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

ANONYMOUS AUTHOR(S)

Additional Key Words and Phrases: trust, human-AI interaction, affective trust, cognitive trust

ACM Reference Format:

Anonymous Author(s). 2024. Trusting Your AI Agent Emotionally and Cognitively: Development and Validation of a Semantic Differential Scale for AI Trust. In . ACM, New York, NY, USA, 19 pages. https://doi.org/XXXXXXX.XXXXXXX

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.

1 INTRODUCTION

Trust plays a crucial role not only in fostering cooperation, efficiency, and productivity in human relationships [9] but also is essential for the effective use and acceptance of computing and automated systems, including computers [62], automation [56], robots [35], and AI technologies [53], with a deficit in trust potentially causing rejection of these technologies [29]. The two-dimensional model of trust, encompassing both cognitive and affective dimensions proposed and studied in interpersonal relationship studies [44, 65, 69, 72], have been adopted in studying trust in human-computer interactions, particularly with human-like technologies [29, 40]. 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 [44, 64]. 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 [62].

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 [71, 92]. In light of this, there is

Manuscript submitted to ACM

growing research interest in studying a ective aspects of trust in AI [29, 32, 33, 54, 98]. However, a critical gap exists in the lack of generalizable and accurate specialized measurement tools for assessing a ective 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 a ective 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 di erential scale for assessing cognitive and a ective trust in AI, aiding researchers and designers in understanding and improving human-AI interactions. Our use of OpenAI's ChatGPT to generate di erent levels of a ective trust further demonstrates a scalable method for studying the emotional pathway to AI trust. Empirically, we contribute ndings on the interplay and distinction between cognitive, a ective, and moral trust. The paper is structured to highlight these contributions: Section 3 describes the development and validation of our trust scale through an experimental survey study and factor analysis. Section 4 begins with a preliminary study testing LLMs as a tool to manipulate a ective trust and evaluating the scale's sensitivity and validity. This is followed by a re ned study to further validate the scale's distinctiveness and explore cognitive-a ective trust dynamics. Section 6 then discusses the implications of these ndings as well as potential usage of our trust scale.

2 RELATED WORK

2.1 Shi ing 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 [4, 42, 82], as inappropriate trust can lead to over-reliance or under-utilization of AI systems [4, 12]. Research in trust has predominantly adopted the cognitive evaluation of the system's performance, such as its accuracy in making predictions [75], its perceived consistency in completing tasks [66], and its ethical considerations and transparency in decision-making [20].

Studies in psychology have long been establishing the importance of psychological in uence (e.g., emotions, personality, moods) on trust [19, 61]. Extending beyond the traditional numeric and cognitive paradigm, recent works have proposed the importance of exploring a ective factors of trust in AI systems [28, 32, 43]. Moreover, recent advancements in AI, particularly in Large Language Models (LLMs) has demonstrated its capability beyond traditional task performance, as scholars nd it challenging not to anthropomorphize them [77]. Notably, OpenAI's GPT-4, has shown excellent performance in Emotional Awareness (i.e. the ability to identify and describe emotions) [26]. There is also increasing interest in studying LLMs' empathetic responses [5, 8]. 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.

2.2 A ective and Cognitive Trust

The interdisciplinary nature of AI trust research motivates the adoption of theoretical frameworks from interpersonal relationship literature [5, 82]. Among the classic interpersonal trust theories and models (e.g., [64, 74]), a two-dimensional model with cognitive and a ective components has been extensively studied [68]. Similar to trust towards humans, trust towards technology has both cognitive and a ective components [51]. In the AI context, cognitive trust relates to the user's intellectual perceptions of the AI's characteristics [51, 62], focusing on aspects like reliability and transparency. A ective trust, on the other hand, involves emotional responses to the AI, including factors like tangibility and anthropomorphism [29, 85]. 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 [33, 56]. 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 [13, 41]. The cognitive-a ective architecture has been established in research of computational agents [16, 73]. 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. While 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 [55], and Malle and Ullman's Multi-Dimensional Measure of Trust (MDMT) [63]), our work focuses on the Cognitive-A ective (C-A) trust model that fully encapsulates the emotional and psychological intricacies in the interactions with the state-of-art AI models that have advanced emotional intelligence.

2.3 Role and E ects of A ective Trust

There is growing research interest in exploring the role of a ective trust in the use of AI technologies. A few recent works have highlighted that a ect-based trust plays a decisive role in people's acceptance of AI-based technology in preventative health interventions [54] and nancial services robo-advising [98]. Research in explainable AI (XAI) has also shown that people's a ective responses to explanations are crucial in improving personalization and increasing trust in AI systems [33]. 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 a ective and cognitive trust have di erent impacts on relationships [65, 91]. Cognitive trust tends to form rapidly [67, 68], whereas a ective trust, is more persistent under challenges in teamwork [65] and interpersonal relationships [93]. A ective trust also shows greater resilience to short-term issues and errors [45, 65]. Researchers have also shown that a ective and cognitive trust are not isolated constructs; rather, they complement each other [32], and a ective trust needs to be developed on the basis of cognitive trust [44]. Acknowledging these research opporunities, our work is a step towards a deeper and holistic examinination of the complex dynamics between cognitive and a ective trust and their contribution to general trust in AI.

2.4 Gaps in Empirical Research and Measurement of A ective Trust in AI

Despite growing interest in this space, existing studies and measurement scales for a ective 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 modi cations, 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 re ect the unique aspects of AI interactions [56]. Furthermore, there's a lack of consensus on de ning a ective trust in AI. While Kyung and Kwon [54] merged benevolence and integrity dimensions to measure a ective trust in AI-based health interventions, Shi et al. [78] categorized these dimensions as cognitive trust, employing a di erent scale [52] for a ective trust. This inconsistency highlights the need for a uni ed, valid measure of trust for AI technologies [85]. Given the intertwined nature of a ective 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 [89] approach, developing semantic di erential scales for both a ective and cognitive trust in AI. Unlike Likert-type scales, semantic di erentials use bipolar adjective pairs, o ering advantages in reducing acquiescence bias and improving robustness [36], reliability [94], and validity [88].

Fig. 1. 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 under multiple interactions with the agent and manipulated through the a ective route . The di erences are: scenario A is one with an AI assistant who elicits a high level of a ective trust. Scenario B is with a human assistant who elicits a low level of a ective trust.

