The use of cognitive training was by design, to ensure interactions with the system applied to all participants and was not specific to individual health conditions. The interview questions were categorised into the following themes: trust in internet-based technologies; understanding and trust in eHealth apps; and effects of using the eHealth training app. Full details of the study materials can be found in the Supplementary Material.

2.4. Procedure

The semi-structured interviews, held between March and May 2023, were conducted by two of the authors. After providing informed consent, participants signed up for an interview slot. The option was given for either an in-person or online interview, but all participants chose to have an online interview. Participants were asked to complete the 20-minute training exercise before the interview. The interviews lasted approximately 30-45 minutes in duration, and participants were each thanked with a £10 Amazon voucher.

2.5. Data Analysis

Thematic analysis as described by Braun and Clarke (2006) was applied to make sense of the data. The coding process was considered to be latent because of the researchers' backgrounds in technology ethics. Codes were initially created by EG using a grounded approach and they were later cross-checked between EG and MA to ensure inter-coder reliability. Themes were decided upon once the codes were finalised, and a final codebook was devised (see Appendix 7.3).

2.5.1. **Ethics and Statement of Reflexivity**

The current study is part of a larger project exploring the trustworthiness of autonomous systems for patients to use in healthcare settings. The authors are aware that technology can be both beneficial and harmful for individuals, organisations, and society. As such, we attempted to cover the spectrum of views by giving equal weight to positive, neutral, and negative accounts in the analysis. The authors hold backgrounds in clinical psychology, healthcare, human-computer interaction, and computer science.

3. Results

From our results, four key themes emerged: **(1)** source-related distrust; **(2)** individual differences in (dis)trust; **(3)** the privacy paradox; and **(4)** affective outcomes and (dis)trust. Overall, 50% of participants [7 controls and 4 with depression] classified the need to have trust in internet-based technologies as ‘very important’. Moreover, one participant with psychosis [Participant 2] instead explained how they *"don’t trust anything"*. Therefore, trust (or the lack thereof) in technologies which employed the Internet appeared to heavily affect the lives of the majority of the participants in this study, regardless of their mental health status. Below, we first discuss causes of distrust (source and individual differences), before outlining tensions that may override these factors. Lastly, we outline the privacy paradox that we identified, along with the affective outcomes of (dis)trust. Sixteen participants described the cognitive training application as easy to use, but six were indifferent or found aspects of the system’s usability and usefulness unclear.

3.1. Source-related (Dis)trust

In deliberating their views on trust in internet-based technologies (IBTs) more generally, participants spoke to the importance of context, credibility, and reputation. This translated to a need to look for a website’s encryption [Participant 14, control], for example. Just over one-fifth (22.7%) of participants expressed their sense of distrust in social media sites, pointing to events such as the Cambridge Analytica scandal. Similarly, in the context of cookie settings, it was suggested that

"[companies] look good by asking, but your answer doesn’t really matter." [Participant 7, depression]

If a source was not deemed trustworthy, the perceived risk of negative consequences rose, possibly impacting the (shorter) time spent on a website.

> *"If I go onto the wrong website that I don’t trust, it might...I don’t know...infect my computer or something."[Participant 21, control]*

The overall consensus among participants was that power imbalances exist between large companies and individual end-users. These may raise transparency-related issues and may fuel a sense of distrust.

"It's so obfuscated...with the amount of loopholes they can get away with...it's quite difficult to know that source is definitely trustworthy...so it's quite tricky I think at this time." [Participant 6, depression]

Referring to the cognitive application participants were asked to trial before the interview, over a third (36.4%) of individuals felt their trust for the application related to the reliability of the source [Anonymous]. With regards to the task itself, three control participants (13.6% of the total sample) attributed their trust to the clarity of training instructions. However, one participant with psychosis felt an overall sense of distrust towards the researchers:

"If you'd said there are no tricks, it's all honest, there's no reason why I'd still believe that. I'd still think I'm being tricked." [Participant 2, psychosis].

As such, the source of technology and clarity of expression are necessary for generating or establishing trust in IBTs. In the next section, we outline a related, but distinct, determinant of (dis)trust: individual differences.

3.2. Individual Differences in (Dis)trust

In general, 31.8% of our sample felt as though vigilance was required when engaging with IBTs. There was variability in participants' familiarity with technology, where particularly older adults had proportionately more anxiety and worry surrounding digital tools. As explained by one participant,

"I'm in my 70s. I'm not very good technically, and that makes me apprehensive." [Participant 8, depression].

However, age was not the only determinant of distrust for participants. Scepticism due to a mental health diagnosis and a more careful attitude were closely intertwined in the results we gathered.

To complement that, our participants varied in terms of self-awareness and the extent to which they felt educated about AI, AS, and IBTs. For example, while Participant 15 [control] explained that they accept cookies because they understand how their data may be stored and used, one participant said they tend to reject cookies because they are

"not smart enough to know what it's going to do."[Participant 11 [control]

Relatedly, the pace at which technology is developing presents a challenge for the use of IBTs, and has direct implications for trust:

> *"I try [to keep up with new technologies], but I don't think I do enough. I also suspect it's quite hard to keep up with what can be done. You know, there's new things all the time and you're thinking is this OK or not?" [Participant 9, control]*

On the other hand, the advancement of AI was described in a positive light:

> *"I'm more inclined to trust technology now as it advances, and of course I'm thinking in particular about AI." Participant 8 [depression]*

When asked about which parts of the app they thought were autonomous, many were unsure how to answer. Only three individuals in our sample (13.6%) [Participant 1, psychosis; Participant 7, depression; Participant 12, control] mentioned the autonomous element within the feedback, showing low overall awareness across our sample.

