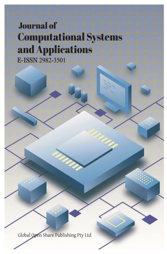
Journal of Computational Systems and Applications
https://jcsa.gospub.com/jcsa
Global Open Share Publishing



Article

Humanoid Robots, Public Trust, and Ethical Algorithm Design: A Multidisciplinary Framework for Societal Integration and Future Adoption

Hafize Nurgul Durmus Senyapar[1,*], Ramazan Bayindir[2]

[1]Faculty of Health Sciences, Gazi University, Ankara, Turkey

[2]Faculty of Technology, Gazi University, Ankara, Turkey

[*]Corresponding Author: Hafize Nurgul Durmus Senyapar, nurguld@gazi.edu.tr

Abstract

This study investigates the complex relationship between public trust and the ethical design of behavioral algorithms in humanoid robots, drawing on a qualitative, interdisciplinary narrative review. By synthesizing research across AI ethics, Human–Robot Interaction (HRI), behavioral science, and cultural studies, it demonstrates that public acceptance depends not only on technical performance, but also on emotional, symbolic, and cultural factors that shape trust and legitimacy. Key findings reveal that ethical principles such as transparency and fairness are interpreted through local social contexts, and current algorithmic frameworks often fail to address the full diversity of user expectations. As an original contribution, the study introduces the Human–Robot Integration Ethics Matrix (HR-IEM), a conceptual tool linking core algorithmic objectives to measurable social trust indicators. The HR-IEM bridges the persistent gap between abstract ethical standards and lived human experience, offering actionable guidance for developers and policymakers to evaluate ethical–social alignment in real-world settings. The analysis also identifies critical methodological limitations, including an overreliance on laboratory studies and a lack of participatory and cross-cultural design practices. To address these gaps, the study advocates for adaptive ethics modules, participatory feedback mechanisms, and context-sensitive frameworks that support the co-production of social legitimacy. Ultimately, this research presents a new interdisciplinary agenda for the responsible integration of humanoid robots, emphasizing cross-sector collaboration and the continuous calibration of ethical design to evolving public values.

Keywords

Humanoid robots, Ethical algorithm design, Public trust, Human–robot interaction, Socio-technical integration, AI ethics in robotics

Article History

Received: 02 July 2025 | Revised: 17 July 2025

Accepted: 15 August 2025 | Available Online: 09 September 2025

1. Introduction

Humanoid robots are increasingly being integrated into a wide range of environments, including social, service-oriented, and domestic settings. Once confined to controlled laboratory conditions or speculative fiction, these robots now appear in eldercare facilities, educational institutions, hotels, and private homes. This transition from abstract innovation to tangible social reality has presented two critical challenges: understanding how different segments of society perceive and accept humanoid robots and ensuring that the algorithms governing their behavior are both ethically sound and socially attuned [1,2]. These concerns are increasingly viewed as interdependent; public acceptance depends on robots behaving in ethically transparent, emotionally intelligible, and culturally sensitive ways—qualities not always present in technically robust but socially detached systems.

Public trust in humanoid robots is shaped by more than technical performance. Factors such as human-likeness, behavior in sensitive situations, communication with vulnerable populations, transparency in decision-making, and perceived alignment with human values all influence acceptance [3]. Concepts like emotional accountability—the extent to which a robot supports human emotional well-being without manipulation—and adaptive ethics modules that respond to cultural cues and evolving norms, now extend ethical design beyond static compliance toward dynamic, context-aware interaction. For example, a medical robot must not only function reliably but also be seen as respectful and compassionate, which underscores the ethical importance of design choices and their adaptation to diverse cultural contexts. Cross-cultural studies highlight that perceptions vary—Shintoist beliefs in Japan foster trust in robots as harmonious partners. At the same time, Western societies often express concerns about dehumanization and loss of agency, emphasizing the need for globally adaptive yet locally sensitive frameworks.

Trust is further influenced by users' subjective experiences and social norms, with regional and demographic factors shaping how robots are received [4]. Recent empirical work shows that robot behaviors, such as agreement or disagreement with users, measurably impact trust, perceived expertise, and willingness to share data, demonstrating the ethical significance of communicative behavior (Figure 1) [3]. This reinforces that trust cannot be reduced to functionality alone; it is constructed through relational and affective dynamics, particularly in emotionally sensitive contexts such as healthcare or education.

