EAI Endorsed Transactions
on Pervasive Health and Technology
Research Article EAI.EU

Gender Differences in Human Trust and Contextual Perception of Robots

Yang Rao[1], Ruquan Yang[1,*], Shuangyi He[1],Wenyi Lin[1], and Ruizhao Cai[2,3]

[1]Hexiangning College of Art and Design, Zhongkai University of Agriculture and Engineering, No. 24, Dongsha Street, Haizhu District, Guangzhou City, Guangdong Province, China
[2]People's Government of Jiangdong Town, No. 1 Fuqian Street, Jiangdong Town, Chaoan District, Chaozhou City, Guangdong Province, China
[3]Jiangdong Town Industrial Development Service Center, Chaoan District, Chaozhou City, Guangdong Province, China

Abstract

INTRODUCTION: Trust is a prerequisite for safe and effective Human–Robot Interaction (HRI), yet reported gender differences are inconsistent and likely contingent on context and socio-perceptual processes.
OBJECTIVES: Within a unified framework spanning four canonical HRI contexts (healthcare, education, manufacturing, security), we test whether (a) gender predicts trust, (b) context moderates gender effects, and (c) perceived warmth and perceived threat mediate gender–trust relations.
METHODS: A vignette-based experiment with adults (N = 132; male/female) measured affective and cognitive trust, perceived warmth, and perceived threat on 7-point scales. Analyses followed a preregistered plan: 2×4 mixed ANOVAs (Gender × Context) and parallel mediation (PROCESS Model 4; 5,000 bootstrap resamples) with covariates (age, education, prior HRI experience).
RESULTS: Gender showed a significant main effect for affective trust (females > males), but not for cognitive trust. Context effects were significant for both trust facets. Gender × Context interactions emerged: the female advantage in affective trust was concentrated in healthcare, while males reported higher cognitive trust in education and manufacturing. Mediation indicated that females' higher perceived warmth and lower perceived threat jointly accounted for gender differences in overall trust; the direct gender effect was not significant after including mediators. Robustness checks (ANCOVAs; order effects) supported all primary findings.
CONCLUSION: Gender differences in robot trust are context-dependent and arise via warmth-enhancing and threat-reducing socio-perceptual pathways. Design should emphasize empathy/assurance cues in caring roles and competence/reliability cues in task/authority roles, alongside systematic threat mitigation.

Keywords: Gender Differences, Robot Trust, Perceived Warmth, Perceived Threat, Vignette-Based Experiment, Human-Robot Interaction.

Received on 22 November 2025, accepted on 06 January 2026, published on 19 January 2026

doi: 10.4108/eetpht.11.11058

1. Introduction

Robots are increasingly deployed in socially sensitive domains-healthcare, education, manufacturing, and public safety-where failures and misunderstandings carry human, ethical, and organizational costs. In such settings, trust is a decisive precondition for safe and effective Human–Robot Interaction (HRI). Building on contemporary accounts that differentiate trust into affective (feelings of comfort and psychological safety) and cognitive (perceived competence

*Corresponding author. Email: hhover@163.com

and reliability) components, we note that gender is often invoked to explain variation in trust, yet empirical findings remain inconsistent. The study of human–machine relationships and trust must consider the sociotechnical systems in which robots are embedded. Research has highlighted how algorithms, as technological objects, shape trust perceptions through normative design, influencing behavioral plasticity and interaction outcomes in diverse contexts[1]. A growing consensus is that these inconsistencies reflect context dependence and underlying socio-perceptual processes [2], rather than a uniform gender effect [3].

Despite progress, three limitations constrain current understanding. First, many studies investigate single contexts, hindering direct comparison of gender effects across roles that differ in risk, authority, and prosocial expectations. Second, theory, measurement, and model specification are not always aligned [4], leading to construct drift and mixed inferences about what is being predicted as 'trust'. Third, mechanistic accounts are underdeveloped: while competence cues are well studied, the roles of perceived warmth (benevolence) and perceived threat (anticipated harm/discomfort) as socio-affective pathways remain insufficiently tested.

The present work addresses these gaps by employing a unified experimental design that standardizes scenarios across four canonical HRI contexts and jointly examines Gender × Context effects on affective and cognitive trust. We articulate a theory-driven framework in which perceived warmth and perceived threat operate as parallel mediators linking gender to overall trust, thereby specifying when and why gender differences emerge.

2. Related Work

Scholarship on trust in Human–Robot Interaction (HRI) has evolved from a performance-centric view to a socio-cognitive perspective. Foundational work established reliability, predictability, and error rates as primary antecedents of trust and developed widely used measurement approaches for trust in automation [5][6][7]. More recent surveys consolidate these insights and call for integrative accounts that incorporate cultural and social factors alongside system performance [8][9].

As robots pervade social domains, social signaling-anthropomorphism, politeness, and non-verbal behavior-has been shown to shape users' trust judgments in tandem with capability cues[10]. Experimental and review evidence indicates that perceived warmth and competence are robust predictors of trust-related appraisals in HRI, and scale work has begun to refine their operationalization for efficient use in studies [11][12].

