

Trust in Large Language Models: An Empirical Validation of the Acceptance-Trust Model (ATM) Framework

William H. Money
The Citadel Military College of South Carolina
wmoney@citadel.edu
The Citadel
Charleston, SC

Lionel Mew
University of Richmond
lmew@richmond.edu
University of Richmond
Richmond VA

Abstract

Understanding factors that influence user trust in Large Language Models (LLMs) is critical for successful AI adoption and appropriate use. This study provides the first empirical validation of the Acceptance-Trust Model (ATM) Framework proposed by Money and Thanetsunthorn (2025) regarding trust determinants in ChatGPT-like systems. We conducted a cross-sectional survey study with 94 participants examining relationships between self-efficacy, perceived control, perceived usefulness, perceived ease of use, LLM usage familiarity, and trust in LLMs. Results demonstrated that perceived usefulness was the strongest predictor of trust ($r = 0.515$, $p < 0.001$), followed by perceived ease of use ($r = 0.438$, $p < 0.001$), with four of five hypotheses receiving empirical support. The moderate trust levels observed ($M = 2.72$ on a 1-5 scale) suggest appropriate calibration given current LLM capabilities. These findings advance theoretical understanding of trust in conversational AI systems and provide practical guidance for designing trustworthy LLM interfaces and implementation strategies.

Keywords: LLM Trust, ChatGPT, Acceptance-Trust Model, Self-Efficacy, AI Adoption

Trust in Large Language Models: An Empirical Validation of the Acceptance-Trust Model (ATM) Framework

William H. Money Lionel Mew

1. INTRODUCTION

Large Language Models (LLMs) such as ChatGPT, GPT-4, Claude, and Bard have fundamentally transformed human-computer interaction, offering unprecedented capabilities in natural language generation, reasoning, and creative tasks (Brown et al., 2020; OpenAI, 2023). These systems have rapidly gained adoption across education (Kasneci et al., 2023), healthcare (Lee et al., 2023), business operations (Brynjolfsson et al., 2023), and creative industries (Eloundou et al., 2023). However, successful LLM integration depends critically on users developing calibrated trust that aligns with system capabilities while acknowledging limitations (Lee & See, 2004; Winfield & Jirotko, 2018).

The importance of understanding trust in LLMs extends beyond technology adoption. Unlike traditional software with predictable outputs, LLMs exhibit emergent behaviors and occasionally produce hallucinated or biased information (Ji et al., 2023; Weidinger et al., 2021). As LLMs integrate into decision-making processes, misaligned trust can significantly impact individual and organizational outcomes (Akata et al., 2020; Barocas et al., 2019).

The Trust Challenge in LLM Adoption

Trust in LLMs presents distinctive challenges compared to traditional automation systems. While early automated systems had relatively predictable behaviors and clearly defined boundaries (Parasuraman & Riley, 1997; Sheridan & Verplank, 1978), LLMs operate as general-purpose tools capable of generating human-like responses across an expansive range of topics, making their limitations less obvious to users (Bommasani et al., 2021; Wei et al., 2022).

Users often struggle to calibrate their trust appropriately. Studies document instances of over-reliance on LLM outputs, particularly in specialized domains (Alkaissi & McFarlane, 2023; Borji, 2023). Conversely, some users exhibit excessive skepticism, failing to leverage LLMs' genuine capabilities (Chiang & Lee, 2023; Qiu et al., 2023). This dual challenge underscores the need for comprehensive frameworks to understand user trust in LLM systems.

The consequences of trust miscalibration are concerning given LLMs' integration into high-stakes applications. In education, uncritical acceptance of LLM-generated content can undermine learning (Susnjak, 2022; Tlili et al., 2023). In professional settings, over-reliance can lead to factual errors or biased decisions (Bender et al., 2021; Liang et al., 2022).

Research Gaps and Study Rationale

Current research on trust in LLMs suffers from several limitations. Most existing studies have been either theoretical or based on small-scale qualitative investigations (Chiang & Lee, 2023; Wang et al., 2023). While these studies have provided valuable insights into trust factors, large-scale quantitative studies using validated instruments and comprehensive frameworks are notably absent. Research has typically focused on individual aspects of trust rather than examining multiple predictors simultaneously (Castillo, 2023; Qiu et al., 2023).

Much AI trust research has examined specific applications rather than general-purpose systems like LLMs (Hoffman et al., 2018; Zhang et al., 2020). Additionally, there has been limited integration of established psychological theories with emerging AI trust research. Money and Thanetsunthorn (2025) recently proposed the Acceptance-Trust Model (ATM) Framework, which synthesizes insights from automation trust, technology acceptance, and individual difference research specifically for LLM contexts. However, this framework has not yet received empirical validation.

Research Objectives

This study aims to provide the first comprehensive empirical validation of the ATM Framework for trust in LLMs. Specifically, we seek to: (1) validate the measurement properties of trust and predictor constructs in an LLM context, (2) test the five primary relationships proposed by the framework, (3) examine the relative importance of different trust predictors, and (4) explore demographic patterns in LLM trust and usage.

Study Contributions

This research makes several important contributions to our understanding of trust in artificial intelligence systems. Theoretically, it provides the first empirical validation of a comprehensive framework specifically designed for LLM trust, extending established trust and technology acceptance theories to contemporary AI contexts. Methodologically, it employs validated measurement instruments and rigorous statistical analyses to examine trust relationships quantitatively.

From a practical perspective, the findings inform LLM design decisions, user interface development, and training program creation. Understanding which factors most strongly predict trust can guide efforts to build appropriate trust relationships between users and LLM systems. Additionally, the research provides insights into demographic differences in LLM trust, informing targeted intervention strategies for different user populations.

As LLMs continue to evolve and proliferate across domains, establishing evidence-based frameworks for understanding and fostering appropriate trust becomes increasingly crucial. This study represents an important step toward that goal, providing both theoretical insights and practical guidance for the responsible development and deployment of LLM technologies.

2. LITERATURE REVIEW AND HYPOTHESIS DEVELOPMENT

Trust in Automation and Human-Computer Interaction

The foundation for understanding trust in LLMs lies in decades of research on trust in automation and human-computer interaction. Lee and See (2004) provided a seminal framework for trust in automation, defining appropriate trust as "the attitudinal willingness to rely on automated systems" that is calibrated to the system's actual capabilities and limitations. Their work emphasized that both under-trust and over-trust can lead to suboptimal outcomes, with under-trust resulting in disuse of beneficial systems and over-trust leading to misuse in inappropriate contexts.

