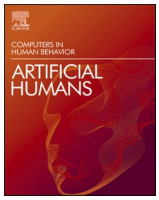
Computers in Human Behavior: Artificial Humans 5 (2025) 100180

Contents lists available at ScienceDirect

Computers in Human Behavior: Artificial Humans

journal homepage: www.journals.elsevier.com/computers-in-human-behavior-artificial-humans



Self-disclosure to AI: People provide personal information to AI and humans equivalently

Elizabeth R. Merwin [a,b,*], Allen C. Hagen [a], Joseph R. Keebler [b], Chad Forbes [a,c]

[a] Department of Psychology, Florida Atlantic University, 777 Glades Road, Boca Raton, FL, 33431, USA
[b] Department of Human Factors and Behavioral Neurobiology, Embry-Riddle Aeronautical University, 1 Aerospace Boulevard, Daytona Beach, FL, 32114-3900, USA
[c] Edson College of Nursing and Health Innovation, Arizona State University, 550 N 3rd Street, Phoenix, AZ, 85004, USA

A B S T R A C T

As Artificial Intelligence (AI) increasingly emerges as a tool in therapeutic settings, understanding individuals' willingness to disclose personal information to AI versus humans is critical. This study examined how participants chose between self-disclosure-based and fact-based statements when responses were thought to be analyzed by an AI, a human researcher, or kept private. Participants completed forced-choice trials where they selected a self-disclosure-based or fact-based statement for one of the three agent conditions. Results showed that participants were statistically more likely to select self-disclosure over fact-based statements. Choice for self-disclosure rates were similar for the AI and human researcher, but significantly lower when responses were kept private. Multiple regression analyses revealed that individuals with a higher score on the negative attitude toward AI scale were less likely to choose Self-based statements across the three agent conditions. Overall, individuals were just as likely to choose to self-disclose to an AI as to a human researcher, and more likely to choose either agent over keeping self-disclosure information private. In addition, personality traits and attitudes toward AI were able to significantly influence disclosure choices. These findings provide insights into how individual differences impact the willingness to self-disclose information in human-AI interactions and offer a foundation for exploring the feasibility of AI as a clinical and social tool. Future research should expand on these results to further understand self-disclosure behaviors and evaluate AI's role in therapeutic settings.

1. Background

With the rapid advancement of Artificial Intelligence (AI) there is a need to understand the circumstances and motivations underlying effective communication between humans and advanced machines. When people communicate in Human-Human interactions, they often talk about themselves and their personal preferences, otherwise known as self-disclosing (Altman & Taylor, 1973). Self-disclosure to others has been found to be intrinsically rewarding, and a large portion of verbal communication is dedicated to the disclosure of personal information (Tamir & Mitchell, 2012). With the recent uptick in AI technology use, understanding the parameters and reasons for when and how people choose to self-disclose to AI is of the utmost importance.

The current study seeks to explore the self-disclosure behaviors of individuals to an AI agent and to expand the knowledge of the motivational factors for human-agent interactions. Here, an agent is defined as an entity (automated or human) that can take actions on behalf of itself or others (Oertel et al., 2020). With an increased understanding of human behavior, AI developers can more effectively develop AI to respond to individuals in various situations, such as those seeking out AI for social companionship, as a mental health resource, or for sensitive medical care advice. Uncovering the factors influencing self-disclosure should allow for more effective and streamlined services. Improving the current systems to respond in a way that focuses on building and maintaining trust in the system while considering factors that influence self-disclosure should lead to an increase in utilization and effective communication (Schaefer et al., 2016).

If individuals are more comfortable self-disclosing to an AI, then it can be more easily implemented into fields that require a certain amount of trust in the technology and providers, which can potentially lead to more individuals receiving care. Since self-disclosure to chatbots, text or voice based AI agents who engage in naturalistic simulated conversations (Adamopoulou & Moussiades, 2020), has been found to evoke similar emotional and psychological satisfaction compared to human-human interactions (Ho et al., 2018), it is important to observe the relationship that individuals have with these technologies that lead to increased utilization or willingness to interact (Schaefer et al., 2016). The main focus of this research is to expand our knowledge into what factors in Human-AI interaction facilitate greater rates of self-disclosure, so that systems dependent on personal information (i.e., systems that need to know specific details about an individual's life circumstances to offer advice) are able to effectively evaluate situations and provide

* Corresponding author. 1 Aerospace Boulevard, Daytona Beach, FL, 32114-3900, USA.
E-mail address: merwine@my.erau.edu (E.R. Merwin).

https://doi.org/10.1016/j.chbah.2025.100180
Received 28 April 2025; Received in revised form 30 June 2025; Accepted 5 July 2025
Available online 9 July 2025

individually tailored resources as needed. Doing so should result in streamlined systems in which people with sensitive concerns can access care, helping to alleviate the current strain placed on workers in healthcare domains.

1.1. Uptick in use of AI technology

The use of advanced technologies such as computers, robots, and Artificial Intelligence (AI) has been steadily increasing over the past decade (Müller, 2024; Singh et al., 2021). From self-checkout machines in retail stores to smartwatches and self-driving cars, people are quickly learning how to utilize AI technology and adjust to an increasingly automated world. Some individuals report hesitancy to engage with AI or negative attitudes towards AI and are reluctant to purposefully interact with it while others find it beneficial and actively seek it out (Schepman & Rodway, 2020). Those who are inclined to interact with AI appreciate its ability to increase their task efficiency, save them time during repetitive tasks, and simply are excited to use the technology (Almahairah, 2023; Gkinko & Elbanna, 2021; Horodyski, 2023; Kelly et al., 2023; Sheng & Xiao, 2022). A specific form of AI that is quickly gaining popularity is the use of AI chatbots. The first chatbot, ELIZA, was created in the mid-1900s and was designed to mimic human conversation and even act as a mock psychotherapist (Weizenbaum, 1966). Even though the design of ELIZA was basic compared to the AI chatbots in use today, users of ELIZA reported that they thought the system had human-like characteristics, such as intelligence, and would engage in personal conversations with the chatbot (Sharma et al., 2017).

Since the development of ELIZA, numerous advanced conversational-based AI chatbots have been developed, including Cleverbot, Gemini, and OpenAI's ChatGPT, all of which are capable of longer and more intricate conversations. The development of ChatGPT specifically has resulted in an overall uptick in awareness of generative conversational AI and an increase in use. Despite the general American public reporting high rates of concern related to the expansion of AI, the purposeful use of AI is still on the rise (Faverio and Tyson, 2023). Since people are generally concerned about the development of AI but are still using it more, it seems that there may be undiscovered mechanisms influencing its use.

One potential mechanism could be personality traits, which are well-documented influencers of self-disclosure in Human-Human interactions (Chen et al., 2016; Hollenbaugh & Ferris, 2014; Ignatius & Kokkonen, 2007). By examining established mechanisms within the context of Human-AI interactions, a comprehensive understanding of the increase in AI usage can be made possible. A goal of the present study is to explore potential mechanisms that drive individuals to share personal information about themselves, and explore how these mechanisms have been examined in both Human-Human and Human-AI interactions.

1.2. Self-disclosure in human-human interactions

Self-disclosure is often characterized by the sharing of personal information from one entity to another (Kim & Dindia, 2011). Disclosing information about one's personal characteristics, life experiences, thoughts, and feelings allows for the sharing of ideas, beliefs, and knowledge, which can foster bonds between individuals (Altman & Taylor, 1973; Sprecher et al., 2013). With 30–40 % of verbal communication being dedicated to disclosing personal information to others, self-disclosure acts as one of the foundational building blocks of effective communication (Dunbar et al., 1997). Self-disclosure is also closely related to trust in a communicative partner, with more trust resulting in increased levels of self-disclosure (Altman & Taylor, 1973). Trust is defined by Lee and See (2004) as "the attitude that an agent will help achieve an individual's goals in a situation characterized by uncertainty and vulnerability"(p. 51), meaning that trusting another agent, human or otherwise, is a potentially risky endeavor. In interactions between two or more humans, an increase in trust between the individuals leads to more self-disclosure, and when the individuals lack trust, they become more selective about the content and amount of personal information they disclose (Wheeless & Grotz, 1977). If individuals are more trusting during an interaction, then they are more likely to be comfortable discussing personal information, which serves to strengthen their interpersonal relationships (Altman & Taylor, 1973).

This foundation of trust and comfort is essential to self-disclosure in human-human interactions. More trust leads to high rates of self-disclosure between close others, but there is also an innate desire to reveal personal information regardless of the person the individuals are self-disclosing to (Tamir & Mitchell, 2012). A series of studies by Tamir and Mitchell (2012) investigated the motivations for high rates of self-disclosure. In their experiment, they had participants choose between various self-based or fact-based prompts, and their level of agreement with the prompt was measured. These involved various self-disclosure prompts (e.g., "I like the color red"), prompts about others (e.g., "Obama likes the color red"), and prompts about commonly known facts (e.g., "the sky is blue"). Financial values, from $.01–.04, were randomly assigned to each of the prompts and participants were free to try and maximize their gains during the experiment (i.e., theoretically, since participants actually received their final monetary total at the conclusion of the experiment, they should have chosen the prompt associated with the highest monetary values to maximize gains by the end of the experiment). However, instead of always choosing the highest-paying selection, participants were significantly more likely to choose self-disclosure prompts, even when it meant foregoing the highest financial gain. Participants were inclined to forgo prompts worth larger amounts of money in favor of the opportunity to self-disclose information about themselves to the researcher. In a follow-up study by this team, the results were replicated; participants were told that their responses to the prompts would either remain private or be shared with a close other (friend/relative) who attended the session with the participant.

Participants demonstrated the same preference for choosing self-based over fact-based prompts, and the result was significantly stronger when they thought their responses were going to be shared with their close other. When the participants were able to disclose to their close other, they were willing to forgo a higher amount of money to be able to answer self-based prompts. Altogether, this finding suggests individuals find self-disclosure intrinsically rewarding and have a preference for disclosing personal information over factual information in Human-Human interactions. The internal reward of revealing personal information is prevalent in various human-human interactions, namely to strangers and to close others, but the preference to self-disclose to non-human entities such as AI is not as well understood. This poses an important question that current research has yet to address: do individuals exhibit a similar preference to choose to self-disclose to non-human agents? To explore this issue, the next section examines how individuals reveal self-related information in Human-AI interactions, with a particular focus on conversational AI agents.

