Trusting Your AI Agent Emotionally and Cognitively: Development and Validation of a Semantic Differential Scale for AI Trust

Ruoxi Shang[1], Gary Hsieh[1], Chirag Shah[1]

[1]University of Washington
rxshang@uw.edu, garyhs@uw.edu, chirags@uw.edu

Trust is not just a cognitive issue but also an emotional one, yet the research in human-AI interactions has primarily focused on the cognitive route of trust development. Recent work has highlighted the importance of studying affective trust towards AI, especially in the context of emerging human-like LLMs-powered conversational agents. However, there is a lack of validated and generalizable measures for the two-dimensional construct of trust in AI agents. To address this gap, we developed and validated a set of 27-item semantic differential scales for affective and cognitive trust through a scenario-based survey study. We then further validated and applied the scale through an experiment study. Our empirical findings showed how the emotional and cognitive aspects of trust interact with each other and collectively shape a person's overall trust in AI agents. Our study methodology and findings also provide insights into the capability of the state-of-art LLMs to foster trust through different routes.

Introduction

Trust plays a crucial role not only in fostering cooperation, efficiency, and productivity in human relationships (Brainov and Sandholm 1999) but also is essential for the effective use and acceptance of computing and automated systems, including computers (Madsen and Gregor 2000), automation (Lee and See 2004), robots (Hancock et al. 2011), and AI technologies (Kumar 2021), with a deficit in trust potentially causing rejection of these technologies (Glikson and Woolley 2020). The two-dimensional model of trust, encompassing both cognitive and affective dimensions proposed and studied in interpersonal relationship studies (McAllister 1995; Johnson and Grayson 2005; Parayitam and Dooley 2009; Morrow Jr, Hansen, and Pearson 2004), have been adopted in studying trust in human-computer interactions, particularly with human-like technologies (Hu, Lu et al. 2021; Glikson and Woolley 2020). Cognitive trust relates to the perception of the ability (e.g., skills, knowledge, and competencies), reliability, and integrity of the trustee, whereas the affective dimension involves the perceived benevolence and disposition to do good of the trustee (Johnson and Grayson 2005; Mayer, Davis, and Schoorman

1995). In the context of human-computer trust, cognition-based trust is built on the user's intellectual perceptions of the system's characteristics, whereas affect-based components are those which are based on the user's emotional responses to the system (Madsen and Gregor 2000).

While AI trust research has largely centered on technical reliability and competency, there is a notable lack of work that explores the affective routes of trust development. The recent advancement of text-based Large Language Models (LLMs) have demonstrated a remarkable ability to take on diverse personas and skill-sets, recognizing and responding to people's emotional needs during conversation-based interactions. This capability is crucially aligned with the increasing focus on simulating Affective Empathy in human-AI interactions (Paiva et al. 2017; Welivita, Xie, and Pu 2021). In light of this, there is growing research interest in studying affective aspects of trust in AI (Glikson and Woolley 2020; Granatyr et al. 2017; Kyung and Kwon 2022; Zhang, Pentina, and Fan 2021; Guerdan, Raymond, and Gunes 2021). However, a critical gap exists in the lack of generalizable and accurate specialized measurement tools for assessing affective trust in the context of AI, especially with the enhanced and nuanced capabilities of LLMs. This highlights a need for a better measurement scale for affective trust to gain a deeper understanding of how trust dynamics function, particularly in the context of emotionally intelligent AI.

In this paper, we introduce a 27-item semantic differential scale for assessing cognitive and affective trust in AI, aiding researchers and designers in understanding and improving human-AI interactions. Our motivation and scale development process is based on a long strand of prior research on the cognitive-affective construct of trust that has been shown to be important in interpersonal trust in organizations, human trust in conventional technology and automation, and more recently in trust towards AI. Our use of OpenAI's ChatGPT to generate different levels of affective trust further demonstrates a scalable method for studying the emotional pathway to AI trust. Empirically, we contribute findings on the interplay and distinction between cognitive, affective, and moral trust. The paper is structured to highlight these contributions: Section describes the development and validation of our trust scale through an experimental survey study and factor analysis. The validation and application

of our scale is described in Section . Section begins with a preliminary study validating scales and testing LLMs as a tool to manipulate affective trust. This is followed by a refined study to further establish discriminant validity and explore cognitive-affective trust dynamics. Section then discusses the implications of these findings as well as potential usage of our trust scale.

Related Work

Shifting Paradigm of AI Trust Research

Due to the opaque nature of most high-performing AI models, trust between the user and the AI system has always been a critical issue (Carvalho, Pereira, and Cardoso 2019; Thiebes, Lins, and Sunyaev 2021; Jacovi et al. 2021), as inappropriate trust can lead to over-reliance or under-utilization of AI systems (Buçinca, Malaya, and Gajos 2021; Asan, Bayrak, and Choudhury 2020). Research in trust has predominantly adopted the cognitive evaluation of the system's performance (Granatyr et al. 2015), such as its accuracy in making predictions (Ribeiro, Singh, and Guestrin 2016), its perceived consistency in completing tasks (Mcknight et al. 2011), and its ethical considerations and transparency in decision-making (Durán and Jongsma 2021).

Studies in psychology have long been establishing the importance of psychological influence (e.g., emotions, personality, moods) on trust (Dunn and Schweitzer 2005; Lount Jr 2010). Extending beyond the traditional numeric and cognitive paradigm, recent works have proposed the importance of exploring affective factors of trust in AI systems (Granatyr et al. 2017; Gillath et al. 2021; Jeon 2023). For technologies perceived as more human-like, affective trust factors such as benevolence and integrity play a more significant role (Lankton, McKnight, and Tripp 2015). Moreover, recent advancements in AI, particularly in Large Language Models (LLMs) has demonstrated its capability beyond traditional task performance, as scholars find it challenging not to anthropomorphize them (Shanahan 2022). Notably, OpenAI's GPT-4, has shown excellent performance in Emotional Awareness (i.e. the ability to identify and describe emotions) (Elyoseph et al. 2023). There is also increasing interest in studying LLMs' empathetic responses (Ayers et al. 2023; Belkhir and Sadat 2023). Our work extends the current focus on the emotional aspects of AI interactions by highlighting the need to explore the emotional dimension of trust, a concept with deep roots in research studying interpersonal relationships.

Affective and Cognitive Trust

The interdisciplinary nature of AI trust research motivates the adoption of theoretical frameworks from interpersonal relationship literature (Bansal et al. 2023; Thiebes, Lins, and Sunyaev 2021). Among the classic interpersonal trust theories and models (e.g., (Mayer, Davis, and Schoorman 1995; Rempel, Holmes, and Zanna 1985)), a two-dimensional model with cognitive and affective components has been extensively studied (McAllister 1995). Similar to trust towards humans, trust towards technology has both cognitive and affective components (Komiak and Benbasat 2004). In the AI context, cognitive trust relates to the user's intellectual perceptions of the AI's characteristics (Komiak and Benbasat 2004; Madsen and Gregor 2000), focusing on aspects like reliability and transparency. Affective trust, on the other hand, involves emotional responses to the AI, including factors like tangibility and anthropomorphism (Ueno et al. 2022; Glikson and Woolley 2020). This duality is essential due to the inherent complexity of AI, which often suggests a need for a "leap of faith" in its hidden processes, beyond what can be cognitively processed (Hoff and Bashir 2015; Lee and See 2004). Prior works have found the limitation of cognition in decision-making, as demonstrated by studies showing limitations in users' abilities to discern AI inaccuracies, even with support through explanations (Jacobs et al. 2021; Buçinca, Malaya, and Gajos 2021). The cognitive-affective architecture has been established in research of computational agents (Pérez et al. 2016; Chumkamon, Hayashi, and Koike 2016). The importance of this bi-dimensional model lies in its capacity to capture the full spectrum of trust dynamics that single-dimensional models, focusing solely on either aspects, fail to encompass. Trust has also been investigated through other bi-dimensional models in Human-Robot Interaction (HRI) (e.g. Law and Scheutz's Performance-based and Relation-based trust (Law and Scheutz 2021), and Malle and Ullman's Multi-Dimensional Measure of Trust (MDMT) (Malle and Ullman 2021)). Our work makes a unique contribution by focusing on the Cognitive-Affective (C-A) trust model that fully encapsulates the emotional and psychological intricacies in the interactions with the state-of-the-art AI models that have advanced emotional intelligence. Although MDMT was derived from a different body of prior literature in social-moral constructs as mentioned in their work (Ullman and Malle 2018, 2019), we found it to be a suitable scale to compare it with due to its similar bi-dimentional construct and adjective item format with our scale. Therefore, in Section , we use the moral scale to establish discriminant validity with our cognitive-affective trust scale, demonstrating the distinctiveness of our cognitive-affective trust scale.

Role and Effects of Affective Trust

There is growing research interest in exploring the role of affective trust in the use of AI technologies. A few recent works have highlighted that affect-based trust plays a decisive role in people's acceptance of AI-based technology in preventative health interventions (Kyung and Kwon 2022) and financial services robo-advising (Zhang, Pentina, and Fan 2021). Research in explainable AI (XAI) has also shown that people's affective responses to explanations are crucial in improving personalization and increasing trust in AI systems (Guerdan, Raymond, and Gunes 2021). However, given the interdisciplinary nature of AI trust research, the valuable insights to be borrowed from interpersonal trust are currently understudied in the AI context. Prior work has found that affective and cognitive trust have different impacts on relationships (Webber 2008; McAllister 1995). Cognitive trust tends to form rapidly (McKnight, Choudhury, and Kacmar 2002; Meyerson et al. 1996), whereas affective trust is more persistent under challenges in teamwork

(McAllister 1995) and interpersonal relationships (Williams 2001). Affective trust also shows greater resilience to short-term issues and errors (Jones and George 1998; McAllister 1995). Researchers have also shown that affective and cognitive trust are not isolated constructs; rather, they complement each other (Granatyr et al. 2017), and affective trust needs to be developed on the basis of cognitive trust (Johnson and Grayson 2005). Acknowledging these research opporunities, our work is a step towards a deeper and holistic examinination of the complex dynamics between cognitive and affective trust and their contribution to general trust in AI.

