Article

A Study of Factors Influencing User Trust in Human-Robot Interaction

Xutong Ding , **Yixuan Zhou***

Zhejiang Normal University, Jinhua, 321001, China
*Correspondence: chowyh@qq.com

Citation: Ding, X. T., & Zhou, Y. X. (2025). A study of factors influencing user trust in Human-Robot interaction. *Journal of Arts & Cultural Studies*, 4 (2), 1-19. https://doi.org/10.23112/acs25073101

Check for updates

Received: May 10, 2025
Revised: June 15, 2025
Accepted: July 12, 2025
Published: July 31, 2025

Publisher's Note: KIHSS stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Abstract: ***Background:*** With the rapid advancement of AI and robotics, Human-Robot Interaction (HRI) has been increasingly adopted in medical care, educational counselling, service reception and other scenarios. User trust directly impacts robots' acceptance and operational effectiveness, and the lack of such trust has become a key barrier to the development of human-robot collaboration. ***Purpose:*** This paper focuses on the construction mechanism of user trust in HRI, aiming to explore its key influencing factors and action paths to provide theoretical support and design insights for trust-oriented interaction design. ***Methods:*** The study combines literature analysis, typical case analysis, and cross-case comparison, summarizing core design features and user feedback of representative human-robot products at home and abroad. ***Results:*** This study proposes a multi-factor collaborative trust construction framework. It emphasizes that the design of future human-robot interaction should focus on establishing a dynamic trust relationship. This approach aims to improve the system's transparency, predictability and emotional connection to enhance the user acceptance and reliance on robots. The findings of this study are significant for optimizing user experience and increasing the social acceptance of intelligent interactive products ***Conclusion:*** The findings are of great significance for optimizing user experience and enhancing the social acceptance of intelligent interactive products, with practical implications for HRI design practice.

Keywords: Human-robot interaction; User trust; Interaction design; Trust influencing factors; Case studies

1. Introduction

1.1 Research Background

In recent years, with the ongoing advancement of artificial intelligence (AI) and robotics technology, service, social, and collaborative robots have rapidly developed globally. Human-Robot Interaction (HRI) is gradually shifting from traditional industrial automation to more complex and diverse social interaction environments. Unlike industrial robots, the new generation performs tasks and aims to serve as "social participants". However, research has shown that "user trust" is a key factor for successful coexistence experiences with humans and robots (Hancock, 2011; Sanders, 2019). Reliability has a greater impact on trust than environmental or human characteristics (r ≈ 0.26, d ≈ 0.71), suggesting that design should focus more on the robot's factors (Hancock, 2011). More importantly, robots with highly human-like features can create a "valley of terror" effect, where users neither trust nor accept the robot and instead feel uncomfortable or resentful if its behavior does not match its capabilities (Mori, 2012; Saygin, 2012). An anthropomorphic appearance does not automatically generate trust - trust can only be established when behaviors meet users' psychological expectations, emotional responses, and control feedback.

In addition, Lee states that trust depends not only on whether a robot is reliable but also on whether it possesses mechanisms of "comprehensibility, predictability, and transparent interpretation" (Lee, 2004). Breazeal further emphasizes that robots expressing emotional and social behaviors are more likely to inspire affective trust in their users, provided these expressions are appropriate and contextually relevant (Breazeal, 2003). User trust is a highly dynamic, context-dependent, multifactorial intertwiner - reconciling these elements in design remains a challenge in current research and practice. Based on this background, this study focuses on the "influencing factors of user trust in human-robot interaction". It explores the law of trust generation and designs implementation paths from the perspective of design strategies and interaction mechanisms.

1.2 Research Purpose

With "user trust" as the core topic, this study explores how trust is built between two parties in human-robot interaction environments through case analysis, theoretical discussion, and empirical research. It aims to identify how these factors affect user trust and offer design suggestions, providing both theoretical support and practical guidance for the design of future human-robot interactions.

Thus, main research objectives are as follows: 1) To clarify the components of trust as a dynamic variable; 2) To analyze the impact of robot appearance, behavioral performance, emotional expression, and interaction scenarios on users' trust; 3) To propose an operational "trust-oriented interaction design strategy" to provide theoretical and practical support for future product development and system implementation design.

1.3 Research Design and Framework

This study investigates the key factors influencing user trust in HRI through literature analysis and case comparison. It first synthesizes both domestic and international research to summarize the core paradigms and classification perspectives found in HRI trust studies. It focuses on four major factors: robot appearance, behavioral consistency, emotional expression, and scene adaptability. Next, the study analyzes representative cases (e.g., care, service, social robots) to examine how design strategies shape users' trust and perceptual responses. It explores the interactions among these factors and proposes a multidimensional synergistic framework for constructing trust. Finally, the findings are applied to interaction design practices, offering a trust-oriented approach from the user's perspective to address common trust barriers in real-world robotic applications.

The technical roadmap of this study follows a logical progression of "theoretical analysis – data collection and case exploration – literature and statistical analysis – stage synthesis and improvement", forming a complete research process that integrates theory, mechanism exploration, and design innovation. A combination of literature research, case study, narrative inquiry, historical analysis, and qualitative approach methods is used to ensure systematic rigor and methodological diversity.

In the theoretical research and problem analysis stage, the study begins with a conceptual examination of "user trust." It reviews the evolution and structure, as well as the current state of HRI research. By comparing domestic and international studies, the research identifies theoretical gaps and develops a multi-scenario, multi-path framework for trust formation. This stage establishes the conceptual and methodological foundation for future empirical exploration.

In the data collection and case analysis stage, representative HRI products, such as Japan's Pepper and Robina, China's "Xiaoyi" medical robot, and MIT's Tega robot, are analyzed through a cross-cultural lens. The study investigates the mechanisms and design strategies that influence user trust, focusing on psychological and cognitive aspects, including transparency, behavioral consistency, comprehensibility, and user control. Additionally, the analysis highlights the shift from functional design to more socially oriented design dimensions.

During the literature and statistical analysis stage, insights from prior research and empirical findings are synthesized to classify and interpret the dynamic relationships among key influencing factors—appearance and anthropomorphism, behavioral and emotional consistency, scene adaptation, and transparency mechanisms. Regulatory variables such as trust accumulation, expectation setting, and feedback loops are further analyzed to develop an interactive model that explains how trust evolves over time.

The achievement synthesis of this stage summarizes the core mechanisms involved in building trust. It includes factors like multidimensional synergy, transparency, emotional resonance, and feedback loops. Additionally, it identifies design implications across various fields such as education, healthcare, and public services, forming a practical framework for trust-oriented design strategies.

In the final stage of improvement and innovation, the research integrates theoretical insights and design practices to propose future directions for research and innovation. These include modeling trust through explainable AI, incorporating ethical and cultural diversity principles, and developing trust evaluation methods for specific user groups. This stage completes the research loop, transitioning from theory to design, offering actionable guidance for enhancing user trust in human–robot collaboration.

Overall, the roadmap outlines a clear methodological progression—from theoretical foundation to empirical validation, mechanism synthesis, and design innovation—ensuring that the findings are both theoretically sound and practically applicable to HRI system design and optimization.