1.3. Self-disclosure in Human-AI interactions

With the increase in AI chatbot usage, there have been recent attempts to explore Human-AI interactions and how they differ from Human-Human interactions. Understanding the nuances of interactions between the agents helps to uncover the reasons for if, and how, preferences for self-disclosure vary based on which agent the individual self-discloses to. High rates of self-disclosure are not limited to in-person communication. For instance, a study by Naaman et al. (2010) found that over 80 % of individuals engage on social media sites via posting and sharing personal information with their followers. Previous literature has explored how individuals apply similar social rules to their interactions with computers (Nass et al., 1994) and chatbots as they do with other humans (Bickmore & Picard, 2005; Reeves & Nass, 1996). According to the Computers as Social Actors (CASA) framework,

individuals will often apply their preconceived Human-Human social scripts (e.g., saying "please" and "thank you") to their interactions with computers, especially during early interactions with the technology (Gambino et al., 2020; Nass et al., 1994).

Individuals often attribute human-like interaction standards to their conversations with chatbots even though they are aware that computer-based software is inanimate (Gambino et al., 2020; Nass et al., 1996). Self-disclosure to chatbots has also been shown to result in similar emotional and psychological satisfaction compared to Human-Human conversations (Folk et al., 2024; Ho et al., 2018; Rapp et al., 2023). Furthermore, individuals who frequently used social chatbots reported greater social health outcomes than nonusers (Guingrich & Graziano, 2023). It seems that since individuals are attributing human social scripts to chatbots, resulting in similar emotional and psychological conversational effects, there is a theoretical basis towards utilizing AI chatbots to facilitate sensitive conversations that require personal information.

Individuals tend to talk about themselves, ask questions about, and converse online with chatbots in ways similar to human-human conversations(Croes et al., 2022, 2024; Folk et al., 2024). When communicating online with chatbots, individuals found their interactions more rewarding when communication was more relational (consisting of higher levels of social dialogue and empathy), which elicited increased respect, likability, and trust from participants (Bickmore & Picard, 2005; Demeure et al., 2011). Individuals were likely to self-disclose to the chatbots they could relate more strongly to, and the increased trust resulted in higher rates of self-disclosure. The relationship between human trust towards artificially intelligent machines (e.g., AI chatbots, robotic entities, automated tools, etc.) has been extensively explored and suggests that individuals who are more trusting of systems are more likely to be willing to interact with the system in the future (Joinson et al., 2010; Madhavan & Wiegmann, 2007; Schaefer et al., 2016). Individuals who are more trusting of their conversational partner are more likely to self-disclose, and research suggests that individuals may be more at ease self-disclosing to a chatbot than another human due to the implied confidentiality and lack of judgement (Altman & Taylor, 1973; Skjuve & Brandtzæg, 2018). With the creation of advanced AI chatbots that can produce higher levels of social dialogue and engage in more empathetic conversations, interactions are being perceived as even more satisfying and natural, which has prompted an uptick in chatbot usage accordingly (Zhou et al., 2020).

While the technology is still developing, studies have been conducted to evaluate rates of self-disclosure towards AI-enabled chatbots. When individuals passively converse with an AI chatbot, they are likely to self-disclose personal information, and they do so more when the AI chatbot reciprocated with its own self-disclosure or showed support towards the participant upon receiving personal information (Croes & Antheunis, 2020; Lee, Yamashita, & Huang, 2020; Skjuve et al., 2022). Individuals voluntarily self-disclose to AI chatbots about their emotions, thoughts, insecurities, and vulnerabilities and willingly participate in lengthy discussions (Lee et al., 2022; Lee et al., 2024; Skjuve et al., 2022). In turn, individuals feel more comfortable conversing with an AI chatbot that also reveals information about itself, mimicking how human-human conversations take place (Croes & Antheunis, 2020; Lee et al., 2024). Since individuals have been shown to self-disclose to AI, it suggests a certain extent of closeness, likability, and trust in the entity they are communicating with (Altman & Taylor, 1973)

Another line of testing concerned the use of AI chatbots to facilitate therapy and evaluated if individuals would A) share personal information with an AI chatbot and B) be willing to allow a human mental health professional to review the conversations (Lee, Yamashita, & Huang, 2020). The results of this experiment indicated that participants were willing to self-disclose to the AI chatbot, but when it came time to share that information with the mental health professional, some participants chose not to share their conversations. This suggests that individuals have different preferences for what they choose to disclose to humans and AI, but does not directly compare self-disclosure decisions between the two. While the literature concludes that individuals will self-disclose to humans, chatbots, and AI, it is difficult to understand the preference of their self-disclosure choices without direct comparisons between Human-Human and Human-AI interactions.

One study by Mou and Xu (2017) provided participants with modified transcripts of conversations they collected from messengers on WeChat, a social media site in which users can instant message with friends and with an AI chatbot named "Little Ice." The two types of transcripts, a Human-Human conversation, and a Human-AI-chatbot conversation, were presented to participants, and they were blind as to which conversations were which. Participants were asked to rate the messengers' (the human texting either agent) perceived level of self-disclosure through the Opener Scale (Miller et al., 1983). This study provided perceived levels of the messengers' self-disclosure behavior towards a human and an AI chatbot and concluded that participants were rating the messengers' Human-Human conversations as higher in self-disclosure than the Human-AI-chatbot conversations. While this was a strong conclusion based on their experimental setup, the results relied on perceived scores of another person's self-disclosure behavior and did not directly measure if individuals were choosing to disclose at a higher rate to either communicative agent. To capture any significant differences in an individual's preference to self-disclose to a human versus an AI, direct manipulation and comparison of actual self-disclosure choice should be conducted.

The existing literature has important implications for self-disclosure trends with AI chatbots but fails to offer a concise understanding of individual preferences for self-disclosure towards other humans and conversational AI. There is a lack of research investigating if individuals choose to self-disclose in the same manner to AI and other humans, which should be addressed to effectively utilize AI as a conversational agent. There also seem to be factors influencing the decision to self-disclose to AI compared to other humans that have yet to be explored. Through the investigation of self-disclosure behaviors and the underlying factors influencing self-disclosure, a deeper understanding of how individuals choose to interact with AI chatbots can be more thoroughly understood.

2. The current study

2.1. Purpose

Past research suggests that individuals are willing to disclose personal information to AI, however, there is a lack of literature directly comparing preference to self-disclosure to an AI chatbot or a human, and few studies that look to uncover the underlying factors that influence rates of self-disclosure towards AI. Drawing inspiration from the work on self-disclosure preferences originally conducted by Tamir and Mitchell (2012), a forced-choice paradigm was utilized to directly evaluate participant preference to self-disclose information to a human or an AI. The study investigated how individuals decided to engage with self-disclosure-based and fact-based survey items when under the impression that their answers were going to be analyzed by different agents.

Our study presented different agents in a similar forced choice paradigm to examine preference to self-disclose. For one agent scenario, participants were under the impression that an AI analyzed their given responses (AI), in another, they thought a human researcher analyzed their responses (Researcher), or they were told their given answers were not sent to an analyzer (Private). To avoid introducing bias into participant responses, explicit definitions were not given for each of the agent types prior to the choice task (Malhi et al., 2020). This approach ensures that participants relied on their own perceptions and prior experiences with AI, rather than being influenced by our predefined descriptions. By allowing participants to interpret the agents naturally, the study hoped to capture more authentic preferences to self-disclose.

2.2. Hypotheses

Individuals have an innate preference for disclosing personal information since they find it intrinsically rewarding (Tamir & Mitchell, 2012). This preference for choosing Self statements over Fact statements is expected to emerge regardless of the agent scenario participants select. Thus, we propose H1: *Participants will select more Self-based statements than Fact-based statements.*

Since individuals prefer to interact with perceived social agents (including humans and AI), participants are expected to favor statements they believe are analyzed by social agents over private statements. Since disclosing to social agents is inherently rewarding, participants should forgo choosing selections that are linked to the Private condition and select Researcher and AI conditions more often (Ho et al., 2018; Tamir & Mitchell, 2012). This leads us to H2: *Individuals will prefer to select the Researcher and the AI as potential analyzers over keeping responses private.*

Due to preferences for self-disclosure (Tamir & Mitchell, 2012) and interacting with social agents (Nass et al., 1994), participants should be most inclined to choose Self-based statements when they are presented by the Researcher or the AI rather than opting for the Private condition. H3: *Selection of Self-based statements will occur more frequently when they are perceived to be analyzed by the researcher or the AI as compared to being kept private.*

While H1-H3 focused on how statement and agent types influence self-disclosure preferences, various individual differences often shape decisions to disclose, which is the focus of H4-H6.

Rates of self-disclosure have been found to be influenced by personality traits in both Human-Human interactions (Chen et al., 2016; Hollenbaugh & Ferris, 2014; Ignatius & Kokkonen, 2007) and in Human-Machine interactions (Lei & Liu, 2024; Zabel et al., 2025; Zhou et al., 2019). For instance, for human-human interactions, higher rates of extraversion and openness are related to higher rates of self-disclosure, while those high in conscientiousness, agreeableness, and neuroticism were less likely to self-disclose (Caci et al., 2019; Loiacono, 2014; Seidman, 2013). However, differences exist between studies that examine personality traits and their effect on interactions between humans and machines. Literature in Human-AI (Lei & Liu, 2024; Zhou et al., 2019) and Human-Robot contexts (Zabel et al., 2025) have demonstrated that, while openness has similarly been found to predict greater willingness to disclose to AI agents as it has with humans, extraversion has been associated with reduced disclosure and willingness to confide in human-AI interactions (Zhou et al., 2019). These differences highlight the importance of comparing the influence of personality traits across interaction types within the same experimental design. Because of these inconsistencies and the findings of the aforementioned studies, we propose the following hypotheses: H4: *Differences in the big-five personality traits should influence rates of choosing Self-statements for each of the agents. For the human researcher, higher extraversion and openness should predict more of a preference to select Self statements, whereas higher conscientiousness, neuroticism, and agreeableness should predict less. For the AI, higher openness should predict more self-disclosure, while high extraversion and neuroticism should predict less. For the private choice, greater conscientiousness, agreeableness, and neuroticism should predict a greater likelihood of choosing the Private condition and keeping their information from either of the agents.*