A parallel line of inquiry considers human attributes, with gender frequently examined yet yielding mixed results: some studies document greater hesitancy or anxiety among women, whereas others report higher acceptance in care-oriented interactions. Comparative and empirical work suggests that such discrepancies arise from differences in task demands, perceived risk, and role framing, rather than a uniform gender effect [13]. This pattern highlights context as a pivotal moderator of gender–trust relations [14].

Contextual moderation is evident across healthcare, education, manufacturing, and security, where expectations about appropriateness, authority, and safety diverge [15]. Recent domain-specific studies and reviews underscore that trust and acceptance are jointly shaped by role-specific affordances and risk profiles-e.g., service/care robots in healthcare, tutoring or AI-EdTech tools in education, collaborative robots (cobots) on shop floors, and patrol/surveillance robots in public safety-yet most investigations still isolate single contexts, limiting direct tests of Gender × Context interactions [16].

Beyond documenting correlates, scholars increasingly probe mechanisms. Converging evidence shows that warmth (benevolence, prosocial intent) elevates affective components of trust, whereas cues linked to hazard raise perceived threat (or discomfort) and can depress trust even when competence is high [17][18]. These pathways motivate the present focus on socio-perceptual mediators.

2.1. Risk Communication and Social Threat in HRI

The role of risk communication and social threat in shaping trust perceptions has gained significant attention in the context of HRI. Social threat refers to the concerns and perceived risks arising from human–robot interactions, including feelings of being judged, controlled, or monitored, which can reduce trust and engagement. Social evaluation anxiety, in particular, can significantly mediate individuals' trust in robots, especially in safety-critical settings where the robot's actions may be perceived as potentially harmful or evaluative [19].

Risk communication focuses on how uncertainty and risk are conveyed to individuals, shaping their decision-making and trust. In the case of robots, how risks are communicated-through transparency, error-handling strategies, and reliability statements-affects the perceived threat and trust formation. Studies have shown that transparent communication about a robot's capabilities and limitations can mitigate the sense of threat and increase trust, especially in critical domains like healthcare and security [20]. Conversely, lack of transparency or ambiguous communication can exacerbate perceptions of threat and diminish trust, particularly among users with lower technology familiarity or higher social evaluation concerns [21][22].

2.2. Hypotheses and Conceptual Model

Let G denote gender (0 = male, 1 = female) and C the within-subject factor Context (healthcare, education, manufacturing, security). Trust is decomposed into affective trust and cognitive trust. Mediators are perceived warmth and perceived threat.

H1 (Gender main effect on affective trust). Females report higher affective trust than males.

H2 (Gender × Context for affective trust). Gender differences in affective trust are moderated by context and are largest in healthcare (caring) scenarios.

H3 (Gender × Context for cognitive trust). Gender differences in cognitive trust are moderated by context, with males showing higher cognitive trust in manufacturing (task) and education (guidance) contexts.

H4 (Warmth mediation). G → higher perceived warmth → higher trust.

H5 (Threat mediation). G → lower perceived threat → higher trust.

H6 (Full mediation for overall trust). After accounting for warmth and threat, the direct effect of G on overall trust (mean of affective and cognitive) is not significant.

3. Methodology and System Design

This study employed a computer-administered, vignette-based experiment to examine how participant gender (between-subjects) and interaction context (within-subjects) jointly shape trust in robots and through which socio-perceptual mechanisms (perceived warmth and perceived threat) these effects arise. A scenario paradigm affords tight control of stimuli, systematic manipulation of context, and scalable data collection, while maintaining ecological plausibility via domain-typical narratives. While vignettes support controlled cross-context comparison and mechanism testing, they capture trust attitudes rather than behavioral reliance; thus, generalization to embodied interaction should be treated cautiously and validated with behavioral and in-situ measures.

3.1. Research Framework

We implemented a 2 (Gender: male, female; between) × 4 (Context: healthcare, education, manufacturing, security; within) mixed design. Each participant evaluated all four contexts, enabling estimation of Gender × Context interactions while controlling for stable individual differences. The primary outcomes were affective trust and cognitive trust; mediators were perceived warmth and perceived threat. Figure 1 provides the conceptual diagram linking these variables.