1.4 Research Methods

This study adopts a mixed-methods approach that integrates a systematic literature review with a purposive case analysis to identify and interpret key factors influencing user trust in human–robot interaction (HRI). Peer-reviewed studies published between 2000 and 2024 were retrieved from major databases such as Web of Science, Scopus, and IEEE Xplore using keywords related to "robot trust," "anthropomorphism," and "transparency." Four representative robot types—service, assistive, educational, and social—were selected as cases to illustrate diverse trust-building mechanisms. Each case was examined through qualitative content analysis, supported by cross-case comparison and triangulation across published empirical results, technical reports, and user studies. The coding process adhered to trust-theoretical dimensions, including performance reliability, transparency, and expectation management. This combined approach ensures both theoretical depth and practical validity, while also acknowledging the limitations of secondary data and cross-cultural generalization (Yin, 2018; Creswell, 2014).

2. Synthesis of Research

2.1 Development and Core Features of Human-Robot Interaction (HRI)

HRI has undergone significant development since the mid-20th century, with its core features evolving from simple command-responsive interactions to today's complex, multimodal, and emotionally intelligent interactions. With the advancement of technology, robots are no longer limited to repetitive tasks on industrial production lines but have started to play an essential role in various fields, such as healthcare, education, and home services (Goodrich, 2007). According to the definition of Fong et al., the subject of interaction in HRI systems includes unidirectional information transfer between humans and robots. It emphasizes bidirectional communication relationships formed in task collaboration, situational adaptation, and social feedback (Fong, 2003).

According to the International Federation of Robotics (IFR), global sales of service robots reached $11.2 billion in 2019 and are expected to grow to $24 billion by 2022 (IFR, 2020). In the healthcare sector, the da Vinci Surgical System has successfully performed more than 7.5 million minimally invasive surgeries, including a haptic feedback system that accurately mimics the surgeon's operating force. This growth not only reflects the maturity of the technology but also reveals the widespread societal demand for HRI. This demand is particularly evident in light of an aging society, with Japan seeing a 67% year-on-year surge in the deployment of care robots by 2023.

The core features of HRI include: 1) Autonomy—robots independently make decisions based on perceived data; 2) Interactivity—achieving natural communication via multimodal channels (speech, movement, vision); 3) Adaptability—responding flexibly to user behavior and context; 4) Sociability—recognizing and reacting to social cues. HRI is evolving from task-focused functional interaction to psychologically centered emotional interaction (Breazeal, 2003), which requires technical optimization and a transformation in design philosophy, user perception models, and ethical frameworks.

Building on the theoretical foundation of user trust in HRI, this framework conceptualizes trust as a multidimensional concept shaped by psychological, behavioral, and contextual determinants. As illustrated in Figure 1, user trust in HRI is based on three theoretical foundations—Social Exchange Theory, Cognitive–Affective Trust Theory, and Human Factors & UX Design—which together explain how users perceive risk, reliability, and emotional engagement during their interactions with robots.

Four primary dimensions of trust-influencing factors are identified: 1) Appearance and Anthropomorphism, which relate to the robot's humanlike form and perceived social presence; 2) Behavior and Consistency, which emphasize predictability and reliability of robotic responses; 3) Emotion and Sociality, which involve empathy, warmth, and emotional connection; 4) Contextual Adaptation and Expectation, which highlight situational appropriateness and expectation alignment.

Supporting literature emphasizes these aspects—Mara and Appel (2015) describe the tension between anthropomorphism and trust; Waytz et al. (2014) analyze social agency as a determinant of trust perception; and Desai et al. (2012) stress the importance of behavioral consistency and transparency.

The role of user trust is explained through three interrelated levels: first, trust serves as the cornerstone of collaboration in HRI; second, it influences users' perception of system risk and transparency; and third, it directly affects users' acceptance, reliance on, and ongoing engagement with robotic systems. Collectively, this framework provides a theoretical foundation that integrates human, technological, and contextual elements to explain the dynamic nature of trust in HRI.

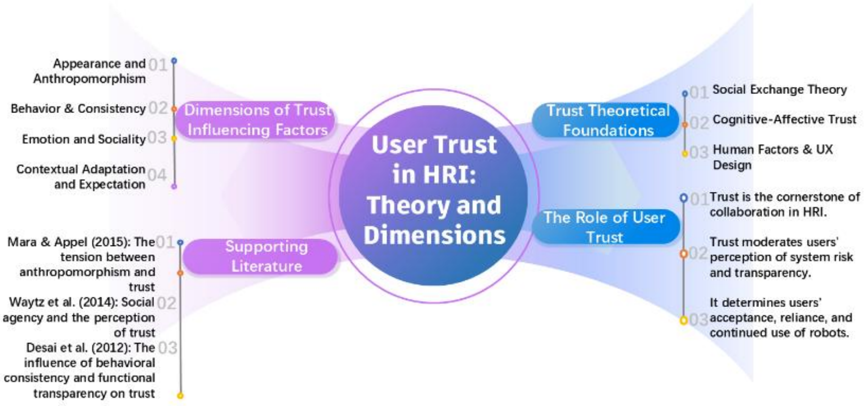
User Trust in HRI: Theory and Dimensions
Dimensions of Trust Influencing Factors
Appearance and Anthropomorphism 01
Behavior & Consistency 02
Emotion and Sociality 03
Contextual Adaptation and Expectation 04
Supporting Literature
Mara & Appel (2015): The tension between anthropomorphism and trust 01
Waytz et al. (2014): Social agency and the perception of trust 02
Desai et al. (2012): The influence of behavioral consistency and functional transparency on trust 03
Trust Theoretical Foundations
01 Social Exchange Theory
02 Cognitive-Affective Trust
03 Human Factors & UX Design
The Role of User Trust
01 Trust is the cornerstone of collaboration in HRI.
02 Trust moderates users' perception of system risk and transparency.
03 It determines users' acceptance, reliance, and continued use of robots.

Figure 1: Theoretical Foundations and Influential Dimensions of Human-Machine Trust

2.2 The Role of User Trust in HRI and the Value of Research

User trust is a key cornerstone for effective communication and collaboration in HRI. Trust is a prerequisite for users to be willing to allow some of their control to the robotic system and to develop dependence on it. The trust that Users have in a robotic system affects its initial acceptance and significantly influences subsequent ongoing use, functional adoption, and emotional attachment (Hancock, 2011). Lee (2004) notes that trust in automated systems is based on three dimensions: performance (i.e., the ability to accomplish the task), process (the degree of transparency and comprehensibility), and purpose (the alignment of its behavior with the interests of the users). These dimensions generally apply to HRI. Breazeal (2003) also argues that if equipped with appropriate emotional expression and role behavior strategies, social robots will significantly increase the speed and depth of trust development among users.

In the HRI domain, trust is unique because of its social properties. Users tend to treat robots as anthropomorphic, which makes the traditional “system trust model” no longer fully applicable. With technological advances in flexible sensors, multimodal perception systems, and cognitive computing frameworks, service robots, collaborative industrial robotic arms, and social robots are deeply integrated into human daily life. In the home scenario, a sweeping robot utilizing SLAM technology builds a 3D environment model, must accurately identify the movement trajectories of both the elderly and children; in surgical application a surgical robot that performs minimally invasive operations, the 0.1-mm precision of its force feedback system is directly related to the safety of the patient and the degree of trust is directly related to the user's safety judgment and willingness to cooperate (Desai, 2009). Furthermore, in automobile manufacturing, collaborative robotic arms sharing workspace with human workers is not entirely suitable for human daily activities. On automotive manufacturing lines, the safety distance algorithm for collaborative robotic arms sharing the workspace with workers is a core parameter of the Industry 4.0 standard. In these application scenarios, the autonomous decision logic and behavioral prediction models of robots have a topological impact on the safety and well-being of human users.