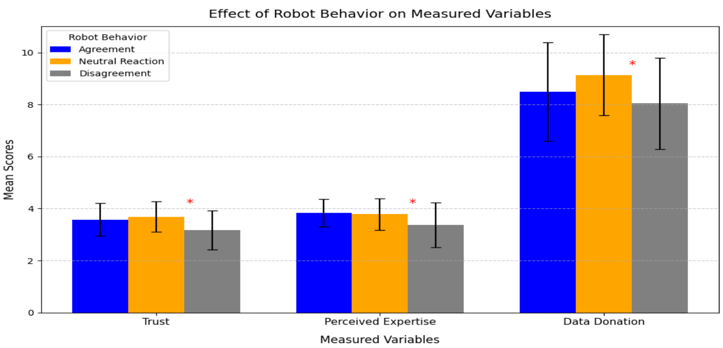


Figure 1. Mean scores for dependent variables across robot behavior conditions [3].

Within this context, ethical design—encompassing transparency, fairness, accountability, and explainability—becomes fundamental for social legitimacy. However, a substantial gap persists between the aspirations of responsible AI frameworks and the practical realities of HRI. Ethical principles are often articulated without considering real user environments, and user perspectives are largely missing from algorithmic development [5]. The siloed nature of academic and industrial research exacerbates this problem: AI ethics literature emphasizes formal principles, while HRI studies focus on usability and engagement, rarely converging in practice. This leads to robots with robust algorithms but poor social alignment, or, conversely, socially engaging robots lacking ethical transparency [6]. Sectoral differences further highlight this challenge: advanced industries adopt AI more readily due to robust digital infrastructure, while others lag, underscoring the need for context-specific approaches [7].

Ultimately, public skepticism toward humanoid robots can hinder deployment, especially in sensitive sectors, and without feedback mechanisms, robots risk reinforcing biases or failing to adapt to societal norms. Addressing these issues requires an integrative approach—one that recognizes the feedback loop between ethical behavior and public trust and builds interdisciplinary collaboration [8]. Progress depends on concrete, participatory models and methodologies that reflect lived realities, including co-design with users, longitudinal studies of trust, and tools that evaluate not only functional success but also ethical and emotional resonance [3,9].

This study conducts a qualitative exploratory investigation at the intersection of public acceptance of humanoid robots and the ethical design of their governing algorithms. Drawing from interdisciplinary literature, the research identifies theoretical gaps, integration challenges, and potential pathways for better aligning technological design with evolving social expectations. The primary objective is to elucidate how ethical algorithmic frameworks influence the social, psychological, and cultural conditions that underpin public trust and acceptance. The analysis explores not only normative concerns and user expectations but also the extent to which current algorithmic designs address these issues and whether existing ethical AI frameworks truly account for public adoption dynamics.

To move this agenda forward, the study introduces the HR-IEM: a conceptual tool that systematically maps core algorithmic objectives, such as transparency, fairness, and emotional sensitivity, to specific social trust indicators and real-world contexts. HR-IEM addresses a central gap in the literature: the persistent disconnect between abstract ethical guidelines and the complex realities of HRI across diverse settings. By explicitly linking algorithmic aims (e.g., fairness to perceptions of justice, transparency to user understanding), the matrix advances both technical rigor and social relevance. This structured approach not only bridges the gap between robotics and sociological perspectives but also provides a practical, actionable framework for designing, evaluating, and governing context-aware robotic systems.

Despite increasing focus on the ethical and social aspects of humanoid robotics, the current scholarship often remains fragmented, treating ethics and public adoption as separate domains. This conceptual and methodological divide hampers our understanding of how ethical design can foster genuine trust and societal acceptance. To bridge this gap, the study is organized around five guiding research questions, each targeting a key aspect of the ethics–acceptance Nexus:

RQ1. What psychological, emotional, cultural, and symbolic factors shape public acceptance of humanoid robots, and how do these multi-layered influences interact beyond the boundaries of technical functionality or mechanical design?