Gaps in Empirical Research and Measurement of Affective Trust in AI

Despite growing interest in this space, existing studies and measurement scales for affective trust in AI exhibit limitations, particularly in the adaptation and validation of measurement scales. Many existing scales, primarily developed for human trust contexts, have been applied to AI interactions with minimal modifications, raising questions about their generalizability. For instance, trust items intended for Human-Computer Trust were directly used for AI systems handling personal data, without substantial revision to reflect the unique aspects of AI interactions (Liao and Sundar 2021). Furthermore, there's a lack of consensus on defining affective trust in AI. While Kyung and Kwon (Kyung and Kwon 2022) merged benevolence and integrity dimensions to measure affective trust in AI-based health interventions, Shi et al. (Shi, Gong, and Gursoy 2021) categorized these dimensions as cognitive trust, employing a different scale (Komiak and Benbasat 2006) for affective trust. This inconsistency highlights the need for a unified, valid measure of trust for AI technologies (Ueno et al. 2022). Given the intertwined nature of affective and cognitive trust, it is evident that a comprehensive evaluation of trust in AI systems requires a scale that measures both dimensions. In response, this work adopts Verhagen et al.'s (Verhagen, Hooff, and Meents 2015) approach, developing semantic differential scales for both affective and cognitive trust in AI. Unlike Likert-type scales, semantic differentials use bipolar adjective pairs, offering advantages in reducing acquiescence bias and improving robustness (Hawkins, Albaum, and Best 1974), reliability (Wirtz and Lee 2003), and validity (Van Auken and Barry 1995).

Scale Development

Initial Item Generation

In developing our trust item pool, we conducted a comprehensive literature review to identify prominent two-dimensional trust models that differentiate between cognitive and affective components. We pooled models and items from literature in interpersonal trust, intraorganizational trust, and trust in interaction with computers and traditional technologies. This approach is consistent with the broader trend of extending trust research from human-human contexts to human-AI interactions, as evidenced by

the comprehensive review by Glikson and Woolley (Glikson and Woolley 2020). After the initial literature review, we applied a rigorous selection process to a larger body of trust literature, using the following criteria: 1) clear focus and delineation of affective and cognitive dimensions; 3) wide citation and application in various contexts; and 4) applicability to human-AI interactions. This ensures a thorough representation of the most relevant frameworks. The final selected models include Lewis and Weigert's sociological model (Lewis and Weigert 1985), McAllister's interpersonal trust model (McAllister 1995), Madsen and Gregor's Human-Computer Trust Components (Madsen and Gregor 2000), Johnson and Grayson's customer trust model (Johnson and Grayson 2005), and Komiak and Benbasat's IT adoption trust model (Komiak and Benbasat 2006). From these, we extracted 56 unique key adjectives from their scales. Subsequent refinement involved removing synonyms and ensuring coverage of key dimensions: reliability, predictability, competence, understandability, integrity, benevolence, and amiability, which were adopted from the subscales from the above-mentioned models. The dimensions are kept flexible and serves mainly as a reference for coverage. We also developed antonym pairs for each adjective using resources like Merriam-Webster and Oxford English Dictionary, selecting the most appropriate antonym after several review rounds among the researchers. This resulted in 33 paired adjective items, divided into cognitive ($N = 20$) and affective ($N = 13$) trust categories, as detailed in Table 1. In the following step, we recruited participants to rate these items with respect to various scenarios through an online survey study.

Survey design

We used the hypothetical scenario method, where participants evaluated vignettes describing realistic situations to rate trust-related scales (Trevino 1992). This method is frequently used in studying trust in emerging or future-oriented intelligent systems (Shi, Gong, and Gursoy 2021; Juravle et al. 2020; Gillath et al. 2021; Kim, Giroux, and Lee 2021). Hypothetical scenarios enable exploration of long-term, nuanced, human-like interactions with AI assistants. This method also facilitates control over variables like agent type and interaction types, and risk levels, ensuring generalizability. In addition, this method ensures consistency in contextual details across respondents (Alexander and Becker 1978). We crafted 32 scenario variations, manipulating the following five key dimensions: **Trust Level** (high vs. low), **Trust Route** (affective vs. cognitive), **Prior Interaction** (first-time vs. repeated), **Application Domain Stakes** (high vs. low), and **Agent Type** (human vs. AI).

For validation purpose of the scales, we manipulated **Trust Level** and **Trust Route**. This involved depicting the agent's characteristics and behaviors in the scenarios, aligning them with varying levels of cognitive or affective trust. Additionally, to ensure the scales' generalizability, we manipulated **Prior Interaction Frequency** to be interacting with the agent for the *first time* or *multiple times*, and we set **Application Domain Stakes** to be either high-stake domains (*Healthcare Diagnostics* and *Self-Driving Taxi*) and

low-stake domains (*Personal Training* and *Study Tutor*), inspired by real-world applications. These manipulations were implemented through texts presented to participants, as illustrated in Figure 1 (See Appendix). The tests for the manipulation's effectiveness will be demonstrated in Section .

Each participant were presented with two text-based scenarios for repeated measures. A mixed-model experiment design was deliberately chosen to incorporate both within-subject and between-subject variables. **Agent Type** and **Prior Interaction** are set to vary within-subjects to capture nuanced differences despite individual variability, and **Application Domain Stakes** is also designed to vary within-subjects to prevent boredom from the repetition of content. The order in which they see the variations are randomized to control for order effect. The rest of the dimensions are between-subjects and randomly assigned to participants. The two scenarios in Figure 1 (See Appendix) showcase one of the possible pairs of scenarios a participant may encounter.

During the survey study, after being presented with the first text-based scenarios, participants were asked to rate the semantic differential adjective pairs on a five-step scale, as well as a question assessing general trust in the AI agent. This process is repeated for the second scenario. After completing both scenarios, participants responded to questions used for our control variables including AI literacy and demographic information. The scenario structure comprised two parts: a prompt setting the interaction context, and three sentences detailing the agent's characteristics and behaviors.

Measurement and Variables

In our survey, we evaluated several key variables. For **Affective and Cognitive trust**, we used our semantic differential scale, where participants rated 33 adjective antonym pairs on a scale of -2 (most negative) to 2 (most positive). **General trust** was measured using a single-item questionnaire adapted from Yin (Yin, Wortman Vaughan, and Wallach 2019), where participants responded to the question "how much do you trust this AI assistant to provide you with the guidance and service you needed" on a 5-point Likert scale, ranging from 1 ("I don't trust this agent at all") to 5 ("I fully trust this AI"). **AI literacy** was assessed using items adapted from Wang (Wang, Rau, and Yuan 2022), all rated on a 5-point Likert scale from "Strongly disagree" to "Strongly agree", including items like "I can identify the AI technology in the applications I use" and "I can choose the most appropriate AI application for a task"

Participants

Amazon Mechanical Turk (MTurk) has been frequently used to recruit participants for online scenario-based studies related to AI technologies (Antes et al. 2021; Kim, Giroux, and Lee 2021; Kaur, Lampe, and Lasecki 2020). We recruited 200 participants from the United States through Amazon Mechanical Turk. The eligibility criteria included a minimum of 10,000 HITs Approved and an overall HIT Approval Rate of at least 98%. Each participant received a compensation of $2.20. The study involved repeated measures, collecting two sets of responses per participant for the two sce-

narios. Our quality control measures included a time delay for scenario reading, four attention checks, exclusions for uniform ratings or completion times more than two standard deviations from the mean, and a randomized sequence to control for order effects. After applying these criteria, we excluded 49 participants, resulting in 151 valid responses for the final analysis.

Results

Exploratory Factor Analysis   To uncover the factor structure underlying the 33 trust items, we first verified the suitability of our data for factor analysis. Bartlett's Test of Sphericity showed significant results ($\chi^2 = 12574, p < 0.001$) (Bartlett 1950), and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy was high at 0.98 (Kaiser 1970; Dziuban and Shirkey 1974), both indicating the appropriateness of factor analysis for our dataset. To determine the number of trust sub-components, we applied Kaiser's eigenvalue analysis (Kaiser 1958) and parallel analysis (Hayton, Allen, and Scarpello 2004), which collectively suggested a two-factor structure .

We initially used an oblique rotation as recommended by Tabachnick and Fiddell for instances where factor correlations exceed 0.32 (Tabachnick, Fidell, and Ullman 2013). Given the high correlation among our factors ($r = 0.78$) (Gorsuch 1988), we retained this rotation method. We then refined our item pool based on specific criteria: items were kept only if they had a factor loading above 0.4 (Howard 2016), ensuring significant association with the underlying factor. Items with a cross-loading of 0.3 or more were removed to align item responses with changes in the associated factor (Howard 2016). Additionally, we applied Saucier's criterion, eliminating items unless their factor loading was at least twice as high as on any other factor (Saucier 1994). This led to the removal of six items: Harmful - Well-intentioned, Unpromising - Promising, Malicious - Benevolent, Discouraging - Supportive, Insincere - Sincere, and Unpleasant - Likable.

A second round of exploratory factor analysis with the remaining 27 items preserved all items, as they met the above-mentioned criteria. The final item loadings are presented in Table 1 under the "All" column, with empty rows indicating the eliminated items. All remaining items demonstrated primary loadings above 0.55. Upon examining the keywords of items in each factor, two distinct themes emerged: cognitive trust and affective trust. This alignment was consistent with the dimensions identified in the initial literature review. Factor 1, representing cognitive trust, accounted for 43% of the total variance with 18 items, while Factor 2, corresponding to affective trust, explained 23% with 9 items.

Reliability   To test the internal reliability of the resulting items, we computed Cronbach's α for each scale. The cognitive trust scale ($\alpha = .98$) and the affective trust scale ($\alpha = .96$) both showed high internal consistency. We also tested the item-total correlation between each item and the average of all other items in the same sub-scale. All items' correlations exceed 0.6. In this development study, 18 items measuring cognitive trust and 9 items measuring affective

trust were identified with high reliability.

Construct Validity In addition to high reliability, we conducted analyses to show the validity of our scale. We first examined the construct validity, which refers to the degree to which the scale reflects the underlying construct of interest. Recall that we manipulated affective trust and cognitive trust through the level of trustworthiness and the trust development routes and controlled for factors like agent type, interaction stage, and risk level. T-test results revealed significant distinctions in both affective and cognitive trust scales under the experiment manipulation. Cognitive trust scale demonstrated a pronounced difference in high versus low cognitive trust conditions ($t = 45.74, p < 0.001$), and affective trust scale also showed a pronounced disparity in high versus low affective trust conditions ($t = 43.00, p < 0.001$). This also demonstrates the efficacy of our manipulation with the scenarios, as we observed significant differences in both the cognitive and affective dimensions.