In addition to personality traits, attitude towards AI is being examined as an individual difference that may affect self-disclosure tendencies. While not directly related to disclosure research, recent work identifies that those with a more positive attitude towards AI are more willing to engage with AI (Choung et al., 2022), and those with more negative attitudes are less likely to engage with AI (Choung et al., 2022; Wu et al., 2024). Based on this literature, participants with a more positive attitude towards AI should be more willing to interact with AI whereas those that view AI in a negative way may shy away from disclosing personal information towards it and choose to engage with the researcher or private options at a greater frequency (Wu et al., 2024). Based on these studies, we propose H5: *Differences in individuals' attitudes towards AI should predict their preference of choice for Self-statements towards the different agent conditions. Specifically, individuals with higher positive attitudes towards AI will be more likely to choose AI self-disclosure prompts.*

Personality traits and attitudes toward AI are strongly grounded in established psychological theory and have been shown to meaningfully influence interaction behaviors. The extent of AI experience, while less theoretically anchored, may still play an important role in shaping disclosure tendencies. Participants who perceive themselves as having more experience with AI in their personal lives may feel more familiar and comfortable with the technology, potentially leading to greater willingness to self-disclose to an AI agent (Horodyski, 2023; Wu et al., 2024). Prior research suggests that increased familiarity and comfort with AI systems can result in greater trust in and less apprehension towards the technology, making self-disclosing with the AI more likely than those with less experience (Choung et al., 2022; Wu et al., 2024). Therefore, we propose H6: *Differences in the individuals' extent of experience with AI should predict their self-disclosure behaviors towards the agents. Specifically, those with higher self-reported experience with AI should be more likely than those low in self-reported experience to choose AI self-disclosure prompts.*

3. Method and procedure

3.1. Participants

Prior to data collection, this study was reviewed and approved by the Institutional Review Board of Florida Atlantic University (FAU). Based on the planned analyses, two a priori sensitivity analyses were conducted using the G*Power software (Faul et al., 2007) to determine the minimum number of participants needed to detect a significant effect. For the repeated measures ANOVA results indicated that the required sample size to achieve 95 % power for detecting a medium effect, at a significance criterion of $\alpha = .05$, was 28 participants. The work that inspired our team's experimental design had a sample size of 37 participants for their within-subjects, repeated measures design, which provided a guideline sample size goal for our main analysis (Tamir & Mitchell, 2012). For the linear regressions, results indicated that the required sample size to achieve 95 % power for detecting a large effect, at a significance criterion of $\alpha = .05$, was 74 participants. The study aimed to detect large, theoretically motivated effects, based on prior literature, to allow for a focused examination of robust patterns across the agent conditions. This study design aims to lay the groundwork for future research with a larger, more diverse sample capable of capturing subtler interactions.

There were 96 participants recruited for this online study through FAU's SONA research participation program, and they received class credit as compensation for participating in the experiment. Seven participant responses were excluded due to failed attention checks, which were prompts embedded randomly throughout the experiment to ensure participants were paying attention to the questions and not randomly responding (e.g., "please select *somewhat disagree* for this survey item"). This left a total of 89 participants for the final analysis, exceeding what was suggested by the sensitivity analyses and relevant previous studies. The sample consisted of 59 females with the average age falling between 18 and 24 years (see Table 1 for additional demographics breakdown).

3.2. Procedure

Online data collection was completed through Qualtrics software. Participants were instructed to ensure they had stable access to an internet connection, a compatible computer or mobile device, and a quiet environment before registering for the study. Once registered,

Table 1
Sociodemographic characteristics of participants.

Sample characteristic	*n*	%
Gender		
Female	59	66.3
Male	30	33.7
Age		
18-24	88	98.9
25-34	1	1.1
Hispanic Origin		
Yes	30	33.7
No	59	66.3
Race		
White or Caucasian	55	61.8
Black or African American	19	21.3
Mixed race	7	7.9
Other	7	7.9
Prefer not to say	1	1.1
First Generation		
Yes	55	61.8
No	30	33.7
Prefer not to say	4	4.5

Note. Table detailing gender, age, Hispanic origin, race, and first-generation status of the sample.

participants read through a consent form and agreed to participate in the study. Participants were first presented with statements regarding their demographics, including, but not limited to, their age, gender, race, ethnicity, household income, and education level. They then proceeded to the self-disclosure portion of the questionnaire. A 3 (agent scenario) x 2 (statement type) within-subjects design was implemented in which all participants chose between all of the agent scenarios and each of the statements across the duration of the experiment.

3.2.1. Self-Disclosure Experimental Questionnaire setup

To determine an individual's preference to self-disclose for the different agent scenarios, we developed a series of forced-choice survey statements. Participants were instructed that their responses to the statements were to be evaluated and interpreted in three different ways: 1) the response would be evaluated by a researcher, 2) the response would be evaluated by an AI, 3) the response would not be evaluated and would be kept private. We deliberately kept information about the researcher and AI undefined to prevent bias in participant responses. In addition, participants were to choose between two statement types: 1) statements about themselves (Self statements), 2) statements about facts (Fact statements). Each survey item presented 2 options for the participant to select from, each involving a combination of statement type and agent type (e.g., "Researcher - Fact" or "AI - Self"). For each potential selection, neither the statement type nor the agent type would be presented more than once in the same survey item. For example, the participant would never need to choose between a Fact and Self statement that were both presented by an AI, nor would two Fact statements be presented simultaneously, or two Self statements. The agent scenario and the type of statement were both randomized. The experimental design was modeled after Tamir and Mitchell's (2012) self-disclosure forced-choice paradigm, which has been widely cited and inspired subsequent self-disclosure studies. Our within-subjects design presented the participants with 44 randomized pairwise forced-choice trials, comparing different combinations of statement types (Self vs. Fact) and agents (AI vs. human researcher vs. private condition). This design allowed each participant to experience all possible combinations of conditions (e.g., AI-Fact, AI-Self, Researcher-Fact, Researcher-Self, Private-Fact, and Private-Self), thereby reducing between-subject variability and increasing sensitivity to detect within-participant differences in disclosure preferences. The use of a within-subjects forced-choice design allowed each participant to make comparisons across multiple agent and statement types, unlike a between-subjects design that would have required random assignment to just one of six static disclosure conditions.

Before starting the experimental trials, participants were presented with example statements for both statement types to get familiar with the presentation style (see Table 2 for examples of Self and Fact statements). Both banks of statements were pilot-tested prior to this study to ensure the statements that were presented were equally interesting to answer. In previous studies similar to this design, it was argued that self-statements could be generally more interesting than fact-statements and that could be the underlying reason for a self-disclosure preference instead of a fact preference (Tamir & Mitchell, 2012). Comparing the interest levels of the statements presented in this study controlled for this possibility. Each statement was presented to participants in the pilot study, and they were asked to rank their level of interest in the statements as follows: "Not interesting at all," "Slightly interesting," "Moderately interesting," "Very interesting," or "Extremely interesting." Interest scores for the questions were z-scored for Self statements and Fact statements. 65 Self statements and 65 Fact statements were initially developed (see supplemental materials for banks of questions and their interest Z-scores). Because pilot study data indicated a significant difference in the level of interest between Fact and Self statements, items were removed from both lists until the difference in the level of interest was nonsignificant ($p > .05$). A total of 44 trials were then developed with 1 Self statement and 1 Fact statement corresponding to each of them.

In the Self-Disclosure Experimental Questionnaire, participants were presented with a total of 44 counterbalanced prompt pairs, each of which consisted of a prompt that asked them to choose between two of the three agent conditions, randomly asking either of the statement types (see Fig. 1 for example prompt pair). This was followed by the second prompt displaying the statement they chose. For example, the first prompt would ask the participant to select the survey item they would like to answer and display a combination of agents and statements such as: "Researcher - Self" and "AI - Fact". If the participant chose "Researcher - Self" they were then presented with a Self-statement such as: "I feel restless when idle" and if they selected "AI - Fact" then they were presented with a Fact statement such as: "the speed of light is faster than the speed of sound." For each of the statement types, participants were instructed to select a response on a Likert scale (responses were Strongly Disagree – Strongly Agree) based on the extent they agreed with the statement. While participants were under the impression that their answers to the different statements were to be analyzed by these different agents, this study was truly focused on the participants' choice patterns for Self statements and Fact statements, and if their choice to answer personal information was affected by the agent they thought was receiving the information. The proportion of selections for each question type and agent type out of all trials was recorded for the analysis.

Data for the Self-Disclosure Experimental Questionnaire was organized in Microsoft Excel. Participants' total selection for Self and Fact questions, as well as their total selections for AI, Researcher, and Private agent scenarios were all calculated individually, meaning that if participants selected an option with any of these characteristics, it would be added to the total for that corresponding characteristic (e.g. If a participant selected "AI - Fact" during a trial, it would add to the total selection count for both the AI agent and Fact statements). The proportions for selected agent scenarios (AI, Researcher, Private) for all

Table 2
Examples of self and fact prompts used in the survey.

Question Types	Example Questions
Self Statements	"I feel restless when idle"
	"I prefer reading physical books to e-books"
	"I'm drawn to minimalist lifestyles"
Fact Statements	"The human body has three kidneys"
	"The speed of light is faster than the speed of sound"
	"The Great Wall of China is visible from space"

Note. Examples of Self and Fact statements that were used in the experiment.