This question seeks to identify the diverse determinants—cognitive, affective, and cultural—that inform individual and collective attitudes toward humanoid robots. It emphasizes the need to move beyond technocentric models of adoption to account for how societal narratives, lived experiences, and emotional responses co-construct trust and legitimacy.

RQ2. What normative expectations do users hold regarding the ethical behavior of humanoid robots, and how are principles such as transparency, fairness, and accountability interpreted across varying social contexts?

RQ2 aims to illuminate how ethical principles are socially constructed, emotionally interpreted, and contextually embedded in HRI. It interrogates the gap between abstract moral imperatives and the culturally situated expectations of real-world users.

RQ3. To what extent do current behavioral algorithms in humanoid robots embody the ethical principles necessary to foster public trust, and where do they fall short in aligning with socially meaningful interpretations of value and justice?

This question evaluates the fidelity of current algorithmic implementations to core ethical ideals, highlighting the limitations of formalistic or technocratic approaches. It calls for a shift toward ethically adaptive, user-centered, and socially resonant algorithmic design.

RQ4. How adequately does the existing literature integrate insights from public adoption research into ethical design frameworks, and what are the consequences of this disciplinary disconnection for human–robot coexistence?

RQ4 problematizes the siloed nature of current academic discourse, wherein technical ethics and HRI studies often operate in parallel rather than in dialogue. It examines how this fragmentation hinders the development of comprehensive, socio-technically integrated ethical standards.

RQ5. What theoretical and methodological limitations persist in the study of ethical behavior modeling and public trust in humanoid robotics, and how can participatory, cross-cultural, and ecologically valid research approaches address these gaps?

This final research question highlights the structural blind spots that limit inclusive and generalizable ethical design, emphasizing the need for methodological pluralism—especially field-based, longitudinal, and culturally sensitive research—to support the development of robots that are both technically sound and socially accepted.

Taken together, the research questions highlight the need to reassess the ethical foundations of humanoid robotics as evolving, socially embedded constructs rather than isolated technical parameters. The study's dual ambition is to reveal the critical disconnections between ethical algorithmic design and public perception and to present the HR-IEM as a practical, interdisciplinary tool for future research, policy, and development. Public trust is both a prerequisite for and a consequence of ethical robotic behavior: technically advanced robots may be rejected without social legitimacy, while ethical design requires continuous public engagement to reflect diverse norms and expectations. By treating algorithmic ethics and public acceptance as mutually constitutive, this study advances a holistic, integrative framework that is often lacking in fragmented AI ethics and HRI discourse.

Importantly, the HR-IEM is positioned as a practical roadmap for developers, policymakers, and regulators. As both a theoretical and applied framework, it maps ethical design objectives to measurable social trust indicators, providing actionable insights for the deployment of adaptive, culturally sensitive robotic systems. This orientation ensures the research's relevance to both academic inquiry and real-world robotics development.

The qualitative exploratory methodology, grounded in interdisciplinary literature analysis, enables the identification of key conceptual blind spots, theoretical tensions, and underexplored intersections across AI ethics, robotics, and HRI.

2. Methodology

This study employs a qualitative exploratory research design, using a narrative literature review to investigate the intersection of ethical algorithmic design and public acceptance of humanoid robots. Unlike systematic reviews, the narrative approach is chosen for its flexibility in integrating conceptual and empirical contributions from fragmented and interdisciplinary fields. This is critical for synthesizing insights across robotics, AI ethics, behavioral science, and cultural studies, where standard protocols often constrain theoretical innovation.

Data Sources and Keywords: The review is based on peer-reviewed literature retrieved from the Web of Science, Scopus, IEEE Xplore, and the ACM Digital Library, covering the period from 2013 to 2025. Key search terms included “humanoid robots,” “ethical algorithm design,” “public trust,” “human–robot interaction (HRI),” “participatory design,” “socio-technical integration,” and “cultural adaptation.” Studies were selected for their conceptual or empirical relevance to ethical, social, and emotional dimensions of humanoid robotics.

Analytical Process: Thematic content analysis was conducted using open and axial coding techniques. Initial open coding identified recurrent constructs, such as ethical expectations, emotional responses, and functional design considerations. These were then refined through axial coding to reveal higher-order thematic clusters, including gaps between normative ethical frameworks and lived user experiences, as well as the methodological limitations of laboratory-based research. This process enabled the synthesis of patterns across technical and social science literatures.