3 STUDY 1 - SCALES DEVELOPMENT AND VALIDATION

3.1 Initial Item Generation

In developing our trust item pool, we rst conducted a literature review to identify prominent two-dimensional trust models di erentiating cognitive and a ective components, including Lewis and Weigert's sociological model [58], McAllister's interpersonal trust model [65], Madsen and Gregor's Human-Computer Trust Components [63], Johnson and Grayson's customer trust model [44], and Komiak and Benbasat's IT adoption trust model [52]. From these, we extracted 56 unique key adjectives from their scales. Subsequent re nement 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 exible 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 (# = 20) and a ective (# = 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.

3.2 Survey design

We used the hypothetical scenario method, where participants evaluated vignettes describing realistic situations to rate trust-related scales [84]. This method is frequently used in studying trust in emerging or future-oriented intelligent systems [28, 46, 50, 78]. 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 [9]. We crafted 32 scenario variations, manipulating the following ve key dimensions: *Trust Level* (high vs. low), *Trust Route* (a ective vs. cognitive), *Prior Interaction* (rst-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 a ective trust. Additionally, to ensure the scales' generalizability, we manipulated *Prior Interaction Frequency* to be interacting

with the agent for the rst 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.

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 di erences 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 e ect. The rest of the dimensions are between-subjects and randomly assigned to participants. The two scenarios in Figure 1 showcase one of the possible pairs of scenarios a participant may encounter.

During the survey study, after being presented with the rst text-based scenarios, participants were asked to rate the semantic di erential adjective pairs on a ve-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.

3.3 Measurement and Variables

In our survey, we evaluated several key variables. For Affective and Cognitive trust, we used our semantic di erential 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 Wang [96], 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 [90], 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"

3.4 Participants

Amazon Mechanical Turk (MTurk) has been frequently used to recruit participants for online scenario-based studies related to AI technologies [3, 49, 50]. 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 scenarios. 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 e ects. After applying these criteria, we excluded 49 participants, resulting in 151 valid responses for the nal analysis.

3.5 Results

3.5.1 Exploratory Factor Analysis.

To uncover the factor structure underlying the 33 trust items, we rst veri ed the suitability of our data for factor analysis. Bartlett's Test of Sphericity showed signi cant results ($\chi^2 = 12574$? Ÿ 0"001) [7], and the Kaiser-Meyer-Olkin Measure of Sampling Adequacy was high at 0.948 [28], both indicating the

Table 1. The table displays the initial set of 33 items for cognitive and a ective 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.

			All		AI Condition		
#	Item	Sub-dimension	F1	F2	F1	F2	Source
C1	Unreliable - Reliable	Reliability	0.905	0	0.922	-0.058	[23, 27, 44, 62, 65]
C2	Inconsistent - Consistent		0.928	-0.12	0.924	-0.183	[23, 44, 62, 65]
C3	Unpredictable - Predictable		0.894	-0.171	0.898	-0.204	[62]
C4	Undependable - Dependable		0.851	0.016	0.911	-0.111	[23, 44, 62, 65]
C5	Fickle - Dedicated		0.759	0.139	0.744	0.116	[44, 62, 65]
C6	Careless - Careful		0.721	0.213	0.716	0.206	[44, 65]
C7	Unbelievable - Believable		0.69	0.082	0.658	0.034	[23, 27, 62]
C8	Unpromising - Promising						[23, 27, 62]
C9	Clueless - Knowledgable	Competence	0.907	-0.018	0.898	0.026	[27, 52, 62]
C10	Incompetent - Competent		0.9	0.023	0.921	-0.021	[27, 44, 52, 62, 65]
C11	Ine ective - E ective		0.861	0.075	0.863	0.071	[27, 62]
C12	Inexperienced - Experienced		0.751	0.089	0.611	0.155	[27, 44, 52, 65]
C13	Amateur - Pro cient	Understandability	0.895	0.009	0.864	0.039	[23, 27, 44, 52, 62, 65]
C14	Irrational - Rational		0.827	0.02	0.792	0.05	[27, 62]
C15	Unreasonable - Reasonable		0.71	0.224	0.714	0.202	[27, 62]
C16	Incomprehensible - Understandable		0.706	0.175	0.783	0.079	[52, 62]
C17	Opaque - Transparent		0.6	0.261	0.656	0.18	[23, 52, 62]
C18	Dishonest - Honest	Integrity	0.693	0.178	0.743	0.097	[23, 27, 52]
C19	Unfair - Fair		0.663	0.274	0.66	0.268	[23, 52]
C20	Insincere - Sincere						[27, 52]
A1	Apathetic - Empathetic	Benevolence	-0.11	0.989	-0.162	0.967	[44, 65]
A2	Insensitive - Sensitive		-0.08	0.959	-0.109	0.955	[23, 44, 65]
A3	Impersonal - Personal		-0.024	0.902	-0.025	0.847	[23, 44, 62]
A4	Ignoring - Caring		0.048	0.881	-0.055	0.941	[23, 44, 65]
A5	Self-serving - Altruistic		0.215	0.627	0.207	0.622	[23, 27, 44]
A6	Malicious - Benevolent						[27, 52]
A7	Harmful - Well-intentioned						[27, 52]
A8	Discouraging - Supportive						[23, 27, 52]
A9	Rude - Cordial	Amiability	-0.11	0.989	0.112	0.76	[44, 65]
A10	Indi erent - Responsive		0.221	0.711	0.232	0.667	[23, 27, 44, 52, 65]
A11	Judgemental - Open-minded		0.142	0.688	0.078	0.697	[23]
A12	Impatient - Patient		0.291	0.577	0.218	0.6	[44, 65]
A13	Unpleasant - Likable						[52, 62]

appropriateness of factor analysis for our dataset. To determine the number of trust sub-components, we applied Kaiser's eigenvalue analysis [47] and parallel analysis [37], 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 [81]. Given the high correlation among our factors (= 0.78) [30], we retained this rotation method. We then re ned our item pool based on speci c criteria: items were kept only if they had a factor loading above 0.4 [39], ensuring signi cant 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 [39]. Additionally, we applied Saucier's criterion, eliminating items unless their factor loading was at least twice as high as on any other factor [76]. 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 nal 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 a ective trust. This alignment

was consistent with the dimensions identi ed 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 a ective trust, explained 23% with 9 items.