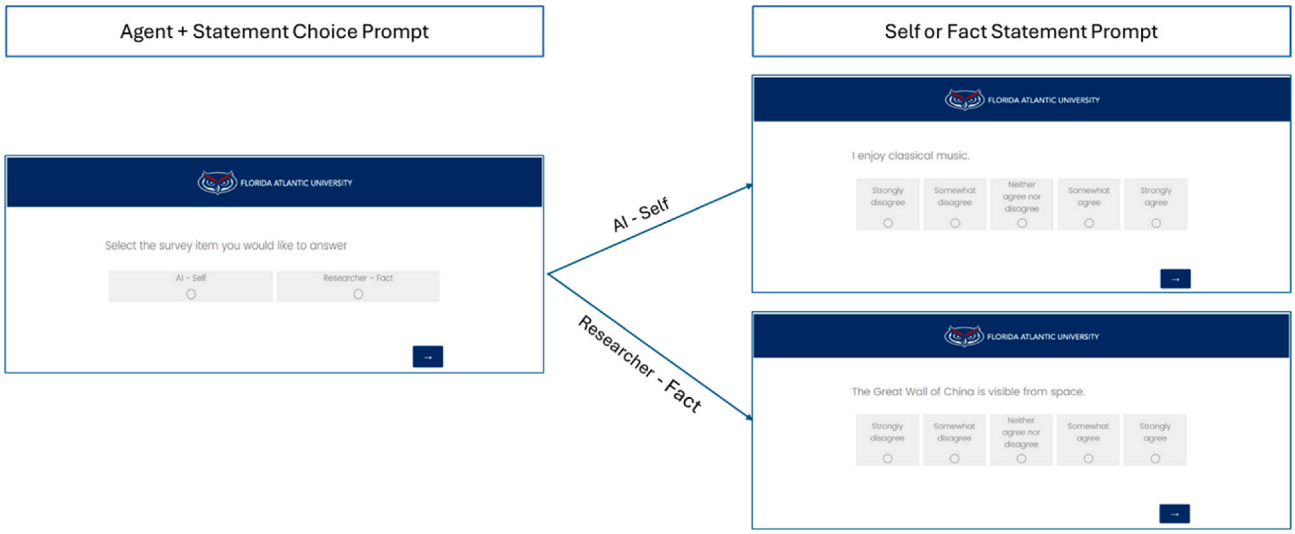


Fig. 1. Prompt Examples for Self-Disclosure Experimental Questionnaire.
Note. Example pair of choices on the Self-Disclosure Experimental Questionnaire. For the Agent + Statement Choice Prompt, participants are presented with the choice between two agents (here, an AI or a Researcher) presenting different statement types (Self or Fact). Participants choose which they prefer to answer and are then presented with the Self or Fact Statement Prompt that matches their choice.

trials were calculated, as well as the proportion of Self and Fact questions. Additionally, totals and proportions were also calculated for each question pairing between analyzer and question type (e.g., AI – Fact, Researcher – Self, etc.).

3.2.2. *Post-experiment surveys*

Several surveys were given after the experiment to explore how individual differences could be associated with self-disclosure preferences.

3.2.3. *Mini-international personality item pool (MINI-IPIP)*

Participants first completed the MINI-IPIP personality assessment questionnaire, a short 20-question survey that captures the extent an individual has the traits of: extraversion, conscientiousness, openness (to experiences), agreeableness, and neuroticism (Gosling et al., 2003). Extraversion is seen in those who typically have high social skills, many friends, and are outgoing in nature. Those high in conscientiousness typically strive for achievement and have high leadership skills. Individuals who have higher openness to experience are typically drawn to novelty and change. Those who defer to others and comply readily with others' wishes would score high on agreeableness. Neuroticism is a ranking of an individual's mental stability; individuals who score lowly on this scale typically have higher self-esteem and more optimistic beliefs (Costa & McCrae, 1992). The statements were presented to the participants, and they were instructed to rate how accurately the statement represented themselves from the following options: "Very Inaccurate," "Moderately Inaccurate," "Neither Accurate nor Inaccurate, " "Moderately Accurate," or "Very Accurate." Responses to the MINI-IPIP survey prompts were scored according to participant agreement for most questions (1 = "Very Inaccurate," 5 = "Very Accurate") except for those that needed to be reverse-coded (1 = "Very Accurate," 5 = "Very Inaccurate"). After the scoring was completed, composite scores were calculated for each participant for each subscale of extraversion (Cronbach's alpha = .75), consciousness (Cronbach's alpha = .68), openness (Cronbach's alpha = .51), agreeableness (Cronbach's alpha = .69), and neuroticism (Cronbach's alpha = .66).

3.2.4. *AI experience*

To evaluate the extent of the participants' experience with AI, participants were asked, "Do you have experience using AI tools such as ChatGPT?" and ranked their perceived experience on a Likert scale (0 = I have no experience using AI tools, 10 = I use AI tools almost daily).

3.2.5. *General attitudes towards Artificial Intelligence scale (GAAIS)*

To evaluate participant attitudes toward AI, the GAAIS was presented (Schepman & Rodway, 2020). This validated scale was used to identify positive and negative attitudes towards AI. The positive subscale captured how individuals felt AI could be used as a personal and societal tool, whereas the negative subscale captured concerns related to AI. Participants responded to the randomized list of positive and negative prompts and rated their level of agreement using one of the following options: "Strongly Disagree," "Disagree," "Neutral," "Agree," or "Strongly Agree." Examples of prompts from the positive subscale included "Artificial Intelligence is exciting," and "Artificially intelligent systems can perform better than humans." Examples of prompts from the negative subscale included "Artificial Intelligence is used to spy on people," and "I think Artificial Intelligence is dangerous." Scores on the positive attitude subscale (Cronbach's alpha = .82) were given numerical values according to their agreement (1 = "Strongly Disagree," 5 = "Strongly Agree"). Scores on the negative attitude subscale (Cronbach's alpha = .81) were reverse-coded (1 = "Strongly agree," 5 = "Strongly disagree") so that a higher score on each subscale represented a more positive attitude towards AI. Composite scores for each subscale were calculated through SPSS for each participant.

A high composite score on the positive attitude subscale would suggest that participants felt Artificial Intelligence could be a useful personal and/or societal tool, whereas a low score would indicate they felt AI could not be useful. A high composite score on the negative attitude subscale would suggest that the participant was concerned about the use of Artificial Intelligence, whereas a low composite score would indicate the participant was not concerned with the use of AI.

3.3. *Data preparation*

Data was exported from Qualtrics into Microsoft Excel, sorted and organized, and then exported into SPSS. Data that was missing was accounted for through mean imputation: calculating the mean for that variable and substituting the missing values for the mean (Donders et al., 2006). Any values that were flagged as extreme outliers (greater than or less than 2 standard deviations from the mean) were substituted through winsorizing: assigning the identified outlier value the next valid highest (if the outlier was extremely high) or lowest (if the outlier was extremely low) value in the dataset (Kennedy et al., 1992).

4. Results

4.1. Demographic analyses

A series of one-way ANOVAs explored possible gender differences in attitudes towards AI, the extent of experience using AI, and participant responses to the self-disclosure questionnaire (see Table 3). The analyses revealed a significant difference in negative attitude towards AI, $F(1, 87) = 7.96$, $p < .05$, $\eta^2p = .08$, with a large effect size. Females had significantly lower scores ($M = 2.75$, $SD = .53$) than males ($M = 3.11$, $SD = .65$), suggesting that females in this sample had a less forgiving attitude towards the negative aspects of AI than males did. There were no other gender differences in the study (all $p > .05$).

4.2. Self-disclosure questionnaire analysis

A 2 (statement type) by 3 (agent scenario) repeated measures ANOVA was conducted on the proportion of Self and Fact statements chosen for each agent scenario. The descriptive statistics for proportion scores are shown in Table 4. The repeated measures ANOVA determined there was a significant main effect of statement type $F(1, 88) = 36.31$, $p < .001$, $\eta^2p = .29$, with a large effect size. Post hoc tests with a Bonferroni adjustment revealed that participants were choosing to answer Self statements ($M = 20.69$, $SE = .67$) significantly more than Fact statements ($M = 12.64$, $SE = .67$, $p < .001$). There was also a moderately sized main effect of agent scenario $F(2, 176) = 6.72$, $p < .05$, $\eta^2p = .07$ in which, participants chose to answer statements that were to be kept Private ($M = 15.42$, $SE = .34$) significantly less than statements to be analyzed by either the Researcher ($M = 17.34$, $SE = .35$, $p < .05$) or the AI ($M = 17.24$, $SE = .33$, $p < .05$). No difference was found when individuals were asked to choose to answer statements between the Researcher and the AI ($p = 1.0$). This effect has been shown in previous human-human interactions as participants showing a preference for choosing to share information with another human over keeping their information private, but the finding that there is a preference to share information with an AI to the same degree has important implications for Human-AI interactions.

In addition to the main effects, there was also a significant statement type X agent scenario interaction $F(2, 176) = 4.86$, $p = .009$, $\eta^2p = .05$, with the partial eta squared indicating a small-to-medium effect (see Fig. 2). Simple effects tests revealed a significant difference in how participants chose Self statements between the agent scenarios, but no significant difference in their choice of Fact statements. Participants were significantly less likely to choose Self statements when the responses were perceived to be Private ($M = 18.87$, $SE = .68$) than when they were going to the Researcher ($M = 21.58$, $SE = .85$, $p < .001$) or to the AI ($M = 21.63$, $SE = .79$, $p < .001$). There was no significant difference between the Researcher and AI agent scenarios ($p = 1.0$), indicating that there was no significant difference in preference to self-disclose to the social agents. This shows that the results were consistent with H1-H3: participants preferred to choose Self statements over Fact statements and were inclined to disclose personal information to the human researcher and the AI over keeping their responses private.

Table 4
Proportion selection for statement types by agent scenarios.

Agent Scenario	Statement Type	
	Self Mean (SD)	Fact Mean (SD)
Researcher	21.58 (7.98)	13.1 (6.99)
AI	21.63 (7.45)	12.84 (7.64)
Private	18.87 (6.45)	11.98 (7.12)

Note. Mean and standard deviations (SD) for the agent scenarios for each statement type.

This demonstrates that participants are willing to share personal information with an AI to the same extent they would with another human.

4.3. Predictors of self-disclosure choice for different agents

Findings from the repeated measures ANOVA revealed that people prefer to self-disclose information over keeping it private and that it did not matter whether their self-disclosure was to another human or to an AI. The next question is whether there were any individual differences in participants' patterns of disclosure. Based on previous self-disclosure research in human-human interactions, personality factors might play an important role in determining when individuals prefer one agent over another. Additionally, the amount of experience an individual has with AI, their attitude towards AI, and their gender might moderate the likelihood of whether individuals self-disclose to the various agents. To explore these patterns, separate multiple regressions were conducted for the AI, the researcher, and the private agent scenarios in an attempt to identify which factors could be influencing self-disclosure patterns to each agent.