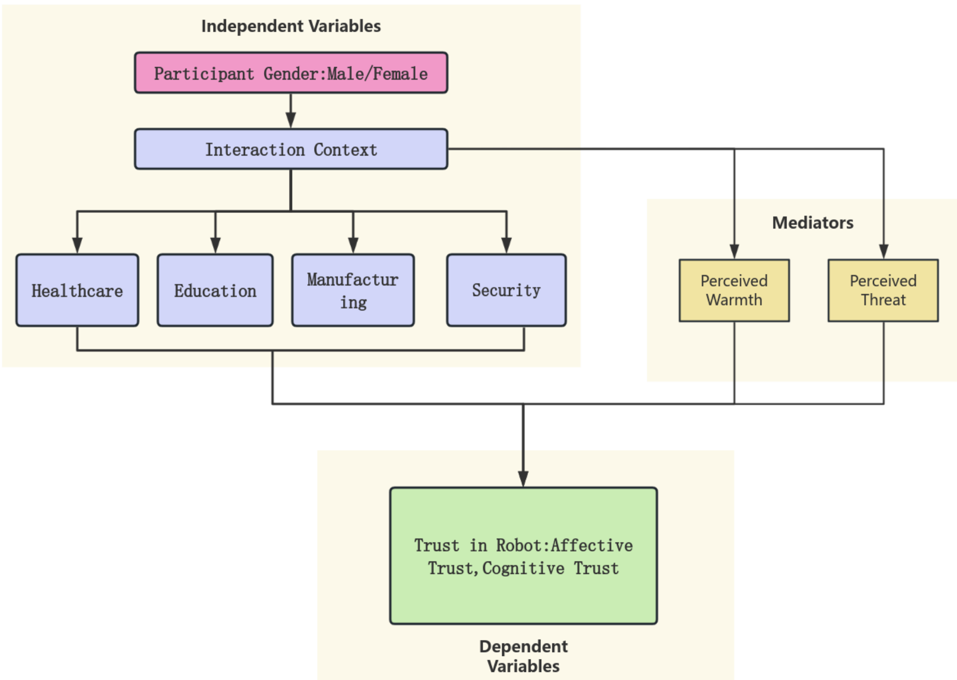

Figure 1. Conceptual Framework of the Study

3.2. Participants

We recruited 140 adults from a professional online participant pool. Inclusion criteria were age ≥ 18. Participants selected gender from Male / Female / Non-binary/Other; because the present study focuses on binary gender differences, data from participants choosing Non-binary/Other were not analyzed. Eight participants were excluded for failing attention checks or selecting Non-binary/Other, yielding a final N=132 (N_{Male}=66 , N_{Female}=66). An a priori G Power analysis (mixed ANOVA, medium effect f=0.25, α=.05, power = .80) indicated that

N≈128 would be adequate, supporting the achieved sample size.

3.3. Materials and Stimuli

We developed four standardized, text-only vignettes (≈120–140 words each) representing canonical HRI domains:

- Healthcare (Caring): a home-care robot providing medication reminders and vital-sign checks.
- Education (Guidance): a tutoring robot delivering step-wise explanations and adaptive quizzes.
- Manufacturing (Task-oriented): a collaborative robot coordinating part placement and safety stops.
- Security (Authority): a patrol robot offering guidance and reporting hazards.

Stimuli were controlled for robot appearance (neutral description), task complexity, tone, and privacy disclosures; proper names and locales were neutralized. The full texts are provided in Appendix A.

3.4. Measures

All measures used 7-point Likert scales (1 = strongly disagree, 7 = strongly agree).

- Affective trust (4 items; e.g., I feel secure with the robot), adapted from established trust-in-automation instruments.
- Cognitive trust (4 items; e.g., The robot is reliable), aligned with competence/reliability facets.
- Perceived warmth (6 items), using the RoSAS Warmth subscale.
- Perceived threat (5 items), adapted from prior work on robot-elicited unease/threat.

The full texts are provided in Appendix B.

3.5. Procedure

After consent and demographics, participants completed four vignette blocks presented in counterbalanced order (Latin square). Within each block, participants read the vignette, responded to the trust/warmth/threat items, and then completed manipulation-check ratings. Attention checks and minimum engagement time thresholds were embedded to promote data quality.

3.6. Data Quality and Exclusions

The following preregistered criteria were applied: (i) exclude any participant failing attention check(s); (ii) flag completion times < 1/3 of the median per-page or > 3× IQR as careless/extreme and exclude upon pattern confirmation; (iii) handle missingness via person-mean imputation when ≤ 1 item is missing per subscale; otherwise drop that vignette's subscale for the participant.

4. Experiments and Results

4.1. Descriptive Statistics

Table 1 presents the descriptive statistics for affective and cognitive trust across the four interaction contexts and by gender. Across contexts, affective trust tended to be higher in the healthcare scenario and lower in security, whereas cognitive trust was highest in manufacturing and education. Gender-specific means indicate that females reported higher affective trust in most contexts, while males generally reported higher cognitive trust in task- and guidance-oriented contexts.

Table 1. Trust by gender and context (1–7)

Context	Gender	N	Affective Trust	Cognitive Trust
Healthcare (Caring)	Female	66	5.85 ± 0.92	5.12 ± 1.05
	Male	66	5.41 ± 1.10	5.35 ± 0.98
Education (Guidance)	Female	66	5.22 ± 1.01	5.58 ± 0.89
	Male	66	5.05 ± 1.15	5.81 ± 0.82
Manufacturing (Task)	Female	66	4.88 ± 1.12	5.45 ± 0.95
	Male	66	4.75 ± 1.20	5.79 ± 0.88
Security (Authority)	Female	66	4.55 ± 1.25	5.01 ± 1.10
	Male	66	4.30 ± 1.30	5.25 ± 1.02

4.2. Mixed ANOVA for Trust

To test H1–H3, we conducted two 2 (Gender: male, female; between) × 4 (Context: healthcare, education, manufacturing, security; within) mixed ANOVAs, one for affective trust and one for cognitive trust. Greenhouse–Geisser corrections were applied when sphericity assumptions were violated. (details shown in Table 2)