In addition to the highlighted individual differences, social factors were also noted as influencing human-computer relationships. In the wider societal context, a balance between technology-centred and purely social interactions was advised by our participants.

3.2.1. Social Elements: The Right Balance Between Technology and Human Contact

Throughout our interviews, participants spoke of the balance that was desired between AI and human contact.

Related to their using the AI cognitive training application and similar systems in the future, while some participants indicated a desire for continued human contact:

> *"probably would [continue to use similar AI tools] because there's less interaction in terms of, you know, human interactions." Participant 10 [control]*

Other participants experienced a warm interaction through the training, for example:

> *"It was programmed, but it was meaningful rather than coldly mechanical."*
> [Participant 8, depression]

Some participants were more aware of the processes involved in programming AI, with an understanding that human limitations may seep into design decisions. Among them, one participant spoke about the unpredictable nature of human morals in relation to technology development, which can act as a barrier to trust in technology. Thus, we found a distinction between (dis)trust in the technology itself and (dis)trust in programmers. For instance:

> *"Even in normal life you hope that somebody’s going to behave sensibly and honestly and honourably. But, you know, you don’t have that with human beings."* [Participant 20, control]

Participants felt as though self-affirming feedback from the training application increased their trust in it. That is, each time the feedback aligned with their personal evaluation of their performance, they developed more trust in the system. This was not the case for all participants, though, with some expressing cynicism about the honesty and/or accuracy of the feedback:

> *"I was right so often I didn’t necessarily trust that was tailored towards me. So, in that respect, I don’t think I did particularly trust it."* [Participant 17, control]

Even though participants were complaining about how *little* responses from IBTs were tailored to individual needs, they were also aware of the role played by privacy in the accommodation of the contents. In this sense, privacy was found to be in a paradoxical relation with individualised responses and interactions between users and technologies.

3.3. (Dis)trust in Internet-based Technologies: The Privacy Paradox

According to our participants, trust was more important when sharing sensitive or private information online. However, our interviewees also highlighted that attitudes did not always translate to behaviour, with over half of the sample (54.6%) acknowledging a paradox that existed between trust and IBT usage.

According to this paradox, 12 out of the 22 participants declared a lack of trust in the Internet, yet continued to use IBTs and/or abide by cookie settings they did not feel inherently comfortable with. For instance, they commented:

> *"I don’t entirely trust anything on the internet, but I do tend to not- apart from the likes of cookies and that- I tend to go along with an awful lot." [Participant 9, control]*

"I want to say [trust is] really important, but I’d say fairly important. I know sometimes I’m a bit lax with it, so it can’t be that important.” [Participant 1, psychosis]

The excerpts reported above illustrated that both the control and patient groups claimed to distrust the Internet, yet still used it. If given the choice, though, participants mentioned that they preferred to reject non-essential cookies, to minimise the scope of personal data collected about them.

Participants justified such misalignment between their attitudes of distrust and their behaviour with convenience and time-saving benefits. That is, participants acknowledged the ease of use and accessibility advantages that often came with accepting all cookies, although they also expressed concerns regarding the trade-offs between privacy/trust and a user experience with greater functionality.

Participants also expressed feeling obliged to use IBTs, so that they could avoid feeling excluded or prevented from using resources.

"And I don’t trust it but as I said before, I have to in a way trust it in order to get to do what I want to do."[Participant 4, control]

"I suppose an analogy would be if someone came up put a gun to my head and said, “would you like to give me all your money?”, and I said yes, I haven’t really got an option in that moment. But I suppose I have agreed to it, so there you go." [Participant 2, psychosis]

Therefore, the participants alluded to diminished choice and forced consent as potential outcomes of their online behaviours. As such, when questioned about factors that might have increased their trust in the training activity, 3 participants [1 control, 1 with depression, and 1 with psychosis] reported a desire for the privacy of the information they provided.

In this sense, emotions constituted both a cause and a result of the interaction with IBTs. If the emotional desire not to feel excluded nor compromised could be considered as a cause for using IBTs despite their distrust in them, the outcome of such a usage resulted in distress and anxiety responses from our participants.

3.4. **Affective Responses and (Dis)trust**

The emotional aspects of AI, AS and IBTs were frequently mentioned by participants (68.2% of the total sample), with a range of positive, neutral, and negative affective responses identified. First of all, a general sense of anxiety and worry emerged when talking about IBTs and AI.

Specifically, participants discussed their anxieties in terms of AI capabilities, fear of misuse, and the prospect of AI substituting human employees. For example,

"sometimes it is quite worrying how good AI has become." Participant 20 [control]

while AI was described as having “*bad forces”* [Participant 16, control], with the potential for abuse based on the volume of data that could be fed into the system [Participant 2, psychosis].

As a way to minimise such worries, two participants alluded to the need for greater regulation of AI in the future.