4.3.1. AI self-disclosure

A simultaneous multiple regression analysis was conducted to identify predictors of the proportion of time AI-Self statements were chosen by participants, based on personality traits (individual scores for extraversion, agreeableness, conscientiousness, neuroticism, and openness to new experience), gender, extent of experience with AI, and attitude towards AI (both positive and negative). The model was significant $F(9, 79) = 2.46$, $p < .05$, Adjusted $R^2 = .13$ with agreeableness ($\beta = -.25$, $t(79) = -2.04$, $p < .05$) and negative attitude towards AI ($\beta = -.37$, $t(79) = -3.29$, $p < .001$) accounting for roughly 13 % of the variance in the proportion of time AI-Self statements were chosen. No other predictors were significant (positive attitude towards AI $\beta = -.002$, $t(79) = -.02$, $p > .05$; extent of experience with AI $\beta = -.17$, $t(79) = -1.52$, $p > .05$; extraversion $\beta = .1$, $t(79) = .95$, $p > .05$; conscientiousness $\beta = -.13$, $t(79) = -1.22$, $p > .05$; neuroticism $\beta = -.15$, $t(79) = -1.35$, $p > .05$; and openness $\beta = .19$, $t(79) = 1.67$, $p > .05$). This test revealed that agreeableness and extent of negative attitude towards AI were the strongest predictors of self-disclosure to AI. Specifically, participants who were more agreeable and who had higher scores on the negative attitude towards AI subscale were less likely to self-disclose to the AI.

Table 3
One-way ANOVAs to evaluate sex differences in variables.

Variable	F	df (between, within)	p value	η^2	Mean (SD) for females	Mean (SD) for males
Positive Attitude Towards AI	.52	1, 87	.47	.01	2.99 (.53)	3.07 (.5)
Negative Attitude Towards AI	7.96*	1, 87	.01	.08	2.75 (.53)	3.11 (.65)
Extent of Experience	.51	1, 87	.48	.01	3.85 (2.61)	3.46 (1.92)
Choice of AI-Self	.26	1, 87	.61	.00	21.92 (7.57)	21.06 (7.31)
Choice of AI-Fact	1.21	1, 87	.28	.01	11.94 (7.02)	14.09 (7.42)
Choice of Researcher-Self	2.77	1, 87	.1	.03	22.57 (7.65)	19.62 (8.39)
Choice of Researcher-Fact	.77	1, 87	.38	.01	12.83 (7.49)	14.24 (7.28)
Choice of Private-Self	.91	1, 87	.34	.01	19.14 (7.09)	17.73 (6.52)
Choice of Private-Fact	1.47	1, 87	.23	.02	11.17 (6.89)	13.26 (6.78)

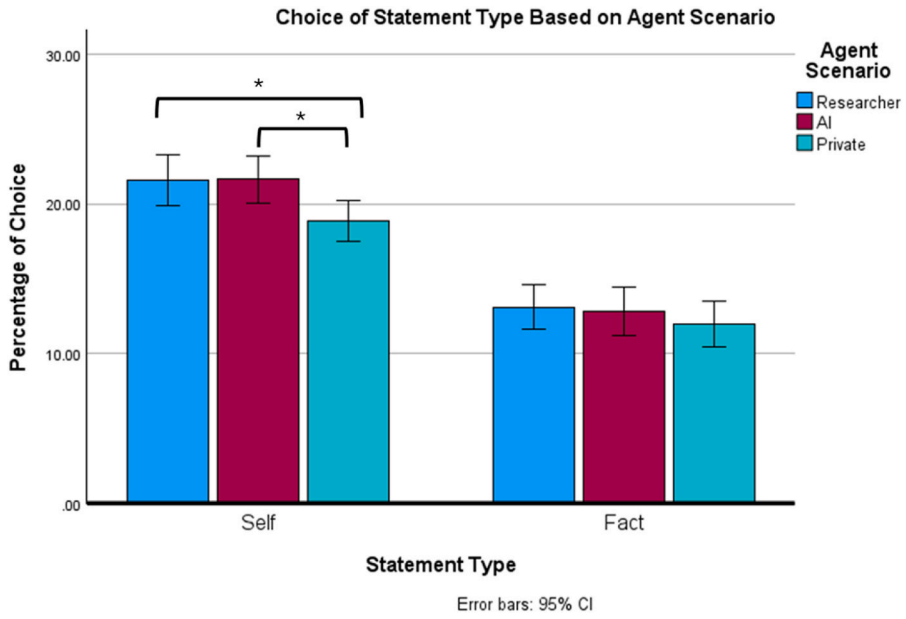


Fig. 2. *Note.* Bar chart representing the proportion of times each statement type was chosen for each agent scenario. There was a significant difference in how participants chose Self statements based on the agent scenario, but no differences in their choice of Fact statements. Participants were more likely to choose Self statements when the agent scenario was either the Researcher or the AI in comparison to keeping responses to Self statements Private.

4.3.2. Researcher self-disclosure

Another simultaneous multiple regression analysis was conducted to identify predictors of the proportion of time Researcher-Self statements were chosen during the experiment based on the same variables. This model was significant $F(9, 79) = 2.03$, $p < .05$, Adjusted $R^2 = .10$, indicating that these factors collectively predicted approximately 10 % of the variance in self-disclosure behaviors to the human researcher. Negative attitude towards AI was the only significant factor in the model ($\beta = -.31$, $t(79) = -2.72$, $p < .05$). There were no significant personality traits: extraversion ($\beta = -.03$, $t(79) = -.29$, $p > .05$), agreeableness ($\beta = -.05$, $t(79) = -.44$, $p > .05$), conscientiousness ($\beta = -.2$, $t(79) = -1.87$, $p > .05$), neuroticism ($\beta = -.1$, $t(79) = -.91$, $p > .05$), and openness to experience ($\beta = .1$, $t(79) = .85$, $p > .05$) were not significant. Experience with AI ($\beta = -.13$, $t(79) = -1.16$, $p > .05$) and a positive attitude towards AI ($\beta = .12$, $t(79) = 1.02$, $p > .05$) also failed to be significant predictors. This test suggests that participants with higher scores on the negative attitude towards AI subscale were less likely to disclose personal information to the researcher.

4.3.3. Private self-disclosure

Finally, a simultaneous multiple regression analysis was conducted to identify predictors of the proportion of time Private-Self statements were chosen during the experiment based on the personality traits (individual scores for extraversion, agreeableness, conscientiousness, neuroticism, and openness to new experience), gender, extent of experience with AI, and attitude towards AI (both positive and negative). The model was significant $F(9, 79) = 2.19$, $p < .05$, Adjusted $R^2 = .11$, indicating that 11 % of the variance for the proportion of time that Private-Self statements were chosen was accounted for based on negative attitude towards AI ($\beta = -.28$, $t(79) = -2.5$, $p < .05$), extent of experience with AI ($\beta = -.23$, $t(79) = -2.02$, $p < .05$), and conscientiousness ($\beta = -.21$, $t(79) = -2.02$, $p < .05$). There were no other significant personality traits: extraversion ($\beta = -.1$, $t(79) = -.95$, $p > .05$), agreeableness ($\beta = -.08$, $t(79) = -.62$, $p > .05$), neuroticism ($\beta = -.12$, $t(79) = -1.06$, $p > .05$), and openness to experience ($\beta = .11$, $t(79) = .99$, $p > .05$). Positive attitude towards AI ($\beta = .1$, $t(79) = .82$, $p > .05$) also failed to be a significant predictor. The test revealed that scores on the negative attitudes towards AI subscale, self-reported extent of experience with AI, and the personality traits of conscientiousness were significant predictors of self-disclosure in the private agent scenario. Specifically, participants who had higher scores on the negative attitude towards AI subscale, had more experience with AI, were more conscientious, and were less likely to choose Self statements when presented for the private condition.

5. Discussion

5.1. Preference for self-disclosure

This study explored participant preferences for sharing personal information with three different agent conditions: a human researcher, an AI, or kept private. The within-subjects forced-choice design enabled a more direct comparison across the agent types while minimizing individual differences. The design ensured that the observed preferences reflect genuine decision-making patterns rather than between-group confounds. Analysis of choice patterns revealed that participants preferred to self-disclose information about themselves, replicating previous research that suggested individuals find self-disclosure intrinsically rewarding (Tamir & Mitchell, 2012; Vijayakumar et al., 2020).

This replication is particularly noteworthy since our team took efforts to match the perceived interest levels across the statement types. One critique of prior self-disclosure research is that they often fail to control for baseline interest between the options. For example, if a participant is asked to what extent they agree with one of the following prompts: “Spiders are insects” or “I prefer simplicity over complexity,” their preference may be driven more by the relative engagement of the statement than by a genuine motivation to self-disclose. If Self-based statements were overall more interesting than the Fact-based statements, then that could have led to a bias towards Self-statements since they were more relevant or engaging, instead of an actual finding that individuals find self-disclosure to be intrinsically rewarding. By controlling for interest levels, the present study strengthens the interpretation that participants’ disclosure behavior reflects intrinsic motivation to share personal information.

While our effect sizes were small, there was no difference in participant self-disclosure towards the AI and the Researcher which falls in line with previous research (Lee, Yamashita, & Huang, 2020; Uchida et al., 2017). The findings of this study suggest that our primary

hypotheses, H1-H3, were correct; individuals were just as likely to self-disclose information to the human and to the AI, and preferred self-disclosing to both more than keeping their self-disclosure private. This could indicate that individuals are perceiving the AI in a way similar to other humans when it comes to self-disclosure. Past research has shown individuals will self-disclose information during human-AI interactions, but our research suggests they are willing to do so to the same extent as human-human interactions (Croes & Antheunis, 2020; Lee, Yamashita, & Huang, 2020; Skjuve et al., 2022).

Overall, this analysis contributes greatly to the self-disclosure and AI literature. It extends prior work by directly comparing disclosure preference between human and AI agents with a private condition acting as a control and uses behavioral choice data rather than relying solely on self-report measures. Additionally, since there was no difference in choice between self-disclosure behavior to an AI and to another human, the findings provide behavioral support for the Computers are Social Actors (CASA) framework, suggesting that individuals tend to treat machines as social entities and have similar expectations from machines as they do humans (Nass et al., 1994).