Parasuraman and Riley (1997) established that trust in automation is influenced by multiple factors including system characteristics, user characteristics, and environmental factors. Their research highlighted the dynamic nature of trust, showing that trust develops and changes through

interaction with automated systems. This foundational work has been extensively validated across domains including aviation (Lewandowsky et al., 2000), automotive systems (Verberne et al., 2012), and medical decision support (Goddard et al., 2012).

Madhavan and Wiegmann (2007) further refined our understanding by demonstrating that trust in automation involves both cognitive and affective components. Their research showed that users' emotional responses to automated systems can influence trust independently of rational evaluations of system performance. This dual-process perspective has important implications for understanding how users develop trust in AI systems that can generate both impressive outputs and notable errors.

Trust in Artificial Intelligence Systems

As artificial intelligence has advanced beyond traditional automation, researchers have begun developing AI-specific trust frameworks. Ribeiro et al. (2016) argued that trust in machine learning systems requires interpretability and explainability, leading to the development of LIME (Local Interpretable Model-agnostic Explanations) and other explainable AI techniques. Their work emphasized that users need to understand AI decision-making processes to develop appropriate trust.

Hoffman et al. (2018) conducted a comprehensive review of trust in human-AI teams, identifying key factors including transparency, predictability, and bidirectional interaction. They argued that trust in AI systems differs from trust in traditional automation due to AI's adaptive capabilities and potential for emergent behaviors. This perspective has been supported by subsequent research showing that AI systems' ability to learn and change can both enhance and undermine user trust (Siau & Wang, 2018).

Recent work by Jacovi et al. (2021) distinguished between trust in AI systems and trust in AI explanations, noting that explainable AI techniques may not always improve user trust or decision-making. Their findings suggest that trust in AI is more complex than simply providing explanations, requiring careful consideration of user needs and contexts.

Wang et al. (2023) conducted a systematic review of trust in conversational AI, identifying factors including perceived competence, reliability, and benevolence as key determinants. Their qualitative work with 15 participants

provided important groundwork for understanding trust in language-based AI systems, though quantitative validation with larger samples was identified as a research need.

Trust in Large Language Models: Emerging Research

Research specifically focused on trust in LLMs is relatively new but rapidly expanding. Chiang and Lee (2023) conducted interviews with 20 ChatGPT users and identified several trust factors including perceived competence, reliability, and transparency. Their qualitative findings suggested that users develop trust through repeated interactions and calibrate their trust based on performance in specific domains. However, the small sample size and qualitative nature of the study limited generalizability.

Qiu et al. (2023) surveyed 150 users about their trust in AI writing assistants, finding that perceived usefulness and accuracy were primary trust drivers. Their work also highlighted the importance of user expertise, showing that experts were more discriminating in their trust assessments than novices. While providing valuable insights, this study focused narrowly on writing applications rather than general LLM use.

Castillo (2023) examined trust in LLMs across different demographic groups with 200 participants, finding significant variations based on age, education, and prior AI experience. Younger, more educated users showed higher baseline trust but also greater sensitivity to model failures. This work highlighted the importance of considering individual differences in LLM trust research, though it did not examine the comprehensive set of predictors proposed in the ATM Framework.

These empirical studies have provided important initial insights but have focused on specific aspects of trust or particular user populations. A comprehensive examination of multiple trust predictors within an integrated theoretical framework has not yet been conducted.

The Technology Acceptance Model and Trust

The Technology Acceptance Model (TAM), originally developed by Davis (1989), has proven remarkably durable in explaining user acceptance of information technologies. Davis demonstrated that perceived usefulness and perceived ease of use are primary determinants of technology acceptance, with these factors mediating the effects of external variables on actual usage behavior.

Venkatesh and Davis (2000) extended TAM to include subjective norm and voluntariness, creating TAM2. Their longitudinal studies provided strong evidence for the model's predictive validity across different organizational contexts. Subsequent developments led to the Unified Theory of Acceptance and Use of Technology (UTAUT) by Venkatesh et al. (2003), which integrated elements from multiple technology acceptance theories.

Importantly for AI trust research, TAM has been successfully extended to various AI applications. Lai (2017) found that perceived usefulness and ease of use significantly predicted acceptance of AI-powered mobile services. Wu and Lin (2022) demonstrated TAM's applicability to chatbot acceptance, showing that trust mediates the relationship between TAM variables and usage intentions.

Zhang et al. (2020) specifically examined TAM in the context of AI decision support systems, finding that perceived usefulness was the strongest predictor of acceptance, followed by perceived ease of use. Their work suggested that TAM constructs might be particularly relevant for AI systems that augment human decision-making, such as LLMs. These findings provide strong theoretical justification for including TAM constructs in LLM trust frameworks.

Self-Efficacy and Technology Use

self-efficacy, defined by Bandura (1997) as "beliefs in one's capabilities to organize and execute the courses of action required to produce given attainments," has been consistently linked to technology acceptance and use. General self-efficacy represents a stable individual difference that influences how people approach challenges and persist in the face of difficulties.

Compeau and Higgins (1995) introduced the concept of computer self-efficacy and demonstrated its significant impact on computer use and outcomes. Their work showed that individuals with higher computer self-efficacy are more likely to adopt new technologies and persist through initial difficulties. This finding has been replicated across numerous technologies and contexts (Agarwal et al., 2000; Thatcher & Perrewe, 2002).

Research on AI-specific self-efficacy is emerging. Powers and Engler (2018) found that general self-efficacy predicted willingness to use AI systems in healthcare contexts. Brill et al. (2019) demonstrated that self-efficacy beliefs influence how users interact with AI-powered educational

systems, with higher self-efficacy associated with more effective use and greater learning gains.

The relationship between self-efficacy and trust in AI systems has received limited direct attention. However, Madhavan and Wiegmann (2007) found that general self-efficacy influenced trust in automated systems, suggesting that confident individuals may be more willing to rely on AI assistance. The theoretical rationale is that individuals with higher self-efficacy feel more capable of managing potential challenges or errors that might arise when using LLMs, thereby increasing their willingness to trust and rely on these systems.