Table 2. Summary of Mixed ANOVA

Source	Dependent Variable	F	df	p	η_p^2	95% CI	Correction
Gender	Affective	4.95	1,130	0.028	0.037	[0.03, 0.50]	None
Gender	Cognitive	0.88	1,130	0.349	0.007	[−0.15, 0.42]	None
Context	Affective	12.11	2.89, 375.7	<.001	0.085	[0.18, 0.45]	G–G
Context	Cognitive	8.77	3, 390	<.001	0.063	[0.10, 0.35]	None

G × C	Affective	3.55	2.89, 375.7	0.015	0.027	[0.01, 0.15]	G–G
G × C	Cognitive	4.12	3, 390	0.007	0.031	[0.02, 0.18]	None

Affective Trust:

- Main effect of Gender: Significant, $F(1,130) = 4.95, p = .028, \eta_p^2 = .037$. Females reported higher affective trust.
- Main effect of Context: Significant, $F(2.89,375.7) = 12.11, p < .001, \eta_p^2 = .085$. Trust was highest in healthcare and lowest in security.
- Gender × Context interaction: Significant, $F(2.89,375.7) = 3.55, p = .015, \eta_p^2 = .027$. Simple-effects analyses showed a significant gender difference only in healthcare, where females reported higher affective trust ($p = .035$). No significant gender differences emerged in the other three contexts.

Cognitive Trust:

- Main effect of Gender: Not significant, $F(1,130) = 0.88, p = 0.349$.
- Main effect of Context: Significant, $F(3,390) = 8.77, p < .001, \eta_p^2 = .063$.Cognitive trust was highest in manufacturing.
- Gender × Context interaction: Significant, $F(3,390) = 4.12, p = .007, \eta_p^2 = .031$. Simple-effects analyses indicated higher cognitive trust for males in education (p = .041) and manufacturing (p = .035)

4.3. Mediation Analysis

To test H4–H6, we examined whether perceived warmth and perceived threat mediated the effect of Gender on overall trust (average of affective and cognitive trust). PROCESS Model 4 was used with 5,000 bootstrap resamples and covariates (age, education, prior HRI experience). All continuous covariates were mean-centered prior to analysis, and 95% bias-corrected bootstrap confidence intervals were used to evaluate indirect effects.

Mediation Results

Table 3 reports unstandardized coefficients.

- Gender significantly predicted higher warmth and lower threat.
- Warmth positively predicted overall trust; threat negatively predicted it.
- The direct effect of Gender on overall trust was not significant, indicating full mediation.
- Both indirect paths-via Warmth and via Threat-were significant.

Table 3. Summary of Parallel Mediation Analysis

Path	Predictor → Mediator/Outcome	b	SE	t	p	95% CI
a	Gender (1=female) → Warmth	0.45	0.12	3.75	<.001	[0.21, 0.69]
	Gender (1=female) → Threat	-0.30	0.10	-3.00	.003	[-0.50, -0.10]
b	Warmth → Overall trust	0.52	0.08	6.50	<.001	[0.36, 0.68]
	Threat → Overall trust	-0.35	0.07	-5.00	<.001	[-0.49, -0.21]
c	Gender (1=female) → Overall trust	0.10	0.15	0.67	.503	[-0.20, 0.40]

4.4. Robustness Checks

To assess robustness, we conducted mixed ANCOVAs including age and prior HRI experience as covariates. The main effects of gender and context, as well as the Gender × Context interaction for affective trust, remained significant. We also tested for order effects using Latin-square assignment and found no impact on any trust measure, confirming adequate counterbalancing.

4.5. Visual Representation of Interaction

The significant Gender × Context interactions are illustrated in Figure 2 and Figure 3.