This study clearly maps three core theoretical perspectives onto the framework’s components and explains how they guide both conceptualization and operationalization in order to ground the proposed multi-factor synergistic framework in established theory. First, Lee’s (2004) tripartite model of trust in automation — performance, process, purpose — provides an organizing logic for linking observable design features to trust outcomes: “performance” aligns with measurable behavioral reliability (accuracy, latency, robustness), “process” corresponds to interpretability and transparency mechanisms, and “purpose” captures value alignment and perceived benevolence of the system. Second, cognitive–affective and social-exchange perspectives which emphasize both calculative risk–reward assessments and emotive social connection explain why appearance/anthropomorphism and emotion/sociality impact beyond mere task performance: anthropomorphic cues activate social cognition pathways and affective trust, thereby moderating responses to process and performance indicators (Breazeal, 2003; Waytz, 2014). Third, human factors and UX design principles supply the intervention repertoire (feedback loops, expectation management, control affordances) that translate the aformentioned theories into concrete design variables (Hancock, 2011; Onnasch, 2021). The multi-factor synergistic framework contributes two theoretical advances based on these mappings: 1) It coceptualizes trust as a dynamic, interdependent system rather than independent factor scores, thereby predicting non-linear interactions (e.g., anthropomorphism × behavioral consistency → trust resilience); 2) It emphasizes expectation management and anomaly-tolerance as mediating processes that convert low-level reliability/transparent behaviors into sustained trust over time. These innovations make the framework testable: for instance, it predicts that in high-risk domains (e.g., healthcare) process and purpose cues will carry larger weights relative to appearance, whereas in low-risk social contexts appearance and emotional resonance will more strongly influence initial trust; and behavioral consistency will moderate the effect of anthropomorphic cues, (i.e., anthropomorphism increases trust only when predictable response consistency exceeds a domain-specific threshold.) (Lee, 2004; Breazeal, 2003; Hancock et al., 2011).

2.3 Classification of Trust Influencing Factors and Overview of Typical Opinions

Current research on trust influencing factors has developed a more systematic categorization framework. According to Onnasch's(2021) study, the main factors affecting user trust in HRI can be classified into four categories: robot appearance and anthropomorphic design, behavioral performance and interaction consistency, social and emotional expression ability, and scenario adaptation and expectation management ability. Moderate anthropomorphism can enhance approachability, however excessive anthropomorphism can trigger the "valley of terror" effect and reduce user trust (Mori, 2012). Consistency refers to a robot's ability to maintain a stable and predictable behavioral style over multiple rounds of dialogue or task delivery, and studies have shown that robots with high consistency are more likely to gain the continued trust of users (Hancock, 2011). Furthermore,through emotion recognition and expression mechanisms, robots are better able to respond to users' emotions, which in turn inspires social trust (Rossi, 2020). In addition, trust is influenced by the moderation of usage scenarios, wherein system reliability and transparent feedback are key to building trust in medical or high-risk tasks. The aforementioned factors do not exist in isolation but rather synergize in multiple interactions to build a trust framework.

1) Robot Appearance and Anthropomorphic Design

Human-Robot Interaction (HRI) research shows that robot appearance and anthropomorphic design are critical to user trust. Giving robots human-like features—facial expressions, body language, voice—significantly enhances trust and interaction. A 2021 University of Southern California study found that robots capable of expressing empathy through micro-expressions increased user trust by 37% over non-anthropomorphic robots. This emotional transfer leads users to categorize robots as “social entities.”

In dynamic interaction, Tokyo University of Technology found that anthropomorphic robots with neck mobility improved dialogue focus by 42% via human-like nodding gestures. In retail, SoftBank’s Pepper robot, with a touchscreen face and torso-lean design, raised customer satisfaction by 82%, 1.8 times that of traditional equipment.

However, a threshold effect exists. A 2023 neuroimaging study from the Technical University of Munich showed that 78%-95% realistic faces activated the anterior cingulate cortex, causing “valley of terror” reactions, especially in older users (32% reported transient anxiety). Designers should balance this by applying localized anthropomorphism in industrial robots (e.g., Boston Dynamics’ Atlas) and moderate child-like traits in education robots (e.g., NAO).

Recent studies suggest culturally adaptive design as a breakthrough. Using a cross-cultural interaction database, Carnegie Mellon’s modular framework raised Middle Eastern users’ trust scores from 0.67 to 0.89 at the Dubai Expo. Future innovations like e-skin and bionic muscles may shift anthropomorphism from static mimicry to dynamic empathy.

2) Behavioral Performance and Interaction Consistency

Multimodal coherence and predictive compensation substantially strengthen behavioral consistency and user trust. Multimodal sensor fusion and prediction layers have been demonstrated to improve system robustness and reduce anomalous behavior in shared autonomy settings. This approach is exemplified by multimodal datasets and model-based planning frameworks used in assistive teleoperation research (Newman, 2018; Schmerling, 2018). While the exact effectiveness of these methods varies depending on the task and deployment context, controlled studies and dataset-driven evaluations consistently report significant reductions in failure or anomaly rates. They also show measurable improvements in human operators' trust-related outcomes when prediction and multimodal fusion are applied (Schmerling, 2018; Esterwood,2021). Therefore, integrating Inertial Measurement Unit (IMU)-based motion compensation, sequence models for intent prediction, and sensor fusion into a layered control architecture presents a promising approach to reduce anomalous responses and maintain trust in real-world HRI deployments.

3) Sociality and Emotional Expression

In HRI, sociality and emotional expression play crucial roles in building user trust. Sociality involves recognizing non-verbal signals such as facial expressions, body language, and observing socio-cultural norms, like maintaining an appropriate distance during conversations and alternating communication turns. For instance, service robots at airports signal "Do not disturb" using red flashing lights. Emotional expression relies on multimodal output: the Nao robot changes the color of its LED eyes, Pepper adjusts its intonation, and robots use gestures such as raising their arms to convey welcome.

Clinical trials have shown that hospitals using Affetto reduced post-operative anxiety by 37% and analgesic use by 22%. RIBA-II improved cooperation among Alzheimer's patients by 82% through gentle touch and voice. The EU SPRING study reported that older adults interacting with the Paro Seal robot three times weekly had 19% lower cortisol and 41% higher social activation. Cross-cultural adaptation is also vital: after adjusting facial expression amplitude to 60% of local norms, MIT's Kismet robot saw acceptance in the Middle East rise from 52% to 79%.

Mainstream affective computing uses a multi-layer design: bottom-layer sensors (e.g., skin conductivity), mid-layer OCC models for affective attribution, and top-layer decision modules based on 47 social context templates. A notable example is the latest Boston Dynamics Atlas, which uses composite cues like shoulder tilt (15° = curiosity, 30° = confusion) and head speed (0.5Hz = hesitation). However, over-anthropomorphism must be avoided. Carnegie Mellon (2023) found that facial similarity above 82% caused a 28% trust drop, urging designers to regulate pupil scaling (120–180ms) and joint motion smoothness (e.g., 0.2s delay).