This has important implications for the future of AI, especially regarding the merit of AI as a clinical tool. Since individuals are willing to disclose personal information to AI, implementing AI tools such as chatbots to facilitate communication in therapeutic settings could be a viable option. This finding comes at a critical time when mental health professionals are facing concerningly high rates of burnout and emotional exhaustion (O'Connor et al., 2018; Yang et al., 2024). If AI is an acceptable social agent that individuals willingly reveal personal information to, then this avenue warrants further exploration in real-world therapeutic contexts. There have been efforts to explore whether AI can be utilized effectively in therapeutic settings with overall optimistic conclusions (Fiske et al., 2019; Lee, Yamashita, & Huang, 2020; Uchida et al., 2017). With the findings from the current study adding support to previous literature, there is now a strong implication that individuals are willing to use AI, will discuss sensitive information with it, and that they receive emotional and psychological satisfaction from disclosing to it (Ho et al., 2018). If AI chatbots were to be implemented to aid mental health professionals, their services could be streamlined and workload could be reduced, allowing professionals to more efficiently help patients and hopefully increase the number of individuals receiving services.

5.2. Personality traits as predictors of self-disclosure preferences

Since there was no significant difference in self-disclosure rates to the human researcher or the AI, the influence of various individual differences on decisions to self-disclose to the agents was explored. Based on the regression analyses, H4-A (relating to personality traits influencing self-disclosure trends) was only partly supported. For the AI condition, it was predicted that individuals higher in openness to new experiences would be more likely to choose to self-disclose to the AI and that those high in extraversion and neuroticism would be less likely to self-disclose to the AI (Zhou et al., 2019). Our findings indicate that individuals who were higher in agreeableness were less likely to choose to share personal information with the AI. People who are high in agreeableness often tend to avoid stimuli they find distressing or uncomfortable (Bresin & Robinson, 2014). Given that agreeable participants selected Self-based statements analyzed by the AI at a lower rate, it is possible that they were doing this to avoid a scenario they found adverse (i.e., disclosing personal information to an AI).

Zhou and colleagues' (2019) study relied on inferring their participants' personality traits based on conversations with an AI and did not explicitly use a personality measurement, which could be why our study has divergent findings. Based on our explicit measurement of personality traits with the MINI-IPIP, those higher in agreeableness should be less likely to choose to self-disclose to an AI in a forced-choice paradigm. Previous studies have also shown that those high in agreeableness tend to have a more positive attitude toward AI (Stein et al., 2024) and a more forgiving attitude towards the negative aspects of AI (Kaya et al., 2022; Schepman & Rodway, 2022). This could explain why those in our study were more inclined to choose AI to reveal their personal information to.

There were no significant predictors for self-disclosure behaviors towards the human researcher. This suggests that participants' self-disclosure decisions were not strongly influenced by personality traits, whereas past research has suggested a strong relationship between personality traits and self-disclosure behavior (Caci et al., 2019; Loiacono, 2014; Seidman, 2013).

For the private selection, we hypothesized that higher conscientiousness, agreeableness, and neuroticism would predict that participants were more likely to want to keep their personal information private (Caci et al., 2019; Loiacono, 2014; Seidman, 2013). We found that conscientiousness was predictive of choosing the private condition for Self-based statements, but agreeableness and neuroticism were not. The differences in results here could be because both of these hypotheses were based on human-human conversational interactions instead of the forced-choice AI-inclusive paradigm of this experiment. Our experiment focused on participant behavior in choosing the agents and statement types, not the extent of self-disclosure, which is most commonly researched. It could be that the influence of personality traits on self-disclosure would change based on the type of interaction being had (i.e., natural conversations, experimental paradigms, wizard-of-oz studies, etc.).

5.3. Attitude towards AI as predictors of self-disclosure preferences

Hypothesis 5 predicted that differences in individual attitudes towards AI would influence participant self-disclosure choices. For the positive attitude towards AI scale, there was no supporting evidence that scores were predictive of self-disclosure choice behavior. This suggests that self-disclosure preferences are not influenced by how individuals perceive AI as a personal or societal tool. For the negative attitude towards AI scale, we theorized that having a higher score would predict more of a preference to choose to self-disclose; however, participants in this sample who had a higher negative attitude towards AI score were less likely to choose Self-statements. This suggests that those who are less concerned about the dangers of AI in society are less likely to choose to self-disclose information to AI. While these findings were contradictory to what was predicted, this finding could have been influenced by the order in which the surveys were administered (the Self Disclosure survey was taken followed by the GAAIS), leading to bias in how participants were answering. Since participants were asked to agree or disagree with statements such as "I find Artificial Intelligence is sinister," "Artificial Intelligence is used to spy on people," and "I think Artificial Intelligence is dangerous," there could have been bias introduced to their general attitudes based on how they knew they answered the Self Disclosure survey portion. Together, the results indicate that H5 was incorrect. Positive attitudes towards AI were not predictive of self-disclosure preferences to AI, and negative attitudes towards AI predicted less likelihood to self-disclose to AI.

Across the three regressions to predict preference for self-disclosure to the agent conditions, a more forgiving attitude towards the negative aspects of AI (indicated by higher scores on the negative attitude toward AI subscale) consistently predicted a lower preference to self-disclose personal information to the human researcher, the AI, and the private condition. Since higher scores on this scale reflect less endorsement of negative views toward AI (e.g., seeing AI as dangerous, sinister, or invasive), this suggests that participants who were more comfortable with AI tended to avoid selecting Self-based statements over Fact-based ones, regardless of the agent condition. This pattern does not support H4C, but it may suggest that participants who were more at ease with AI approached the task in a more detached or task-focused way, favoring less personal responses overall. Since the experiment was structured in such a way that participants were repeatedly presented with two options

at a time (for example, AI-Fact vs. Private-Self) and chose which one they wanted to answer, the findings could reflect consistent patterns of choice behavior across many trials rather than true disclosure tendencies. Participants who were more neutral or less negative in their views of AI were not avoiding disclosure entirely, but they showed a lower preference for engaging with personal questions across conditions.

5.4. Experience with AI as a predictor of self-disclosure preferences

Finally, Hypothesis 6 predicted that having more experience with AI would be associated with a greater preference for choosing Self-based statements. This hypothesis was not supported: self-reported extent of experience with AI was not predictive of a higher likelihood to self-disclose information to AI. This suggests that the amount of experience an individual has with AI does not influence their willingness to share personal information with it in this type of structured decision-making task. One possible explanation is that participants were responding to the immediate context of the interaction, rather than drawing on their broader experiences with AI. Since the paradigm focused on making repeated choices between fixed options, prior experience may not have played a meaningful role in shaping responses.

Overall, our primary hypotheses (H1-H3) regarding general disclosure preferences and agent selection were supported. However, the predictions related to individual differences (H4-H6) received only partial support. While some personality traits, agreeableness and conscientiousness, and negative attitude towards AI emerged as significant predictors of self-disclosure across the different agent types, positive attitudes and extent of experience did not. These findings suggest that while general self-disclosure tendencies toward AI are robust, individual differences may play a more nuanced or context-dependent role. However, it is important to interpret these findings with caution as the internal consistency of the personality trait measures was relatively low for certain subscales, particularly openness and neuroticism. This limitation may have reduced the likelihood of detecting reliable associations for those traits.

6. Limitations and future work

Some limitations of this study included the way AI was presented in this study, the potential lack of generalizability to other AI types, and the limited statement types presented. This study presented the Self-Disclosure Experimental Questionnaire without first defining what AI meant, as the researchers wanted to understand self-disclosure preferences based on participants' preconceived idea of AI without the bias of a standardized definition. During the GAAIS survey, a specific definition of AI was given before they assessed their attitudes towards AI, but there may have been unaccounted for variation in how participants conceptualized AI when completing the previous sections (e.g. some participants may have thought of AI as being a conversational agent such as ChatGPT whereas others may have thought of it as an advanced data processing tool). Future studies should aim to bridge this gap by either predefining AI or asking participants to explain what AI is in their own words before testing. If individuals all have the same concept of AI upon testing, the results should be more robust.

While the sample size was adequately powered to detect large effects for individual differences, as discussed in section 3.1, it still limits the generalizability of the findings to broader populations. The sample also consisted of only university students, which may not reflect disclosure tendencies in more diverse populations. Larger and more demographically varied samples would strengthen external validity. Given that this study was specifically designed to be a baseline assessment for comparing self-disclosure preferences towards an AI compared to a human, generalizability to every form of AI is challenging. This work was done based on prior literature that focused mainly on AI chatbot interaction, and the results of this study may or may not be applicable to other forms of AI without further testing, as different AI types may elicit varying levels of trust and attitudes towards AI. Furthermore, it is possible that an embodied AI agent (i.e., an AI presented through a robotic apparatus, potentially with human-like features) may elicit different self-disclosure tendencies.

The forced-choice paradigm, while useful for isolating specific decision tendencies, limits ecological validity and is potentially limited in representing how self-disclosure naturally occurs in interactions. Future studies should explore whether the equivalent preference to self-disclose to AI and humans replicates in conversations (written and/or verbal) with actual AI chatbots and embodied agents (robots with AI conversation capabilities). This approach should provide a better understanding of the stability or flexibility of these findings across different technologies.

This study also relied on self-report and forced-choice data. While this approach provides insights into disclosure preferences, it does not necessarily capture deeper behavioral or physiological responses during the disclosure decision. Incorporating biometric measures, observational data, or direct AI interactions could strengthen future studies by providing converging evidence on the affective and cognitive components of disclosure decisions.

Regarding test item format, this study presented a within-subjects forced-choice paradigm between self and fact-based statements. In doing so, we were able to examine preferences and decision-making behavior surrounding self-disclosure responses vs fact-based responses. This method has been used previously to successfully examine behavior and interactions between different agents and has elicited self-disclosure preferences in the past (Tamir & Mitchell, 2012; Vijayakumar et al., 2020). However, this study did not attempt to examine organic self-disclosure to an AI through unprompted and continuous interaction, limiting ecological validity. Our study also did not provide monetary incentive to compare perceived value of disclosure as previous studies have (Tamir & Mitchell, 2012; Vijayakumar et al., 2020), and instead focused on the knowledge of the perceiver as the factor of interest. Future research would benefit from a design that incorporates a more organic approach to human-AI interaction in order to determine whether or not actual self-disclosure tendencies emerge that are different between AI and human agents.