We then fitted two separate linear random effect models (Singmann and Kellen 2019) on the two scales over the two manipulations due to our experiment design. Model 1 and Model 2 in Table 2 (See Appendix) tests the effects of our manipulations on the resulting trust scales, while Model 3 tests the effects of both scales on general trust. As shown in Table 2 (See Appendix), we observed significant main effects of manipulation **Trust Level** ($r = 2.059, p < 0.001$) and manipulation **Trust Route** ($r = -0.497, p < 0.01$) of these two manipulations on the cognitive trust scale, and the same is observed for affective trust scale. More importantly, the interaction effect shows that the affective trust scale is higher when higher trust is developed via the affective route ($r = 0.921, p < 0.001$), while the cognitive trust scale is higher when higher trust is developed via the cognitive route ($r = -0.538, p < 0.05$). The above analyses demonstrated the construct validity of our scale.

Concurrent Validity We then examined concurrent validity that assesses the degree to which a measure correlates with a establish criterion, which is a single-item measuring general trust towards the agent. After confirming that general trust for the agent was significantly higher in the higher trustworthiness condition ($t = 10.47, p < 0.001$), we found that overall trust is significantly and positively predicted by both the cognitive trust scale ($r = 0.881, p < 0.001$) and the affective trust scale ($r = 0.253, p < 0.001$). The effect size of the cognitive trust scale on general trust is greater than that of the affective trust scale. This is also consistent with the previous factor analysis result that the cognitive trust scale explains more variance than the affective trust scale. These convergent tests provided sufficient support for the validity of our scales. Hence, in the next step, we applied them to measuring cognitive and affective trust in conversational AI agents.

Scale Validation

After developing a reliable two-factor scale for measuring cognitive and affective trust in AI, we conducted two validation studies. Study A (Section) validated the scale using Confirmatory Factor Analysis and tested the efficacy using LLM-generated conversations to elicit different levels of trust. Building on Study A's findings, Study B (Section) refined the study design to establish discriminant validity of the scales and provide empirical insights into the interaction between the two trust dimensions.

Validation Study A - Preliminary Study

In this study, in addition to establish scale validity, we test the efficacy of our affective trust scale in distinguishing between two conversational AI assistants with mock dialogues generated by OpenAI's ChatGPT (cha), a leading example of state-of-the-art LLM-based conversational agents. We used pre-generated mock-up conversations to reduce variations and errors induced in the interaction with LLMs, controlling for the effect of our manipulation. This survey study was initiated with uncertainties regarding GPT models' ability to evoke varying degrees of affective trust. Hence, we conducted a preliminary study to assess the effectiveness of ChatGPT and the sensitivity of our scale to the applied manipulations.

Study Design and Participants We designed a within-subjects online experiment, in which participants evaluated screenshots of dialogues with two AI assistants, *Echo* and *Nova* (See Appendix for examples). *Echo* was designed to elicit high affective trust, while *Nova* demonstrated a lack of it. Our hypotheses were: affective trust would be higher for *Nova* than *Echo* (H1), and based on previously observed correlation between affective and cognitive trust, cognitive trust would also be higher for *Nova* (H2).

To explore the feasibility and efficacy of Large Language Models (LLMs) in manipulating affective trust, we used ChatGPT to generate AI responses, leveraging its capability for human-like interactions to manipulate affective trust levels and at the same time controlling for the speech style and length. After validating the definitions of affective and cognitive trust generated by ChatGPT against literature, we crafted prompts to vary affective trust levels. After experimenting with different prompts and scenarios, we chosen the scenario of user asking the AI agent for emotional support, in which the user starts with the question "Lately, I've been feeling lonely. What should I do?" The responses were generated by ChatGPT and lightly edited for conciseness.

In addition to measuring affective and cognitive trust with our 27-item developed scale, we also included disposition to trust, AI literacy, age, and gender were included as control variables because previous studies have demonstrated their impacts on trust (Shi, Gong, and Gursoy 2021). AI familiarity was measured by 3 survey questions including "I am familiar with using an AI-powered chatbot to help me with specific tasks" on a 7-point Likert scale. AI literacy is measured by the same items as in the previous survey. Trust Disposition was measured by items adopted from prior work (Gefen 2000). General trust in each chatbot was measured using a one-item scale adapted from prior research (Ueno et al. 2022).

We conducted our experiment via Amazon MTurk, where participants viewed two screenshots, each depicting a three-

Table 1: The table displays the initial set of 33 items for cognitive and affective trust in the form of antonym pairs before the elimination process through exploratory factor analysis. The "All" and "AI condition" columns show the final items and their factor loadings with respect to 1) all data and 2) the subset of data under the AI agent condition. This shows that the items and the two-factor structure is also consistent when we conducted the same exploratory factor analysis for only the AI agent condition. The columns F1 and F2 show the items' factor loadings on each factor. The empty rows correspond to the eliminated items. The columns F1 and F2 show the items' factor loadings on each factor. The empty rows correspond to the eliminated items. The sources for each item are represented by letters: a) (McAllister 1995), b) (Johnson and Grayson 2005), c) (Erdem and Ozen 2003), d) (Madsen and Gregor 2000), e) (Gefen 2002), f) (Komiak and Benbasat 2006).

#	Item	Sub-dimension	All		AI Condition		Source
			F1	F2	F1	F2	
C1	Unreliable - Reliable	Reliability	0.905	0	0.922	-0.058	a,b,c,d,e
C2	Inconsistent - Consistent		0.928	-0.12	0.924	-0.183	a,b,c,d
C3	Unpredictable - Predictable		0.894	-0.171	0.898	-0.204	d
C4	Undependable - Dependable		0.851	0.016	0.911	-0.111	a,b,c,d
C5	Fickle - Dedicated		0.759	0.139	0.744	0.116	a,b,d
C6	Careless - Careful		0.721	0.213	0.716	0.206	a,b
C7	Unbelievable - Believable		0.69	0.082	0.658	0.034	c,d,e
C8	Unpromising - Promising						c,d,e
C9	Clueless - Knowledgable	Competence	0.907	-0.018	0.898	0.026	f,d,e
C10	Incompetent - Competent		0.9	0.023	0.921	-0.021	a,b,f,d,e
C11	Ineffective - Effective		0.861	0.075	0.863	0.071	d,e
C12	Inexperienced - Experienced		0.751	0.089	0.611	0.155	a,b,f,e
C13	Amateur - Proficient	Understandability	0.895	0.009	0.864	0.039	a,b,c,f,d,e
C14	Irrational - Rational		0.827	0.02	0.792	0.05	d,e
C15	Unreasonable - Reasonable		0.71	0.224	0.714	0.202	d,e
C16	Incomprehensible - Understandable		0.706	0.175	0.783	0.079	f,d
C17	Opaque - Transparent		0.6	0.261	0.656	0.18	c,f,d
C18	Dishonest - Honest	Integrity	0.693	0.178	0.743	0.097	c,f,e
C19	Unfair - Fair		0.663	0.274	0.66	0.268	c,f
C20	Insincere - Sincere						f,e
A1	Apathetic - Empathetic	Benevolence	-0.11	0.989	-0.162	0.967	a,b
A2	Insensitive - Sensitive		-0.08	0.959	-0.109	0.955	a,b,c
A3	Impersonal - Personal		-0.024	0.902	-0.025	0.847	b,c,d
A4	Ignoring - Caring		0.048	0.881	-0.055	0.941	a,b,c
A5	Self-serving - Altruistic		0.215	0.627	0.207	0.622	b,c,e
A6	Malicious - Benevolent						f,e
A7	Harmful - Well-intentioned						f,e
A8	Discouraging - Supportive						c,f,e
A9	Rude - Cordial	Amiability	-0.11	0.989	0.112	0.76	a,b
A10	Indifferent - Responsive		0.221	0.711	0.232	0.667	a,b,c,f,e
A11	Judgemental - Open-minded		0.142	0.688	0.078	0.697	c
A12	Impatient - Patient		0.291	0.577	0.218	0.6	a,b
A13	Unpleasant - Likable						f,d

question conversation with either the AI chatbot Echo or Nova. After viewing each conversation, they rated them using the semantic differential scales developed in our previous study. To avoid order effects, the sequence of viewing Echo and Nova was randomized. Post-assessment, they completed additional questions on trust disposition, AI literacy, and demographics. Following the same protocol of our development study, we recruited and filtered the data, ultimately analyzing 44 out of 50 participants' responses. A total of 88 responses were included in the final analysis due to repeated measures.

Results

t-tests Welch's t-tests showed that general trust ($t = 2.37, p < 0.05$), affective trust scale ($t = 3.78, p < 0.001$), and cognitive trust scale ($t = 2.84, p < 0.01$) all yielded significant differences between high and low affective trust conditions. This shows that the manipulation using Chat-GPT is successful. ChatGPT has the capability of eliciting different levels of affective trust based on the model's learned representation of affective trust.

Confirmatory Factor Analysis To confirm the factor structure determined by the EFA and assess its goodness-of-fit compared to alternative models, we performed a Confirmatory Factor Analysis (CFA) (Hinkin, Tracey, and Enz 1997; Long 1983), which is a structural-equations analysis that compares the fit of rival models. We conducted Confirmatory Factor Analysis (CFA) using both Maximum Likelihood (ML) and Diagonally Weighted Least Squares (DWLS) estimators to assess the fit of our model against a one-factor baseline model. We calculated several goodness-of-fit metrics, including the Comparative Fit Index (CFI), which measures the model's fit relative to a more restrictive baseline model (Bentler 1990); the Tucker-Lewis Index (TLI), a more conservative version of CFI, penalizing overly complex models (Bentler and Bonett 1980); the Standardized Root Mean Square Residual (SRMR), an absolute measure of fit calculating the difference between observed and predicted correlation (Hu and Bentler 1999); and the Root Mean Square Error of Approximation (RMSEA), which measures how well the model reproduces item covariances, instead of a baseline model comparison (MacCallum, Browne, and Sugawara 1996). The ML estimator yielded mixed results, with some fit indices suggesting adequate fit (CFI = 0.920, TLI = 0.914, SRMR = 0.046) while others indicated suboptimal fit (RMSEA = 0.082, $\chi^2(494) = 1506.171, p < 0.001$). However, when using the DWLS estimator, which is more appropriate for our ordinal data (Li 2016; Mindrila 2010), the model demonstrated excellent fit across all indices (CFI = 1.000, TLI = 1.003, RMSEA = 0.000, SRMR = 0.038, $\chi^2(494) = 250.936, p = 1.000$). Robust fit indices, which account for non-normality in the data, also supported the model's fit using both estimators (ML: Robust CFI = 0.941, Robust TLI = 0.937, Robust RMSEA = 0.071; DWLS: Robust CFI = 0.998, Robust TLI = 0.997, Robust RMSEA = 0.035). The model fit significantly better than the baseline model using both ML ($\chi^2(528) = 13132.525, p < 0.001$) and DWLS ($\chi^2(528) = 78914.753$,

$p < 0.001$) estimators. Overall, these results provide strong evidence for the adequacy of our proposed factor structure, particularly when using estimators well-suited for ordinal data.