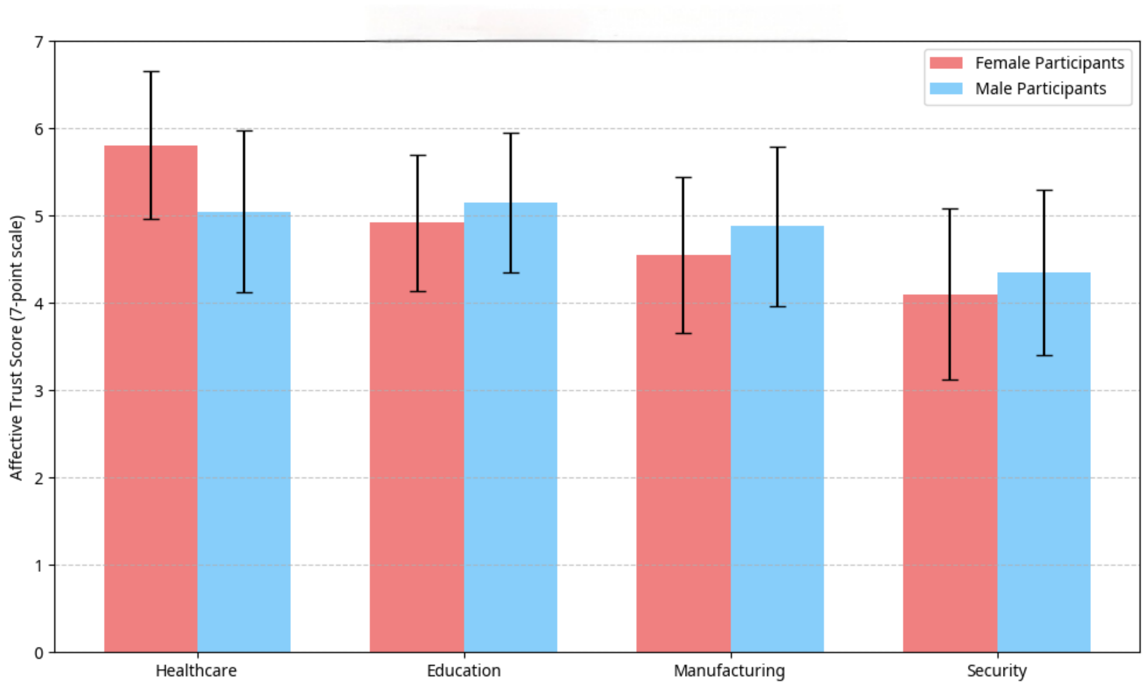

Figure 2. Gender × Context Interaction on Affective Trust

Figure 2. Gender × Context Interaction on Affective Trust. Error bars represent the Standard Error of the Mean (SE). The asterisk indicates a significant difference between genders within the Healthcare context ($p < .05$). $N = 132$.

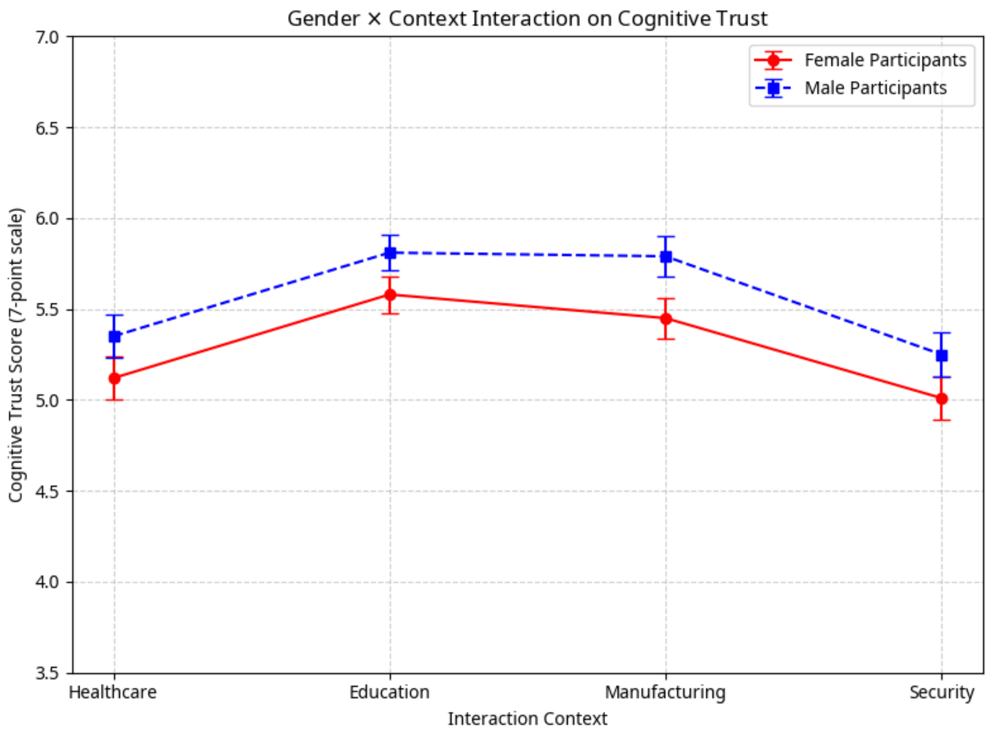

Figure 3. Gender × Context Interaction on Cognitive Trust

Figure 3. Gender × Context Interaction on Cognitive Trust. Error bars represent the Standard Error of the Mean (SE). Asterisks () indicate significant differences between genders within the Education and Manufacturing contexts ($p < .05$). $N = 132$.

5. Analysis and Discussion

This section interprets the empirical findings in light of the conceptual model and prior work on human–robot trust, gender, and social perception. We first summarize how the results map onto the preregistered hypotheses (H1–H6), then discuss the implications for theory and design in Human–Robot Interaction (HRI).

5.1. Summary of Findings Relative to the Hypotheses

The mixed-design ANOVAs and mediation analyses provide convergent support for the proposed framework. First, H1 predicted a main effect of gender on affective trust, with females reporting higher levels than males. This hypothesis was supported: across contexts, females exhibited significantly higher affective trust, whereas no corresponding main effect of gender emerged for cognitive trust. This pattern indicates that gender differences are more pronounced for trust as a feeling of comfort and psychological safety than for trust as perceived competence or reliability.

H2 and H3 posited that gender effects on trust would be context-dependent. The results align closely with these expectations. For affective trust, a significant Gender × Context interaction showed that the female advantage was concentrated in the healthcare (caring) context, with no reliable gender differences in education, manufacturing, or security. For cognitive trust, gender did not show a main effect, but the Gender × Context interaction revealed that males reported higher cognitive trust than females in education (guidance) and manufacturing (task) contexts, consistent with the idea that these scenarios foreground competence- and performance-related expectations.