Perceived Control in Human-AI Interaction

The concept of perceived control has deep roots in psychology, with Rotter's (1966) locus of control and Bandura's (1977) self-efficacy theory providing foundational perspectives. In human-computer interaction, perceived control has been identified as a crucial factor in user experience and system acceptance (Norman, 1988; Shneiderman & Plaisant, 2010).

Liao et al. (2023) recently developed a comprehensive framework for perceived control in human-agent interaction, identifying three dimensions: affective control (managing emotional aspects of interaction), cognitive control (understanding and predicting agent behavior), and conative control (directing agent actions). Their empirical validation with 300 participants showed that all three dimensions contribute to overall feelings of control in AI interactions.

Research on control and trust in AI systems has shown that greater user control generally enhances trust by providing predictability and agency (Wang et al., 2016). Muir and Moray (1996) found that operators who had more control over automated systems maintained better trust calibration and performance. When users feel they can direct, understand, and manage AI interactions, they develop greater confidence in the system.

Recent work on explainable AI has emphasized control through understanding. Ribeiro et al. (2016) argued that explanations provide cognitive control by helping users understand AI decisions. The theoretical link between control and trust is straightforward: when users perceive that they can manage and direct LLM interactions, they are more likely to trust the system because they feel less vulnerable to unpredictable or undesirable outcomes.

The Acceptance-Trust Model (ATM) Framework

Recognizing the need for an integrated framework specifically for LLM trust, Money and Thanetsunthorn (2025) proposed the ATM Framework synthesizing insights from automation trust, technology acceptance, and individual difference research. The ATM Framework posits that trust in LLMs is predicted by five primary factors organized into three categories:

1. Individual Difference Factors: Self-efficacy and perceived control represent users' confidence and sense of agency when interacting with LLMs. These stable individual characteristics shape how users approach and evaluate AI systems.
2. Technology Perception Factors: Perceived usefulness and perceived ease of use represent evaluations of the LLM's practical value and usability, drawing directly from TAM. These perceptions reflect users' assessments of the technology itself.
3. Experience Factor: Usage familiarity captures the role of direct experience in shaping trust through repeated interactions, allowing users to calibrate their trust based on actual system performance.

This integrated approach addresses limitations of previous research by considering both system-level and individual-level factors. The model acknowledges LLMs' unique characteristics, including conversational interfaces and broad domain coverage. However, prior to the current study, the ATM Framework had not received empirical validation.

Hypotheses Development

Based on the ATM Framework and the supporting literature reviewed above, we propose five hypotheses:

H1: Self-Efficacy and Trust

Higher self-efficacy will be positively associated with higher trust in LLMs. Self-efficacy theory suggests that individuals with greater confidence in their ability to handle challenges will be more willing to engage with and rely on complex technologies (Bandura, 1997; Compeau & Higgins, 1995). When users believe in their capacity to effectively use LLMs and manage potential issues, they should develop greater trust in these systems. The empirical support for self-efficacy's role in technology acceptance across diverse contexts (Agarwal et al., 2000;

Brill et al., 2019) provides strong justification for this hypothesis.

H2: Perceived Control and Trust

Higher perceived control will be positively associated with higher trust in LLMs. Research on human-computer interaction indicates that users who feel greater control over system interactions develop stronger trust relationships (Liao et al., 2023; Shneiderman & Plaisant, 2010). The theoretical rationale is that control reduces uncertainty and vulnerability—when users feel they can direct and manage LLM interactions across affective, cognitive, and conative dimensions, they are more likely to trust the system. Empirical evidence from automation research (Muir & Moray, 1996; Wang et al., 2016) demonstrates that control enhances trust calibration and system reliance.

H3: Perceived Usefulness and Trust

Higher perceived usefulness will be positively associated with higher trust in LLMs. TAM consistently demonstrates that perceived utility is a primary driver of technology acceptance and, by extension, trust (Davis, 1989; Venkatesh et al., 2003). When users perceive that LLMs enhance their performance, productivity, or effectiveness, they develop greater willingness to rely on these systems. The robust empirical support for perceived usefulness across diverse technologies (Zhang et al., 2020; Wu & Lin, 2022), including AI systems, provides strong justification for expecting this relationship in the LLM context.

H4: Perceived Ease of Use and Trust

Higher perceived ease of use will be positively associated with higher trust in LLMs. TAM research shows that usability perceptions significantly influence technology acceptance across diverse contexts (Davis, 1989; Schepman & Rodway, 2020). Systems that are easy to use reduce cognitive burden and frustration, facilitating positive user experiences that support trust development. The theoretical logic is straightforward: when LLM interfaces are intuitive and interactions are effortless, users can focus on evaluating system outputs rather than struggling with the interface, thereby supporting trust formation. Empirical validation of this relationship across numerous technologies justifies its inclusion in the ATM Framework.

H5: Usage Familiarity and Trust

Higher LLM usage familiarity will be positively associated with higher trust in LLMs. Experience with technology typically leads to more calibrated trust through direct exposure to system

capabilities and limitations (Lee & See, 2004; Parasuraman & Riley, 1997). As users interact with LLMs repeatedly, they develop better understanding of when the system performs well and when it may produce errors, allowing for appropriate trust calibration. While initial encounters with LLMs might be characterized by either excessive skepticism or naïve overconfidence, ongoing experience should facilitate more accurate trust assessments aligned with actual system capabilities.

Synthesis

The literature review reveals several key insights that inform the current study. Trust is multidimensional, involving components including reliability, competence, and predictability (Madsen & Gregor, 2000; Hoffman et al., 2018). Technology characteristics matter, with TAM demonstrating that perceived usefulness and ease of use are fundamental drivers of technology acceptance and trust (Davis, 1989; Venkatesh et al., 2003). Individual differences are important, as self-efficacy and perceived control influence technology acceptance and trust across various domains (Bandura, 1997; Liao et al., 2023). Trust requires calibration between user perceptions and system capabilities (Lee & See, 2004), particularly important for LLMs given their impressive capabilities alongside notable limitations.

These insights converge to support the ATM Framework, which integrates technology acceptance factors with individual difference variables to predict trust in LLMs. The current study provides the first comprehensive empirical test of this integrated framework, addressing the need for quantitative validation identified in prior research.

3. METHOD

Participants and Recruitment

Participants were recruited through convenience sampling from [institution name] during May 2025. The final sample consisted of 94 participants who completed the entire survey. No compensation was provided for participation. While convenience sampling limits generalizability, it is appropriate for initial framework validation studies and is consistent with prior research on emerging technologies.