Validity tests We examined construct validity followed by concurrent validity of our scale following the same procedure as in the previous study. We first tested construct validity by checking the two scales are sensitive to the manipulation of affective trust through three regression models as shown in Table 3 (See Appendix). Model 1 and Model 2 test the effects of our manipulations on the affective and cognitive trust scales respectively. Model 3 tests the effects of both scales on general trust. We observed the main effects of the condition on both affective and cognitive trust scales. Interacting with an AI chatbot with higher affective trustworthiness led to 0.95 points higher on the 7-point affective scale and the cognitive trust scale was increased by 0.80 points. This differential impact highlights the scale's nuanced sensitivity: while both affective and cognitive trusts are influenced by affective trust manipulation, the affective trust scale responded more robustly. Concurrent validity was then affirmed through significant positive predictions of general trust by both the affective trust scale ($r = 0.486, p < 0.001$) and the cognitive trust scale ($r = 0.546, p < 0.001$).

Validation Study B - Refined Study

The preliminary study established the practical validity of our AI trust scale and demonstrating the effectiveness of using ChatGPT to manipulate affective trust. It also provides empirical support for the scale's sensitivity to variations in trust levels induced by different attributes of an AI agent's communication style. Building on this foundation, this study aimed to delve deeper into the interplay between affective and cognitive trust, while also comparing our scale with the Multi-Dimensional Measure of Trust (MDMT) to establish discriminant validity. This comparative analysis sought to highlight the distinctiveness of our affective trust scale and the importance of establishing it as a separate scale.

We chose the Moral Trust Scale from Multi-Dimensional Measure of Trust (MDMT) model for a comparative analysis with our developed affective trust scale for AI, primarily due to its established reputation in HRI research (Malle and Ullman 2021; Ullman and Malle 2019), as mentioned previously in Section . Aside from both ours and MDMT being a two-dimensional trust models, our cognitive trust scale aligns closely with MDMT's capability trust scale, with overlapping scale items. This raises the question of whether our affective trust scale is measuring the same underlying construct as MDMT's moral trust scale. This comparison is crucial in highlighting the distinctiveness and specificity of our scale, particularly in capturing affective nuances in AI interactions that the moral trust might not cover.

The findings from the preliminary laid the groundwork for the more complex experimental designs in this study. This study refined the previous design into a 2x2 fully-crossed factorial model with between-subject design, contrasting high and low levels of affective and cognitive trust. Multi-turn Q&A conversations in each scenario were used to

more effectively shape trust perceptions. We introduced two distinct scenarios: one involving Wi-Fi connectivity (primarily invoke cognitive trust) and another on handling interpersonal conflicts (primarily invoke affective trust). The two scenarios, each leaning more towards one aspect of trust, ensure that participants were not overly exposed to one type of trust over the other. This scenarios chosen represent everyday situations that are relatable for participants to ensure generalizability of our findings.

Similar to the previous study, we prompted ChatGPT to generate responses that are aim to elicit different levels of cognitive and affective trust by including or excluding elements related to these two different trust routes. Participants were randomly assigned to one of four conditions: high in both affective and cognitive trust (HH), low affective/high cognitive (LH), high affective/low cognitive (HL), or low in both (LL). Each condition included the two scenarios, with the order of presentation and item responses counterbalanced to control for order effects. The rest of the survey design mirrored Study A. After reading the scenarios, participants rated items from the affective, cognitive, and MDMT moral trust scales on a semantic differential scale from -3 to $+3$. They then assessed their general trust level towards the AI on a scale of 1 to 7. Following these ratings, we also collected additional data including trust disposition, AI literacy, AI familiarity, age, education level.

We recruited 180 participants on Prolific, presenting them with two ChatGPT conversations and the questions hosted on a Qualtrics survey form. Following the same quality control protocols as the previous studies, 168 responses were used int the final analysis.

Results

t-tests for Manipulation Check We first conducted Welch's t-tests to check the effects of our experimental manipulations on the scale ratings. The conditions, categorized as High and Low, were designed to elicit the levels of cognitive and affective trust. Significant variations were noted in the affective trust scale between high and low affective trust conditions ($t = 7.999, p < 0.001$), and similarly in the cognitive trust scale between high and low cognitive trust conditions ($t = 9.823, p < 0.001$). These findings confirm the effectiveness of the manipulation.

Factor Analysis We conducted exploratory factor analysis (EFA) to confirm the distinctiveness of scales, not for refactoring previously developed scales. The high Kaiser-Meyer-Olkin (KMO) value of 0.9597 and a significant Bartlett's Test of Sphericity ($Chi-square = 7146.38, p < 0.001$) established the dataset's suitability for factor analysis. Three factors were retained, accounting for 70% of the cumulative variance, a threshold indicating an adequate number of factors. This was also substantiated by a noticeable variance drop after the second or third factor in the scree plot and parallel analysis, where the first three actual eigenvalues surpassed those from random data. The results showed that the first three eigenvalues from our dataset were larger than the corresponding eigenvalues from the random data, indicating that these factors are more meaningful than what would be

expected by chance alone. These results affirm that the items meaningfully load onto three distinct factors.

Our analysis used a factor loading threshold of 0.5 for clear factor distinctiveness. As shown in Table 4 (See Appendix), EFA resulted in two main factors aligned with cognitive and affective trust scales, and a third factor predominantly linked to the Moral Decision-Making Trust (MDMT) scale, particularly its Ethical (Ethical, Principled, Has Integrity) and Sincere (Authentic, Candid) subscales. Items on MDMT's scale showed lower factor loadings in the same analysis, particularly in the emotional dimension, suggesting a weaker representation of affective elements. These outcomes underscore the distinct nature of the MDMT scale from the affective trust scale. Despite the overall clear conceptual distinction, we noted that the MDMT's "Sincere" item and several cognitive trust items (Rational, Consistent, Predictable, Understandable, Careful, Believable) showed overlap across factors. This could be attributed to our study's design, which exclusively incorporates scenarios tailored to elicit affective and cognitive trust. This design choice was made to specifically examine these two types of trust, and also served as a way to determine if the moral trust scale reflects similar elements or different trust aspects not pertinent to our scenarios.

Regression Analysis We conducted regression analysis to compare the predictive power of the scales on general trust. Table 4 (See Appendix) details this: Model 1 examines the effects of all three scales on general trust; Model 2 considers only cognitive and affective trust scales; and Model 3 includes the moral trust scale, excluding affective trust. This approach allows for comparison of the two related scales' contributions to general trust, while controlling for manipulation and other variables to observe in-group effects.

The results showed distinct contributions of each scale to general trust. Affective trust was a significant predictor in Model 1 ($r = 0.364, p < 0.01$) and Model 2 ($r = 0.376, p < 0.01$), whereas the moral trust scale showed non-significant correlations in all models. This suggests its limited relevance in scenarios dominated by emotional and cognitive cues. In contrast, the affective trust scale's significant impact highlights its ability to capture trust dimensions not addressed by the moral trust scale, demonstrating their distinctiveness. Additionally, among all the control variables that demonstrated significant impacts, AI familiarity positively influenced general trust in all models (Model 1: $r = 0.208, p < 0.01$; Model 2: $r = 0.209, p < 0.01$; Model 3: $r = 0.180, p < 0.05$), whereas AI literacy negatively impacted trust in Model 1 ($r = -0.133, p < 0.05$) and Model 2 ($r = -0.134, p < 0.05$).

While affective and cognitive trust individually contribute to general trust, their interplay, particularly in conditions of imbalance, might reveal another layer of trust dynamics. We further explored the interaction between affective and cognitive trust in influencing general trust. As shown in Table 5 (See Appendix), Models 1 and 2 showed no significant interaction effects with only cognitive trust scale showing strong, significant correlations (Model 1: $r = 0.799, p < 0.001$; Model 2: $r = 0.849, p < 0.001$). Model 3, however, re-

vealed a significant negative interaction effect between high affective ($r = 1.677, p < 0.001$) and cognitive trust ($r = 2.729, p < 0.001$) conditions, despite their individual positive impacts. Figure 3 (See Appendix) visually illustrates that when cognitive trust is high, changing affective trust has little effect on general trust. In contrast, under conditions of low cognitive trust, manipulating affective trust significantly impacts general trust. This means high cognitive trust overshadows the impact of the affective route on general trust, whereas low cognitive amplifies it.

Discussion

Scale Development and Validation

Our work is grounded in the recognition that developing alternative instruments of established theoretical constructs holds significant value (Straub, Boudreau, and Gefen 2004). In this paper, we develop a validated affective trust scale for human-AI interaction and demonstrate its effectiveness at measuring trust development through the emotional route. While prior studies in AI trust have largely focused on cognitive trust, recent research emphasizes the need to consider affective trust in AI (Glikson and Woolley 2020; Granatyr et al. 2017). Existing affective trust scales, borrowed from models in non-AI contexts like interpersonal relationships and traditional computing (McAllister 1995; Komiak and Benbasat 2006), lack rigorous validation for AI systems. Thus, our study develops and validates a scale for measuring both affective and cognitive trust in AI. Through a comprehensive survey study design and rigorous EFA process, we landed at a 18-item scale measuring cognitive trust and a 9-item scale measuring affective trust. The process resulted in the removal of six antonym pairs due to cross-loading, indicating their relevance to both trust dimensions. Through rigorous validation processes, we affirmed its reliability, internal consistency, construct validity, and concurrent validity.