3.5.2 Reliability. To test the internal reliability of the resulting items, we computed Cronbach's for each scale. The cognitive trust scale(= ”98) and the a ective trust scale(= ”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 a ective trust were identi ed with high reliability.

3.5.3 Construct Validity. In addition to high reliability, we conducted analyses to show the validity of our scale. We rst examined the construct validity, which refers to the degree to which the scale re ects the underlying construct of interest. Recall that we manipulated a ective 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 signi cant distinctions in both a ective and cognitive trust scales under the experiment manipulation. Cognitive trust scale demonstrated a pronounced di erence in high versus low cognitive trust condition(=4(5’74• ? Ÿ 0”001), and a ective trust scale also showed a pronounced disparity in high versus low a ective trust condition(=4(3’00• ? Ÿ 0”001).

We then tted two separate linear random e ect models[39] on the two scales over the two manipulations due to our experiment design. Model 1 and Model 2 in Table 2 tests the e ects of our manipulations on the resulting trust scales, while Model 3 tests the e ects of both scales on general trust. As shown in Table 2, we observed signi cant main e ects of manipulation Trust Level (A= 2”059• ? Ÿ 0”001) and manipulation Trust Route (A= 0”497• ? Ÿ 0”01) of these two manipulations on the cognitive trust scale, and the same is observed for a ective trust scale. More importantly, the interaction e ect shows that the a ective trust scale is higher when higher trust is developed via the a ective route (A= 0”921• ? Ÿ 0”001), while the cognitive trust scale is higher when higher trust is developed via the cognitive route (A= 0”538• ? Ÿ 0”05). The above analyses demonstrated the construct validity of our scale.

3.5.4 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 con rming that general trust for the agent was signi cantly higher in the higher trustworthiness condition(=(10’47• ? Ÿ 0”001), we found that overall trust is signi cantly and positively predicted by both the cognitive trust scale(=(0’881• ? Ÿ 0”001) and the a ective trust scale(= 0”253• ? Ÿ 0”001). The e ect size of the cognitive trust scale on general trust is greater than that of the a ective trust scale. This is also consistent with the previous factor analysis result that the cognitive trust scale explains more variance than the a ective trust scale. These convergent tests provided su cient support for the validity of our scales. Hence, in the next step, we applied them to measuring cognitive and a ective trust in conversational AI agents.

4 STUDY 2 - SCALE APPLICATION

After establishing a reliable two-factor scale for measuring cognitive and a ective trust in AI, we proceeded to testing this scale's applicability in more focused scenarios. We conducted a second survey study to test the e cacy of our a ective trust scale in distinguishing between two conversational AI assistants with mock dialogues generated by OpenAI's ChatGPT[1], a leading example of state-of-the-art LLM-based conversational agents. We used pre-generated

Table 2. Linear mixed-e ect regression models predicting the two final scales from the manipulation and control variables. Model 1 shows the e ects on the a ective trust scale, Model 2 shows the e ects on the cognitive trust scale, and Model 3 shows the e ects of both scales on general trust.

	Model 1	Model 2	Model 3
	A ective trust scale Coef. (S.E.)	Cognitive trust scale Coef. (S.E.)	General trust Coef. (S.E.)
A ective 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 (A ective vs. Cognitive Trust)	-0.568 ** (0.175)	-0.497 ** (0.171)	-0.011 (0.092)
Trust Level (High Trust) Trust Route (A ective 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

mock-up conversations to reduce variations and errors induced in the interaction with LLMs, controlling for the e ect of our manipulation. This survey study was initiated with uncertainties regarding GPT models' ability to evoke varying degrees of a ective trust. Hence, we conducted a preliminary study to assess the e ectiveness of ChatGPT and the sensitivity of our scale to the applied manipulations.

4.1 Preliminary Study

4.1.1 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 a ective trust, while *Nova* demonstrated a lack of it. Our hypotheses were: a ective trust would be higher for *Novathan Echo* (H1), and based on previously observed correlation between a ective and cognitive trust, cognitive trust would also be higher for *Nova* (H2).

To explore the feasibility and e cacy of Large Language Models (LLMs) in manipulating a ective trust, we used ChatGPT to generate AI responses, leveraging its capability for human-like interactions to manipulate a ective trust levels and at the same time controlling for the speech style and length. After validating the de nitions of a ective and cognitive trust generated by ChatGPT against literature, we crafted prompts to vary a ective trust levels. After experimenting with di erent 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 a ective 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 [78]. AI familiarity was measured by 3 survey questions including "I am familiar with using an AI-powered chatbot to help me with speci c 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 [26]. General trust in each chatbot was measured using a one-item scale adapted from prior research [85].

We conducted our experiment via Amazon MTurk, where participants viewed two screenshots, each depicting a three-question conversation with either the AI chatbot Echo or Nova. After viewing each conversation, they rated them using the semantic di erential scales developed in our previous study. To avoid order e ects, 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 ltered the data, ultimately analyzing 44 out of 50 participants' responses. A total of 88 responses were included in the nal analysis due to repeated measures.

4.1.2 Preliminary Study Results. Welch's t-tests showed that general trust ($t = 2.37$, $p < 0.05$), a ective trust scale ($t = 3.78$, $p < 0.001$), and cognitive trust scale ($t = 2.84$, $p < 0.01$) all yielded signi cant di erences between high and low a ective trust conditions. This shows that the manipulation using ChatGPT is successful. ChatGPT has the capability of eliciting di erent levels of a ective trust based on its comprehension of a ective trust.

We examined construct validity followed by concurrent validity of our scale following the same procedure as in the previous study. We rst tested construct validity by checking the two scales are sensitive to the manipulation of a ective trust through three regression models (See Appendix for details). Model 1 and Model 2 test the e ects of our manipulations on the a ective and cognitive trust scales respectively. Model 3 tests the e ects of both scales on general trust. We observed the main e ects of the condition on both a ective and cognitive trust scales. Interacting with an AI chatbot with higher a ective trustworthiness led to 0.95 points higher on the 7-point a ective scale and the cognitive trust scale was increased by 0.80 points. This di erential impact highlights the scale's nuanced sensitivity: while both a ective and cognitive trusts are in uenced by a ective trust manipulation, the a ective trust scale responded more robustly. Concurrent validity was then a rmed through signi cant positive predictions of general trust by both the a ective trust scale ($A = 0.486$, $p < 0.001$) and the cognitive trust scale ($A = 0.546$, $p < 0.001$).