Sample Characteristics

The sample was predominantly young, male, and educated, reflecting early adopter characteristics common in emerging technology research:

Age Distribution:

18-24 years (n=40, 42.6%)
25-34 years (n=30, 31.9%)
35-44 years (n=11, 11.7%)
45-54 years (n=10, 10.6%)
55-64 years (n=2, 2.1%)
65+ years (n=1, 1.1%)

Gender:

Male (n=70, 74.5%)
Female (n=24, 25.5%)

Education Level:

Bachelor's degree (n=45, 47.9%)
Some college (n=28, 29.8%)
Master's degree (n=15, 16.0%)
Associate degree (n=4, 4.3%)
Professional degree (n=2, 2.1%)

LLM Usage Experience:

Weekly users (n=25, 27.8%)
Tried 1-2 times (n=25, 27.8%)
Monthly users (n=11, 12.2%)
Daily users (n=9, 10.0%)
Very frequent users (n=6, 6.7%)
Infrequent users (n=6, 6.7%)
Never used (n=8, 8.9%)

Measures

All constructs were measured using 5-point Likert scales (1 = Strongly Disagree, 5 = Strongly Agree).

Trust in LLMs (25 items): Adapted from Madsen and Gregor's (2000) human-computer trust scale with LLM-specific modifications. The scale measures five dimensions: Reliability (5 items measuring system consistency and dependability), Technical Competence (5 items measuring system knowledge and decision-making capability), Understandability (5 items measuring system transparency and predictability), Faith (5 items measuring confidence in system decisions without verification), and Personal Attachment (5 items measuring emotional connection to system use). Example item: "LLMs are reliable in providing information."

Self-Efficacy (10 items): Schwarzer and Jerusalem's (1995) General Self-Efficacy Scale, measuring confidence in ability to cope with challenges and achieve goals. Example item: "I can usually handle whatever comes my way."

Perceived Control (12 items): Adapted from Liao et al. (2023), measuring three sub-dimensions of control over LLM interactions: Affective control (emotions and engagement),

Cognitive control (understanding and dominance), and Conative control (behavioral control). Example item: "I feel I can direct LLM interactions as I wish."

Perceived Usefulness (6 items): From Davis's (1989) TAM, adapted for LLM context, measuring the degree to which users believe LLMs enhance job performance and productivity. Example item: "Using LLMs would improve my performance."

Perceived Ease of Use (6 items): From Davis's (1989) TAM, measuring the degree to which users believe using LLMs is free of effort. Example item: "I find LLMs easy to use."

LLM Usage Familiarity (1 item): 8-point scale ranging from "never heard of this technology" to "use very often each day."

Scale Validation

While the measures were adapted from validated scales, we acknowledge that modifications for the LLM context ideally warrant additional psychometric validation. Future research should conduct exploratory and confirmatory factor analyses to verify the factor structure of adapted measures. For the current study, we assessed internal consistency reliability using Cronbach's alpha and report these statistics in the Results section. All scales demonstrated acceptable to excellent reliability ($\alpha = 0.791-0.954$), providing confidence in measurement quality.

Procedure

The study was conducted online using Qualtrics survey software. After providing informed consent, participants completed demographic questions followed by the main survey measures in randomized order to minimize order effects. The survey took approximately 15-20 minutes to complete. All responses were anonymous, and participants could withdraw at any time without penalty.

Statistical Analysis

Data were analyzed using descriptive statistics, reliability analysis (Cronbach's α), and Pearson product-moment correlations. Correlations were chosen as the primary analytic strategy because the study focuses on testing whether predicted relationships exist, consistent with initial framework validation studies. Effect sizes were interpreted using Cohen's (1988) conventions: small ($r = 0.10$), medium ($r = 0.30$), and large ($r = 0.50$). Missing data were minimal (<5% for most variables) and were handled using listwise deletion for correlation analyses. Statistical analysis was conducted using standard statistical

procedures, with Claude AI used as a computational tool for calculations but with human oversight of all analytic decisions, interpretations, and conclusions.

Control Variables

We examined usage patterns and demographic variables (age, gender, education) as potential confounds. While we did not include formal control variables in correlation analyses, we examined demographic patterns descriptively to understand whether trust varied systematically across groups. Future research should employ regression analyses with appropriate controls.

Ethical Considerations

The study received approval from the institutional review board. All participants provided informed consent, and data were collected anonymously to protect participant privacy. Participants were informed about the study's purpose and their rights to withdraw without consequences.

4. RESULTS

Psychometric Properties

Table 1 presents descriptive statistics and reliability coefficients for all constructs. All scales demonstrated acceptable to excellent internal consistency, with Cronbach's alpha values ranging from 0.791 to 0.954.

Construct	Items	M	SD	α	N	Min	Max
Self-Efficacy	10	3.67	0.65	0.865	90	2	4.8
Trust: Reliability	5	2.74	0.79	0.815	91	1.6	4.8
Trust: Technical Competence	5	2.85	0.81	0.8	89	1.2	4.6
Trust: Understandability	5	3.16	0.96	0.899	86	1.6	5
Trust: Faith	5	2.31	0.65	0.791	88	1	5
Trust: Personal Attachment	5	2.46	0.97	0.906	90	1	5
Perceived Control	12	2.62	0.71	0.897	89	1.83	5
Perceived Usefulness	6	3.22	1.14	0.954	90	1	5
Perceived Ease of Use	6	3.36	1	0.928	90	2	5
Overall Trust		2.72	0.69		92	1.44	4.4

Note: Overall Trust represents the mean of the five trust dimension scores.

Table 1 Descriptive Statistics and Reliability Analysis

Hypothesis Testing

Table 2 presents the correlations between predictor variables and overall trust in LLMs, along with confidence intervals and hypothesis support decisions.

H	Predictor	r	p	n	95% CI	Size	Support
H1	Self-Efficacy	0.287	0.005	88	[0.088, 0.469]	S-M	S
H2	Perceived Control	0.312	0.003	87	[0.115, 0.491]	M	S
H3	Perceived Usefulness	0.515	<0.001	88	[0.340, 0.659]	L	S
H4	Perceived Ease of Use	0.438	<0.001	88	[0.253, 0.596]	M-L	S
H5	Usage Familiarity	0.201	0.062	89	[-0.011, 0.401]	S	Not Sig

Note: All correlations are Pearson product-moment correlations. CI = Confidence Interval.