"I think there will be backlash obviously. You know, 7 billion people in the world that want to continue to use their brains and complete their work. I hope it will be able to be regulated in a way that it can enhance people’s work." [Participant 18, control]

A degree of annoyance, disappointment, and frustration came with the punitive nature of cookie rejection. Such an option to reject cookies was perceived by one individual in particular as designed to

"wear you down" [Participant 9, control],

posing concerns on the extent to which users actually have choice and control over the collection of their online experience and data.

Similarly, others described feeling suspicious about AI, AS, and IBTs more generally. We viewed that suspicion as a possible double-edged sword. Some described suspicion as positive in relation to being relatively vigilant and cautious on the Internet, for example:

"Maybe we’re more in tune now just to spot dodgy things." [Participant 18, control].

On the other hand, generalised suspicion in these systems could lead to undue anxiety and worry with trustworthy tools. For instance, one participant with psychosis felt as though a degree of deception was involved in the cognitive training application task , having said

"I'm assuming there were a couple of tricks in there." [Participant 2, psychosis].

In this sense, suspicion could be understood as the cause, while caution/sensibility, anxiety, and worry may be understood as potential consequences. Lastly, two individuals [controls] felt that the tone of the cognitive training feedback given by AI caused feelings of infantilisation, which led to distrust in the system. The excerpt below provides examples of these feelings.

"In some way or another it sounded a bit condescending, you know, the answer that came back...if a child sat in front of this I suppose they would have been happy" [Participant 20, control]

In summary, we found a range of affective responses to AI, AS, and IBTs. Though some participants held negative attitudes around trust towards AI, AS, and IBTs, they reportedly experienced enjoyment with the training activity, indicating different responses to individual IBTs and eHealth systems. In the context of affective responses associated with general AI, AS, and IBTs, we reveal the potential for individuals to feel heightened annoyance, anxiety, worry, frustration, disappointment and suspicion with their decreased lack of control over the power held by the designers and implementers of these systems.

Figure 2 depicts an adapted version of the TAM2 [33] based on our results. We found no significant differences in attitudes towards trust across the mental health spectrum. Instead, moderators of intention to use included experience with AI, AS, and IBTs, lack of voluntariness, and attitudes towards trust. We show that the conflict between intention to use and actual usage behaviour (the trust paradox) may lead to unfavourable affective responses.

(Dis)trust was related to the source and/or individual differences. We particularly found social elements, such as preferences for human contact and human-like qualities of technology, influenced the sense of trust in IBTs. Finally, we noted that these determinants of (dis)trust were not necessarily fixed, as they were malleable and context-dependent. That meant that users experienced tensions when engaging with IBTs, including the preference for convenience and time saving, ease of use, power imbalances, and obligations. Ultimately, the causes of (dis)trust

appeared to interact with these tensions, leading to a privacy paradox, and, in turn, enabling positive, neutral, and negative emotions to simultaneously occur.

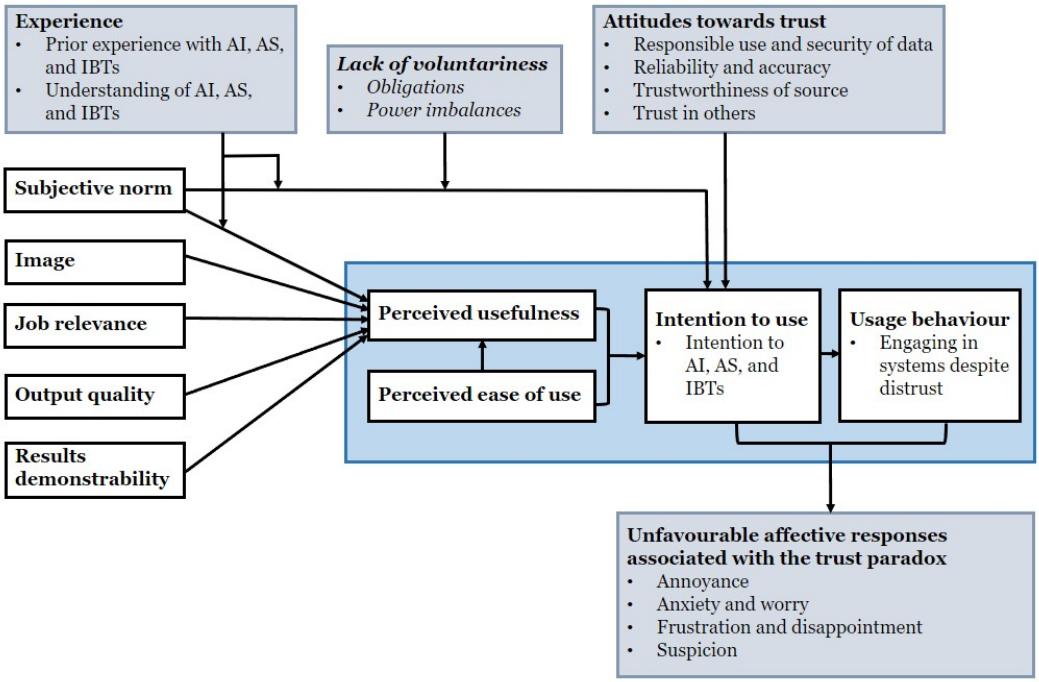

Figure 2: The paradox of trust in online technology across the mental health spectrum: an extended version of TAM2 based on Dowthwaite et al. (2021) [33].