Strengths and Limitations: This narrative methodology offers conceptual depth and cross-disciplinary integration, yet lacks the reproducibility and protocol transparency characteristic of systematic reviews. Thematic relevance, rather than pre-registered protocols, guided the inclusion and exclusion criteria; therefore, the review is more interpretive than exhaustive. The HR-IEM framework, developed through this synthesis, remains conceptual and calls for empirical validation in future research.

By integrating perspectives from engineering, behavioral sciences, and applied ethics, the methodology supports a critical appraisal of current scholarship, informing the design of participatory, adaptive, and socially aligned robotic systems. This approach directly addresses concerns about methodological inclusivity and ecological validity that have been raised by reviewers and in the literature.

3. Findings

3.1 Social Dynamics of Public Adoption

The public adoption of humanoid robots is shaped by a complex interplay of cognitive, emotional, cultural, and contextual factors that extend far beyond assessments of technical capability or design aesthetics. These dynamics include psychological associations, cultural values, historical precedents, and emotional reactions that collectively define societal engagement with robotic technologies [10].

Three key thematic domains–public trust, emotional responses, and ethical algorithm design–emerge as pivotal to understanding the social integration of humanoid robots. Public trust is primarily driven by transparency, perceived fairness, and culturally appropriate conduct. Emotional responses, including empathy and alienation, are closely linked to both robot design and contextual deployment. Finally, ethical algorithm design must be adaptive and participatory, addressing both universal standards and local value systems. These insights jointly inform the operationalization of the HR-IEM across diverse contexts.

Three interrelated domains–public trust, emotional responses, and ethical algorithm design–are pivotal for understanding the social integration of humanoid robots. Trust is grounded in transparency, fairness, and contextually appropriate behavior; emotional responses, such as empathy or alienation, are influenced by robot design and the deployment setting. Ethical algorithm design requires adaptability to both universal and local value systems. These domains underpin the HR-IEM’s operational relevance across settings. Perceived usefulness is central to public acceptance, especially when robots fulfill socially valued roles, such as eldercare or educational support, rather than replace existing human relationships [11,12].

Anthropomorphism, by making robots more human-like, can facilitate intuitive interaction but also trigger ambivalence, as highlighted by the “uncanny valley” effect, where near-human likeness causes discomfort or eeriness [13,14]. Trust is multidimensional, shaped not only by expectations of reliability and ethical conduct but also by anxieties over agency, privacy, and potential manipulation–concerns heightened in environments where robots are equipped with sensors and cameras [15]. Beyond individual responses, public perception is strongly influenced by media and cultural narratives. While dystopian depictions in popular culture can reinforce fears, portrayals of robots as empathetic caregivers or companions foster openness and curiosity, particularly where direct real-life exposure is limited [16-18].

Cultural context further modulates adoption. For example, Japan’s traditions promote harmonious coexistence between humans and robots, while Western societies emphasize autonomy and privacy, and emerging economies may require simplified, community-oriented solutions. This diversity underscores the need for the HR-IEM to serve as a flexible framework, guiding ethical adaptation to local norms and expectations rather than imposing a one-size-fits-all model [12,19].

A practical example illustrates the application of HR-IEM: In a hospital, a humanoid robot utilizing HR-IEM would communicate transparently in the patient's native language, monitor fairness across demographic groups, and support emotional well-being through adaptive dialogue and real-time sentiment analysis. This operationalizes abstract ethical principles as concrete design and feedback mechanisms, aiding both developers and decision-makers.

Ultimately, emotional and symbolic factors—such as feelings of alienation or distrust—cannot be fully captured by usability metrics alone. Robots designed with expressive features can foster companionship, especially among vulnerable populations, but also introduce ethical concerns around overdependence or emotional manipulation [20]. Public attitudes evolve with experience and exposure, highlighting the need for longitudinal, adaptive research and participatory design. Genuine trust and ethical alignment emerge not through technical engineering alone, but via ongoing engagement with diverse stakeholders—end-users, caregivers, educators, and community leaders—who mediate the realities of HRI.