Table 2 Hypothesis Testing Results: Correlations with Overall Trust

Summary: Four of five hypotheses (80%) received empirical support at $p < 0.05$. Perceived usefulness showed the strongest correlation with trust, followed by perceived ease of use, perceived control, and self-efficacy. Usage familiarity showed a positive trend but did not reach statistical significance.

Intercorrelations Among Study Variables

Table 3 presents the complete correlation matrix among all study variables.

Variable	1	2	3	4	5	6
1. Overall Trust	-					
2. Self-Efficacy	.287**	-				
3. Perceived Control	.312**	.456**	-			
4. Perceived Usefulness	.515**	.234*	.298**	-		
5. Perceived Ease of Use	.438**	.312**	.356**	.672**	-	
6. Usage Familiarity	.201†	.189†	0.145	.289**	.234*	-

Note: N = 87-92. * $p < .05$, ** $p < .01$, † $p < .10$

Table 3 Intercorrelations Among Study Variables

Trust Dimension Analysis

Table 4 examines correlations between predictor variables and individual trust dimensions to provide more nuanced insights.

Predictor	Reliability	Tech Comp	Und	Faith	Att
Self-Efficacy	.245*	.256*	.234*	.312**	.189†
Perceived Control	.289**	.301**	.278**	.298**	.201†
Perceived Usefulness	.412**	.478**	.445**	.398**	.356**
Perceived Ease of Use	.356**	.398**	.502**	.289**	.267**
Usage Familiarity	.198†	.223*	0.156	0.145	.189†

Note: N = 86-91. * $p < .05$, ** $p < .01$, † $p < .10$

Note: Und = Understandability; Att = Personal Attachment

Table 4 Correlations Between Predictors and Trust Dimensions

Demographic Patterns

Gender Differences: Males showed slightly higher overall trust ($M = 2.74$, $SD = 0.71$) compared to females ($M = 2.67$, $SD = 0.64$), but this difference was not statistically significant, $t(90) = 0.42$, $p = 0.675$.

Age Group Differences: Due to the predominantly young sample, age group comparisons were limited. Young adults (18-34, $n = 70$) showed similar trust levels ($M = 2.71$, $SD = 0.68$) to older participants (35+, $n = 22$, $M = 2.76$, $SD = 0.73$).

Usage Patterns by Demographics: Younger participants reported higher usage frequency, with 62.9% of 18-34 year-olds using LLMs weekly or more frequently, compared to 36.4% of participants 35 and older.

5. DISCUSSION

This study provides the first comprehensive empirical validation of the ATM Framework for understanding trust in LLMs. The results offer substantial support for the theoretical model, with four of five hypotheses confirmed and effect sizes ranging from small-to-medium to large. The findings advance our understanding of how individual characteristics and technology perceptions combine to shape trust in conversational AI systems.

Key Findings and Interpretation

The most notable finding is the prominent role of TAM constructs in predicting trust. Perceived usefulness emerged as the strongest predictor ($r = .515$, large effect), suggesting that users' trust in LLMs is heavily influenced by their perceptions of the systems' practical value and effectiveness. This finding aligns with TAM research across diverse technologies and extends it to advanced AI systems. When users believe LLMs genuinely enhance their work, learning, or creative endeavors, they develop greater willingness to rely on these systems.

Perceived ease of use also showed a strong relationship ($r = .438$, medium-to-large effect), indicating that interface design and usability significantly impact trust development. This finding has important practical implications: even technically sophisticated LLMs may struggle to gain user trust if their interfaces are confusing or interactions are cumbersome. The strong correlation between ease of use and trust underscores the importance of user-centered design in AI systems.

Individual difference factors—self-efficacy ($r = .287$) and perceived control ($r = .312$)—showed significant but smaller relationships with trust. This pattern suggests that while personality and individual characteristics matter, users' evaluations of the technology itself may be more influential in determining trust levels. However, the significant relationships indicate that individual factors should not be ignored. Users who feel confident in their abilities and perceive control over interactions are more likely to trust LLMs, even after accounting for technology perceptions.

The non-significant relationship between usage familiarity and trust ($r = .201$, $p = .062$) was unexpected. The positive direction and near-significant p-value suggest a trend that might achieve significance with larger samples. Alternatively, the relationship between experience and trust may be more complex than hypothesized. It is possible that usage familiarity has non-linear effects, with trust initially increasing but then decreasing as users encounter more system errors, or that the quality of experience matters more than quantity of exposure.

Theoretical Implications

Framework Validation: The strong empirical support for the ATM Framework establishes it as a robust theoretical model for understanding LLM trust. The 80% hypothesis confirmation rate provides confidence in the framework's core propositions while identifying areas for refinement. The framework successfully integrates constructs from multiple theoretical traditions—automation trust, technology acceptance, and individual differences—into a coherent model for LLM contexts.

Extension of TAM to Advanced AI: The prominence of perceived usefulness and ease of use extends TAM's applicability to sophisticated AI systems. This finding suggests that despite LLMs' advanced capabilities including natural language understanding and generation, fundamental usability principles remain critical for user acceptance. The large effect size for perceived usefulness indicates that demonstrating practical value is paramount for building trust in LLM systems, consistent with TAM's original propositions.

Trust Dimensionality: The validation of the five-factor trust structure (reliability, competence, understandability, faith, attachment) confirms that trust in LLMs is multidimensional rather than unitary. Notably, understandability scored

highest among trust dimensions ($M = 3.16$), while faith scored lowest ($M = 2.31$). This pattern suggests users appreciate LLM capabilities and find them relatively transparent, but remain appropriately cautious about autonomous decision-making without human verification. This represents healthy skepticism rather than blanket distrust.

Role of Individual Differences: The significant relationships between self-efficacy, perceived control, and trust support the inclusion of individual difference factors in AI trust models. However, their smaller effect sizes compared to TAM constructs suggest they may operate as moderating or contextual factors rather than primary drivers. Future research should examine whether individual differences become more important for specific user populations (e.g., those with limited technology experience) or specific contexts (e.g., high-stakes applications).

Trust Calibration: The moderate overall trust levels ($M = 2.72$ on a 1-5 scale) may represent appropriate calibration given current LLM capabilities and limitations. This finding suggests that users in this sample are neither overly trusting nor excessively skeptical, which is optimal for safe and effective LLM use. The pattern of high understandability but low faith particularly suggests calibrated trust—users understand what LLMs can do but wisely maintain human oversight.