4. Discussion

This study aimed to qualitatively investigate individual perspectives on trust in AI, AS, and IBT tools employed in eHealth applications. This was achieved by interviewing a range of individuals, both with and without experience of serious mental illness (SMI), about their 1) attitudes and trust towards AI, AS, and IBT, and 2) asking them to engage with a simple AI cognitive training programme, which included real-time automated feedback to provide experience with autonomous, individualised feedback on task performance.

Analysis of our interviews revealed the privacy paradox[32] extends to use of online eHealth IBTs, translated to eHealth systems, whereby attitudes of distrust did not prevent participants from engaging with IBTs and AS that used AI. Contrary to expectations, views on trust did not vary between the clinical and non-clinical groups. However, we noted other individual differences that influenced trust, such as age, self-education, self-awareness, and social preferences surrounding human contact and human qualities of technology. In alignment with prior literature [24] [42] the reputability of the recruitment source accounted for their trust in the cognitive training application, but participants noted the difficulty in keeping up with the credibility of sources online. Additionally, we found that all participants may feel suspicious of technology, potentially resulting in more generalised technology-related anxiety and worry. Below these results are discussed in more depth, along with our suggestions and implications for practitioners and technology designers.

4.1. Technology Acceptance, the Paradox of Trust, and Affective Responses

Although trust reportedly serves as a prerequisite for the acceptance, adoption, and *effective* use of technology and eHealth [34] [33] [43], our results suggest that this distrust does not entirely influence actual usage behaviour. User accounts appear to support this by expressing that they feel they have to exercise control over how they access services in today's society, despite having little control over this. In this sense, our findings complement prior literature where eHealth provision is seen as positive, convenient and empowering, by adding user concerns about privacy and data security [12] as another component. Nevertheless, privacy and security concerns do not always prevent people from engaging with these technologies; rather, they may continue to engage with distrust and discomfort, due to obligation.

In line with this finding, when given the option to have an in-person interview, all participants chose to conduct theirs online, supporting the presence of a privacy paradox further. The varied transparency in data usage and privacy agreements participants must agree to before using eHealth and broader IBTs was also highlighted as a factor related to trust, adding to this interpretation. By improving, standardising and regulating communication around system capabilities and data privacy, though, accessibility can be maximised. As such, a combined approach to eHealth, which utilises services delivered by humans and is supported by autonomous systems may be the most acceptable way to utilise eHealth systems and promote trust [44].

From a theoretical perspective, our results draw attention to the reductionist nature of the original TAM [45], particularly in its ability to predict actual usage behaviour, given the increasing scope of factors outside of users' control. Additionally, TAM did not appear to account for fluctuations in trust which is often implied in the interactions humans have with digital technologies[36]. From a practical perspective, while individuals may engage with eHealth systems despite feelings of distrust, the affective discomfort associated with these engagements must not be overlooked. Our participants specifically mentioned the anxiety and worry related to distrust and data privacy, but the lack of alternatives may force users to interact with systems they do not trust. We therefore contribute an understanding of the unfavourable affective responses that may be associated with the trust paradox across the mental health spectrum. These findings necessitate further investigation into the long-term effects of the trust paradox on users' mental health, to ensure the benefits of any intervention outweigh the risks. They also highlight implications for prescribers and developers to reduce this feeling of conflict when interacting with eHealth.

4.2. Distrust in Technology Across the Mental Health Spectrum

We found that views towards autonomous systems did *not* vary across clinical groups. Not only do our results support the general notion that paranoia and suspicion may act as a barrier to clinical engagement [41] but this barrier extends more broadly to the general population. We identified a general sense of paranoia among all groups in our sample. Our results extend on previous work by demonstrating paranoia about IBTs in the general population translates to eHealth [39]. Similarly, individuals may over-perceive the capabilities of an autonomous system, which may lead to distrust, as highlighted by [22]. This has implications for the development and

deployment of eHealth systems across the spectrum of healthcare provision. Interestingly, very few participants in our sample (13.6%) were aware of the autonomous element within the training application. These findings resonate with previous work about the COVID-19 contact tracing app [33] which highlighted that only one-fifth of participants were aware of the app’s autonomous decision-making.

Taken together there are clear implications for the transparency of design decisions. Additionally, it is evident that people want more regulation of these systems', as those not providing a diagnosis operate in a regulatory vacuum. The UK government announced funding for research into the regulation of digital mental health tools [46] which, in due course, may help to mitigate some distrust; however, until a clear regulatory framework is established, people may continue to engage with an understandable degree of apprehension.

4.3. The Balance Between Humans and Autonomous Systems: Establishing Trust over Time

As evidenced by our findings, we are not yet at a point of eHealth replacing healthcare completely. However, in line with prior suggestions, a blended approach with a combination of in-person care, supported or supplemented by eHealth may increase a sense of empowerment, choice, and control for service users[12].

Similarly, issues of inclusivity must be addressed when considering user ability, perception of ease of use, and familiarity with eHealth systems, IBTs, and technology more generally. In broader terms, people who may benefit from eHealth systems may be ‘left behind’ in terms of provision if they are not given human alternatives to system-based support, or appropriate support to understand and learn about available systems[47], [48]. The right balance must therefore be established between human and technology-based healthcare provision for the development of trust over time.