3.2 Ethical Concerns in Algorithmic Design

The ethical design of behavioral algorithms for humanoid robots is essential for fostering public trust and enabling responsible deployment. Ethical concerns in algorithmic design can be grouped into three core domains: technical transparency, socio-emotional accountability, and participatory value alignment. Socially accepted ethical behavior must not only meet technical criteria but also be emotionally resonant and responsive to community input. Core public expectations—transparency, fairness, predictability, and controllability—define the ethical baseline for robotic behavior, especially in sensitive domains such as healthcare, education, and public service [21-23]. Transparency involves not just explainable outputs, but clarity in how a robot's logic is communicated to users. Fairness requires the avoidance of bias and the promotion of equity and dignity across all contexts. Predictability and controllability are crucial for maintaining user trust and ensuring appropriate human oversight.

However, many AI ethics frameworks remain narrowly focused on algorithmic explainability and performance, overlooking relational ethics, contextual appropriateness, and socio-emotional consequences [6,24]. This technocentric focus often overlooks the lived realities and evolving expectations of end-users, particularly in fields where emotional resonance and cultural sensitivity are crucial for establishing trust. As a result, algorithmic designs can become detached from local norms and interpersonal dynamics.

Operationalizing the HR-IEM framework addresses these issues by directly linking ethical principles to practical outcomes. For instance, in healthcare, an ethical algorithm should make triage decisions transparent and fair for both staff and patients from diverse backgrounds. In education, robots must adapt their behavior based on feedback regarding inclusivity and cultural sensitivity, ensuring transparent and adjustable decision-making processes.

Cross-cultural application of HR-IEM highlights further complexities. In Japan, group harmony may shape algorithmic fairness through collective consultation, while in Western societies, individual consent and privacy are prioritized. In low- and middle-income countries, value-sensitive design may require simplified interfaces and strong community engagement. These variations emphasize the need for adaptive, context-sensitive governance and flexible ethical calibration.

A persistent gap remains in integrating societal feedback into the development of algorithms. Value-sensitive design—actively incorporating diverse stakeholder input—remains underutilized, resulting in algorithms that risk perpetuating bias or missing complex ethical realities of HRI. Institutionalizing participatory feedback loops and iterative co-design is critical for ongoing recalibration of ethical standards, ensuring that robots remain aligned with evolving social norms and expectations [25,26].

3.3 Disconnection Between Ethical Frameworks and Public Adoption

A central finding of this review is the persistent disconnection between abstract ethical frameworks in AI and the lived realities of public adoption for humanoid robots. While AI ethics has developed robust normative principles—autonomy, beneficence, non-maleficence, and justice—these are often articulated in decontextualized terms, with limited consideration of how robots are perceived and accepted by diverse user groups in everyday settings [27]. Prevailing engineering-centric approaches emphasize formal logic, risk mitigation, and algorithmic explainability, yet often neglect emotional responses, cultural narratives, and lived user experiences that fundamentally shape public attitudes [2,28]. This gap is especially pronounced in sensitive application areas such as healthcare and education. Compliance with formal ethical guidelines does not guarantee user acceptance if robots fail to align with local caregiving norms, emotional expectations, or patient values. Conversely, robots seen as emotionally attuned by users may still present ethical challenges if their algorithms remain opaque to regulators or professionals.

Socio-technical misalignment is further amplified in cross-cultural contexts. In Japan, acceptance is closely tied to social harmony, whereas in Western societies, concerns about privacy and autonomy are more prevalent. In resource-constrained settings, the lack of participatory design may lead to resistance, even if technical standards are met. These realities underscore the need for ethics frameworks to be continuously recalibrated through dialogue with affected communities.

Ultimately, ethical compliance alone is insufficient for meaningful public acceptance. Bridging this gap requires cross-disciplinary collaboration among engineers, ethicists, and social scientists, informed by empirical insights into public perception, behavior, and values [25,28]. The HR-IEM proposed in this study addresses this challenge by mapping formal ethical objectives to operational trust indicators, enabling systematic evaluation of both technical compliance and public legitimacy across diverse environments. In doing so, the HR-IEM moves ethical design beyond theory, offering developers and policymakers a practical roadmap for aligning algorithmic principles with the realities of social integration.