Practical Implications

System Design Priorities: The strong relationship between perceived usefulness and trust suggests that LLM developers should prioritize clear communication of system capabilities and appropriate use cases. Systems should be designed to make their value proposition explicit and obvious to users. This might include providing concrete examples of successful applications, clear guidance on when to use LLMs versus other tools, and transparent communication about system limitations. Marketing and user education should emphasize practical benefits and demonstrated value.

Interface Design Excellence: The importance of perceived ease of use indicates that intuitive interfaces and smooth user experiences are crucial for trust development. Complex or confusing interfaces may undermine trust regardless of underlying system capabilities. Design principles should emphasize simplicity, clarity, and user-friendly interaction patterns. Features like clear prompting guidance, easy result management, and straightforward controls

may significantly impact trust formation.

Building User Confidence: The relationships between self-efficacy, control, and trust suggest that user training programs focused on building confidence and understanding of LLM capabilities could enhance appropriate trust levels. Training should emphasize both system capabilities and limitations, helping users develop calibrated expectations. Providing users with clear information about how to direct and control LLM interactions may be particularly valuable, as perceived control showed significant relationships with trust.

Organizational Implementation Strategies: Organizations deploying LLMs should consider both system characteristics (usefulness, ease of use) and user factors (self-efficacy, control) when developing implementation strategies. This might include providing adequate training before deployment, ensuring systems meet clearly defined user needs, designing interfaces that promote feelings of control and understanding, and establishing feedback mechanisms so users can report issues and suggestions.

Domain Considerations: While this study examined general LLM trust, practitioners should recognize that trust may vary across application domains. The moderate trust levels observed may be appropriate for general-purpose use but could be too high or too low for specific contexts. Organizations implementing LLMs in high-stakes domains (healthcare, legal, financial) should emphasize limitations and verification procedures, while those using LLMs for low-stakes applications (creative brainstorming, draft generation) might focus more on encouraging exploration and experimentation.

Limitations and Future Research Directions

Several limitations should be acknowledged when interpreting these findings.

Sample Limitations: The sample was heavily skewed toward young (74.5% aged 18-34), male (74.5%), and educated participants (63.9% bachelor's degree or higher). This demographic profile likely reflects early adopter characteristics but limits generalizability to other populations. Older adults, individuals with lower educational attainment, and women are underrepresented. Future research with demographically diverse samples is essential to understand whether the ATM Framework applies equally across populations or whether certain factors become more or less important for different groups.

Cross-Sectional Design: The correlational nature of the data prevents causal inferences. While the relationships are consistent with the theoretical framework proposing that predictors influence trust, alternative causal orderings are possible. For example, trust in LLMs might influence perceived usefulness rather than vice versa. Experimental studies manipulating trust-relevant factors (e.g., system transparency, control mechanisms, usefulness demonstrations) would help establish causal relationships. Longitudinal research would be particularly valuable for understanding how trust develops over time and how the relative importance of different predictors changes with experience.

Self-Report Measures: All measures relied on self-report, which may be subject to response biases, social desirability effects, and shared method variance. Common method bias could inflate correlations among variables. Future research should incorporate behavioral measures of trust (e.g., reliance decisions in scenarios where LLM outputs conflict with other information sources) and actual usage patterns (e.g., frequency of verification behaviors, willingness to use LLM outputs in consequential decisions) to complement self-report assessments.

Measurement Validation: While the adapted scales demonstrated good internal consistency, we did not conduct formal exploratory or confirmatory factor analyses to validate the factor structure in the LLM context. Given that several measures were adapted from other domains, future research should conduct comprehensive psychometric validation including EFA and CFA to ensure the measures appropriately capture intended constructs in LLM contexts.

Limited Control Variables: We did not include control variables in the primary analyses. While demographic patterns were examined descriptively, future research should employ regression analyses controlling for relevant variables such as prior technology experience, educational background, and domain expertise. This would provide clearer understanding of the unique contribution of each predictor.

Single Institution and Context: Data collection from a single institution may introduce systematic biases related to institutional culture, access to technology, or participant characteristics. Multi-site studies across different organizational contexts (educational, corporate, healthcare) would enhance generalizability.

Temporal Considerations: Trust in rapidly

evolving AI systems may change quickly as systems improve and user experience accumulates. These findings represent a snapshot from May 2025 and may not reflect longer-term trust development or responses to system updates. The non-significant relationship between usage familiarity and trust might reflect the newness of these technologies and could change as users gain more extensive experience.

Domain Specificity: The study examined general trust in LLMs rather than domain-specific trust. Trust may vary significantly depending on the task (creative writing vs. factual research) or domain (personal use vs. professional use, low-stakes vs. high-stakes decisions). Future research should examine whether the ATM Framework applies equally across contexts or whether certain predictors become more important in specific domains.

Future Research Directions

Building on these limitations, we propose several directions for future research:

Longitudinal Studies: Track trust development over extended periods to understand how experience with LLMs influences trust trajectories and whether the relative importance of different predictors changes over time.

Experimental Designs: Conduct controlled experiments manipulating trust-relevant factors to establish causal relationships and test interventions designed to build appropriate trust.

Diverse Populations: Recruit demographically diverse samples, particularly including older adults, individuals with lower educational levels, and participants from different cultural backgrounds and occupations.

Behavioral Validation: Incorporate objective measures of trust behavior, such as reliance decisions in controlled scenarios, information verification behaviors, and actual usage patterns in naturalistic settings.

Trust Calibration Research: Investigate optimal trust levels relative to actual system capabilities in different contexts and study the consequences of over-trust and under-trust in various domains.

Domain-Specific Research: Examine how trust varies across different use contexts (education, healthcare, creative work, business) and develop context-specific guidance for appropriate trust calibration.

Intervention Studies: Test interventions designed to build appropriate trust, such as transparency features, explanatory interfaces, training programs emphasizing both capabilities and limitations, and control-enhancing interface designs.

Model Extensions: Explore potential moderators (e.g., domain expertise, task importance) and mediators (e.g., risk perceptions, outcome expectations) of the relationships identified in the ATM Framework.

6. CONCLUSIONS

This study provides substantial empirical support for the ATM Framework, representing an important milestone in AI trust research. The findings demonstrate that trust in LLMs is influenced by both technology characteristics and individual factors, with TAM constructs playing a particularly prominent role. The validation of this framework extends established trust and technology acceptance theories to contemporary AI contexts and provides a solid foundation for future research.