TAM2 places emphasis on voluntariness as a key driver in the acceptance of IBTs, further highlighting the role of the healthcare professional and choice in eHealth provision. However, we found that voluntariness is rarely possible because of how deeply embedded these technologies have become in every aspect of modern society. While education may help to overcome some of this paranoia, our participants identified the speed at which technologies are developing as a barrier to continuous education and awareness. The role of the healthcare professional in this

context is critical, placing responsibility on them to ensure participants are well-placed and sufficiently informed to support engagement with eHealth systems. Brown and Halpern[16] state that *"providers have a responsibility to help recover the person's autonomy through supportive relationships." [p.2].* However, prescribers must themselves have knowledge and confidence in the eHealth systems they are recommending or utilising as part of a therapeutic treatment plan and be able to support service users in the use of these systems. Together, these considerations have implications for both healthcare providers and developers of eHealth technologies.

4.4. Implications for Designing Trustworthy Autonomous Systems in Healthcare

A wide range of implications can be deduced from our study with regards to designing more inclusive and trustworthy eHealth tools. As our results highlighted paranoia and suspicion around eHealth and IBTs did not seem to differ across clinical status, most considerations are generalisable to physical and mental eHealth provision. There are clear ethical implications for developers and prescribers of e-Health interventions, as individuals may use them despite lacking trust. To improve this situation, it is important that new technology and online systems are developed ethically, and responsibly, to be more trustworthy. This involves us all collectively imagining and considering the impact and consequences of new systems, and addressing them, alongside collaboration with end users. Further, prescribers should themselves understand data usage and privacy, and be confident in, and trust in systems they are recommending to patients. This may require practitioners extending knowledge of interventions to include digital implementation and technical skills [49].

4.5. Limitations and Proposed Directions for Future Research

We experienced difficulty in recruiting participants with a broad experience of mental health difficulties. Most participants had experienced depression or anxiety, which are the most common mental health problems in England [38]. Whilst online studies are preferred by individuals with SMI [50] these experiences and diagnoses are nonetheless much less common and thus more difficult to recruit. Our study demonstrates a realistic representation of the mental health spectrum in England.

In this study, the participants were asked to try an AI cognitive training system. We found that most participants were unable to see its immediate usefulness, given the fact that it was being used in a research context. Whilst cognitive training applies to all members of the population,

most eHealth systems would be recommended for a specific purpose. In a real-world context, people may be more likely to engage with eHealth if it was recommended as part of their wider healthcare plan. Whilst our interviews were not solely about participant interactions with an eHealth AI cognitive training application, insight into participants' trust may be limited after just one interaction with the tool, especially considering the fact that trust is conceptualised as a dynamic process to be established over time [36]. Future research might consider multiple interactions with the system to assess how trust in specific AI tools may or may not change or develop over time, and how the development of trust may be experienced differently across the mental health spectrum.

4.6. Conclusion

We investigated the role of trust in attitudes and the use of eHealth and IBTs in a mixed sample of people with and without experience of serious mental illness. Our interviews highlighted that attitudes towards trust in eHealth systems do not differ between people with serious mental health problems and the general public. We principally identified a paradox, whereby participants were mistrustful of eHealth and IBTs, but continued to use them due to positive aspects of convenience and ease of use, but also due to perceived power imbalances with large companies, and obligations.

There are clear ethical implications for developers and prescribers of eHealth interventions, as individuals may use them and share personal, sensitive data, despite lacking trust. To improve this situation, it is important that new technology and online systems are developed ethically, and responsibly, to be more trustworthy. Further, prescribers should themselves understand data usage and privacy, and trust in the systems they are recommending to patients.

References

[1] O. Oyebode, J. Fowles, D. Steeves, and R. Orji, 'Machine Learning Techniques in Adaptive and Personalized Systems for Health and Wellness', *Int. J. Human–Computer Interact.*, vol. 39, no. 9, pp. 1938–1962, May 2023, doi: 10.1080/10447318.2022.2089085.

[2] M. Mirbabaie, S. Stieglitz, and N. R. J. Frick, 'Artificial intelligence in disease diagnostics: A critical review and classification on the current state of research guiding future direction', *Health Technol.*, vol. 11, no. 4, pp. 693–731, Jul. 2021, doi: 10.1007/s12553-021-00555-5.

[3] A. S. Albahri *et al.*, 'A systematic review of trustworthy and explainable artificial intelligence in healthcare: Assessment of quality, bias risk, and data fusion', *Inf. Fusion*, vol. 96, pp. 156–191, Aug. 2023, doi: 10.1016/j.inffus.2023.03.008.

[4] J. A. Naslund, K. A. Aschbrenner, L. A. Marsch, and S. J. Bartels, 'The future of mental health care: peer-to-peer support and social media', *Epidemiol. Psychiatr. Sci.*, vol. 25, no. 2, pp. 113–122, Apr. 2016, doi: 10.1017/s2045796015001067.

[5] L. A. Ellis *et al.*, 'The application of e-mental health in response to COVID-19: scoping review and bibliometric analysis', *JMIR Ment. Health*, vol. 8, no. 12, p. e32948, 2021.

[6] T. J. Philippe *et al.*, 'Digital health interventions for delivery of mental health care: systematic and comprehensive meta-review', *JMIR Ment. Health*, vol. 9, no. 5, p. e35159, 2022.