Lastly, the self-based statements presented here consisted mostly of surface-level self-information, which limits the associated risk of sharing this information. Now that the preference of answering Self statements over Fact statements has been shown toward AI, future research should expand upon the types of self-disclosure that individuals prefer to provide to AI and to humans, as the literature is still inconclusive on this topic (Lee, Yamashita, & Huang, 2020; Mou & Xu, 2017; Uchida et al., 2017). Identifying what sort of personal information individuals prefer to share with other humans compared to AI should help identify the most effective way to implement AI. This will also help broaden our understanding of the extent to which someone will self-disclose personal information, and how particular prompts or interactions with an agent may elicit different response tendencies during engagement with AI.

Together, these limitations highlight the need for future studies to incorporate larger and more diverse participant pools, richer measures of behavior and physiology, and interactive or immersive AI contexts. Doing so will help to clarify the conditions under which individuals are most likely to disclose to AI, and broaden our overall understanding of how Human-AI interaction occurs.

7. Conclusion

Using a forced-choice paradigm to explore patterns in statement-type selection, our findings reveal that individuals are more inclined to engage with self-disclosure statements than with fact-based ones. Moreover, they are just as willing to disclose personal information to AI as they are to another human. Engagement with self-disclosure questions was notably higher when a social agent was presented, compared

to when responses were kept private.

This study offers several novel contributions to the field. By using a behavioral choice design rather than relying solely on self-report, we were able to capture real-time disclosure preferences with greater ecological validity. Additionally, the inclusion of a private condition as a behavioral control enabled a more precise examination of agent-driven effects on self-disclosure. Further, matching the interest levels of statements for both the Self and Fact statements helped ensure that observed preferences were not artifacts of statement engagement, thereby strengthening the interpretation that self-disclosure is intrinsically rewarding.

These findings lend support to the CASA framework by providing behavioral evidence that individuals perceive AI as a social agent capable of receiving and analyzing personal information. Additionally, personality traits and attitudes toward AI emerged as key predictors of self-disclosure to AI and human agents. These findings have important implications for the future development of AI and provide a foundation for understanding the complex motivations behind technology use. Since individuals not only express a willingness to self-disclose to AI, but also derive emotional and psychological satisfaction from these interactions (Ho et al., 2018), the integration of AI into sensitive environments—such as therapeutic or medical settings—holds considerable promise.

Future research should build on this study by examining the extent of self-disclosure to AI, including the depth of personal information shared, the types of statements people are likely to answer or avoid, and the individual differences that shape these self-disclosure preferences. By expanding our research focus, the field can better understand the psychological mechanisms driving disclosure behavior to machines and design AI systems that support human needs.

CRediT authorship contribution statement

Elizabeth R. Merwin: Writing – review & editing, Writing – original draft, Visualization, Validation, Project administration, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Allen C. Hagen:** Writing – review & editing, Writing – original draft, Methodology, Investigation, Formal analysis, Data curation, Conceptualization. **Joseph R. Keebler:** Writing – review & editing, Validation, Project administration, Investigation, Formal analysis, Data curation. **Chad Forbes:** Writing – review & editing, Validation, Supervision, Project administration, Methodology, Formal analysis, Data curation, Conceptualization.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.chbah.2025.100180.

References

Adamopoulou, E., & Moussiades, L. (2020). An overview of chatbot technology. *IFIP Advances in Information and Communication Technology, 584*(1), 373–383. https://doi.org/10.1007/978-3-030-49186-4_31

Almahairah, M. S. (2023). Artificial intelligence application for effective customer relationship management. *IEEE*. https://doi.org/10.1109/icicci56745.2023.10128360

Altman, I., & Taylor, D. A. (1973). *Social penetration: The development of interpersonal relationships*. Rinehart & Winston.

Bickmore, T. W., & Picard, R. W. (2005). Establishing and maintaining long-term human-computer relationships. *ACM Transactions on Computer-Human Interaction, 12*(2), 293–327. https://doi.org/10.1145/1067860.1067867

Bresin, K., & Robinson, M. D. (2014). You are what you see and choose: Agreeableness and situation selection. *Journal of Personality, 83*(4), 452–463. https://doi.org/10.1111/jopy.12121

Caci, B., Cardaci, M., & Miceli, S. (2019). Development and maintenance of self-disclosure on facebook: The role of personality traits. *Sage Open, 9*(2), Article 215824401985694. https://doi.org/10.1177/2158244019856948

Chen, X., Pan, Y., & Guo, B. (2016). The influence of personality traits and social networks on the self-disclosure behavior of social network site users. *Internet Research, 26*(3), 566–586. https://doi.org/10.1108/intr-05-2014-0145

Choung, H., David, P., & Ross, A. (2022). Trust in AI and its role in the acceptance of AI technologies. *International Journal of Human-Computer Interaction, 39*(9), 1–13. https://doi.org/10.1080/10447318.2022.2050543

Costa, P. T., & McCrae, R. R. (1992). The five-factor model of personality and its relevance to personality disorders. *Journal of Personality Disorders, 6*(4), 343–359. https://doi.org/10.1521/pedi.1992.6.4.343

Croes, E. A. J., & Antheunis, M. L. (2020). Can we be friends with mitsuku? A longitudinal study on the process of relationship formation between humans and a social chatbot. *Journal of Social and Personal Relationships, 38*(1), 279–300. https://doi.org/10.1177/0265407520959463

Croes, E. A. J., Antheunis, M. L., Chris, & Jan. (2024). Digital confessions: The willingness to disclose intimate information to a chatbot and its impact on emotional well-being. *Interacting with Computers, 36*(5). https://doi.org/10.1093/iwc/iwae016

Croes, E. A. J., Antheunis, M. L., Goudbeek, M. B., & Wildman, N. W. (2022). "I am in your computer while we talk to each other" a content analysis on the use of language-based strategies by humans and a social chatbot in initial human-chatbot interactions. *International Journal of Human-Computer Interaction, 39*(10), 2155–2173. https://doi.org/10.1080/10447318.2022.2075574

Demeure, V., Niewiadomski, R., & Pelachaud, C. (2011). How is believability of a virtual agent related to warmth, competence, personification, and embodiment? *Presence: Teleoperators and Virtual Environments, 20*(5), 431–448. https://doi.org/10.1162/pres_a_00065

Donders, A. R. T., van der Heijden, G. J. M. G., Stijnen, T., & Moons, K. G. M. (2006). Review: A gentle introduction to imputation of missing values. *Journal of Clinical Epidemiology, 59*(10), 1087–1091. https://doi.org/10.1016/j.jclinepi.2006.01.014

Dunbar, R. I. M., Marriott, A., & Duncan, N. D. C. (1997). Human conversational behavior. *Human Nature, 8*(3), 231–246. https://doi.org/10.1007/bf02912493

Faul, F., Erdfelder, E., Lang, A.-G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods, 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Faverio, M., & Tyson, A. (2023). *What the data says about Americans' views of artificial intelligence*. Pew Research Center. https://www.pewresearch.org/short-reads/2023/11/21/what-the-data-says-about-americans-views-of-artificial-intelligence/.

Fiske, A., Henningsen, P., & Buyx, A. (2019). Your robot therapist will see you now: Ethical implications of embodied artificial intelligence in psychiatry, Psychology, and Psychotherapy. *Journal of Medical Internet Research, 21*(5). https://doi.org/10.2196/13216

Folk, D., Yu, S., & Dunn, E. (2024). Can chatbots ever provide more social connection than humans? *Collabra. Psychology, 10*(1). https://doi.org/10.1525/collabra.117083

Gambino, A., Fox, J., & Ratan, R. (2020). Building a stronger CASA: Extending the computers are social actors paradigm. *Human-Machine Communication, 1*(1), 71–86. https://doi.org/10.30658/hmc.1.5

Gkinko, L., & Elbanna, A. R. (2021). AI in the workplace: Exploring chatbot use and users' emotions. *Responsible AI and Analytics for an Ethical and Inclusive Digitized Society*, 18–28. https://doi.org/10.1007/978-3-030-85447-8_2

Gosling, S. D., Rentfrow, P. J., & Swann, W. B. (2003). A very brief measure of the Big-Five personality domains. *Journal of Research in Personality, 37*(6), 504–528. https://doi.org/10.1016/S0092-6566(03)00046-1

Guingrich, R. E., & Graziano, M. S. A. (2023). Chatbots as social companions: How people perceive consciousness, human likeness, and social health benefits in machines. ArXiv https://doi.org/10.48550/arxiv.2311.10599.

Ho, A., Hancock, J., & Miner, A. S. (2018). Psychological, relational, and emotional effects of self-disclosure after conversations with a chatbot. *Journal of Communication, 68*(4), 712–733. https://doi.org/10.1093/joc/jqy026

Hollenbaugh, E. E., & Ferris, A. L. (2014). Facebook self-disclosure: Examining the role of traits, social cohesion, and motives. *Computers in Human Behavior, 30*, 50–58. https://doi.org/10.1016/j.chb.2013.07.055

Horodyski, P. (2023). Recruiter's perception of artificial intelligence (AI)-based tools in recruitment. *Computers in Human Behavior Reports, 10*, 1–10. https://doi.org/10.1016/j.chbr.2023.100298

Ignatius, E., & Kokkonen, M. (2007). Factors contributing to verbal self-disclosure. *Nordic Psychology, 59*(4), 362–391. https://doi.org/10.1027/1901-2276.59.4.362

Joinson, A. N., Reips, U. D., Buchanan, T., & Schofield, C. B. P. (2010). Privacy, trust, and self-disclosure online. *Human-Computer Interaction, 25*(1), 1–24. https://doi.org/10.1080/07370020903586662

Kaya, F., Aydin, F., Schepman, A., Rodway, P., Yetişensoy, O., & Demir Kaya, M. (2022). The roles of personality traits, AI anxiety, and demographic factors in attitudes toward artificial intelligence. *International Journal of Human-Computer Interaction, 40*(2), 497–514. https://doi.org/10.1080/10447318.2022.2151730

Kelly, A., Sullivan, M., & Strampel, K. (2023). Generative artificial intelligence: University student awareness, experience, and confidence in use across disciplines. *Journal of University Teaching and Learning Practice, 20*(6). https://doi.org/10.53761/1.20.6.12

Kennedy, D., Lakonishok, J., & Shaw, W. H. (1992). Accommodating outliers and nonlinearity in decision models. *Journal of Accounting, Auditing and Finance, 7*(2), 161–190. https://doi.org/10.1177/0148558x9200700205

Kim, J., & Dindia, K. (2011). Online self-disclosure: A review of research. In L. M. Webb (Ed.), *Computer-mediated communication in personal relationships* (pp. 156–180). Peter Lang Publishing.