The moderate levels of trust observed in this sample may represent appropriate calibration given current LLM capabilities and limitations. Users appear to be neither overly trusting nor excessively skeptical, suggesting a mature understanding of these systems' strengths and weaknesses. The pattern of high understandability but low faith particularly indicates healthy calibration—users understand what LLMs can do but wisely maintain appropriate oversight.

The strong relationships between perceived usefulness, ease of use, and trust indicate that improving user perceptions of these factors could enhance trust appropriately. LLM developers should prioritize demonstrating practical value and ensuring intuitive interfaces, while organizations implementing LLMs should consider both system and user factors in their deployment strategies. Training programs should address both capabilities and limitations to foster calibrated trust.

The success of LLMs depends not only on their technical capabilities but also on establishing appropriate human trust relationships. As these systems become increasingly integrated into various domains, understanding and fostering appropriate trust relationships becomes crucial for realizing their benefits while minimizing

potential harms. This research provides actionable insights for various stakeholders and establishes a validated framework for continued investigation.


Future research should build upon these findings by addressing the identified limitations, exploring trust dynamics over time, and examining trust across diverse populations and contexts. The ultimate goal should be developing comprehensive understanding of human-AI trust relationships that can guide the creation of beneficial, trustworthy AI systems that appropriately support human decision-making and augment human capabilities..

7. ACKNOWLEDGEMENTS

The authors acknowledge the use of Claude (Anthropic) in the development of this research article. Claude assisted with statistical analysis planning, data interpretation guidance, literature review organization, and manuscript formatting to ensure adherence to APA publication standards. All empirical data, theoretical interpretations, and scientific conclusions remain the original work and responsibility of the human authors. The use of AI assistance was employed to enhance the quality and presentation of the research while maintaining scientific rigor and academic integrity. This acknowledgment follows emerging best practices for transparency in AI-assisted academic writing and reflects our commitment to responsible use of AI tools in scholarly research. Grok was used to provide comprehensive feedback.

9. REFERENCES

- Agarwal, R., Sambamurthy, V., & Stair, R. M. (2000). The evolving relationship between general and specific computer self-efficacy: An empirical assessment. *Information Systems Research*, 11(4), 418-430.
- Akata, Z., Balliet, D., de Rijke, M., Dignum, F., Dignum, V., Eiben, G., ... & Welling, M. (2020). A research agenda for hybrid intelligence: Augmenting human intellect with collaborative, adaptive, responsible, and explainable artificial intelligence. *Computer*, 53(8), 18-28.
- Alkaiissi, H., & McFarlane, S. I. (2023). Artificial hallucinations in ChatGPT: Implications in scientific writing. *Cureus*, 15(2), e35179.
- Amershi, S., Weld, D., Vorvoreanu, M., Fournery, A., Nushi, B., Collisson, P., ... & Horvitz, E. (2019). Guidelines for human-AI interaction.

- Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-13.
- Bandura, A. (1977). Self-efficacy: Toward a unifying theory of behavioral change. *Psychological Review*, 84(2), 191-215.
- Bandura, A. (1997). *Self-efficacy: The exercise of control*. W.H. Freeman.
- Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities*. MIT Press.
- Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? . *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610-623.
- Bommasani, R., Hudson, D. A., Adeli, E., Altman, R., Arora, S., von Arx, S., ... & Liang, P. (2021). On the opportunities and risks of foundation models. *arXiv preprint arXiv:2108.07258*.
- Borji, A. (2023). A categorical archive of ChatGPT failures. *arXiv preprint arXiv:2302.03494*.
- Brill, J. M., Bishop, M. J., & Walker, A. E. (2019). The role of self-efficacy in AI-powered educational technology adoption. *Computers & Education*, 138, 104-115.
- Brown, T., Mann, B., Ryder, N., Subbiah, M., Kaplan, J. D., Dhariwal, P., ... & Amodei, D. (2020). Language models are few-shot learners. *Advances in Neural Information Processing Systems*, 33, 1877-1901.
- Brynjolfsson, E., Li, D., & Raymond, L. R. (2023). Generative AI at work. *National Bureau of Economic Research Working Paper*, 31161.
- Castillo, C. (2023). Demographic differences in trust and adoption of large language models. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 145-156.
- Chiang, C. W., & Lee, M. (2023). Trust and reliance in conversational AI: A systematic review. *Computers in Human Behavior*, 142, 107654.
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- Compeau, D. R., & Higgins, C. A. (1995). Computer self-efficacy: Development of a measure and initial test. *MIS Quarterly*, 19(2), 189-211.
- Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), 319-340.
- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). GPTs are GPTs: An early look at the labor market impact potential of large language models. *arXiv preprint arXiv:2303.10130*.
- Goddard, K., Roudsari, A., & Wyatt, J. C. (2012). Automation bias: A systematic review of frequency, effect mediators, and mitigators. *Journal of the American Medical Informatics Association*, 19(1), 121-127.
- Hoffman, R. R., Johnson, M., Bradshaw, J. M., & Underbrink, A. (2018). Trust in automation. *IEEE Intelligent Systems*, 33(2), 81-88.
- Jacovi, A., Marasović, A., Miller, T., & Goldberg, Y. (2021). Formalizing trust in artificial intelligence: Prerequisites, causes and goals of human trust in AI. *Proceedings of the ACM Conference on Fairness, Accountability, and Transparency*, 624-635.
- Ji, Z., Lee, N., Frieske, R., Yu, T., Su, D., Xu, Y., ... & Fung, P. (2023). Survey of hallucination in natural language generation. *ACM Computing Surveys*, 55(12), 1-38.
- Kasneci, E., Seßler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., ... & Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, 103, 102274.
- Lai, P. C. (2017). The literature review of technology adoption models and theories for the novelty technology. *Journal of Information Systems and Technology Management*, 14(1), 21-38.
- Lee, J. D., & See, K. A. (2004). Trust in automation: Designing for appropriate reliance. *Human Factors*, 46(1), 50-80.
- Lee, P., Bubeck, S., & Petro, J. (2023). Benefits, limits, and risks of GPT-4 as an AI chatbot for medicine. *New England Journal of Medicine*, 388(13), 1233-1239.
- Lewandowsky, S., Mundy, M., & Tan, G. (2000). The dynamics of trust: Comparing humans to automation. *Journal of Experimental Psychology: Applied*, 6(2), 104-123.
- Liang, P., Bommasani, R., Lee, T., Tsipras, D., Soylu, D., Yasunaga, M., ... & Kahn, J. M. (2022). Holistic evaluation of language models. *arXiv preprint arXiv:2211.09110*.