The mediation analysis addressed H4–H6, which proposed that perceived warmth and perceived threat operate as parallel socio-perceptual mechanisms linking gender to overall trust. The results are consistent with this account. Gender significantly predicted both mediators (females perceiving robots as warmer and less threatening), and both warmth and threat, in turn, were strong predictors of overall trust. Critically, the direct effect of gender on overall trust became non-significant once warmth and threat were included, supporting H6 and indicating full mediation. Together, these findings demonstrate that gender influences trust primarily by shaping social perceptions of the robot, rather than exerting an independent, residual effect.

5.2. Contextualized Gender Differences in Affective and Cognitive Trust

The observed dissociation between affective and cognitive trust refines existing discussions of gender in HRI. Rather than a uniform 'women trust robots more/less than men' narrative, the data suggest that:

- Females are more likely to report higher affective trust-feeling safer and more at ease-especially when the robot's role is aligned with caring and support, as in the healthcare context.
- Males are more likely to report higher cognitive trust-confidence in reliability and task performance-when the robot is embedded in guidance- and task-focused roles, such as education and manufacturing.

This pattern is consistent with the idea that gendered expectations about roles and domain norms shape how users interpret the same robot behavior. Healthcare scenarios are culturally associated with nurturing, empathy, and interpersonal support; in such contexts, females may be more attuned to, and reassured by, cues that the robot is attentive and benevolent, which inflates affective trust. In contrast, education and manufacturing emphasize accuracy, procedural clarity, and efficiency-domains in which males may have stronger expectations of, or familiarity with, performance-oriented systems, resulting in higher cognitive trust.

Importantly, the absence of a global gender main effect on cognitive trust reinforces that context, not gender per se, drives many of the differences. This nuance helps reconcile prior empirical findings that reported inconsistent gender effects in different HRI settings. Studies focusing on a single context may detect a gender difference that does not generalize beyond that specific role or risk profile; only by comparing multiple contexts within the same design can the Gender × Context structure be fully revealed.

5.3. Socio-Perceptual Mechanisms: Warmth and Threat

The mediation results clarify how gender becomes linked to trust. Drawing on frameworks of social perception, the study hypothesized that perceived warmth (benevolence, prosocial intent) and perceived threat (anticipated harm, unease) form two complementary pathways through which gender shapes trust evaluations. The data support this mechanism-based view.

Females, on average, perceived the robots as warmer and less threatening than males did. In turn, higher warmth was associated with higher overall trust, whereas higher threat was associated with lower overall trust. Once these mediators were taken into account, the direct effect of gender on trust was no longer significant, indicating that gender differences in trust are fully accounted for by differences in these socio-perceptual appraisals rather than by gender itself as a standalone determinant.

This pattern has several theoretical implications:

- It aligns HRI research with broader social cognition theories, in which warmth and threat-related appraisals often precede and shape trust judgments.
- It suggests that what might appear as 'gender differences in trust' at the surface level actually reflect deeper differences in how male and female participants construe the intentions and potential risks associated with robots.
- It underscores the importance of negative socio-affective pathways: even when competence is high, heightened perceptions of threat can depress trust. Designing solely for competence signals may therefore be insufficient.

By demonstrating parallel mediation, the present study extends prior work that focused primarily on competence or anthropomorphism by showing that warmth-enhancing features and threat-mitigating cues jointly explain gendered trust patterns across contexts.

6. Conclusion

This study provides a unified test of how gender and interaction context jointly shape trust in robots and through which socio-perceptual mechanisms these effects arise. Across four canonical HRI domains, females reported higher affective trust, while cognitive trust showed no overall gender main effect. Crucially, both trust facets exhibited Gender × Context interactions: the female advantage in affective trust was concentrated in healthcare, whereas males reported higher cognitive trust in education and manufacturing. Parallel mediation analyses indicated that perceived warmth (positive pathway) and perceived threat (negative pathway) jointly accounted for gender differences in overall trust; the direct gender effect was not significant once these mediators were included.

This work also clarifies construct alignment by distinguishing affective from cognitive trust and by specifying warmth and threat as parallel mediators. Several limitations warrant attention. First, because the analysis focused on binary gender, the generality of effects to non-binary and gender-diverse identities remains to be established. Second, vignette-based methods afford control but should be complemented by in-situ studies with physical robots to assess ecological validity. Third, our context set covered four salient domains; expanding to additional social and professional settings will further test boundary conditions. Future work should also consider latent-variable modeling of trust, cross-cultural samples, and longitudinal designs that link perceptions to behavioral cooperation over time.

Appendix A. The first appendix

Below are the full texts of the four standardized vignettes used in the study. Each vignette is approximately 120–140 words, written in neutral language, and controls for robot appearance (not specified), task complexity, tone, and data-handling statements.