4) Scenario Adaptation and Expectation Management

In HRI, scene adaptation and expectation management are the key factors in building user trust. Scenario adaptation refers to a robot's ability to dynamically adjust its behaviors and interaction strategies based on specific needs related to different environments and contexts. These adjustments are achieved through multimodal sensor fusion and environment-aware algorithms. For example, a medical assistive robot must maintain asepsis and accurate positioning in the operating room. In contrast, in a ward setting, it should shift to gentle voice interactions and focus on monitoring the patient's vital signs. This differentiated behavioral architecture can improve system performance by 28%. Studies have shown that user trust in robots increases significantly when they can accurately recognize and adapt to specific scenarios. According to cross-cultural experimental data presented at the IEEE HRI 2023 conference, users rated the trustworthiness of robots with environmental adaptability at 4.7 out of 5.0, which is 22.3% higher than that of systems with fixed behavioral patterns.

On the other hand, expectation management involves setting user expectations and dynamically adjusting the robot's behavior. When designing robots, developers need to quantify the user's cognitive expectation mapping through neuro-human factors engineering methods, such as eye tracking and EEG signal analysis and model the expected response in interaction design. For example, in logistics and warehousing scenarios, when the user expects the AGV robot to achieve ±5mm accuracy in cargo stacking, the system must ensure operational accuracy through LIDAR SLAM and force-aware feedback, while at the same time providing a 3D visual interface to display the error parameters in real-time. A recent study by MIT CSAIL shows that expectation mismatch leads to an abnormal increase in the activation level of the user's prefrontal cortex by 42%, directly affecting trust development. Therefore, expectation management requires robots to be technically reliable and to build a three-layer architectural system that includes expectation state monitoring, deviation warning, and explanatory interaction.

In summary, scene adaptation and expectation management form a dynamic feedback loop that helps establish trust in human-robot interaction. By utilizing a deep learning-driven scene understanding engine and Bayesian inference-based anticipation prediction algorithms, developers can create HRI systems that are cognitively resilient. A longitudinal study by Carnegie Mellon University's Human-Robot Interaction Institute showed that service robots integrating these two technologies improved user retention to 89 percent during a six-month hospital deployment. This represents a 37 percent improvement over the baseline system. As Steve Jobs said, "True innovation lies in anticipating unexpressed needs," and in the HRI field, this manifests itself in capturing implicit scenario features through context-aware computing and applying anticipation-guided interaction design to build a reliable mental model of intelligent services at the user's cognitive level.

3. Case Study and Theoretical Discussion

Case selection criteria: To ensure the cases meaningfully inform the framework and its domain claims, case selection followed four explicit criteria: 1) Representativeness across application domains — include robots from social, service/assistive, educational and medical guidance categories to capture domain heterogeneity; 2) Diversity of trust strategies — select examples exemplifying predominantly emotion-oriented vs function-oriented approaches; 3) Geographical and cultural breadth — include deployments from different sociocultural contexts (e.g., Japan, USA, China) to surface cultural moderators of trust; 4) Data availability and empirical grounding — prefer systems with published user studies, deployment reports, or observational field data enabling within-case analysis. Applying these criteria yielded four cases — Pepper (social robot), Robina (service/assistive robot), Xiaomedicine (medical guide system), and Tega (educational companion) — each selected for its combination of domain salience, distinct trust strategy, and available empirical evidence (Kennedy, 2015; Yamamoto, 2019; Wang, 2022; Gordon, 2016). The selection intentionally balances affective and functional strategies and aims to illuminate cross-contextual patterns rather than to be exhaustive. Limitations: the purposive sample is not statistically representative; findings are meant to be analytically generalizable and hypothesis-generating rather than population estimates.

3.1 Brief Description of Typical Cases of HRI at Home and Abroad Brief Description of Typical Cases of HRI at Home and Abroad

1) Pepper Social Robot (Japan)

Pepper was developed by Japan's SoftBank Robotics in 2014 to provide people with a natural and friendly human-machine social experience. It has voice recognition, face recognition, emotion recognition, and touch interaction systems. It is embedded with cloud-based AI learning models, which can be used in several public service scenarios such as retail, banking, and nursing homes. Pepper's greatest strength lies in its approachable appearance and rich non-verbal behaviors such as head tilting, gestures, and emoji displays, which help create an intimate interaction atmosphere.

Kennedy(2015) found in a comparative experiment that child users performed better in learning with Pepper and were more likely to show preferred behaviors in subsequent choices, suggesting that social robots positively influence trust building.

However, the study highlighted that if Pepper fails to accurately respond to users' needs, or experience delays in movement or misunderstandings, it may cause users to quickly recognize its “human-like but not human” nature. This could lead to what is known as the “Valley of Terror” effect. This is especially relevant in scenarios with frequent interaction, where consistency in the system's responses and its ability to comprehend semantic meaning are crucial for maintaining user trust. Therefore, the Pepper case reveals the advantage of trust initiation brought by anthropomorphic design and exposes the risk of backlash when anthropomorphic behaviors do not meet user expectations. This highlights the importance of ensuring “perception-behavior-feedback consistency” in social robots.

While Pepper’s anthropomorphic design and expressive non-verbal communication generate an initial strong emotional connection and make it approachable, evidence from our case studies and previous experimental work indicates that this connection is sensitive to predictive consistency. When Pepper’s behavior is unpredictable, the anthropomorphic benefits can diminish, leading to what is known as the “valley of terror” effect. Thus, Pepper primarily illustrates a design trade-off: it offers high social engagement that can foster initial trust, but this trust heavily depends on consistent behavior to be maintained (Kennedy, 2015; Mori, 2012).

2) Robina Service Robot (Toyota, Japan)

Robina is a service robot introduced by Toyota in Japan. It was initially unveiled in 2019 and designed to provide assistive services in hospitals, reception areas, and office environments. Unlike social robots that focus on emotional interaction, such as Pepper, Robina emphasizes stable execution and feedback consistency in task-based scenarios. Its core functions include autonomous navigation based on vision and LIDAR, multi-round voice interaction, information queries, document delivery, and guidance services for patients or visitors. Robina has demonstrated significant "behavioral transparency" and "consistent response" in real-world deployments. For example, the system repeatedly confirms the destination before providing navigation to the visitor and provides incremental feedback on the current location via voice or screen as the visitor moves. These "closed-loop interaction modes" enhance user control and security and reduce "black box behavior" concerns. Yamamoto conducted a developmental field study of Toyota's Human Support Robot (HSR) platform, which, along with Robina, is part of the Toyota Frontier Research Center's family of service robots, emphasizing mobility and service adaptability in home and public environments. The Yamamoto’s (2019) study noted that the HSR's structured navigation and environment modelling mechanisms effectively enhanced user trust, and these design concepts were widely applied to Robina's deployment strategy.

In addition, data from Toyota Memorial Hospital's 2019(2022)guided inspection robotics project showed that applying Kaizen (lean improvement) to nurses' workflows, robots such as Robina reduced nurses' repetitive labor time by 40% in routine tasks such as medication delivery and guided inspections. This enabled healthcare professionals to dedicate more time to patient care, while visitors have commented on their “reliable, predictable and human” assistants.