3.4 Theoretical and Methodological Gaps

A review of the literature highlights persistent theoretical and methodological shortcomings in the study of ethical behavior modeling and public adoption of humanoid robots. Chief among these is the limited use of participatory design and user-in-the-loop evaluation. Existing frameworks are predominantly shaped by top-down, expert-driven processes that often overlook the perspectives of end users and marginalized groups directly affected by robotic deployment [29], [30]. This exclusion constrains the ability to integrate nuanced expectations, ethical concerns, and culturally specific values into algorithmic design.

A further limitation is the lack of cross-cultural analysis, as most studies are based in Western contexts, thereby neglecting the variation in ethical norms, social practices, and user perceptions that occur globally. This narrow focus undermines the generalizability and applicability of ethical design principles for diverse environments [31,32]. Moreover, the methodological reliance on laboratory-based studies fails to capture the dynamic complexity of real-world HRI, restricting insight into how robots are interpreted and trusted in daily life. There is a clear need for more field-based, longitudinal research to evaluate ethical behavior across varied, evolving social contexts [33,34]. To address these gaps, future research should systematically incorporate co-design workshops, participatory prototyping, and iterative feedback from diverse stakeholders to improve ecological validity and societal relevance. The HR-IEM framework introduced in this study addresses these challenges by promoting interdisciplinary, participatory, and context-sensitive research strategies that are adaptable to real-world applications.

4. Discussion

4.1 Multilayered Determinants of Public Acceptance Beyond Utility and Form

Findings related to RQ1 indicate that public acceptance of humanoid robots is not solely determined by technical capability or design, but rather by a complex interplay of psychological, emotional, and cultural factors. Perceived usefulness and human-likeness can increase openness, yet are inseparable from emotional reactions, such as empathy, unease, or alienation, that are filtered through societal narratives and local values. The "uncanny valley" effect illustrates that even minor discrepancies in appearance or behavior can provoke discomfort, underscoring the importance of sensitive, context-aware design. The cultural setting further modulates these reactions: societies accustomed to harmonious human–robot relations are generally more accepting, whereas those with histories of disruption or technological anxiety are more skeptical. These findings confirm that attitudes toward robots are highly context-dependent, highlighting the limitations of universal design approaches. Effective integration thus requires frameworks that account for this diversity, prioritizing participatory and culturally sensitive engagement with stakeholders throughout the development process. In summary, public acceptance hinges on robots' ability to align with the emotional and cultural expectations of the communities they serve. The main determinants of acceptance are outlined in Table 1.

Table 1. Summary of factors influencing public acceptance of humanoid robots (RQ1).

Factor Category	Key Elements	Implications for Design and Policy
Psychological	Perceived usefulness, human-likeness, familiarity, discomfort	Avoid uncanny valley triggers; calibrate anthropomorphic features
Emotional	Empathy, fear, alienation, and companionship	Design for socio-emotional intelligence; prevent manipulation risks
Cultural	Religious beliefs, historical context, media representations, and national values	Customize design and deployment strategies across cultural contexts.
Symbolic	Representations of progress, control, disruption, or care	Engage in value-sensitive communication strategies
Temporal (Longitudinal)	Evolving trust based on exposure, learning, and consistency	Incorporate adaptive feedback loops; monitor sentiment over time

4.2 Ethical Expectations as Drivers of Trust, Resistance, and Social Legitimacy

Findings related to RQ2 demonstrate that ethical expectations shaped by lived experience, social context, and culture are fundamental to public trust, resistance, and the legitimacy of humanoid robots. Principles like transparency and fairness are meaningful to users only when enacted through emotionally resonant and contextually appropriate behavior. True transparency involves robots clearly and empathetically explaining their actions, while fairness requires inclusivity and responsiveness to diverse social needs, particularly in sensitive settings such as caregiving or education. A major weakness in current practice is the lack of participatory feedback and emotional accountability. When robots appear manipulative or unresponsive, public trust is eroded—even when ethical codes are formally met. This highlights the danger of static, checkbox approaches to ethics. To address this, robot design and deployment must embrace adaptive, user-centered, and culturally responsive ethical frameworks. Such flexibility is vital for building sustainable trust and social legitimacy. The principal ethical drivers of public trust and legitimacy are summarized in Table 2.