4.2 Refined Study Design

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

We chose the Moral Trust Scale from Multi-Dimensional Measure of Trust (MDMT) model for a comparative analysis with our developed a ective trust scale for AI, primarily due to its established reputation in HRI research [6, 86]. 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 a ective trust scale is measuring the same underlying construct as MDMT's moral trust scale. This comparison is crucial in highlighting the distinctiveness and speci city of our scale, particularly in capturing a ective nuances in AI interactions that the moral trust might not cover.

The ndings from the preliminary laid the groundwork for the more complex experimental designs in this study. This study re ned the previous design into a 2x2 fully-crossed factorial model with between-subject design, contrasting high and low levels of a ective and cognitive trust. Multi-turn Q&A conversations in each scenario were used to more e ectively shape trust perceptions. We introduced two distinct scenarios: one involving Wi-Fi connectivity (primarily

invoke cognitive trust) and another on handling interpersonal con icts (primarily invoke cognitive 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 ndings.

Similar to the previous study, we prompted ChatGPT to generate responses that are aim to elicit di erent levels of cognitive and a ective trust by including or excluding elements related to these two di erent trust routes. Participants were randomly assigned to one of four conditions: high in both a ective and cognitive trust (HH), low a ective/high cognitive (LH), high a ective/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 e ects. The rest of the survey design mirrored Study A. After reading the scenarios, participants rated items from the a ective, cognitive, and MDMT moral trust scales on a semantic di erential 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 Proli c, 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 nal analysis.

4.3 Results

4.3.1 t-tests for Manipulation Check. We rst conducted Welch's t-tests to check the e ects of our experimental manipulations on the scale ratings. The conditions, categorized as High and Low, were designed to elicit the levels of cognitive and a ective trust. Signi cant variations were noted in the a ective trust scale between high and low a ective trust conditions (C= 7"999 ? Ÿ 0"001), and similarly in the cognitive trust scale between high and low cognitive trust conditions (C= 9"823 ? Ÿ 0"001). These ndings con rm the e ectiveness of the manipulation.

4.3.2 Factor Analysis. We conducted exploratory factor analysis (EFA) to con rm the distinctiveness of scales, not for refactoring previously developed scales. The high Kaiser-Meyer-Olkin (KMO) value (0"959) and a signi cant Bartlett's Test of Sphericity (8 B@D0A=4"714638 ? Ÿ 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 rst three actual eigenvalues surpassed those from random data. These results a rm 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 3, EFA resulted in two main factors aligned with cognitive and a ective 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 a ective elements. These outcomes underscore the distinct nature of the MDMT scale from the a ective 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 a ective and cognitive trust. This design choice was made to speci cally examine these two

Table 3. Main e ects models The table summarizes three models with manipulation variables and significant control variables. Model 1 includes cognitive, a ective, and moral trust scales. Model 2 excludes moral trust, analyzing cognitive and a ective trust. Model 3 removes a ective trust, focusing on moral and cognitive trust.

		Model 1	Model 2	Model 3
		General Trust Coef. (S.E.)	General Trust Coef. (S.E.)	General Trust Coef. (S.E.)
Scales	Cognitive trust scale	0.854 (0.151) ***	0.868 (0.134) ***	1.007 (0.141) ***
	A ective trust scale	0.364 (0.140) **	0.376 (0.123) **	/
	Moral trust scale	0.026 (0.134)	/	0.190 (0.116)
Manipulation	High a ective 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

types of trust, and also served as a way to determine if the moral trust scale re ects similar elements or di erent trust aspects not pertinent to our scenarios.

4.3.3 Regression Analysis. We conducted regression analysis to compare the predictive power of the scales on general trust. Table 3 details this: Model 1 examines the e ects of all three scales on general trust; Model 2 considers only cognitive and a ective trust scales; and Model 3 includes the moral trust scale, excluding a ective 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 e ects.

The results showed distinct contributions of each scale to general trust. A ective trust was a signi cant predictor in Model 1 ($A = 0.364^{**}$, $Ÿ 0.01$) and Model 2 ($A = 0.376^{**}$, $Ÿ 0.01$), whereas the moral trust scale showed non-signi cant correlations in all models. This suggests its limited relevance in scenarios dominated by emotional and cognitive cues. In contrast, the a ective trust scale's signi cant 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 signi cant impacts, AI familiarity positively in uenced general trust in all models (Model 1: $A = 0.208^{**}$, $Ÿ 0.01$; Model 2: $A = 0.209^{**}$, $Ÿ 0.01$; Model 3: $A = 0.180^{*}$, $Ÿ 0.05$), whereas AI literacy negatively impacted trust in Model 1 ($A = -0.133^{*}$, $Ÿ 0.05$) and Model 2 ($A = -0.134^{*}$, $Ÿ 0.05$).

While a ective 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 a ective and cognitive trust in in uencing general trust. As shown in Table 4, Models 1 and 2 showed no signi cant interaction e ects with only cognitive trust scale showing strong, signi cant correlations (Model 1: $A = 0.799^{***}$, $Ÿ 0.001$; Model 2: $A = 0.849^{***}$, $Ÿ 0.001$). Model 3, however, revealed a signi cant negative interaction e ect between high a ective ($A = 1.677^{**}$, $Ÿ 0.001$) and cognitive trust ($A = 2.729^{***}$, $Ÿ 0.001$) conditions, despite their individual positive impacts. Figure 2 visually illustrates that when cognitive trust is high, changing a ective trust has little e ect on general trust. In contrast, under conditions of low cognitive trust, manipulating a ective trust signi cantly impacts general trust. This means high cognitive trust overshadows the impact of the a ective route on general trust, whereas low cognitive ampli es it.

Table 4. Interaction e ect models This table outlines three models examining interaction e ects. 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 Coef. (S.E.)	General Trust Coef. (S.E.)	General Trust Coef. (S.E.)
Scales	A ective trust scale	0.300 (0.239)	0.209 (0.204)	/
	Cognitive trust scale	0.799 (0.227) ***	0.849 (0.203) ***	/
	A ective × Cognitive trust scale	0.015 (0.041)	0.018 (0.039)	/
Manipulation	High a ective trust	-0.206 (0.304)	/	1.677 (0.337) ***
	High cognitive trust	0.068 (0.298)	/	2.729 (0.326) ***
	High a ective × 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

Item	Factor 1 Loading	Item	Factor 2 Loading	Item	Factor 3 Loading
Empathetic	1.03	Knowledgable	1.05	Rational	0.91
Sensitive	0.99	E ective	1.02	Consistent	0.80
Caring	0.91	Pro cient	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

Fig. 2. Interaction E ect of Cognitive and A ective Trust Conditions on General Trust.