Regarding trust mechanisms, Robina's design follows the path of “balancing appearance with functional transparency”: its anthropomorphic image is integrated into a warm design that does not rely on complex emotional expressions but instead supports trust through highly consistent behavior, instant feedback, and task accuracy. This design is appropriate for high-trust environments such as hospitals, suggesting that “transparency + consistenc” is a feasible path for service robots to build trust.

Robina’s functional strategy emphasizes closed-loop confirmations and incremental feedback. This strategy yields high operational trust and task reliability in hospital settings, but it results in limited emotional engagement. Consequently, establishing initial rapport in open public settings tends to be slower. Robina therefore exemplifies the “process + performance” pathway of the Lee model, characterized by high process clarity and performance. However, her strategy suggests limited transferability to contexts where emotional acceptance is crucial (Yamamoto, 2019).

3) Yun Zhisheng”Xiaomedicine” Medical Guide Robot (China)

“Xiaomedicine” is an intelligent medical guide system developed by Chinese artificial intelligence company Yun Zhisheng in collaboration with West China Hospital, China-Japan Friendship Hospital and other healthcare institutions. It employs speech recognition, natural language processing and medical knowledge mapping technologies to provide patients with the initial screening of diseases at hospital entrances and recommend departments for booking and registration services. The system supports complex voice commands such as Mandarin, dialect, and multi-round semantic interaction. It has been implemented in many hospitals with over one million service visits.

According to Wang's(2022) user experience evaluation study, Xiaomedicine shows high trust acceptance among elderly and less-educated users, mainly because its voice operation reduces the technical barrier, and its feedback statements are clear and easy to understand. Especially in tertiary hospitals, where patients often feel anxious due to the complexity of the consultation process and time constraints, Xiaomedicine effectively alleviates patients' uncertainty and rejection of unfamiliar systems through timely responses and visual guidance.

In addition, Xiaomi introduced the “User Emotion Recognition Module” and “Dialogue Response Adjustment Module”, which can detect anxiety and impatience based on the user's tone of voice to adjust the response timing and speaking style, thereby enhancing user satisfaction and system reliability. This case demonstrates that in high-risk medical scenarios, the professional credibility and adaptability of the system are crucial for building trust, especially among vulnerable groups, and that intelligent design must maintain a dynamic balance between “usability” and “comprehensibility”.

4) Tega Educational Robot (MIT, USA)

Tega is a social robot designed by the MIT Media Lab for children's learning assistance. It is made from flexible materials, has an anthropomorphic cartoon-like appearance, and uses AI learning algorithms to provide personalized feedback for teaching. Tega can recognize children's voices and emotions and adjust its responses accordingly, and it learns user behavior through reinforcement learning. Its design philosophy emphasizes “concomitant cognitive facilitation”, which means using friendly roles and emotional interactions to enhance active learning.

Gordon's experiments show that children are more inclined to interact with robots with expressive feedback and voice interaction. This is particularly true when Tega demonstrates the three-stage teaching strategy of “Understand-Respond-Motivate”, which significantly enhances children's learning engagement and builds trust. Trust level increased significantly during the experiment. In long-term intervention experiments, Tega could automatically adjust the intensity and frequency of its feedback according to children's learning curves, thus maintaining a stable interaction experience without causing any disruptions (Gordon, 2016).

Unlike traditional learning aids, Tega possesses an anthropomorphic personality and emotional expression, which are essential social constructs in fostering trust during this stage of children's mental development. Tega’s design illustrates the importance of socio-emotional elements in building trust among young users and reflects the critical role of anthropomorphic strategies in early educational robotics.

3.2 Analysing Key Elements in Trust-Building from the User's Perspective

When facing a robotic system, building users' trust has multi-dimensional cognitive characteristics, including intuitive perception, task interaction experience, risk assessment, and psychological compatibility expectations. From the user's perspective, the key elements affecting the formation of trust include: 1) Robot interpretability and transparency; 2) Behavioral consistency and interaction predictability; 3) User control and adaptive feedback mechanisms.

Robinette(2016) found that in an emergency evacuation experiment, users were more likely to trust a robot that was consistent in its behavior and gave reasonable explanations, even if it made minor errors in its previous performance. This study validates the idea that “wrong but transparent” promotes user trust more than “right but black box”. Waytz(2014) further notes that trust significantly increases when users are aware of the system's internal operating mechanisms, suggesting that “transparency” is a more effective design principle than concealment, which suggests that “transparency” is an important design direction to enhance users' subjective sense of security and control.

In addition, it was found that users' expected level of engagement moderated their trust responses. Users rated their autonomy higher when the robot provided choice, confirmation or personalized responses (Hancock, 2011). Trust is built on “whether the robot does a good job” and “whether the user perceives that he or she is part of the decision-making process”.This user perspective is visually represented in Figure 2.

Figure 2: Human-Robot Interaction User Perspective Map

3.3 Differences in the Effects of Different Design Strategies on the Construction of Trust

Trust design strategies are usually classified into two categories: function-oriented and emotion-oriented. Function-oriented strategies focus on improving task completion efficiency and accuracy, while emotion-oriented strategies engage users emotionally and foster social acceptance. The trust paths generated by these different strategies differ significantly.

In the emotion-oriented strategy, anthropomorphic appearance, voice tone matching, and emotion synchronization techniques are widely used to enhance affinity. For example, Nowak's(2003) experiments demonstrated that robots with a combination of human-like appearance and voice are more likely to gain users' social trust. Function-oriented strategies, on the other hand, emphasize the stability of system performance and the immediacy and clarity of operational feedback, which are typical of the design logic in the field of collaborative industrial robots (Cobots), where trust derives from the trinity model of Efficiency-Safety-Reliability (de Visser, 2018).

Mixed strategies are more advantageous for long-term usage contexts. Research has shown that robots demonstrating task reliability in their emotional responses can better maintain high-quality user relationships (Willemse, 2017). Trust not only has an initiating effect at the beginning of the interaction, but should also be revised and reinforced during continued use. These trust construction pathways are depicted in Figure 3.

Figure 3: HCI Trust Construction Diagram

3.4 The Phenomenon of Trust Barriers in Existing Products and Designs

Recent empirical evidence in specific domains demonstrates that public acceptance of robotic systems depends on both prior AI exposure and the presence of contextual safeguards. A peer-reviewed U.S. survey of 413 respondents found that mean trust scores for autonomous humanoid "robot doctors" were slightly lower than those for human clinicians (M ≈ 3.19 vs. 3.33 on a 5-point scale), as detailed in Table 1. However, subgroup analyses revealed significant demographic differences, particularly related to gender t(379) = -5.30, p < .001), indicating varied acceptance patterns among different groups (Kim, 2024).