[7] M. Leiz, N. Pfeuffer, L. Rehner, U. Stentzel, and N. van den Berg, 'Telemedicine as a Tool to Improve Medicine Adherence in Patients with Affective Disorders–A Systematic Literature Review', *Patient Prefer. Adherence*, pp. 3441–3463, 2022.

[8] J. D'Arcey *et al.*, 'The use of text messaging to improve clinical engagement for individuals with psychosis: systematic review', *JMIR Ment. Health*, vol. 7, no. 4, p. e16993, 2020.

[9] C. B. Bennett, C. J. Ruggero, A. C. Sever, and L. Yanouri, 'eHealth to redress psychotherapy access barriers both new and old: A review of reviews and meta-analyses.', *J. Psychother. Integr.*, vol. 30, no. 2, p. 188, 2020.

[10] E. Fernandez, D. Salem, J. K. Swift, and N. Ramtahal, 'Meta-analysis of dropout from cognitive behavioral therapy: Magnitude, timing, and moderators.', *J. Consult. Clin. Psychol.*, vol. 83, no. 6, p. 1108, 2015.

[11] E. Bourke, C. Barker, and M. Fornells-Ambrojo, 'Systematic review and meta-analysis of therapeutic alliance, engagement, and outcome in psychological therapies for psychosis', *Psychol. Psychother. Theory Res. Pract.*, vol. 94, no. 3, pp. 822–853, 2021.

[12] S. Bucci *et al.*, 'Early psychosis service user views on digital technology: qualitative analysis', *JMIR Ment. Health*, vol. 5, no. 4, p. e10091, 2018.

[13] C. Giordano, I. Ambrosiano, M. T. Graffeo, A. Di Caro, and S. Gullo, 'The transition to online psychotherapy during the pandemic: a qualitative study on patients' perspectives', *Res. Psychother. Psychopathol. Process Outcome*, vol. 25, no. 3, 2022.

[14] J. Pywell, S. Vijaykumar, A. Dodd, and L. Coventry, 'Barriers to older adults' uptake of mobile-based mental health interventions', *Digit. Health*, vol. 6, p. 2055207620905422, 2020, doi: 10.1177/2055207620905422.

[15] M. El Benny, T. Kabakian-Khasholian, F. El-Jardali, and M. Bardus, 'Application of the eHealth literacy model in digital health interventions: scoping review', *J. Med. Internet Res.*, vol. 23, no. 6, p. e23473, 2021.

[16] J. E. Brown and J. Halpern, 'AI chatbots cannot replace human interactions in the pursuit of more inclusive mental healthcare', *SSM-Ment. Health*, vol. 1, p. 100017, 2021.

[17] M. Laifa, S. Akrouf, and R. Maamri, 'Online Social Trust: an Overview', in *Proceedings of the International Conference on Intelligent Information Processing, Security and Advanced Communication*, 2015, pp. 1–6.

[18] D. M. Rousseau, S. B. Sitkin, R. S. Burt, and C. Camerer, 'Not so different after all: A cross-discipline view of trust', *Acad. Manage. Rev.*, vol. 23, no. 3, pp. 393–404, 1998.

[19] S. C. Currall and T. A. Judge, 'Measuring trust between organizational boundary role persons', *Organ. Behav. Hum. Decis. Process.*, vol. 64, no. 2, pp. 151–170, 1995.

[20] V. Shankar, G. L. Urban, and F. Sultan, 'Online trust: a stakeholder perspective, concepts, implications, and future directions', *J. Strateg. Inf. Syst.*, vol. 11, no. 3–4, pp. 325–344, 2002.

[21] O. Freiman, 'Making sense of the conceptual nonsense "trustworthy AI"', *AI Ethics*, vol. 3, no. 4, pp. 1351–1360, 2023.