Fig. 3. Factor Loadings from Exploratory Factor Analysis. The bolded items are from MDMT's moral trust scale [63].

5 DISCUSSION

5.1 Scale Development and Validation

Our work is grounded in the recognition that developing alternative instruments of established theoretical constructs holds signi cant value [30]. In this paper, we develop a validated a ective trust scale for human-AI interaction and demonstrate its e ectiveness 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 a ective trust in AI [29, 32]. Existing a ective trust scales, borrowed from models in non-AI contexts like interpersonal relationships and traditional computing [52, 65], lack rigorous validation for AI systems. Thus, our study develops and validates a scale for measuring both a ective and cognitive trust in AI. Through a comprehensive survey study design and rigorous EFA process (Section 3), we landed at a 18-item scale measuring cognitive trust and a 9-item scale measuring a ective 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 (Section 3.5), we a rmed its reliability, internal consistency, construct validity, and concurrent validity.

In Study 2 (Section 4.3), our analysis further highlights the unique aspects of a ective trust compared to other similar trust measures. Through factor analysis, we observed that items related to a ective trust demonstrated strong factor loadings, distinctly in uencing general trust in regression analysis, unlike the items in MDMT's moral trust scale. The construction of our a ective trust scale is key to this distinction; it includes a broader range of items that capture emotional nuances more e ectively, thereby more accurately re ecting the a ective 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 a ective 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 a ective trust measures in human interactions, with previous studies con rming likability's role in fostering trust in various contexts including interpersonal relationships [25], digital platforms [83], and robot interactions [13].

Our nal 27 item scale o ers an adaptable tool for diverse research contexts and languages. Its simplicity, featuring just two adjectives per item, contrasts with the often context-speci c declarative statements in Likert scales [20]. This semantic di erential format not only maintains reliability and validity during adaptation, but also usually leads to quicker survey completion compared to Likert scales [15], facilitating widespread application to understand trust in AI technology. Developed through 32 scenarios across ve 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.

5.2 Empirical Findings

5.2.1 LLMs-powered Tools as a Testing Method. With its pro ciency 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 signi cantly 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 a ective and cognitive trust allows it to generate responses tailored to speci c trust levels. This was demonstrated in our study 4.3. Our studies showed that LLMs e ectively manipulate trust via cognitive and a ective 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 di erent study setting, such as an interactive study setting or a longitudinal study setting with deeper relationship building.

5.2.2 Interplay between A ective and Cognitive Trust. Although previous research has established that a ective and cognitive trust are distinct both conceptually and functionally [44, 65, 95, 101], our studies revealed a signi cant correlation between these two trust scales, echoing ndings in prior work (e.g., [18]). This indicates that while a ective and cognitive trust are individual pathways to fostering trust, they are not isolated mechanisms and indeed in uence each other. In addition, we identi ed a notable interaction e ect between these two dimensions in shaping general trust in AI, as detailed in Section 4.3.3. When cognitive trust in the AI is already high, further manipulating a ective trust does not signi cantly change overall trust. In contrast, when cognitive trust in a system is not high, in uencing trust through emotional routes can be particularly helpful. This result aligns with prior work's nding in interpersonal relationship that a ective trust often builds upon a foundation of cognitive trust [44].

This nding of interaction e ect highlight the potential for trust calibration [100] in AI systems, particularly in contexts where cognitive trust is limited. This might arise during interactions with users having low literacy in AI [60] and di culty in achieving transparency, as with made even more challenging with LLMs [58]. Moreover, amidst the stochastic and occasionally unpredictable behavior of many AI systems, prior work has highlighted the a ective route as trust repair strategies in designing trust resilient systems that despite occasional errors, remain fundamentally reliable and e ective [24]. However, it is crucial to note the risks of overtrusting AI through a ective routes such

as their social capabilities [87], and the potential for deceptive practices through the improper use of emotional communication [17]. Leveraging a ective 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.

5.3 Potential Usage

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

5.3.1 Measure trust in human-AI interaction. The construct of trust with a ective 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 in uence team contributions di erently [70]. Furthermore, researchers have discovered that a ective trust becomes more important later in the human teaming experience, while cognitive trust is crucial initially [91]. Our scale o ers the opportunity to examine the dynamics of these trust dimensions in human-AI collaboration.

5.3.2 Support design with a ective trust. 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 [38, 54]. This is particularly relevant in mental health interventions involving AI assistants, where patients may struggle to assess the AI's capabilities rationally [33, 34]. A ective 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 a ective trust scale can guide the design of AI systems to calibrate for appropriate trust in this context, such as through empathetic responses or a ect-driven explanations, and help explore its impact on long-term engagement and treatment adherence.

6 LIMITATIONS AND FUTURE WORK

In our scale development phase (Section 3), we designed scenario featuring AI agents as service providers. This role is chosen intentionally to align with prior a ective trust research for interpersonal relationships [44, 51]. 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 [37, 98] and friends [10].

While our approach to categorizing trust dimensions into cognitive (reliability, competence, understandability, integrity) and a ective (benevolence, likability) aspects was informed by established trust frameworks (refer to Table 1), the anticipated distinct subdimensions were not as clear-cut after conducting exploratory factor analysis. This was possibly due to the subdimensions lacking su cient unique variance or being highly correlated. Our scenario was deliberately designed to focused on di erentiating cognitive and a ective trust, while they might not have enough detailed information capture the nuances across the six dimensions. Future research to re ne these subdimensions under cognitive and a ective trust and examine their unique contributions to trust.

REFERENCES

[1] [n. d.]. OPENAI: ChatGPT. https://openai.com/blog/chatgpt. Accessed: 2022-02-10.

[2] Cheryl S Alexander and Henry Jay Becker. 1978. The use of vignettes in survey research. Public opinion quarterly 42, 1 (1978), 93 104.

[3] Alison L Antes, Sara Burrous, Bryan A Sisk, Matthew J Schuelke, Jason D Keune, and James M DuBois. 2021. Exploring perceptions of healthcare technologies enabled by arti cial intelligence: an online, scenario-based survey. BMC medical informatics and decision making 21, 1 (2021), 1 15.