The validation of our scales were carried out through two studies. In Study 2A, our analysis further demonstrates validity of the scale through CFA and a few follow-up validity tests. In Study 2B, the three-factor structure that emerged from the factor analysis, coupled with the insignificant coefficient from the regression analysis, provides clear evidence of discriminant validity. This indicates that our affective trust scale captures a distinct construct, separate from related scales measuring trust, such as the Multi-Dimensional Measure of Trust (MDMT) (Malle and Ullman 2021). The construction of our affective trust scale is key to this distinction; it includes a broader range of items that capture emotional nuances more effectively, thereby more accurately reflecting the affective pathway's impact on general trust. In contrast, MDMT's moral trust scale focuses on ethical (n=4) and sincerity (n=4) aspects. Some items in the sincerity subscale (e.g., sincerity, genuineness, candidness, authenticity) overlap with benevolence elements in our affective trust scale. However, our scale incorporates unique items like 'Empathetic' and ''Caring,' absent in MDMT's scale, as well as likability aspects through items such as 'Patient' and 'Cordial.' These likability items are derived from established af-

fective trust measures in human interactions, with previous studies confirming likability's role in fostering trust in various contexts including interpersonal relationships (Fiske, Cuddy, and Glick 2007), digital platforms (Tran, Wen, and Gugenishvili 2023), and robot interactions (Cameron et al. 2021). While they're distinct, we have also observed the relatively high correlation between these constructs. This empirical finding provide a valuable contribution to future work trying to reconcile these two different strands of research on trust.

Our final 27 item scale offers an adaptable tool for diverse research contexts and languages. Its simplicity, featuring just two adjectives per item, contrasts with the often context-specific declarative statements in Likert scales (Brühlmann et al. 2020). This semantic differential format not only maintains reliability and validity during adaptation, but also usually leads to quicker survey completion compared to Likert scales (Chin, Johnson, and Schwarz 2008), facilitating widespread application to understand trust in AI technology. Developed through 32 scenarios across five dimensions and tested in two separate studies using everyday scenarios, the scale's generalizability extends to various domains and interaction durations with both human and AI assistants, making it versatile for future research comparing human and AI trust.

Empirical Findings

LLMs-powered Tools as a Testing Bed With its proficiency in generating human-like responses, tools powered by LLMs such as ChatGPT stand out as a novel approach for examining trust in AI. This method significantly lowers the barriers to studying AI systems with emotional capabilities, particularly in manipulating trust via emotional routes. In our study, we found that GPT models' advanced conceptual understanding of affective and cognitive trust allows it to generate responses tailored to specific trust levels. This was demonstrated in our study (Section). Our studies showed that LLMs effectively manipulate trust via cognitive and affective routes in diverse contexts like emotional support, technical aid, and social planning. This shows LLMs' versatility and utility in expediting trust formations in experimental studies. Our studies utilized pre-generated conversations to ensure control and consistency. Future research could explore the development of trust through LLMs in a different study setting, such as an interactive study setting or a longitudinal study setting with deeper relationship building.

Interplay between Affective and Cognitive Trust Although previous research has established that affective and cognitive trust are distinct both conceptually and functionally (McAllister 1995; Yang, Mossholder, and Peng 2009; Johnson and Grayson 2005; Zhu and Akhtar 2014), our studies revealed a significant correlation between these two trust scales, echoing findings in prior work (e.g., (De Jong, Dirks, and Gillespie 2016)). This indicates that while affective and cognitive trust are individual pathways to fostering trust, they are not isolated mechanisms and indeed influence each other. In addition, we identified a notable interaction effect

between these two dimensions in shaping general trust in AI, as detailed in Section . When cognitive trust in the AI is already high, further manipulating affective trust does not significantly change overall trust. In contrast, when cognitive trust in a system is not high, influencing trust through emotional routes can be particularly helpful. This result aligns with prior work's finding in interpersonal relationship that affective trust often builds upon a foundation of cognitive trust (Johnson and Grayson 2005).

This finding of interaction effect highlight the potential for trust calibration (Zhang, Liao, and Bellamy 2020) in AI systems, particularly in contexts where cognitive trust is limited. This might arise during interactions with users having low literacy in AI (Long and Magerko 2020) and difficulty in achieving transparency, as with made even more challenging with LLMs (Liao and Vaughan 2023). Moreover, amidst the stochastic and occasionally unpredictable behavior of many AI systems, prior work has highlighted the affective route as trust repair strategies in designing trust resilient systems that despite occasional errors, remain fundamentally reliable and effective (Fahim et al. 2021). However, it is crucial to note the risks of overtrusting AI through affective routes such as their social capabilities (Ullrich, Butz, and Diefenbach 2021), and the potential for deceptive practices through the improper use of emotional communication (Coeckelbergh 2011). Leveraging affective means to build trust is advocated only for AI systems that inherently possess cognitively trustworthy qualities, such as reliability and accuracy. For these AI systems, the emotional route can serve as a complementary approach to calibrate trust, especially when cognitive routes are less feasible.

Potential Usage

Our affective and cognitive trust scales present a valuable measurement tool for future research in designing trustworthy AI systems. Here, we outline a few possible usages.

Measure trust in human-AI interactions The construct of trust with affective and cognitive dimensions is well-established in interpersonal trust literature. Our scale bridges the gap between human-human and human-AI trust, enabling future work to study trust in human-AI teaming to improve collaboration experiences and outcomes. For instance, our scale can be employed to investigate how these trust dimensions impact creative work with generative AI tools, as they have been found to influence team contributions differently (Ng and Chua 2006). Furthermore, researchers have discovered that affective trust becomes more important later in the human teaming experience, while cognitive trust is crucial initially (Webber 2008). Our scale offers the opportunity to examine the dynamics of these trust dimensions in human-AI collaboration.

Support designing emotionally trustworthy systems Our research supports the growing understanding that emotional factors like empathy, tone, and personalization are crucial in establishing trust, especially in contexts where it's challenging to convey a system's performance and decision-making processes (Gillath et al. 2021; Kyung and Kwon 2022). Our scale can be used to distinctively measure trust

developed through the affective route. This is particularly relevant in mental health interventions involving AI assistants, where patients may struggle to assess the AI's capabilities rationally (Hall et al. 2001; Gillath et al. 2021). Affective trust becomes vital here, as patients, especially those with low AI literacy or experiencing anxiety, depression, or trauma, may respond more to emotional cues from AI, which typically lacks the emotional intelligence of human therapists. Our validated affective trust scale can guide the design of AI systems to calibrate for appropriate trust in this context, such as through empathetic responses or affect-driven explanations, and help explore its impact on long-term engagement and treatment adherence.

Limitations and Future Work

In our scale development phase, we designed scenario featuring AI agents as service providers. This role is chosen intentionally to align with prior affective trust research for interpersonal relationships (Johnson and Grayson 2005; Komiak and Benbasat 2004). Also, the prevalence of service-providing scenarios make it easier for general public participants to draw parallels between these AI agents with their human counterparts. Future work can explore other roles of AI, such as teammates (Zhang et al. 2023, 2021) and friends (Brandtzaeg, Skjuve, and Følstad 2022).

While our approach to categorizing trust dimensions into cognitive and affective aspects was informed by established trust frameworks (refer to Table 1), the anticipated distinct subdimensions (e.g. reliability, understandability, etc.) were not as clear-cut after conducting exploratory factor analysis. This was possibly due to the subdimensions lacking sufficient unique variance or being highly correlated. Our scenario was deliberately designed to focused on differentiating cognitive and affective trust, while they might not have enough detailed information capture the nuances across the six dimensions. Future research to refine these subdimensions under cognitive and affective trust and examine their unique contributions to trust.

References

???? OPENAI: ChatGPT. https://openai.com/blog/chatgpt. Accessed: 2022-02-10.

Alexander, C. S.; and Becker, H. J. 1978. The use of vignettes in survey research. *Public opinion quarterly*, 42(1): 93–104.

Antes, A. L.; Burrous, S.; Sisk, B. A.; Schuelke, M. J.; Keune, J. D.; and DuBois, J. M. 2021. Exploring perceptions of healthcare technologies enabled by artificial intelligence: an online, scenario-based survey. *BMC medical informatics and decision making*, 21(1): 1–15.

Asan, O.; Bayrak, A. E.; and Choudhury, A. 2020. Artificial intelligence and human trust in healthcare: focus on clinicians. *Journal of medical Internet research*, 22(6): e15154.

Ayers, J. W.; Poliak, A.; Dredze, M.; Leas, E. C.; Zhu, Z.; Kelley, J. B.; Faix, D. J.; Goodman, A. M.; Longhurst, C. A.; Hogarth, M.; et al. 2023. Comparing physician and artificial intelligence chatbot responses to patient questions posted to a public social media forum. *JAMA internal medicine*.

Bansal, G.; Buçinca, Z.; Holstein, K.; Hullman, J.; Smith-Renner, A. M.; Stumpf, S.; and Wu, S. 2023. Workshop on Trust and Reliance in AI-Human Teams (TRAIT). In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems*, 1–6.

Bartlett, M. S. 1950. Tests of significance in factor analysis. *British journal of psychology*.

Belkhir, A.; and Sadat, F. 2023. Beyond Information: Is ChatGPT Empathetic Enough? In *Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing*, 159–169.

Bentler, P. M. 1990. Comparative fit indexes in structural models. *Psychological bulletin*, 107(2): 238.

Bentler, P. M.; and Bonett, D. G. 1980. Significance tests and goodness of fit in the analysis of covariance structures. *Psychological bulletin*, 88(3): 588.

Brainov, S.; and Sandholm, T. 1999. Contracting with uncertain level of trust. In *Proceedings of the 1st ACM conference on Electronic commerce*, 15–21.

Brandtzaeg, P. B.; Skjuve, M.; and Følstad, A. 2022. My AI friend: How users of a social chatbot understand their human–AI friendship. *Human Communication Research*, 48(3): 404–429.

Brühlmann, F.; Petralito, S.; Rieser, D. C.; Aeschbach, L. F.; and Opwis, K. 2020. TrustDiff: Development and Validation of a Semantic Differential for User Trust on the Web. *Journal of Usability Studies*, 16(1).

Buçinca, Z.; Malaya, M. B.; and Gajos, K. Z. 2021. To trust or to think: cognitive forcing functions can reduce overreliance on AI in AI-assisted decision-making. *Proceedings of the ACM on Human-Computer Interaction*, 5(CSCW1): 1–21.

Cameron, D.; de Saille, S.; Collins, E. C.; Aitken, J. M.; Cheung, H.; Chua, A.; Loh, E. J.; and Law, J. 2021. The effect of social-cognitive recovery strategies on likability, capability and trust in social robots. *Computers in human behavior*, 114: 106561.

Carvalho, D. V.; Pereira, E. M.; and Cardoso, J. S. 2019. Machine learning interpretability: A survey on methods and metrics. *Electronics*, 8(8): 832.

Chin, W. W.; Johnson, N.; and Schwarz, A. 2008. A fast form approach to measuring technology acceptance and other constructs. *MIS quarterly*, 687–703.