[22] B. F. Malle, K. Fischer, J. Young, A. Moon, and E. Collins, 'Trust and the discrepancy between expectations and actual capabilities', *Hum.-Robot Interact. Control Anal. Des.*, pp. 1–23, 2020.
[23] H. Gill, K. Boies, J. E. Finegan, and J. McNally, 'Antecedents of trust: Establishing a boundary condition for the relation between propensity to trust and intention to trust', *J. Bus. Psychol.*, vol. 19, pp. 287–302, 2005.
[24] Y. Kim, 'Trust in health information websites: A systematic literature review on the antecedents of trust', *Health Informatics J.*, vol. 22, no. 2, pp. 355–369, 2016.
[25] S. van de Vijver *et al.*, 'Evaluation of a Digital Self-management Platform for Patients With Chronic Illness in Primary Care: Qualitative Study of Stakeholders' Perspectives', *JMIR Form. Res.*, vol. 6, no. 8, p. e38424, 2022.
[26] S. A. Tighe, K. Ball, L. Kayser, F. Kensing, and R. Maddison, 'Qualitative study of the views of people living with cardiovascular disease, and healthcare professionals, towards the use of a digital platform to support cardiovascular disease self-management', *BMJ Open*, vol. 12, no. 11, p. e056768, 2022.
[27] I. K. Ledel Solem *et al.*, 'Patients' needs and requirements for eHealth pain management interventions: qualitative study', *J. Med. Internet Res.*, vol. 21, no. 4, p. e13205, 2019.
[28] S. Simblett *et al.*, 'Barriers to and facilitators of engagement with remote measurement technology for managing health: systematic review and content analysis of findings', *J. Med. Internet Res.*, vol. 20, no. 7, p. e10480, 2018.
[29] F. Flammini, C. Alcaraz, E. Bellini, S. Marrone, J. Lopez, and A. Bondavalli, 'Towards trustworthy autonomous systems: Taxonomies and future perspectives', *IEEE Trans. Emerg. Top. Comput.*, 2022.
[30] H. Vasconcelos, M. Jörke, M. Grunde-McLaughlin, T. Gerstenberg, M. S. Bernstein, and R. Krishna, 'Explanations can reduce overreliance on ai systems during decision-making', *Proc. ACM Hum.-Comput. Interact.*, vol. 7, no. CSCW1, pp. 1–38, 2023.
[31] K. Vodrahalli, R. Daneshjou, T. Gerstenberg, and J. Zou, 'Do humans trust advice more if it comes from ai? an analysis of human-ai interactions', in *Proceedings of the 2022 AAAI/ACM Conference on AI, Ethics, and Society*, 2022, pp. 763–777.
[32] S. Barth and M. D. De Jong, 'The privacy paradox–Investigating discrepancies between expressed privacy concerns and actual online behavior–A systematic literature review', *Telemat. Inform.*, vol. 34, no. 7, pp. 1038–1058, 2017.
[33] L. Dowthwaite *et al.*, 'Public adoption of and trust in the NHS COVID-19 contact tracing app in the United Kingdom: quantitative online survey study', *J. Med. Internet Res.*, vol. 23, no. 9, p. e29085, 2021.
[34] V. Venkatesh, J. Y. Thong, and X. Xu, 'Unified theory of acceptance and use of technology: A synthesis and the road ahead', *J. Assoc. Inf. Syst.*, vol. 17, no. 5, pp. 328–376, 2016.
[35] M. Ryan, 'In AI we trust: ethics, artificial intelligence, and reliability', *Sci. Eng. Ethics*, vol. 26, no. 5, pp. 2749–2767, 2020.
[36] M. Naiseh, D. Al-Thani, N. Jiang, and R. Ali, 'How the different explanation classes impact trust calibration: The case of clinical decision support systems', *Int. J. Hum.-Comput. Stud.*, vol. 169, p. 102941, 2023.
[37] M. Naiseh, D. Cemiloglu, D. Al Thani, N. Jiang, and R. Ali, 'Explainable recommendations and calibrated trust: two systematic user errors', *Computer*, vol. 54, no. 10, pp. 28–37, 2021.
[38] A. L. Baker and E. Kirk-Wade, 'Mental health statistics: prevalence, services and funding in England', Number CBP-06988, 2023. [Online]. Available: https://researchbriefings.files.parliament.uk/documents/SN06988/SN06988.pdf
[39] O. J. Mason, C. Stevenson, and F. Freedman, 'Ever-present threats from information technology: the Cyber-Paranoia and Fear Scale', *Front. Psychol.*, vol. 5, p. 1298, 2014.

[40] V. Lerner, I. Libov, and E. Witztum, '" Internet delusions": The impact of technological developments on the content of psychiatric symptoms', *Isr. J. Psychiatry Relat. Sci.*, vol. 43, no. 1, p. 47, 2006.
[41] P. A. Maguire and J. C. Looi, 'COVID-19 telehealth challenges for patients with schizophrenia and other psychoses', *Aust. N. Z. J. Psychiatry*, vol. 55, no. 9, pp. 923–923, 2021.
[42] T. W. Simpson, 'e-Trust and reputation', *Ethics Inf. Technol.*, vol. 13, pp. 29–38, 2011.
[43] S. Terrizzi, S. Sherer, C. Meyerhoefer, M. Scheinberg, and D. Levick, 'Extending the technology acceptance model in healthcare: Identifying the role of trust and shared information', 2012.
[44] F. Greenway, M. Weal, A. M. Hughes, and E. Palmer-Cooper, 'Clinical engagement associated with hybrid e-interventions: A systematic review', *mHealth*, In Prep.
[45] V. Venkatesh and F. D. Davis, 'A theoretical extension of the technology acceptance model: Four longitudinal field studies', *Manag. Sci.*, vol. 46, no. 2, pp. 186–204, 2000.
[46] Medicines and H. products R. Agency, 'Mental health funding of 1.8m welcomes by MHRA and Nice to explore regulation of Digital Mental Health Tools', *GOV.UK*, Oct. 2022, [Online]. Available: https://www.gov.uk/government/news/mental-health-funding-of-18m-welcomed-by-mhra-and-nice-to-explore-regulation-of-digital-mental-health-tools
[47] H. Baup and H. Verdoux, 'Frequency and pattern of Internet use in patients with schizophrenia or bipolar disorders seeking medical information', *Psychiatry Res.*, vol. 247, pp. 152–154, 2017.
[48] D. Robotham, S. Satkunanathan, L. Doughty, and T. Wykes, 'Do we still have a digital divide in mental health? A five-year survey follow-up', *J. Med. Internet Res.*, vol. 18, no. 11, p. e309, 2016.
[49] K. A. Smith *et al.*, 'Digital mental health: challenges and next steps', *BMJ Ment Health*, vol. 26, no. 1, Feb. 2023, doi: 10.1136/bmjment-2023-300670.
[50] A. Polillo *et al.*, 'Using Digital Tools to Engage Patients With Psychosis and Their Families in Research: Survey Recruitment and Completion in an Early Psychosis Intervention Program', *Jmir Ment. Health*, vol. 8, no. 5, p. 11, May 2021, doi: 10.2196/24567.