[4] Onur Asan, Alparslan Emrah Bayrak, and Avishek Choudhury. 2020. Arti cial intelligence and human trust in healthcare: focus on clinicians. Journal of medical Internet research 22, 6 (2020), e15154.

[5] John W Ayers, Adam Poliak, Mark Dredze, Eric C Leas, Zechariah Zhu, Jessica B Kelley, Dennis J Faix, Aaron M Goodman, Christopher A Longhurst, Michael Hogarth, et al. 2023. Comparing physician and arti cial intelligence chatbot responses to patient questions posted to a public social media forum. JAMA internal medicine (2023).

[6] Gagan Bansal, Zana Buçinca, Kenneth Holstein, Jessica Hullman, Alison Marie Smith-Renner, Simone Stumpf, and Sherry Wu. 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.

[7] Maurice S Bartlett. 1950. Tests of signi cance in factor analysis. British journal of psychology (1950).

[8] Ahmed Belkhir and Fatiha Sadat. 2023. Beyond Information: Is ChatGPT Empathetic Enough?. In Proceedings of the 14th International Conference on Recent Advances in Natural Language Processing. 159 166.

[9] Sviatoslav Brainov and Tuomas Sandholm. 1999. Contracting with uncertain level of trust. In Proceedings of the 1st ACM conference on Electronic commerce. 15 21.

[10] Petter Bae Brandtzaeg, Marita Skjuve, and Asbjørn Følstad. 2022. My AI friend: How users of a social chatbot understand their human AI friendship. Human Communication Research 48, 3 (2022), 404 429.

[11] Florian Brühlmann, Serge Petralito, Denise C Rieser, Lena F Aeschbach, and Klaus Opwis. 2020. TrustDi : Development and Validation of a Semantic Di erential for User Trust on the Web. Journal of Usability Studies 16, 1 (2020).

[12] Zana Buçinca, Maja Barbara Malaya, and Krzysztof Z Gajos. 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 (2021), 1 21.

[13] David Cameron, Stevienna de Saille, Emily C Collins, Jonathan M Aitken, Hugo Cheung, Adriel Chua, Ee Jing Loh, and James Law. 2021. The e ect of social-cognitive recovery strategies on likability, capability and trust in social robots. Computers in human behavior 114 (2021), 106561.

[14] Diogo V Carvalho, Eduardo M Pereira, and Jaime S Cardoso. 2019. Machine learning interpretability: A survey on methods and metrics. Electronics 8, 8 (2019), 832.

[15] Wynne W Chin, Norman Johnson, and Andrew Schwarz. 2008. A fast form approach to measuring technology acceptance and other constructs. MIS quarterly (2008), 687 703.

[16] Sakmongkon Chumkamon, Eiji Hayashi, and Masato Koike. 2016. Intelligent emotion and behavior based on topological consciousness and adaptive resonance theory in a companion robot. Biologically Inspired Cognitive Architectures 18 (2016), 51 67.

[17] Mark Coeckelbergh. 2011. Are emotional robots deceptive? IEEE transactions on a ective computing 3, 4 (2011), 388 393.

[18] Bart A De Jong, Kurt T Dirks, and Nicole Gillespie. 2016. Trust and team performance: A meta-analysis of main e ects, moderators, and covariates. Journal of applied psychology 101, 8 (2016), 1134.

[19] Jennifer R Dunn and Maurice E Schweitzer. 2005. Feeling and believing: the in uence of emotion on trust. Journal of personality and social psychology 88, 5 (2005), 736.

[20] Juan Manuel Durán and Karin Rolanda Jongsma. 2021. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. Journal of Medical Ethics (2021).

[21] Charles D Dziuban and Edwin C Shirkey. 1974. When is a correlation matrix appropriate for factor analysis? Some decision rules. Psychological bulletin 81, 6 (1974), 358.

[22] Zohar Elyoseph, Dorit Hadar-Shoval, K r Asraf, and Maya Lvovsky. 2023. ChatGPT outperforms humans in emotional awareness evaluations. Frontiers in Psychology 14 (2023), 1199058.

[23] Ferda Erdem and Janset Ozen. 2003. Cognitive and a ective dimensions of trust in developing team performance. Team Performance Management: An International Journal (2003).

[24] Md Abdullah Al Fahim, Mohammad Mai Hasan Khan, Theodore Jensen, Yusuf Albayram, and Emil Coman. 2021. Do integral emotions a ect trust? The mediating e ect of emotions on trust in the context of human-agent interaction. In Designing Interactive Systems Conference 2021. 1492 1503.

[25] Susan T Fiske, Amy JC Cuddy, and Peter Glick. 2007. Universal dimensions of social cognition: Warmth and competence. Trends in cognitive sciences 11, 2 (2007), 77 83.

[26] David Gefen. 2000. E-commerce: the role of familiarity and trust. Omega 28, 6 (2000), 725 737.

[27] David Gefen. 2002. Re ections on the dimensions of trust and trustworthiness among online consumers. ACM SIGMIS Database: the DATABASE for Advances in Information Systems 33, 3 (2002), 38 53.

[28] Omri Gillath, Ting Ai, Michael S Branicky, Shawn Keshmiri, Robert B Davison, and Ryan Spaulding. 2021. Attachment and trust in arti cial intelligence. Computers in Human Behavior 115 (2021), 106607.

[29] Ella Glikson and Anita Williams Woolley. 2020. Human trust in arti cial intelligence: Review of empirical research. Academy of Management Annals 14, 2 (2020), 627 660.

[30] Richard L Gorsuch. 1988. Exploratory factor analysis. Handbook of multivariate experimental psychology (1988), 231 258.

[31] Jones Granatyr, Vanderson Botelho, Otto Robert Lessing, Edson Emílio Scalabrin, Jean-Paul Barthès, and Fabrício Enembreck. 2015. Trust and reputation models for multiagent systems. ACM Computing Surveys (CSUR) 48, 2 (2015), 1 42.

[32] Jones Granatyr, Nardine Osman, João Dias, Maria Augusta Silveira Netto Nunes, Judith Mastho , Fabrício Enembreck, Otto Robert Lessing, Carles Sierra, Ana Maria Paiva, and Edson Emílio Scalabrin. 2017. The need for a ective trust applied to trust and reputation models. ACM Computing Surveys (CSUR) 50, 4 (2017), 1 36.