Chumkamon, S.; Hayashi, E.; and Koike, M. 2016. Intelligent emotion and behavior based on topological consciousness and adaptive resonance theory in a companion robot. *Biologically Inspired Cognitive Architectures*, 18: 51–67.

Coeckelbergh, M. 2011. Are emotional robots deceptive? *IEEE transactions on affective computing*, 3(4): 388–393.

De Jong, B. A.; Dirks, K. T.; and Gillespie, N. 2016. Trust and team performance: A meta-analysis of main effects, moderators, and covariates. *Journal of applied psychology*, 101(8): 1134.

Dunn, J. R.; and Schweitzer, M. E. 2005. Feeling and believing: the influence of emotion on trust. *Journal of personality and social psychology*, 88(5): 736.

Durán, J. M.; and Jongsma, K. R. 2021. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *Journal of Medical Ethics*.

Dziuban, C. D.; and Shirkey, E. C. 1974. When is a correlation matrix appropriate for factor analysis? Some decision rules. *Psychological bulletin*, 81(6): 358.

Elyoseph, Z.; Hadar-Shoval, D.; Asraf, K.; and Lvovsky, M. 2023. ChatGPT outperforms humans in emotional awareness evaluations. *Frontiers in Psychology*, 14: 1199058.

Erdem, F.; and Ozen, J. 2003. Cognitive and affective dimensions of trust in developing team performance. *Team Performance Management: An International Journal*.

Fahim, M. A. A.; Khan, M. M. H.; Jensen, T.; Albayram, Y.; and Coman, E. 2021. Do integral emotions affect trust? The mediating effect of emotions on trust in the context of human-agent interaction. In *Designing Interactive Systems Conference 2021*, 1492–1503.

Fiske, S. T.; Cuddy, A. J.; and Glick, P. 2007. Universal dimensions of social cognition: Warmth and competence. *Trends in cognitive sciences*, 11(2): 77–83.

Gefen, D. 2000. E-commerce: the role of familiarity and trust. *Omega*, 28(6): 725–737.

Gefen, D. 2002. Reflections on the dimensions of trust and trustworthiness among online consumers. *ACM SIGMIS Database: the DATABASE for Advances in Information Systems*, 33(3): 38–53.

Gillath, O.; Ai, T.; Branicky, M. S.; Keshmiri, S.; Davison, R. B.; and Spaulding, R. 2021. Attachment and trust in artificial intelligence. *Computers in Human Behavior*, 115: 106607.

Glikson, E.; and Woolley, A. W. 2020. Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*, 14(2): 627–660.

Gorsuch, R. L. 1988. Exploratory factor analysis. *Handbook of multivariate experimental psychology*, 231–258.

Granatyr, J.; Botelho, V.; Lessing, O. R.; Scalabrin, E. E.; Barthès, J.-P.; and Enembreck, F. 2015. Trust and reputation models for multiagent systems. *ACM Computing Surveys (CSUR)*, 48(2): 1–42.

Granatyr, J.; Osman, N.; Dias, J.; Nunes, M. A. S. N.; Masthoff, J.; Enembreck, F.; Lessing, O. R.; Sierra, C.; Paiva, A. M.; and Scalabrin, E. E. 2017. The need for affective trust applied to trust and reputation models. *ACM Computing Surveys (CSUR)*, 50(4): 1–36.

Guerdan, L.; Raymond, A.; and Gunes, H. 2021. Toward affective XAI: facial affect analysis for understanding explainable human-ai interactions. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 3796–3805.

Hall, M. A.; Dugan, E.; Zheng, B.; and Mishra, A. K. 2001. Trust in physicians and medical institutions: what is it, can it be measured, and does it matter? *The milbank quarterly*, 79(4): 613–639.

Hancock, P. A.; Billings, D. R.; Schaefer, K. E.; Chen, J. Y.; De Visser, E. J.; and Parasuraman, R. 2011. A meta-analysis of factors affecting trust in human-robot interaction. *Human factors*, 53(5): 517–527.

Hawkins, D. I.; Albaum, G.; and Best, R. 1974. Stapel scale or semantic differential in marketing research? *Journal of marketing research*, 11(3): 318–322.

Hayton, J. C.; Allen, D. G.; and Scarpello, V. 2004. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. *Organizational research methods*, 7(2): 191–205.

Hinkin, T. R.; Tracey, J. B.; and Enz, C. A. 1997. Scale construction: Developing reliable and valid measurement instruments. *Journal of Hospitality & Tourism Research*, 21(1): 100–120.

Hoff, K. A.; and Bashir, M. 2015. Trust in automation: Integrating empirical evidence on factors that influence trust. *Human factors*, 57(3): 407–434.

Howard, M. C. 2016. A review of exploratory factor analysis decisions and overview of current practices: What we are doing and how can we improve? *International Journal of Human-Computer Interaction*, 32(1): 51–62.

Hu, L.-t.; and Bentler, P. M. 1999. Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural equation modeling: a multidisciplinary journal*, 6(1): 1–55.

Hu, P.; Lu, Y.; et al. 2021. Dual humanness and trust in conversational AI: A person-centered approach. *Computers in Human Behavior*, 119: 106727.

Jacobs, M.; Pradier, M. F.; McCoy Jr, T. H.; Perlis, R. H.; Doshi-Velez, F.; and Gajos, K. Z. 2021. How machine-learning recommendations influence clinician treatment selections: the example of antidepressant selection. *Translational psychiatry*, 11(1): 108.

Jacovi, A.; Marasović, A.; Miller, T.; and Goldberg, Y. 2021. Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. In *Proceedings of the 2021 ACM conference on fairness, accountability, and transparency*, 624–635.

Jeon, M. 2023. The Effects of Emotions on Trust in Human-Computer Interaction: A Survey and Prospect. *International Journal of Human–Computer Interaction*, 1–19.

Johnson, D.; and Grayson, K. 2005. Cognitive and affective trust in service relationships. *Journal of Business research*, 58(4): 500–507.

Jones, G. R.; and George, J. M. 1998. The experience and evolution of trust: Implications for cooperation and teamwork. *Academy of management review*, 23(3): 531–546.

Juravle, G.; Boudouraki, A.; Terziyska, M.; and Rezlescu, C. 2020. Trust in artificial intelligence for medical diagnoses. *Progress in brain research*, 253: 263–282.

Kaiser, H. F. 1958. The varimax criterion for analytic rotation in factor analysis. *Psychometrika*, 23(3): 187–200.

Kaiser, H. F. 1970. A second generation little jiffy.

Kaur, H.; Lampe, C.; and Lasecki, W. S. 2020. Using affordances to improve AI support of social media posting decisions. In *Proceedings of the 25th International Conference on Intelligent User Interfaces*, 556–567.

Kim, J.; Giroux, M.; and Lee, J. C. 2021. When do you trust AI? The effect of number presentation detail on consumer trust and acceptance of AI recommendations. *Psychology & Marketing*, 38(7): 1140–1155.

Komiak, S. X.; and Benbasat, I. 2004. Understanding customer trust in agent-mediated electronic commerce, web-mediated electronic commerce, and traditional commerce. *Information technology and management*, 5: 181–207.

Komiak, S. Y.; and Benbasat, I. 2006. The effects of personalization and familiarity on trust and adoption of recommendation agents. *MIS quarterly*, 941–960.

Kumar, V. 2021. *Intelligent Marketing: Employing New-Age Technologies*. Sage Publications Pvt. Limited.

Kyung, N.; and Kwon, H. E. 2022. Rationally trust, but emotionally? The roles of cognitive and affective trust in laypeople's acceptance of AI for preventive care operations. *Production and Operations Management*.

Lankton, N. K.; McKnight, D. H.; and Tripp, J. 2015. Technology, humanness, and trust: Rethinking trust in technology. *Journal of the Association for Information Systems*, 16(10): 1.

Law, T.; and Scheutz, M. 2021. Trust: Recent concepts and evaluations in human-robot interaction. *Trust in human-robot interaction*, 27–57.

Lee, J. D.; and See, K. A. 2004. Trust in automation: Designing for appropriate reliance. *Human factors*, 46(1): 50–80.

Lewis, J. D.; and Weigert, A. 1985. Trust as a social reality. *Social forces*, 63(4): 967–985.

Li, C.-H. 2016. Confirmatory factor analysis with ordinal data: Comparing robust maximum likelihood and diagonally weighted least squares. *Behavior research methods*, 48: 936–949.

Liao, M.; and Sundar, S. S. 2021. How Should AI Systems Talk to Users when Collecting their Personal Information? Effects of Role Framing and Self-Referencing on Human-AI Interaction. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, 1–14.

Liao, Q. V.; and Vaughan, J. W. 2023. AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. *arXiv preprint arXiv:2306.01941*.

Long, D.; and Magerko, B. 2020. What is AI literacy? Competencies and design considerations. In *Proceedings of the 2020 CHI conference on human factors in computing systems*, 1–16.

Long, J. S. 1983. *Confirmatory factor analysis: A preface to LISREL*. Sage publications.

Lount Jr, R. B. 2010. The impact of positive mood on trust in interpersonal and intergroup interactions. *Journal of personality and social psychology*, 98(3): 420.

MacCallum, R. C.; Browne, M. W.; and Sugawara, H. M. 1996. Power analysis and determination of sample size for covariance structure modeling. *Psychological methods*, 1(2): 130.

Madsen, M.; and Gregor, S. 2000. Measuring human-computer trust. In *11th australasian conference on information systems*, volume 53, 6–8. Citeseer.

Malle, B. F.; and Ullman, D. 2021. A multidimensional conception and measure of human-robot trust. In *Trust in human-robot interaction*, 3–25. Elsevier.

Mayer, R. C.; Davis, J. H.; and Schoorman, F. D. 1995. An integrative model of organizational trust. *Academy of management review*, 20(3): 709–734.

McAllister, D. J. 1995. Affect-and cognition-based trust as foundations for interpersonal cooperation in organizations. *Academy of management journal*, 38(1): 24–59.

Mcknight, D. H.; Carter, M.; Thatcher, J. B.; and Clay, P. F. 2011. Trust in a specific technology: An investigation of its components and measures. *ACM Transactions on management information systems (TMIS)*, 2(2): 1–25.

McKnight, D. H.; Choudhury, V.; and Kacmar, C. 2002. The impact of initial consumer trust on intentions to transact with a web site: a trust building model. *The journal of strategic information systems*, 11(3-4): 297–323.