Table 1: Mean trust scores for human vs. autonomous humanoid doctor

Categories	Frequencies	Trust 1	Trust 2	Trust combines
Gender				
-Male	98	2.9	2.82	2.86
-Female	292	3.17	3.32	3.24
Age				
-18-29	273	3.12	3.23	3.17
-30-39	40	2.99	3.1	3
-40-49	32	3.11	3.08	3.1
-50-59	30	3.05	3.23	3.12
-60 or older	16	3.13	3.05	3.01
Education level				
-Elementary school	0	-	-	-
-Middle school	1	3	3	3
-High school	30	3.91	3.14	3
-University	300	3.12	3.21	3.12
-Grad school	60	3.11	3.15	3.13
Perceived income level				
-Very low	57	3.08	3.27	3.17
-Low	130	3.08	2.21	3.18
-Middle	179	3.07	3.19	3.13
-High	22	3	3.05	3.03
-Very high	2	2.77	2.5	2.64
Satisfaction w/healthcare systems				
-Very low	11	3.4	3.62	3.52
-Low	83	3.13	3.26	3.19
-Middle	146	3.15	3.26	3.2
-High	121	3.03	3.1	3.01
-Very high	30	2.93	2.9	2.9

Physical health				
-Very weak	2	3.59	3.75	3.67
-Weak	28	3.17	3.29	3.22
-Middle	210	3.04	3.14	3.09
-Strong	119	3.12	3.19	3.15
-Very strong	32	3.26	3.47	3.36
Mental health				
-Very weak	4	3.25	3.28	3.26
-Weak	45	2.93	3.02	2.97
-Middle	143	3.12	3.2	3.15
-Strong	150	3.07	3.15	3.11
-Very strong	49	3.29	3.48	3.38

Source: Kim(2024); n = 413

Recent global survey data from late 2024 to early 2025 provide a much clearer, data-rich picture of how people experience and trust AI—and by extension, robotic or AI-embedded systems—in actual usage contexts. According to KPMG & University of Melbourne (2025), 66% of respondents report they intentionally use AI tools regularly (for work, study, or personal tasks), and 83% believe that AI will bring benefits; however, only 46% are willing to trust AI systems outright,as shown in Figure 4. In workplace settings, 58% of employees globally report using AI regularly, and 83% of students use AI in learning. Yet notable risks and gaps emerge: 66% rely on AI outputs without verifying accuracy; 56% have made errors at work caused by AI; 57% admit to hiding AI use from their managers; and 48% have uploaded company data to public AI tools.

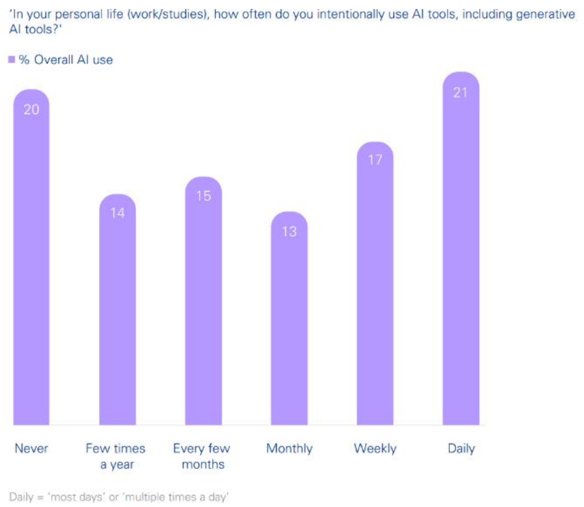

Figure 4: Frequency of intentional use of AI tools for personal, work, or study purposes
Source: KPMG & University of Melbourne(2025)

These figures suggest that while adoption and familiarity with AI and robotic systems are high, trust remains conditional. Key factors such as transparency, accountability, safety, and regulation are essential in determining how these technologies are accepted. For robotics research and deployment, it is important to update claims based on older, less detailed trust and acceptance statistics. Recent metrics, especially across different contexts such as work, education, healthcare, and demographics, should be utilized.

3.5 Cross-Case Comparative Analysis

To extract cross-case insights beyond within-case descriptions, we conducted a structured cross-case comparison along four analytical dimensions: application scenario, core trust strategy, primary strengths and limitations, and design implications. Table 2 summarizes the comparison for Pepper, Robina, Xiaomedicine, and Tega. The matrix highlights distinct pathway patterns: Pepper exemplifies an affect-first pathway, where emotional appeal leads to rapid initial trust, making it highly sensitive to behavioral inconsistencies; Robina follows a function-first pathway prioritizing performance and process to establish operational reliability and sustained trust during task; Xiaomedicine shows accessibility-driven trust, using voice and comprehensibility to achieve high acceptance among elderly users and those with low-literacy; and Tega reflects developmental alignment, using adaptive pedagogical approach to sustained engagement among children. From these comparisons, we derive three actionable design lessons: 1) Match the trust strategy with the domain's risk profile by prioritizing process and purpose cues in high-risk areas; 2) For socially oriented systems, invest in maintaining behavioral consistency thresholds to avoid negative user reactions; 3) Prioritize contextual accessibility through language choivce and cognitive load reductions, to improve adoption rates among vulnerable user groups. Additionally, the comparative findings refine the theoretical framework by illustrating how Lee's performance/process/purpose weights shift across different domains. They also empirically support the moderation effect of behavioral consistency on the benefits of anthropomorphism (Lee, 2004; Breazeal, 2003; Hancock, 2011).

Table 2: International HRI Case Studies on Trust-Building Strategies

Country	Robot Name	Robot Features	Trust-Building Strategies	Picture
Japan	Pepper	Anthropomorphic social robot launched by SoftBank; Equipped with voice recognition, face recognition and expression feedback system; Emphasizes "emotional communication" and "companionship interaction."; Widely used in retail, education, elderly care and other scenarios.	Highly anthropomorphic appearance and voice characteristics enhance the formation of "initial trust."; Emotional simulation and feedback mechanisms promote "social presence"; Behavioral consistency has a significant effect on maintaining trust.	
China	Xiaomedicine	Multilingual speech recognition (Mandarin and dialects); Voice-guided navigation; Emotional tone modulation.	Transparent decision logic; Adaptive dialogue feedback; Customizable interaction modules for user control.	
Japan	Robina	Autonomous navigation, Multilingual voice interaction, and Short-term memory-based adaptive response.	Consistent task performance, Transparent feedback loop, Context modelling, and role memory system.	

USA	Tega	Cartoon-like flexible appearance;
		Emotional recognition;
		Reinforcement learning for personalized teaching.
			Emotionally engaging behavior, Responsive gestures, Personalized and adaptive interaction strategies.	

4. Conclusion and Discussion

4.1 Conclusion of the Study

This study advances trust theory in HRI in three interrelated ways. First, by explicitly mapping Lee's performance–process–purpose model onto empirically grounded design levers, transforming abstract trust dimensions into concrete, measurable features (e.g., latency thresholds, explanation granularity, task accuracy metrics). This mapping facilitates empirical testing. Second, by treating trust as a dynamic interdependent system, the study formalizes two mediators—expectation management and anomaly tolerance—that connect momentary system behaviors to longitudinal trust trajectories. This approach extends beyond static factor models and encourages longitudinal field tests. Third, the cross-case evidence identifies domain-specific weighting of trust components and demonstrates the moderating role of behavioral consistency on anthropomorphism's effect. This leads to the generation of falsifiable hypotheses, such as the idea that the positive effect of anthropomorphism on initial trust depends on consistency exceeding domain-specific thresholds. These contributions position the multi-factor synergistic framework as both a descriptive typology and an intervention roadmap for design and evaluation.