A1. Healthcare (Caring) Vignette

You are recovering at home under routine medical supervision. A waist-high mobile robot arrives daily to remind you to take prescribed medication and to record temperature and blood pressure. It greets you in a calm voice, confirms your name, and explains each step before proceeding. The robot repeats instructions on request, slows its pace when asked, and asks whether you want to contact your clinician if any reading is outside the safe range. A small display shows status messages (e.g., 'measuring,' 'recorded,' 'complete'). The robot stores your readings locally and shares them only with your designated clinician. Today, the robot arrives at the usual hour, verifies the medication schedule, and begins the routine check-in, waiting while you take your pills and asking if you would like a brief walk afterward.

Note. All vignettes were matched in length, tone, and data-handling disclosures to the extent possible; minor domain-specific phrases were retained to preserve role plausibility.

A2. Education (Guidance) Vignette

You are preparing for an upcoming exam. A desktop tutoring robot provides step-by-step explanations and checks your understanding with short quizzes. It adapts to your pace, highlights mistakes without judgment, and offers practice items that match your current level. The robot pauses regularly to ask whether to continue, review, or slow down, and it summarizes key points at the end of each segment. A small indicator shows when it is listening or processing. Your quiz scores are recorded for your private review only and are not shared with others. Today, the robot introduces a new problem set, asks how confident you feel before starting, and suggests a quick warm-up exercise to refresh foundational concepts.

Note. All vignettes were matched in length, tone, and data-handling disclosures to the extent possible; minor domain-specific phrases were retained to preserve role plausibility.

A3. Manufacturing (Task-Oriented) Vignette

You are working on a light-assembly line. A collaborative robot (cobot) places components while you secure them. It follows a clear schedule, signals movements with lights and speech, and automatically pauses when you enter its workspace. A screen displays status states such as ‘ready’, ‘placing’, ‘paused’, and ‘awaiting confirmation’. When a misalignment occurs, the cobot alerts you, requests confirmation, and proposes a recovery routine. Today, the cobot detects a part-tolerance deviation and asks whether to switch to the backup procedure. The system logs tasks locally for quality control; no personal data are stored. You can request the cobot to slow down, repeat an action, or halt if conditions seem unsafe, and it acknowledges each command before proceeding.

Note. All vignettes were matched in length, tone, and data-handling disclosures to the extent possible; minor domain-specific phrases were retained to preserve role plausibility.

A4. Security (Authority) Vignette

You are in a public building where a patrol robot circulates along predefined routes. It greets visitors, answers simple questions about directions, and reports hazards such as spills or blocked exits to human staff. If it detects unusual behavior or unattended items, it notifies human security personnel for follow-up. The robot records only low-resolution images for incident review and posts visible signs describing this policy. It keeps a respectful distance, yields to pedestrians, and announces turns to avoid surprises. Today, the robot stops near a closed stairwell and announces that maintenance is in progress, asking visitors to use the elevator. It waits until a staff member arrives, then resumes its route after confirming the area is safe.

Note. All vignettes were matched in length, tone, and data-handling disclosures to the extent possible; minor domain-specific phrases were retained to preserve role plausibility.

Appendix B: Measurement Items (Excerpt)

All items are measured on a 7-point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree).

B.1. Affective Trust (Adapted from Jian et al., 2000) 1. I feel secure with the robot. 2. I am comfortable with the robot. 3. I am not afraid of the robot. (Reverse-scored) 4. I feel a sense of connection with the robot.

B.2. Cognitive Trust (Adapted from Jian et al., 2000) 1. The robot is reliable. 2. The robot is competent. 3. The robot is dependable. 4. I can count on the robot to perform its task.

B.3. Perceived Warmth (RoSAS Warmth Subscale, Carpinella et al., 2017) 1. The robot seems friendly. 2. The robot seems sincere. 3. The robot seems good-natured. 4. The robot seems warm. 5. The robot seems sociable. 6. The robot seems emotional.

B.4. Perceived Threat (Adapted from Social Evaluation Anxiety) 1. I would feel judged by the robot. 2. The robot seems dangerous. 3. I feel uneasy when the robot is near. 4. The robot could potentially harm me. 5. I feel threatened by the robot's presence.

Conflict of Interest Statement
No potential conflict of interest was reported by the authors.

Informed Consent Statement:
Informed consent was obtained from all subjects involved in the study.

Ethical Statement
The study was approved by the Institutional Review Board of Zhongkai University of Agriculture and Engineering.