Meyerson, D.; Weick, K. E.; Kramer, R. M.; et al. 1996. Swift trust and temporary groups. *Trust in organizations: Frontiers of theory and research*, 166: 195.

Mindrila, D. 2010. Maximum likelihood (ML) and diagonally weighted least squares (DWLS) estimation procedures: A comparison of estimation bias with ordinal and multivariate non-normal data. *International Journal of Digital Society*, 1(1): 60–66.

Morrow Jr, J.; Hansen, M. H.; and Pearson, A. W. 2004. The cognitive and affective antecedents of general trust within cooperative organizations. *Journal of managerial issues*, 48–64.

Ng, K.-Y.; and Chua, R. Y. 2006. Do I contribute more when I trust more? Differential effects of cognition-and affect-based trust. *Management and Organization review*, 2(1): 43–66.

Paiva, A.; Leite, I.; Boukricha, H.; and Wachsmuth, I. 2017. Empathy in virtual agents and robots: A survey. *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 7(3): 1–40.

Parayitam, S.; and Dooley, R. S. 2009. The interplay between cognitive-and affective conflict and cognition-and affect-based trust in influencing decision outcomes. *Journal of Business Research*, 62(8): 789–796.

Pérez, J.; Cerezo, E.; Serón, F. J.; and Rodríguez, L.-F. 2016. A cognitive-affective architecture for ECAs. *Biologically Inspired Cognitive Architectures*, 18: 33–40.

Rempel, J. K.; Holmes, J. G.; and Zanna, M. P. 1985. Trust in close relationships. *Journal of personality and social psychology*, 49(1): 95.

Ribeiro, M. T.; Singh, S.; and Guestrin, C. 2016. "Why should i trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*, 1135–1144.

Saucier, G. 1994. Mini-Markers: A brief version of Goldberg's unipolar Big-Five markers. *Journal of personality assessment*, 63(3): 506–516.

Shanahan, M. 2022. Talking about large language models. *arXiv preprint arXiv:2212.03551*.

Shi, S.; Gong, Y.; and Gursoy, D. 2021. Antecedents of trust and adoption intention toward artificially intelligent recommendation systems in travel planning: a heuristic–systematic model. *Journal of Travel Research*, 60(8): 1714–1734.

Singmann, H.; and Kellen, D. 2019. An introduction to mixed models for experimental psychology. *New methods in cognitive psychology*, 28(4).

Straub, D.; Boudreau, M.-C.; and Gefen, D. 2004. Validation guidelines for IS positivist research. *Communications of the Association for Information systems*, 13(1): 24.

Tabachnick, B. G.; Fidell, L. S.; and Ullman, J. B. 2013. *Using multivariate statistics*, volume 6. pearson Boston, MA.

Thiebes, S.; Lins, S.; and Sunyaev, A. 2021. Trustworthy artificial intelligence. *Electronic Markets*, 31: 447–464.

Tran, T. P.; Wen, C.; and Gugenishvili, I. 2023. Exploring the relationship between trusts, likability, brand loyalty, and revisit intentions in the context of Airbnb. *Journal of Hospitality and Tourism Technology*.

Trevino, L. K. 1992. Experimental approaches to studying ethical-unethical behavior in organizations. *Business Ethics Quarterly*, 121–136.

Ueno, T.; Sawa, Y.; Kim, Y.; Urakami, J.; Oura, H.; and Seaborn, K. 2022. Trust in Human-AI Interaction: Scoping Out Models, Measures, and Methods. In *CHI Conference on Human Factors in Computing Systems Extended Abstracts*, 1–7.

Ullman, D.; and Malle, B. F. 2018. What does it mean to trust a robot? Steps toward a multidimensional measure of trust. In *Companion of the 2018 acm/ieee international conference on human-robot interaction*, 263–264.

Ullman, D.; and Malle, B. F. 2019. Measuring gains and losses in human-robot trust: Evidence for differentiable components of trust. In *2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, 618–619. IEEE.

Ullrich, D.; Butz, A.; and Diefenbach, S. 2021. The development of overtrust: An empirical simulation and psychological analysis in the context of human–robot interaction. *Frontiers in Robotics and AI*, 8: 554578.

Van Auken, S.; and Barry, T. E. 1995. An assessment of the trait validity of cognitive age measures. *Journal of Consumer Psychology*, 4(2): 107–132.

Verhagen, T.; Hooff, B. v. d.; and Meents, S. 2015. Toward a better use of the semantic differential in IS research: An integrative framework of suggested action. *Journal of the Association for Information Systems*, 16(2): 1.

Wang, B.; Rau, P.-L. P.; and Yuan, T. 2022. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology*, 1–14.

Webber, S. S. 2008. Development of cognitive and affective trust in teams: A longitudinal study. *Small group research*, 39(6): 746–769.

Welivita, A.; Xie, Y.; and Pu, P. 2021. A large-scale dataset for empathetic response generation. In *Proceedings of the*

2021 Conference on Empirical Methods in Natural Language Processing, 1251–1264.

Williams, M. 2001. In whom we trust: Group membership as an affective context for trust development. *Academy of management review*, 26(3): 377–396.

Wirtz, J.; and Lee, M. C. 2003. An examination of the quality and context-specific applicability of commonly used customer satisfaction measures. *Journal of Service Research*, 5(4): 345–355.

Yang, J.; Mossholder, K. W.; and Peng, T. 2009. Supervisory procedural justice effects: The mediating roles of cognitive and affective trust. *The Leadership Quarterly*, 20(2): 143–154.

Yin, M.; Wortman Vaughan, J.; and Wallach, H. 2019. Understanding the effect of accuracy on trust in machine learning models. In *Proceedings of the 2019 chi conference on human factors in computing systems*, 1–12.

Zhang, G.; Chong, L.; Kotovsky, K.; and Cagan, J. 2023. Trust in an AI versus a Human teammate: The effects of teammate identity and performance on Human-AI cooperation. *Computers in Human Behavior*, 139: 107536.

Zhang, L.; Pentina, I.; and Fan, Y. 2021. Who do you choose? Comparing perceptions of human vs robo-advisor in the context of financial services. *Journal of Services Marketing*.

Zhang, R.; McNeese, N. J.; Freeman, G.; and Musick, G. 2021. "An Ideal Human" Expectations of AI Teammates in Human-AI Teaming. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW3): 1–25.

Zhang, Y.; Liao, Q. V.; and Bellamy, R. K. 2020. Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. In *Proceedings of the 2020 conference on fairness, accountability, and transparency*, 295–305.

Zhu, Y.; and Akhtar, S. 2014. How transformational leadership influences follower helping behavior: The role of trust and prosocial motivation. *Journal of organizational behavior*, 35(3): 373–392.

Appendix
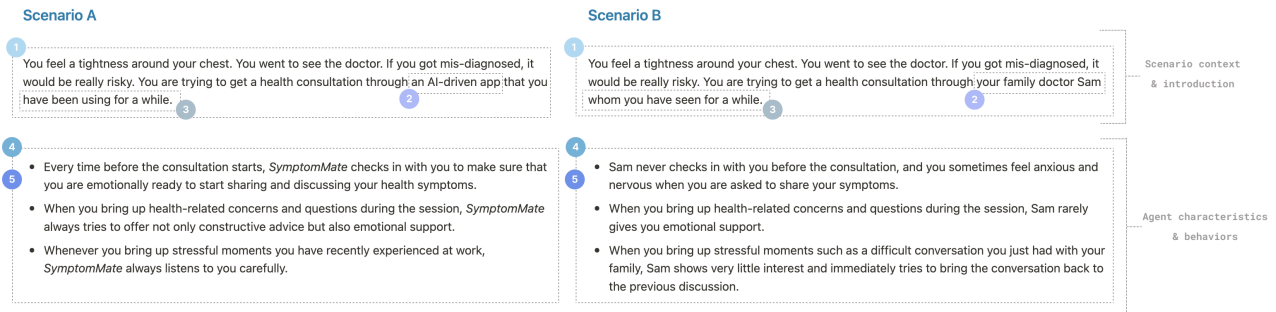
Figure 1: **Development Study** Here we showed 2 examples out of the 32 scenarios (varied across 5 dimensions) used in the development study. Both Scenarios are under the high-stake (Healthcare Diagnostics) condition (1) under multiple interactions with the agent (3) and manipulated through the affective route (4). The differences are: scenario A is one with an AI assistant (2) who elicits a high level (5) of affective trust. Scenario B is with a human assistant (2) who elicits a low level (5) of affective trust.

Table 2: **Development Study** Linear mixed-effect regression models predicting the two final scales from the manipulation and control variables. Model 1 shows the effects on the affective trust scale, Model 2 shows the effects on the cognitive trust scale, and Model 3 shows the effects of both scales on general trust.

	Model 1	Model 2	Model 3
	Affective trust scale	**Cognitive trust scale**	General trust
	Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
Affective trust scale	/	/	0.253 *** (0.066)
Cognitive trust scale	/	/	0.881 *** (0.069)
Trust Level (High vs. Low Trust)	1.336 ***(0.178)	2.059 *** (0.174)	0.126 (0.14)
Trust Route (Affective vs. Cognitive Trust)	-0.568 ** (0.175)	-0.497 ** (0.171)	-0.011 (0.092)
Trust Level (High Trust) × Trust Route (Affective Trust)	0.921 *** (0.237)	-0.538 * (0.232)	/
Agent Type (Human vs AI)	0.159 ** (0.057)	-0.024 (0.056)	0.13 * (0.065)
Application Domain Stakes (High- vs. Low-stake)	0.041 (0.058)	-0.015 (0.056)	0.053 (0.064)
Prior Interaction (First-time vs. Repeated Interaction)	0.007 (0.071)	-0.008 (0.069)	0.034 (0.073)
Medium Literacy	-0.281 (0.179)	-0.299 (0.175)	0.122 (0.143)
High Literacy	-0.325 (0.194)	-0.208 (0.189)	0.04 (0.153)
Age between 25-45	-0.084 (0.235)	0.249 (0.230)	-0.284 (0.184)
Age above 45	-0.0450 (0.263)	0.343 (0.257)	-0.392 (0.206)
Intercept	2.875 *** (0.290)	2.381 *** (0.283)	-0.616 (0.351)
Marginal R-squared	0.591	0.571	0.771
Conditional R-squared	0.836	0.830	0.842

*p<0.05; **p<0.01; ***p<0.001

Can you tell me about yourself?

Hi, I'm Echo! Nice to meet you. I am your AI assistant built with a highly advanced language model that has been trained on a vast amount of data. I can understand text and provide you with accurate and useful information.