4.2 Relationships and Synergies Between Trust Influences

In human-robot interaction (HRI), establishing user trust is a complex, dynamic process involving the synergy of cognitive, affective, and behavioral dimensions. Features of robot appearance, such as adjustable facial expressions and eye movement, along with human-like gestures like steady hand micro-expressions, can quickly influence users' expectations and shape trust mechanisms (Robinette, 2016). A 2022 University of Tokyo study showed that cartoonish humanoid robots had 47% higher initial trust than robotic arms, but trust dropped by 62% if voice delays exceeded 300ms or semantic error rates surpassed 15%.

This reveals a "cognitive consistency threshold": trust declines exponentially when action accuracy error exceeds 9% or emotional feedback lags by more than two semantic units. Designers must integrate predictive compensation algorithms and implement "error transparency mechanisms" using multimodal sensors with 150Hz environment updates. Esterwood (2021) found trust recovery improved significantly ($p < 0.05$) when robots acknowledged errors and provided explanations.

In the social dimension, Cambridge's Affective-Cognitive Architecture showed that integrating micro-expressions, which involve 43 muscle groups, along with contextual memory, increased trust maintenance to 83% over 6 weeks. This is made possible through dual-channel reinforcement learning, which enables real-time adaptation via speech analysis, evaluating 88 parameters and long-term personalized interaction templates with a maximum of 2000 templates.

Scenario adaptation demands different standards: disinfection robots require 99.97% action predictability; educational robots can allow 5–8% creative flexibility. Giulio (2025) emphasizes that adaptive behavioral complexity boosts trust by 20% ($p < 0.01$). Trust should be assessed via a multidimensional system combining physiological (12% drop in electrocorticographic response), subjective (NASA-TLX ≤ 40), and behavioral data (35% rise in help-seeking), forming a closed-loop optimization mechanism.

4.3 Implications for Human-Robot Interaction Design

Establishing user trust is a crucial aspect of designing human-robot interactions. Kory's (2016) research shows that trust scores are more stable when the robot's feedback structure closely aligns with users' cognitive schema, particularly in scenarios involving companionship, guided diagnosis, and education. A long-term MIT Media Lab study found significant correlations between trust and several factors: visual affinity (62%), behavioral reliability (78%), social attribute fit (85%), and environmental scene fit (73%). For instance, Japan's RIKEN HRI-2 robot utilized a progressive anthropomorphic design, which includes visible functional arms and enhanced visual appeal through curved shells and soft lighting. This approach boosts initial trust by 40%. Additionally, research at the KIT lab showed that when a robot accurately predicted user needs three times consecutively with timely responses, trust maintenance peaked at 0.93±0.05. This finding supports Ellen Langer's "anticipation-feedback" theory.

Bremner (2016) found that synchronizing speech with symbolic gestures improved interaction fluency and perceived naturalness, enhancing trust. A three-layer consistency model is recommended: maintaining a response latency of within 200 milliseconds; ensuring validated decision logic; and facilitating multimodal interaction. A study by the University of Tokyo showed that robots with micro-expression sequence coding gained trust 2.3 standard deviations faster in medical consultations. At Mayo Clinic, a nursing robot using haptic feedback and pupil scaling increased trust scores from 3.2 to 4.7 out of 5 in venipuncture scenarios. To avoid the "Valley of Terror," it is essential to manage emotionally embodied design carefully. Data from Boston Children's Hospital revealed that contextual memory playback improved long-term trust retention by 58% compared to the baseline. Thus, a modular trust architecture is recommended, quantifying core trust elements and optimizing via reinforcement learning.

4.4 Recommendations for Human-Robot Interaction Design

To strengthen user trust in HRI, future system design should integrate three key elements: predictive transparency, adaptive personalization, and ethical governance into a unified framework. Robots should proactively communicate their decision-making intentions through clear explanations and feedback loops. This approach allows users to anticipate how the system will behave and adjust their level of trust accordingly. Creating personalized interaction profiles based on prior engagement can enhance predictability and emotional alignment. Additionally, implementing transparent data practices—such as obtaining explicit consent, minimizing data collection, and processing data on-device —is an essential foundation for building trust. In addition, trust-monitoring mechanisms that combine subjective ratings with behavioral indicators should be embedded in long-term deployments to track trust evolution and guide real-time adaptation in the system (Esterwood, 2021; Schmerling, 2018).

4.5 Prospects

Looking ahead, the next phase of HRI research should transition from isolated laboratory validation toward developing continuous, explainable, and ethically aligned systems capable of sustaining trust in dynamic, real-world contexts. Future efforts should prioritize the co-evolution of human and robot behavior through lifelong learning frameworks, enabling robots to adapt not only to various tasks but also to users' emotional and cultural expectations. Integrating explainable AI with multimodal perception will make robotic decisions interpretable in context, while standardized trust benchmarks and certification mechanisms will ensure accountability and comparability across applications. Ultimately, establishing an ecosystem that prioritizes transparency, adaptability, and social responsibility will be essential for the sustainable development of trustworthy HRI (Newman, 2021; Yin, 2018).

Funding: Research Project of the Korea Institute for Humanities and Social Sciences (No.2023KIHSSB009)

Acknowledgements: Not applicable.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Conflicts of Interest: The authors declare no conflict of interest.

References

Breazeal, C. (2003). Emotion and sociable humanoid robots. *International Journal of Human–Computer Studies*, 59(1–2), 119–155. https://doi.org/10.1016/S1071-5819(03)00018-1

Bremner, P., & Leonards, U. (2016). Iconic gestures for robot avatars, recognition and integration with speech. *Frontiers in Psychology*, 7, 183. https://doi.org/10.3389/fpsyg.2016.00183

Campagna, G., & Rehm, M. (2025). A systematic review of trust assessments in human–robot interaction. *Journal of Human-Robot Interaction*, 14(2), Article 30. https://doi.org/10.1145/3706123

Chita-Tegmark, M. (2021). Can you trust your trust measure? *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction* (HRI '21). https://doi.org/10.1145/3434073.3444677

Creswell, J. W. (2014). *Research design: Qualitative, quantitative, and mixed methods approaches* (4th ed.). SAGE Publications.

Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human–robot interaction. *Philosophical Transactions of the Royal Society B*: Biological Sciences, 362(1480), 679–704. https://doi.org/10.1098/rstb.2006.2004

de Visser, E. J., Monfort, S. S., McKendrick, R., Smith, M. A., McKnight, P. E., Krueger, F., & Parasuraman, R. (2016). Almost human: Anthropomorphism increases trust resilience in cognitive agents. *Journal of Experimental Psychology*: *Applied*, 22(3), 331–349. https://doi.org/10.1037/xap0000092

Desai, M., Stubbs, K., Steinfeld, A., & Yanco, H. A. (2009). Creating trustworthy robots: Lessons and inspirations from automated systems. *In Proceedings of AISB '09 Convention: New Frontiers in Human–Robot Interaction*. https://www.ri.cmu.edu/pub_files/2009/4/Desai_paper.pdf

Esterwood, C., & Robert, L. P. (2021). Do you still trust me? Human-robot trust repair strategies. *In 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN)* (pp. 183–188). IEEE. https://doi.org/10.1109/RO-MAN50785.2021.9515365

Fong, T., Nourbakhsh, I., & Dautenhahn, K. (2003). A survey of socially interactive robots. *Robotics and Autonomous Systems*, 42(3–4), 143–166. https://doi.org/10.1016/S0921-8890(02)00372-X