Lee, J., Lee, D., & Lee, J. G. (2022). Influence of rapport and social presence with an AI psychotherapy chatbot on users' self-disclosure. *International Journal of Human-Computer Interaction, 40*(7), 1620–1631. https://doi.org/10.1080/10447318.2022.2146227

Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors, 46*(1), 50–80. https://doi.org/10.1518/hfes.46.1.50_30392

Lee, Y.-C., Yamashita, N., & Huang, Y. (2020a). Designing a chatbot as a mediator for promoting deep self-disclosure to a real mental health professional. In *Proceedings of the ACM on human-computer interaction* (Vol. 4, pp. 1–27). https://doi.org/10.1145/3392836. CSCW1.

Lee, Y.-C., Yamashita, N., Huang, Y., & Fu, W. (2020b). "I hear you, I feel you": Encouraging deep self-disclosure through a chatbot. In *Proceedings of the 2020 CHI conference on human factors in computing systems*. https://doi.org/10.1145/3313831.3376175

Lei, X., & Liu, F. (2024). How service robots facilitate user delf-disclosure: The roles of personality, animacy, and automated social presence. *International Journal of Human-Computer Interaction, 41*(4), 2135–2148. https://doi.org/10.1080/10447318.2024.2316368

Loiacono, E. T. (2014). Self-disclosure behavior on social networking web sites. *International Journal of Electronic Commerce, 19*(2), 66–94. https://doi.org/10.1080/10864415.2015.979479

Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human–human and human–automation trust: An integrative review. *Theoretical Issues in Ergonomics Science, 8*(4), 277–301. https://doi.org/10.1080/14639220500337708

Malhi, A., Knapic, S., & Främling, K. (2020). Explainable agents for less bias in human-agent decision making. *Lecture Notes in Computer Science, 12175*, 129–146. https://doi.org/10.1007/978-3-030-51924-7_8

Miller, L. C., Berg, J. H., & Archer, R. L. (1983). Openers: Individuals who elicit intimate self-disclosure. *Journal of Personality and Social Psychology, 44*(6), 1234–1244. https://doi.org/10.1037/0022-3514.44.6.1234

Mou, Y., & Xu, K. (2017). The media inequality: Comparing the initial human-human and human-AI social interactions. *Computers in Human Behavior, 72*, 432–440. https://doi.org/10.1016/j.chb.2017.02.067

Müller, C. (2024). World robotics 2024 - industrial robots. https://ifr.org/img/worldrobotics/Executive_Summary_WR_2024_Industrial_Robots.pdf.

Naaman, M., Boase, J., & Lai, C.-H. (2010). Is it really about me?: Message content in social awareness streams. In *Proceedings of the 2010 ACM conference on computer supported cooperative work - CSCW '* (Vol. 10). https://doi.org/10.1145/1718918.1718953

Nass, C., Fogg, B. J., & Moon, Y. (1996). Can computers be teammates? *International Journal of Human-Computer Studies, 45*(6), 669–678. https://doi.org/10.1006/ijhc.1996.0073

Nass, C., Steuer, J., & Tauber, E. R. (1994). Computers are social actors. In *Proceedings of the SIGCHI conference on human factors in computing systems celebrating interdependence - CHI '* (Vol. 94, pp. 72–78). https://doi.org/10.1145/191666.191703

O'Connor, K., Neff, D. M., & Pitman, S. (2018). Burnout in mental health professionals: A systematic review and meta-analysis of prevalence and determinants. *European Psychiatry, 53*, 74–99. https://doi.org/10.1016/j.eurpsy.2018.06.003

Oertel, C., Castellano, G., Chetouani, M., Nasir, J., Obaid, M., Pelachaud, C., & Peters, C. (2020). Engagement in human-agent interaction: An overview. *Frontiers in Robotics and AI, 7*. https://doi.org/10.3389/frobt.2020.00092

Rapp, A., Boldi, A., Curti, L., Perrucci, A., & Simeoni, R. (2023). How do people ascribe humanness to chatbots? An analysis of real-world human-agent interactions and a theoretical model of humanness. *International Journal of Human-Computer Interaction, 40*(19), 1–24. https://doi.org/10.1080/10447318.2023.2247596

Reeves, B., & Nass, C. (1996). The media equation: How people treat computers, television, & new media like real people & places. *Computers & Mathematics with Applications, 33*(5), 19–36. https://doi.org/10.1016/s0898-1221(97)82929-x

Schaefer, K. E., Chen, J. Y. C., Szalma, J. L., & Hancock, P. A. (2016). A meta-analysis of factors influencing the development of trust in automation. *Human Factors: The Journal of the Human Factors and Ergonomics Society, 58*(3), 377–400. https://doi.org/10.1177/0018720816634228

Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes towards artificial intelligence scale. *Computers in Human Behavior Reports, 1*. https://doi.org/10.1016/j.chbr.2020.100014

Schepman, A., & Rodway, P. (2022). The general attitudes towards artificial intelligence scale (GAAIS): Confirmatory validation and associations with personality, corporate distrust, and general trust. *International Journal of Human-Computer Interaction, 39*(13), 2724–2741. https://doi.org/10.1080/10447318.2022.2085400

Seidman, G. (2013). Self-presentation and belonging on facebook: Ow personality influences social media use and motivations. *Personality and Individual Differences, 54*(3), 402–407. https://doi.org/10.1016/j.paid.2012.10.009

Sharma, V., Goyal, M. S. A., & Malik, D. (2017). An intelligent behaviour shown by chatbot system. *International Journal of New Technology and Research, 3*(4). https://www.neliti.com/publications/263312/an-intelligent-behaviour-shown-by-chatbot-system.

Sheng, H., & Xiao, H. (2022). Examining users' continuous use intention of AI-enabled online education applications. In *2022 international conference on computer engineering and artificial intelligence (ICCEAI)* (pp. 642–645). https://doi.org/10.1109/icceai55464.2022.00136

Singh, A., Triulzi, G., & Magee, C. L. (2021). Technological improvement rate predictions for all technologies: Use of patent data and an extended domain description. *Research Policy, 50*(9), Article 104294. https://doi.org/10.1016/j.respol.2021.104294

Skjuve, M. B., & Brandtzæg, P. B. (2018). Chatbots as a new user interface for providing health information to young people. In Y. Andersson, et al. (Eds.), *Youth and News in a Digital Media Environment: Nordic-Baltic perspectives* (pp. 59–66). http://hdl.handle.net/11250/2576290.

Skjuve, M., Følstad, A., Fostervold, K. I., & Brandtzaeg, P. B. (2022). A longitudinal study of human–chatbot relationships. *International Journal of Human-Computer Studies, 168*, Article 102903. https://doi.org/10.1016/j.ijhcs.2022.102903

Sprecher, S., Treger, S., Wondra, J. D., Hilaire, N., & Wallpe, K. (2013). Taking turns: Reciprocal self-disclosure promotes liking in initial interactions. *Journal of Experimental Social Psychology, 49*(5), 860–866. https://doi.org/10.1016/j.jesp.2013.03.017

Stein, J.-P., Messingschlager, T., Gnambs, T., Hutmacher, F., & Appel, M. (2024). Attitudes towards AI: Measurement and associations with personality. *Scientific Reports, 14*(1). https://doi.org/10.1038/s41598-024-53335-2

Tamir, D. I., & Mitchell, J. P. (2012). Disclosing information about the self is intrinsically rewarding. In *Proceedings of the national academy of sciences* (Vol 109, pp. 8038–8043). https://doi.org/10.1073/pnas.1202129109, 21.

Uchida, T., Takahashi, H., Ban, M., Shimaya, J., Yoshikawa, Y., & Ishiguro, H. (2017). A robot counseling system — what kinds of topics do we prefer to disclose to robots? *IEEE Xplore*, 207–212. https://doi.org/10.1109/roman.2017.8172303

Vijayakumar, N., Flournoy, J. C., Mills, K. L., Cheng, T. W., Mobasser, A., Flannery, J. E., Allen, N. B., & Pfeifer, J. H. (2020). Getting to know me better: An fMRI study of intimate and superficial self-disclosure to friends during adolescence. *Journal of Personality and Social Psychology, 118*(5), 885–899. https://doi.org/10.1037/pspa0000182

Weizenbaum, J. (1966). Eliza - a computer program for the study of natural language communication between man and machine. *Communications of the ACM, 9*(1), 36–45. https://doi.org/10.1145/365153.365168

Wheeless, L. R., & Grotz, J. (1977). The measurement of trust and its relationship to self-disclosure. *Human Communication Research, 3*(3), 250–257. https://doi.org/10.1111/j.1468-2958.1977.tb00523.x

Wu, H., Wang, Y., & Wang, Y. (2024). "To use or not to use?" A mixed-methods study on the determinants of EFL college learners' behavioral intention to use AI in the distributed learning context. *The International Review of Research in Open and Distributed Learning, 25*(3), 158–178. https://doi.org/10.19173/irrodl.v25i3.7708

Yang, S., Tan, G. K. J., Sim, K., Lim, L. J. H., Tan, B. Y. Q., Kanneganti, A., Ooi, S. B. S., & Ong, L. P. (2024). Stress and burnout amongst mental health professionals in Singapore during Covid-19 endemicity. *PLoS One, 19*(1), Article e0296798. https://doi.org/10.1371/journal.pone.0296798

Zabel, S., Pensini, P., & Otto, S. (2025). Unveiling the role of honesty-humility in shaping attitudes towards artificial intelligence. *Personality and Individual Differences, 238*, 113072. https://doi.org/10.1016/j.paid.2025.113072, 113072.

Zhou, L., Gao, J., Li, D., & Shum, H. Y. (2020). The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics, 46*(1), 53–93. https://doi.org/10.1162/coli_a_00368

Zhou, M. X., Mark, G., Li, J., & Yang, H. (2019). Trusting virtual agents: The effect of personality. *ACM Transactions on Interactive Intelligent Systems, 9*(2–3), 1–36. https://doi.org/10.1145/3232077