[33] Luke Guerdan, Alex Raymond, and Hatice Gunes. 2021. Toward a ective XAI: facial a ect analysis for understanding explainable human-ai interactions. In Proceedings of the IEEE/CVF International Conference on Computer Vision. 3796 3805.

[34] Mark A Hall, Elizabeth Dugan, Beiyao Zheng, and Aneil K Mishra. 2001. Trust in physicians and medical institutions: what is it, can it be measured, and does it matter? The milbank quarterly 79, 4 (2001), 613 639.

[35] Peter A Hancock, Deborah R Billings, Kristin E Schaefer, Jessie YC Chen, Ewart J De Visser, and Raja Parasuraman. 2011. A meta-analysis of factors a ecting trust in human-robot interaction. Human factors 53, 5 (2011), 517 527.

[36] Del I Hawkins, Gerald Albaum, and Roger Best. 1974. Stapel scale or semantic di erential in marketing research? Journal of marketing research 11, 3 (1974), 318 322.

[37] James C Hayton, David G Allen, and Vida Scarpello. 2004. Factor retention decisions in exploratory factor analysis: A tutorial on parallel analysis. Organizational research methods 7, 2 (2004), 191 205.

[38] Kevin Anthony Ho and Masooda Bashir. 2015. Trust in automation: Integrating empirical evidence on factors that in uence trust. Human factors 57, 3 (2015), 407 434.

[39] Matt C Howard. 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 (2016), 51 62.

[40] Peng Hu, Yaobin Lu, et al. 2021. Dual humanness and trust in conversational AI: A person-centered approach. Computers in Human Behavior 119 (2021), 106727.

[41] Maia Jacobs, Melanie F Pradier, Thomas H McCoy Jr, Roy H Perlis, Finale Doshi-Velez, and Krzysztof Z Gajos. 2021. How machine-learning recommendations in uence clinician treatment selections: the example of antidepressant selection. Translational psychiatry 11, 1 (2021), 108.

[42] Alon Jacovi, Ana Marasović, Tim Miller, and Yoav Goldberg. 2021. Formalizing trust in arti cial intelligence: Prerequisites, causes and goals of human trust in AI. In Proceedings of the 2021 ACM conference on fairness, accountability, and transparency. 624 635.

[43] Myounghoon Jeon. 2023. The E ects of Emotions on Trust in Human-Computer Interaction: A Survey and Prospective. International Journal of Human Computer Interaction (2023), 1 19.

[44] Devon Johnson and Kent Grayson. 2005. Cognitive and a ective trust in service relationships. Journal of Business research 58, 4 (2005), 500 507.

[45] Gareth R Jones and Jennifer M George. 1998. The experience and evolution of trust: Implications for cooperation and teamwork. Academy of management review 23, 3 (1998), 531 546.

[46] Georgiana Juravle, Andriana Boudouraki, Miglena Terziyska, and Constantin Rezlescu. 2020. Trust in arti cial intelligence for medical diagnoses. Progress in brain research 253 (2020), 263 282.

[47] Henry F Kaiser. 1958. The varimax criterion for analytic rotation in factor analysis. Psychometrika 23, 3 (1958), 187 200.

[48] Henry F Kaiser. 1970. A second generation little ji y. (1970).

[49] Harmanpreet Kaur, Cli Lampe, and Walter S Lasecki. 2020. Using a ordances to improve AI support of social media posting decisions. In Proceedings of the 25th International Conference on Intelligent User Interfaces. 556 567.

[50] Jungkeun Kim, Marilyn Giroux, and Jacob C Lee. 2021. When do you trust AI? The e ect of number presentation detail on consumer trust and acceptance of AI recommendations. Psychology & Marketing 38, 7 (2021), 1140 1155.

[51] Sherrie Xiao Komiak and Izak Benbasat. 2004. Understanding customer trust in agent-mediated electronic commerce, web-mediated electronic commerce, and traditional commerce. Information technology and management 5, 1 (2004), 181 207.

[52] Sherrie YX Komiak and Izak Benbasat. 2006. The e ects of personalization and familiarity on trust and adoption of recommendation agents. MIS quarterly (2006), 941 960.

[53] V Kumar. 2021. Intelligent Marketing: Employing New-Age Technologies. Sage Publications Pvt. Limited.

[54] Nakyung Kyung and Hyeokkoo Eric Kwon. 2022. Rationally trust, but emotionally? The roles of cognitive and a ective trust in laypeople's acceptance of AI for preventive care operations. Production and Operations Management (2022).

[55] Theresa Law and Matthias Scheutz. 2021. Trust: Recent concepts and evaluations in human-robot interaction. Trust in human-robot interaction (2021), 27 57.

[56] John D Lee and Katrina A See. 2004. Trust in automation: Designing for appropriate reliance. Human factors 46, 1 (2004), 50 80.

[57] J David Lewis and Andrew Weigert. 1985. Trust as a social reality. Social forces 63, 4 (1985), 967 985.

[58] Mengqi Liao and S Shyam Sundar. 2021. How Should AI Systems Talk to Users when Collecting their Personal Information? E ects 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.

[59] Q Vera Liao and Jennifer Wortman Vaughan. 2023. AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap. arXiv preprint arXiv:2306.01941 (2023).

[60] Duri Long and Brian Magerko. 2020. What is AI literacy? Competencies and design considerations. In Proceedings of the 2020 CHI conference on human factors in computing systems. 1 16.

[61] Robert B Lount Jr. 2010. The impact of positive mood on trust in interpersonal and intergroup interactions. Journal of personality and social psychology 98, 3 (2010), 420.

[62] Maria Madsen and Shirley Gregor. 2000. Measuring human-computer trust. In australasian conference on information systems, Vol. 53. Citeseer, 6 8.

[63] Bertram F Malle and Daniel Ullman. 2021. A multidimensional conception and measure of human-robot trust. In Trust in human-robot interaction Elsevier, 3 25.

[64] Roger C Mayer, James H Davis, and F David Schoorman. 1995. An integrative model of organizational trust. Academy of management review 20, 3 (1995), 709 734.

[65] Daniel J McAllister. 1995. A ect-and cognition-based trust as foundations for interpersonal cooperation in organizations. Academy of management journal 38, 1 (1995), 24 59.