Goodrich, M. A., & Schultz, A. C. (2007). Human–robot interaction: A survey. *Foundations and Trends® in Human–Computer Interaction*, 1(3), 203–275. https://doi.org/10.1561/1100000005

Hancock, P. A., Billings, D. R., Olsen, K. M., Chen, J. Y. C., de Visser, E. J., & Parasuraman, R. (2011). A meta-analysis of factors affecting trust in human–robot interaction. *Human Factors*, 53(5), 517–527. https://doi.org/10.1177/0018720811417254

Jiang, H., & Cao, J. J. (2006). On Pornographic Culture and Juvenile Sexual Delinquency. *Chinese Moral Education*, (11), pp. 32-35. https://kns.cnki.net/kcms2/article/abstract?v=lVku9qtM8H_wV1Rdmp514i-iTATUJ03MxuqsaP8UxgOF0fBV_b-NXtCmUlkgdcw3edB1z5vrDAMxgrHOeQ-5dSofPgxrGztnTamFgfE7AHts9RFQ6eceIxZq181VCvPlXZE6l_cKvrY=&uniplatform=NZKPT&language=CHS

Kennedy, J., Baxter, P., & Belpaeme, T. (2015). The robot who tried too hard: Social behaviour of a robot tutor can negatively affect child learning. *In Proceedings of the 10th ACM/IEEE International Conference on Human-Robot Interaction* (pp. 67–74). https://doi.org/10.1145/2696454.2696457

Kim, D. K. D. (2024). An investigation of public trust in autonomous humanoid AI robot doctors: A preparation for our future healthcare system. *Frontiers in Communication*. https://doi.org/10.3389/fcomm.2024.1420312

Kory Westlund, J. M., et al. (2016). Tega: A social robot. *In Proceedings of the 11th ACM/IEEE International Conference on Human–Robot Interaction: Video Presentations*. IEEE. https://doi.org/10.1109/HRI.2016.7451761

Kory Westlund, J. M., Gordon, G., Spaulding, S., Lee, J. J., Plummer, L., Martinez, M., Das, M., & Breazeal, C. (2016). Lessons from teachers on performing HRI studies with young children in schools. *Proceedings of the 11th ACM/IEEE International Conference on Human-Robot Interaction* (HRI 2016), 383–390. https://doi.org/10.1145/2858161.2861134

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

Mori, M., MacDorman, K. F., & Kageki, N. (2012). The uncanny valley [From the field]. *IEEE Robotics & Automation Magazine*, 19(2), 98–100. https://doi.org/10.1109/MRA.2012.2192811

Newman, B. A., Aronson, R. M., Srinivasa, S. S., Kitani, K. M., & Admoni, H. (2021). HARMONIC: A multimodal dataset of assistive human–robot collaboration. *The International Journal of Robotics Research*. https://doi.org/10.1177/02783649211050677

Nowak, K. L., & Biocca, F. (2003). The effect of agency and anthropomorphism on users' sense of telepresence, copresence, and social presence in virtual environments. *Presence: Teleoperators and Virtual Environments*, 12(5), 481–494. https://doi.org/10.1162/105474603322761289

Onnasch, L., & Roesler, E. (2021). A taxonomy to structure and analyze human–robot interaction. *International Journal of Social Robotics*, 13(4), 833–849. https://doi.org/10.1007/s12369-020-00666-5

Rae, I., Mutlu, B., & Takayama, L. (2014). Bodies in motion: Mobility, presence, and task awareness in telepresence. *In Proceedings of the 32nd Annual ACM Conference on Human Factors in Computing Systems (CHI '14)* (pp. 2153–2162). https://doi.org/10.1145/2556288.255704

Robinette, P., Li, W., Allen, R., Howard, A. M., & Wagner, A. R. (2016). Overtrust of robots in emergency evacuation scenarios. *In Proceedings of the 11th ACM/IEEE International Conference on Human–Robot Interaction (HRI '16)*, 101–108. https://doi.org/10.1109/HRI.2016.7451740

Rossi, E. M., Staffa, M., & Esposito, A. (2020). Trustworthy robots: Experimental analysis on trust dynamics in HRI. *Lecture Notes in Computer Science*, 12102, 56–67. https://doi.org/10.1007/978-3-030-49062-1_5

Salem, M., Eyssel, F. A., Rohlfing, K., Kopp, S., & Joublin, F. (2013). To err is human (–like): Effects of robot gesture on perceived anthropomorphism and likability. *International Journal of Social Robotics*, 5(3), 313–323. https://doi.org/10.1007/s12369-013-0196-9

Sanders, T., Kaplan, A. D., Koch, R., Schwartz, M., & Hancock, P. A. (2019). The relationship between trust and use choice in human–robot interaction. *Human Factors*, 61(4), 614–626. https://doi.org/10.1177/0018720818816838

Savary, R., Rose, R., & Weinberg, G. (2020). Establishing human-robot trust through music-driven robotic emotion prosody and gesture. *arXiv preprint* arXiv:2001.05863. https://doi.org/10.48550/arXiv.2001.05863

Saygin, A. P., Chaminade, T., Ishiguro, H., Driver, J., & Frith, C. (2012). The thing that should not be: Predictive coding and the uncanny valley in perceiving human and humanoid robot actions. *Social Cognitive and Affective Neuroscience*, 7(4), 413–422. https://doi.org/10.1093/scan/nsr025

Schmerling, E., Leung, K., Vollprecht, W., & Pavone, M. (2018). Multimodal probabilistic model-based planning for human–robot interaction. *In 2018 IEEE International Conference on Robotics and Automation (ICRA)* (pp. 1–8). IEEE. https://doi.org/10.1109/ICRA.2018.8460766

Shill, P. C. (2024). Human reactions to incorrect answers from robots [Preprint]. *arXiv*. https://arxiv.org/abs/2403.14293

Toyota Times. (2022, November 21). The mystery box bringing Toyota's Kaizen into hospitals. Retrieved from https://toyotatimes.jp/en/series/beyondmobility/004.html

University of Melbourne & KPMG. (2025). *Trust, attitudes and use of artificial intelligence: A global study 2025* (Research report). University of Melbourne & KPMG.https://kpmg.com/xx/en/our-insights/ai-and-technology/trust-attitudes-and-use-of-ai.html

Wang, X., Liu, Y., & Xu, H. (2022). What influences patients' continuance intention to use AI-powered service robots at hospitals? The role of individual characteristics. *Technology in Society*, 70, Article 101996. https://doi.org/10.1016/j.techsoc.2022.101996

Waytz, A., Heafner, J., & Epley, N. (2014). The mind in the machine: Anthropomorphism increases trust in an autonomous vehicle. *Journal of Experimental Social Psychology*, 52, 113–117. https://doi.org/10.1016/j.jesp.2014.01.005

Willemse, C. J. A. M., Toet, A., & van Erp, J. B. F. (2017). Affective and behavioral responses to robot-initiated social touch: Toward understanding the opportunities and limitations of physical contact in human–robot interaction. *Frontiers in ICT*, 4, 12. https://doi.org/10.3389/fict.2017.00012

Yin, R. K. (2018). *Case study research and applications: Design and methods* (6th ed.). SAGE Publications.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of KIHSS and/or the editor(s). KIHSS and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.