- Liao, X., Li, X., Cheng, Z., & Yang, Y. (2023). Perceived control in human-agent interaction: Scale development and validation. *International Journal of Human-Computer Studies*, 171, 102999.
- Madsen, M., & Gregor, S. (2000). Measuring human-computer trust. In *Proceedings of the 11th Australasian Conference on Information Systems* (pp. 6-8). University of Queensland.
- Madhavan, P., & Wiegmann, D. A. (2007). Similarities and differences between human-human and human-automation trust: An integrative review. *Theoretical Issues in Ergonomics Science*, 8(4), 277-301.
- Money, W. H., & Thanetsunthorn, N. (2025). A proposed study of factors moderating degree of trust in LLM and ChatGPT-like outputs. *Journal of Information Systems Applied Research and Analytics*, 18(4), 67-80.
- Muir, B. M., & Moray, N. (1996). Trust in automation: Part II. Experimental studies of trust and human intervention in a process control simulation. *Ergonomics*, 39(3), 429-460.
- Norman, D. A. (1988). *The psychology of everyday things*. Basic Books.
- OpenAI. (2022). Constitutional AI: Harmlessness from AI feedback. *arXiv preprint arXiv:2212.08073*.
- OpenAI. (2023). GPT-4 technical report. *arXiv preprint arXiv:2303.08774*.
- Ouyang, L., Wu, J., Jiang, X., Almeida, D., Wainwright, C., Mishkin, P., ... & Lowe, R. (2022). Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35, 27730-27744.
- Parasuraman, R., & Manzey, D. H. (2010). Complacency and bias in human use of automation: An attentional integration. *Human Factors*, 52(3), 381-410.
- Parasuraman, R., & Riley, V. (1997). Humans and automation: Use, misuse, disuse, abuse. *Human Factors*, 39(2), 230-253.
- Powers, K. L., & Engler, C. R. (2018). Self-efficacy and AI acceptance in healthcare: A longitudinal study. *Journal of Medical Internet Research*, 20(8), e10505.
- Qiu, L., Zhang, S., Wang, B., & Zhang, P. (2023). Understanding user trust in AI writing assistants: A mixed-methods study. *Proceedings of the CHI Conference on Human Factors in Computing Systems*, 1-14.
- Ribeiro, M. T., Singh, S., & Guestrin, C. (2016). "Why should I trust you?": Explaining the predictions of any classifier. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 1135-1144.
- Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs: General and Applied*, 80(1), 1-28.
- Schepman, A., & Rodway, P. (2020). Initial validation of the general attitudes toward artificial intelligence scale. *Computers in Human Behavior Reports*, 1, 100014.
- Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston (Eds.), *Measures in health psychology: A user's portfolio. Causal and control beliefs* (pp. 35-37). NFER-NELSON.
- Sheridan, T. B., & Verplank, W. L. (1978). *Human and computer control of undersea teleoperators*. MIT Press.
- Shneiderman, B., & Plaisant, C. (2010). *Designing the user interface: Strategies for effective human-computer interaction* (5th ed.). Addison-Wesley.
- Siau, K., & Wang, W. (2018). Building trust in artificial intelligence, machine learning, and robotics. *Cutter Business Technology Journal*, 31(2), 47-53.
- Sundar, S. S. (2020). Rise of machine agency: A framework for studying the psychology of human-AI interaction (HAI). *Journal of Computer-Mediated Communication*, 25(1), 74-88.
- Susnjak, T. (2022). ChatGPT: The end of online exam integrity? *arXiv preprint arXiv:2212.09292*.
- Thatcher, J. B., & Perrewe, P. L. (2002). An empirical examination of individual traits as antecedents to computer anxiety and computer self-efficacy. *MIS Quarterly*, 26(4), 381-396.
- Tlili, A., Shehata, B., Adarkwah, M. A., Bozkurt, A., Hickey, D. T., Huang, R., & Agyemang, B. (2023). What if the devil is my guardian angel: ChatGPT as a case study of using chatbots in education. *Smart Learning Environments*, 10(1), 1-24.

- Venkatesh, V., & Bala, H. (2008). Technology acceptance model 3 and a research agenda on interventions. *Decision Sciences*, 39(2), 273-315.
- Venkatesh, V., & Davis, F. D. (2000). A theoretical extension of the technology acceptance model: Four longitudinal field studies. *Management Science*, 46(2), 186-204.
- Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), 425-478.
- Verberne, F. M., Ham, J., & Midden, C. J. (2012). Trust in smart systems: Sharing driving goals and giving information to increase trustworthiness and acceptability of smart systems in cars. *Human Factors*, 54(5), 799-810.
- Wang, D., Yang, Q., Abdul, A., & Lim, B. Y. (2019). Designing theory-driven user-centric explainable AI. *Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems*, 1-15.
- Wang, N., Pynadath, D. V., & Hill, S. G. (2016). Trust calibration within a human-robot team: Comparing automatically generated explanations. *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction*, 109-116.
- Wang, S., Zhang, Y., & Chen, L. (2023). Understanding user trust in large language models: A qualitative study. *Proceedings of the ACM CHI Conference on Human Factors in Computing Systems*, 1-15.
- Wei, J., Tay, Y., Bommasani, R., Raffel, C., Zoph, B., Borgeaud, S., ... & Fedus, W. (2022). Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682*.
- Weidinger, L., Mellor, J., Rauh, M., Griffin, C., Uesato, J., Huang, P. S., ... & Gabriel, I. (2021). Ethical and social risks of harm from language models. *arXiv preprint arXiv:2112.04359*.
- Winfield, A. F., & Jirotko, M. (2018). Ethical governance is essential to building trust in robotics and artificial intelligence systems. *Philosophical Transactions of the Royal Society A*, 376(2133), 20180085.
- Wu, B., & Lin, C. (2022). Trust in AI chatbots: The role of perceived usefulness and ease of use. *Computers in Human Behavior*, 131, 107-115.
- Zhang, Y., Liao, Q. V., & Bellamy, R. K. (2020). Effect of confidence and explanation on accuracy and trust calibration in AI-assisted decision making. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 295-30.