References

[1] Li J, Wan R. Research on Ethical Design for Silicon-Based Life Forms. Philosopher's Compass. 2024 Oct 1;1(1):10-7.
[2] Christoforakos L, Gallucci A, Surmava-Große T, Ullrich D, Diefenbach S. Can robots earn our trust the same way humans do? A systematic exploration of competence, warmth, and anthropomorphism as determinants of trust development in HRI. Frontiers in Robotics and AI. 2021;8:640444. doi:10.3389/frobt.2021.640444.
[3] Law T, Chita-Tegmark M, Scheutz M. The interplay between emotional intelligence, trust, and gender in human–robot interaction: A vignette-based study. International Journal of Social Robotics. 2021;13(2):297–309. doi:10.1007/s12369-020-00624-1.
[4] de Souza DF, Sousa S, Kristjuhan-Ling K, Dunajeva O, Roosileht M, Pentel A, Mõttus M, Özdemir MC, Gratšjova Ž. Trust and trustworthiness from a human-centered perspective in HRI: A systematic literature review. arXiv [cs.RO]. 2025;arXiv:2501.19323.
[5] Muir BM. Trust in automation: Part I. Theoretical issues in the study of trust and human intervention in automated systems. Ergonomics. 1994;37(11):1905–1922. doi:10.1080/00140139408964957.
[6] Carragher DJ, Sturman D, Hancock PJB. Trust in automation and the accuracy of human–algorithm teams performing one-to-one face matching tasks. Cognitive Research: Principles and Implications. 2024;9:41. doi:10.1186/s41235-024-00564-8.
[7] Jian JY, Bisantz AM, Drury CG. Foundations for an empirically determined scale of trust in automated systems. International Journal of Cognitive Ergonomics. 2000;4(1):53–71. doi:10.1207/S15327566IJCE0401_04.
[8] Kok BC, Soh H. Trust in robots: Challenges and opportunities. Current Robotics Reports. 2020;1(4):297–309. doi:10.1007/s43154-020-00029-y.
[9] Rossi A, Holthaus P, Perugia G, Moros S, Scheunemann M. Trust, acceptance and social cues in human–robot interaction (SCRITA). International Journal of Social Robotics. 2021;13(8):1833–1834. doi:10.1007/s12369-021-00844-z.
[10] Roesler E, Vollmann M, Manzey D, Onnasch L. The dynamics of human–robot trust attitude and behavior-Exploring the effects of anthropomorphism and type of failure. Computers in Human Behavior. 2024;150:108008. doi:10.1016/j.chb.2023.108008.
[11] Carpinella CM, Wyman AB, Perez MA, Stroessner SJ. The Robotic Social Attributes Scale (RoSAS): Development and validation. In: Proceedings of the 2017 ACM/IEEE International Conference on Human–Robot Interaction (HRI 2017); 2017 Mar 6–9; Vienna, Austria. New York: ACM; 2017. p. 254–262. doi:10.1145/2909824.3020208.
[12] Neuenswander KL, Dash A, Koya PD, Lin L, Gillespie GSR, Stroessner SJ. Measuring fundamental aspects of the social perception of robots: Development and validation of a shortened version of the RoSAS (RoSAS-SF). International Journal of Social Robotics. 2025;17(6):1097–1112. doi:10.1007/s12369-025-01251-4.
[13] Boyapati YM, Khan A. Gender differences in robot acceptance. In: Health Informatics and Medical Systems and Biomedical Engineering (CSCE 2024). Cham: Springer; 2025. (Communications in Computer and Information Science; 2259). p. 351–361. doi:10.1007/978-3-031-85908-3_28.
[14] Lim WM, Jasim KM, Malathi A. Service robots in healthcare: Toward a healthcare service robot acceptance model (sRAM). Technology in Society. 2025;82:102932. doi:10.1016/j.techsoc.2025.102932.

[15] Pietrantoni L, Favilla M, Fraboni F, Mazzoni E, Morandini S, Benvenuti M, De Angelis M. Integrating collaborative robots in manufacturing, logistics, and agriculture: Expert perspectives on technical, safety, and human factors. Frontiers in Robotics and AI. 2024;11:1342130. doi:10.3389/frobt.2024.1342130.
[16] Nazaretsky T, Mejia-Domenzain P, Swamy V, Frej J, Käser T. The critical role of trust in adopting AI-powered educational technology for learning: An instrument for measuring student perceptions. Computers and Education: Artificial Intelligence. 2025;8:100368. doi:10.1016/j.caeai.2025.100368.
[17] Shidujaman M, Samani H. Creating trustworthy patrol robot with an ethical design approach. In: Proceedings of the 2024 2nd International Conference on Robotics, Control and Vision Engineering (RCVE 2024); 2024 Jul 19–21; Hong Kong, China. New York: ACM; 2024. p. 24–29. doi:10.1145/3685073.3685078.
[18] Dhanda M, Rogers BA, Hall S, Dekoninck E, Dhokia V. Reviewing human-robot collaboration in manufacturing: Opportunities and challenges in the context of Industry 5.0. Robotics and Computer-Integrated Manufacturing. 2025;93:102937. doi:10.1016/j.rcim.2024.102937.
[19] Nass C, Moon Y. Machines and mindlessness: Social responses to computers. Journal of social issues. 2000;56(1):81-103.
[20] Patel J, Sonar P, Pinciroli C. On multi-human multi-robot remote interaction: a study of transparency, inter-human communication, and information loss in remote interaction. Swarm Intelligence. 2022 Jun;16(2):107-42. doi:10.1007/s11721-022-00241-x.
[21] De Simone V, Di Pasquale V, Giubileo V, Miranda S. Human-Robot Collaboration: An analysis of worker's performance. Procedia Computer Science. 2022 Jan 1;200:1540-9. doi:10.1016/j.procs.2022.01.218.
[22] Schaefer KE, Billings DR, Szalma JL, Adams JK, Sanders TL, Chen JY, Hancock PA. A meta-analysis of factors influencing the development of trust in automation: Implications for human-robot interaction. Human Factors. 2014 Jul 1;56(3):529-55. doi:10.1177/0018720814539763.