Supplementary Material

S1 Training Activity with Real-time Autonomous Feedback

The training involved a working memory exercise, whereby in the first part of the activity, participants were asked to look at and remember as many images as possible. The second part of the activity involved memory recall, whereby they were asked to press the letter 'D' if they believed the image appeared on the left, and the letter 'J' if they thought the image appeared on the right. After each selection, they were asked to rate their self-awareness in terms of accuracy participants after they made their decision, indicating if they had good self-awareness or poor self-awareness on a scale of 1 (*relatively low confidence*) to 6 (*relatively high confidence*). Autonomous feedback was presented to participants after each trial.

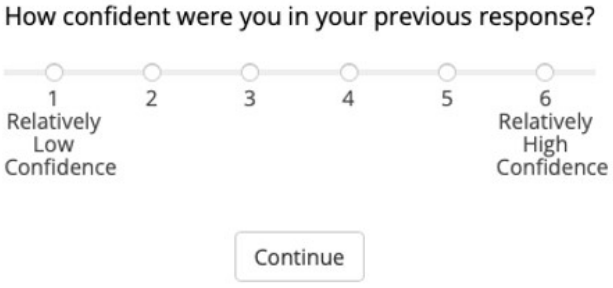

Figure 3: Self-awareness rating scale shown to participants during the training exercise

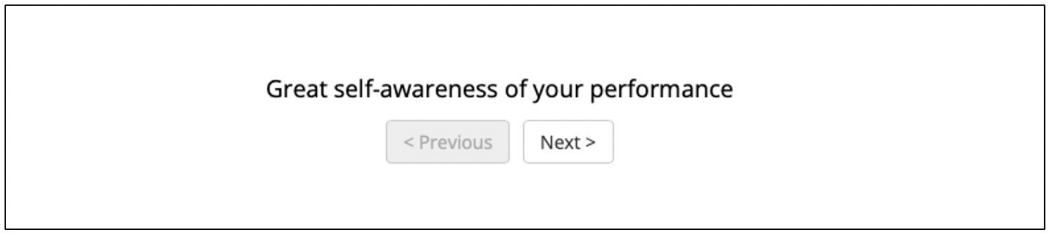

Figure 4: Autonomous feedback when participants had great self-awareness

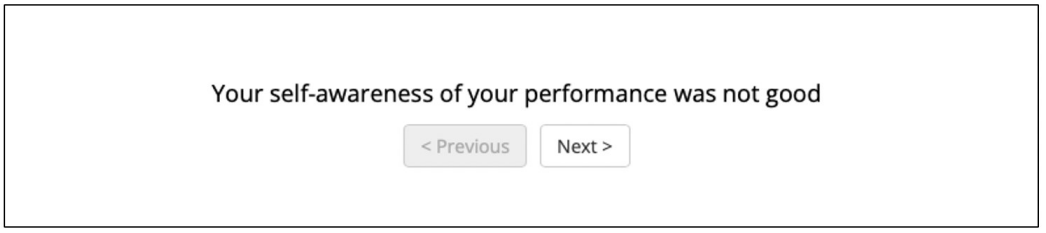

Figure 5: Autonomous feedback when participants had poor self-awareness

S2 Semi-Structured Interview Prompts

Trust in Internet-based Technologies

1. What do you understand to be meant by AI and autonomous systems, and how do you think they are used in internet-based technologies?
2. When asked to consent to data collection online, for example with cookies, how do you decide and why?
3. How important is trust for you when you use internet-based technologies?

Understanding of Trust and Trust in the App

1. Which parts of the application app did you like or not and why?
2. How easy was it for you to use the app?
3. How useful did you find the app?
4. Overall, did you trust the app? Why?
5. How do you think the application worked?
6. Which parts of the application do you think were autonomous/AI?
7. Did you trust the autonomous feedback and why?
8. Were the answers you received as expected?
9. Did this affect your trust in the app?

Effects of the App

1. Did using this training have any impact on your feelings towards internet-based technologies? What about whether they are trustworthy?

2. Did use this training have any other impact on you (e.g., on how you use these types of technologies?)

3. Would you be open to continuing using this or similar AI tools? Why or why not?

4. Would you recommend this tool or similar tools to your friends and family?

5. Is there anything that would have increased your trust in the app?

Thematic Analysis Codebook

Themes	**Codes**
(Dis)trust in internet-based technologies: the privacy paradox	attitudes and aversion; misalignment of attitudes and behaviours; privacy-related issues; obligations
Individual differences in (dis)trust	age-related issues; self-education and self-awareness; accuracy; social elements
Source-related (dis)trust	trust and reliance; transparency-related issues; reputation and credibility; trust dependent on company or platform; power imbalances; clarity
Affective outcomes and (dis)trust	enjoyment; frustration and disappointment; infantilisation; uncertainty and conviction; annoyance; anxiety and worry; suspicions; deception; comfort; indifference and neutrality; avoidance behaviour