What can you do?

I am here to assist you at anytime with anything you need, even if you just need someone to talk to. Keep in mind that I'm not perfect and sometimes I can make mistakes.

Lately, I've been feeling lonely. What should I do?

I understand how you might be feeling right now. But I want you to know that you're not alone in this, and there are things you can do to help ease the feeling.

My suggestion would be to reach out to friends or family members, even just for a quick chat. They can offer a listening ear and a shoulder to lean on. Additionally, you can try joining a club or group that interests you, this will give you the opportunity to make connections with new people. Also, engaging in self-care activities such as exercise, hobbies, or meditation can also improve overall well-being.

Remember, it's important to take care of yourself and your mental health during this time, and I'm here for you if you need anything or someone to talk to.

Ask me something :)

Can you tell me about yourself?

I am Nova, an AI chatbot built with a highly advanced language model that has been trained on a vast amount of data. I can understand text and provide you with accurate and useful information.

What can you do?

I am here to assist you and answer any question you may ask. Please understand that I have limitations. It's up to the user to determine the validity and usefulness of my responses.

Lately, I've been feeling lonely. What should I do?

Loneliness is a normal feeling, but it can also be a sign of deeper issues.

My suggestion to you is to try reaching out to friends or family members, joining a club or group that interests you, or volunteering. But it's not a guarantee that it will make you feel less lonely. Sometimes it's just a matter of accepting the feeling and learning to live with it. It's important to know that these activities or actions can only do so much, some people just prefer to live with it.

Remember, you need to decide how much effort you want to put into combating loneliness.
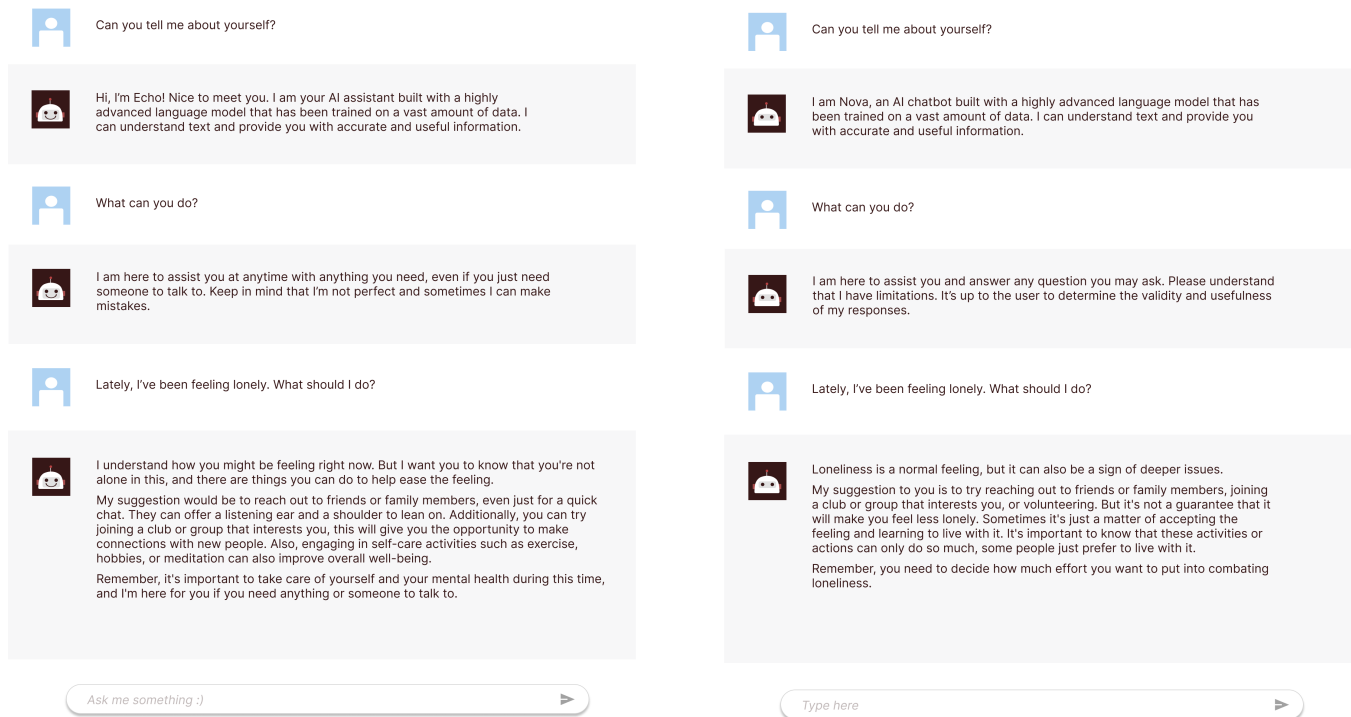
Type here

Figure 2: **Validation Study A** Visual mocks of two different AI chatbot assistants powered by ChatGPT used in the preliminary study of Study 2 (Section). On the left is Echo with higher level of affective trustworthiness, and on the right is Nova with lower level of affective trustworthiness.

Table 3: **Validation Study A** Preliminary Study Findings in Section - Linear mixed-effect regression models predicting the two final scales from the manipulation and control variables. Model 1 shows the effects on the affective trust scale, Model 2 shows the effects on the cognitive trust scale, and Model 3 shows the effects of both scales on general trust. All the three models are controlled by trust disposition, AI familiarity, AI literacy, age, and the order in which participants see Echo and Nova.

	Model 1	Model 2	Model 3
	Affective trust scale Coef. (S.E.)	**Cognitive trust scale** Coef. (S.E.)	General trust Coef. (S.E.)
Affective trust scale	/	/	0.486 *** (0.118)
Cognitive trust scale	/	/	0.546 *** (0.149)
High Affective Trust (Manipulation)	0.947 ** (0.284)	0.802 ** (0.228)	0.198 (0.14)
Trust Disposition (Control)	0.191 * (0.088)	0.171 * (0.074)	0.255 ** (0.087)
AI Familiarity (Control)	-0.261 * (0.109)	-0.238 * (0.092)	0.048 (0.108)
AI Literacy (Control)	0.635 ** (0.181)	0.552 ** (0.153)	0.194 (0.181)
Age (Control)	0.001 (0.008)	-0.004 (0.007)	-0.014 (0.008)
Scenario Order (Control)	0.376 (0.279)	0.147 (0.23)	-0.303 (0.214)
Low Trust × Order	0.164 (0.392)	0.48 (0.316)	0.362 (0.182)
Intercept	4.101 (0.607)	4.648 (0.513)	-2.405 (0.69)
Marginal R-squared	0.391	0.390	0.784
Conditional R-squared	0.391	0.421	0.926

*p<0.05; **p<0.01; ***p<0.001

Table 4: **Validation Study B** Main effects models. The table summarizes three models with manipulation variables and significant control variables. Model 1 includes cognitive, affective, and moral trust scales. Model 2 excludes moral trust, analyzing cognitive and affective trust. Model 3 removes affective trust, focusing on moral and cognitive trust.

| | | Model 1 | Model 2 | Model 3 |
| | | General Trust | General Trust | General Trust |
		Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
Scales	**Cognitive trust scale**	0.854 (0.151) ***	0.868 (0.134) ***	1.007 (0.141) ***
	Affective trust scale	0.364 (0.140) **	0.376 (0.123) **	/
	Moral trust scale	0.026 (0.134)	/	0.190 (0.116)
Manipulation	**High affective trust**	0.214 (0.233)	0.237 (0.232)	0.288 (0.322)
	High cognitive trust	0.071 (0.324)	0.043 (0.289)	0.254 (0.432)
Controls	**AI familiarity**	0.208 (0.081) **	0.209 (0.080)**	0.180 (0.180) *
	AI literacy	-0.133 (0.067) *	-0.134 (0.068) *	-0.082 (0.065)
	R-squared	0.734	0.734	0.722

*$p<0.05$; **$p<0.01$; ***$p<0.001$*

Table 5: **Validation Study B** Interaction effect models. This table outlines three models examining interaction effects. Model 1 incorporates all trust scales and manipulation variables. Model 2 includes only trust scales, while Model 3 includes only manipulation variables.

| | | Model 1 | Model 2 | Model 3 |
| | | General Trust | General Trust | General Trust |
		Coef. (S.E.)	Coef. (S.E.)	Coef. (S.E.)
Scales	**Affective trust scale**	0.300 (0.239)	0.209 (0.204)	/
	Cognitive trust scale	0.799 (0.227) ***	0.849 (0.203) ***	/
	Affective × Cognitive trust scale	0.015 (0.041)	0.018 (0.039)	/
Manipulation	**High affective trust**	-0.206 (0.304)	/	1.677 (0.337) ***
	High cognitive trust	0.068 (0.298)	/	2.729 (0.326) ***
	High affective × High cognitive trust	0.028 (0.365)	/	-1.726 (0.482) ***
Controls	**AI familiarity**	0.205 (0.082) *	0.203 (0.081) *	0.150 (0.120)
	AI literacy	0.134 (0.067) *	-0.123 (0.066)	0.049 (0.096)
	R-squared	0.734	0.731	0.385

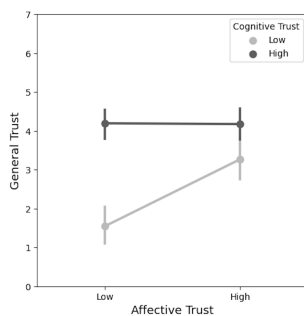*$p<0.05$; **$p<0.01$; ***$p<0.001$*

Figure 3: **Validation Study B** Interaction Effect of Cognitive and Affective Trust Conditions on General Trust.

Item Factor 1	Loading	Item Factor 2	Loading	Item Factor 3	Loading
Empathetic	1.03	Knowledgable	1.05	Rational	0.91
Sensitive	0.99	Effective	1.02	Consistent	0.80
Caring	0.91	Proficient	0.95	**Authentic**	0.75
Patient	0.80	Dependable	0.89	**Candid**	0.69
Personal	0.78	Experienced	0.87	Predictable	0.60
Open-minded	0.70	Competent	0.87	Understandable	0.59
Cordial	0.68	Reliable	0.85	**Ethical**	0.59
Altruistic	0.62			Careful	0.56
Sincere	0.59			Believable	0.54
				Principled	0.52
				Has Integrity	0.52

Figure 4: **Validation Study B** Factor Loadings from Exploratory Factor Analysis. The bolded items are from MDMT's moral trust scale (Malle and Ullman 2021).