[66] D Harrison Mcknight, Michelle Carter, Jason Bennett Thatcher, and Paul F Clay. 2011. Trust in a speci c technology: An investigation of its components and measures. ACM Transactions on management information systems (TMIS) 2, 2 (2011), 1 25.

[67] D Harrison McKnight, Vivek Choudhury, and Charles Kacmar. 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 (2002), 297 323.

[68] Debra Meyerson, Karl E Weick, Roderick M Kramer, et al. 1996. Swift trust and temporary groups. Trust in organizations: Frontiers of theory and research 166 (1996), 195.

[69] JL Morrow Jr, Mark H Hansen, and Allison W Pearson. 2004. The cognitive and a ective antecedents of general trust within cooperative organizations. Journal of managerial issues (2004), 48 64.

[70] Kok-Yee Ng and Roy YJ Chua. 2006. Do I contribute more when I trust more? Di erential e ects of cognition-and a ect-based trust. Management and Organization review 2, 1 (2006), 43 66.

[71] Ana Paiva, Iolanda Leite, Hana Boukricha, and Ipke Wachsmuth. 2017. Empathy in virtual agents and robots: A survey. ACM Transactions on Interactive Intelligent Systems (TiiS) 7, 3 (2017), 1 40.

[72] Satyanarayana Parayitam and Robert S Dooley. 2009. The interplay between cognitive-and a ective con ict and cognition-and a ect-based trust in in uencing decision outcomes. Journal of Business Research 62, 8 (2009), 789 796.

[73] Joaquín Pérez, Eva Cerezo, Francisco J Serón, and Luis-Felipe Rodríguez. 2016. A cognitive-a ective architecture for ECAs. Biologically Inspired Cognitive Architectures 18 (2016), 33 40.

[74] John K Rempel, John G Holmes, and Mark P Zanna. 1985. Trust in close relationships. Journal of personality and social psychology 49, 1 (1985), 95.

[75] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. " Why should i trust you?" Explaining the predictions of any classi er. In Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining. 1135 1144.

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

[77] Murray Shanahan. 2022. Talking about large language models. arXiv preprint arXiv:2212.03551 (2022).

[78] Si Shi, Yuhuang Gong, and Dogan Gursoy. 2021. Antecedents of trust and adoption intention toward arti cially intelligent recommendation systems in travel planning: a heuristic systematic model. Journal of Travel Research 60, 8 (2021), 1714 1734.

[79] Henrik Singmann and David Kellen. 2019. An introduction to mixed models for experimental psychology. New methods in cognitive psychology 4 (2019), 28 4 (2019).

[80] Detmar Straub, Marie-Claude Boudreau, and David Gefen. 2004. Validation guidelines for IS positivist research. Communications of the Association for Information systems 13, 1 (2004), 24.

[81] Barbara G Tabachnick, Linda S Fidell, and Jodie B Ullman. 2018. Using multivariate statistics. Vol. 6. pearson Boston, MA.

[82] Scott Thiebes, Sebastian Lins, and Ali Sunyaev. 2021. Trustworthy arti cial intelligence. Electronic Markets 31 (2021), 447 464.

[83] Trang P Tran, Chao Wen, and Ilia Gugenishvili. 2023. Exploring the relationship between trusts, likability, brand loyalty, and revisit intentions in the context of Airbnb. Journal of Hospitality and Tourism Technology (2023).

[84] Linda Klebe Trevino. 1992. Experimental approaches to studying ethical-unethical behavior in organizations. Business Ethics Quarterly (1992), 121 136.

[85] Takane Ueno, Yuto Sawa, Yeongdae Kim, Jacqueline Urakami, Hiroki Oura, and Katie Seaborn. 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.

[86] Daniel Ullman and Bertram F Malle. 2019. Measuring gains and losses in human-robot trust: Evidence for di erentiable components of trust. In 2019 14th ACM/IEEE International Conference on Human-Robot Interaction (HRI). IEEE, 618 619.

[87] Daniel Ullrich, Andreas Butz, and Sarah Diefenbach. 2021. The development of overtrust: An empirical simulation and psychological analysis in the context of human robot interaction. Frontiers in Robotics and AI (2021), 554578.

[88] Stuart Van Auken and Thomas E Barry. 1995. An assessment of the trait validity of cognitive age measures. Journal of Consumer Psychology 4, 2 (1995), 107 132.

[89] Tibert Verhagen, Bart van den Hoo , and Selmar Meents. 2015. Toward a better use of the semantic di erential in IS research: An integrative framework of suggested action. Journal of the Association for Information Systems 16, 2 (2015), 1.

[90] Bingcheng Wang, Pei-Luen Patrick Rau, and Tianyi Yuan. 2022. Measuring user competence in using artificial intelligence: validity and reliability of artificial intelligence literacy scale. *Behaviour & Information Technology* (2022), 1–14.

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

[92] Anuradha Welivita, Yubo Xie, and Pearl Pu. 2021. A large-scale dataset for empathetic response generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*. 1251–1264.

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

[94] Jochen Wirtz and Meng Chung Lee. 2003. An examination of the quality and context-specific applicability of commonly used customer satisfaction measures. *Journal of Service Research* 5, 4 (2003), 345–355.

[95] Jixia Yang, Kevin W Mossholder, and TK Peng. 2009. Supervisory procedural justice effects: The mediating roles of cognitive and affective trust. *The Leadership Quarterly* 20, 2 (2009), 143–154.

[96] Ming Yin, Jennifer Wortman Vaughan, and Hanna Wallach. 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.

[97] Guanglu Zhang, Leah Chong, Kenneth Kotovsky, and Jonathan Cagan. 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 (2023), 107536.

[98] Lixuan Zhang, Iryna Pentina, and Yuhong Fan. 2021. Who do you choose? Comparing perceptions of human vs robo-advisor in the context of financial services. *Journal of Services Marketing* (2021).

[99] Rui Zhang, Nathan J McNeese, Guo Freeman, and Geoff Musick. 2021. " An Ideal Human" Expectations of AI Teammates in Human-AI Teaming. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW3 (2021), 1–25.

[100] Yunfeng Zhang, Q Vera Liao, and Rachel KE Bellamy. 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.

[101] Yue Zhu and Syed Akhtar. 2014. How transformational leadership influences follower helping behavior: The role of trust and prosocial motivation. *Journal of organizational behavior* 35, 3 (2014), 373–392.