Table 2. Core ethical expectations shaping public attitudes toward humanoid robots (RQ2).

Ethical Principle	Public Interpretation	Design and Policy Implications
Transparency	Clarity in actions, visibility of intent, and explainability in lay terms	Design robots that communicate decisions clearly and narrate their reasoning
Fairness	Avoidance of discrimination, inclusive behavior, and equitable treatment across demographics	Include fairness metrics in algorithm testing and deploy bias-monitoring tools to ensure fairness and equity.
Predictability	Behavioral consistency, avoidance of surprises, or erratic actions	Establish clear interaction patterns and context-aware behavioral protocols
Controllability	User agency, override capability, assurance of human oversight	Ensure manual intervention possibilities and user-informed consent mechanisms
Emotional Accountability	Avoidance of manipulation, support for emotional well-being	Limit persuasive technologies; integrate affective ethics safeguards
Cultural Sensitivity	Respect for norms, values, taboos, and localized moral codes	Localize ethical parameters; embed cultural calibration modules into behavior
Participation	Voice in design, representation in feedback processes	Implement co-design strategies and stakeholder consultations

4.3 The Disjunction Between Technical Ethics and Socially Meaningful Behavior

Findings related to RQ3 reveal a persistent gap between established ethical principles—such as transparency, fairness, contextual responsiveness, and accountability—and their practical application in HRI. While these principles are foundational in AI ethics, their implementation in humanoid robotics often remains technocentric, focusing on algorithmic compliance rather than providing a meaningful user experience. Technical measures for transparency and fairness frequently fall short of public expectations for clear, intuitive, and culturally attuned communication. As a result, formal compliance does not guarantee trust or acceptance; users judge robots by their ability to respond to emotional, cultural, and situational nuances. The limited adoption of participatory, adaptive design processes worsens this disconnect. Static behaviors undermine both ethical legitimacy and user confidence, especially in diverse or changing settings. Bridging this gap requires iterative, user-centered development cycles that actively incorporate community feedback. In doing so, ethical frameworks become practical tools for building the trust and legitimacy required for the real-world integration of robots. The main operational gaps between ethical principles and practice are summarized in Table 3.

Table 3. Gaps between ethical principles and current algorithmic practices in humanoid robots (RQ3).

Ethical Principle	Ideal Implementation	Observed Deficiencies in Practice	Bridging Strategy
Transparency	Intuitive, user-friendly explanations of decision-making	Technocratic explainability that excludes lay users	Employ natural language rationale generators; embed affective signaling
Fairness	Sensitivity to structural biases and interpersonal justice	Statistical fairness metrics detached from social context	Introduce community-based fairness calibration
Value Alignment	Integration of pluralistic human values via participatory design	Top-down encoding of values; insufficient stakeholder involvement	Apply iterative, co-creative design with real users
Contextual Responsiveness	Adaptation to varying social settings, emotional tones, and cultural cues	Uniform algorithmic behavior across contexts	Implement adaptive ethics modules and situational modeling
Accountability	Mechanisms for oversight, traceability, and user recourse	Opaque decision chains; lack of explainable justification mechanisms	Build audit-friendly decision trails and user-facing feedback channels

4.4 The Persistent Disjunction Between Ethical Frameworks and Lived Social Realities

Findings related to RQ4 reveal a fundamental gap between formal AI ethical frameworks and the realities of public adoption in humanoid robotics. While principles such as autonomy, justice, and beneficence are well-articulated, they are often detached from the emotional and cultural dynamics that shape real-world HRI. As a result, robots that meet technical ethical standards may still encounter public resistance if they fail to consider local norms, emotional expectations, or cultural values. This disconnect underscores the limitations of engineering-centric ethics, which often prioritize universal logic over relational and contextual understanding. Evidence from domains such as healthcare and education suggests that formal compliance is insufficient for building trust, particularly when participatory feedback and community engagement are lacking.

Moreover, the lack of integration between ethical design and social science insights leads to frameworks that are externally imposed rather than co-created with users. Bridging this divide requires a socio-technical, co-design approach that makes trust, emotion, and local context central to ethical development. Only through genuine interdisciplinary collaboration can ethical frameworks become credible and adaptable in diverse human–robot settings. Table 4 outlines the main disconnects between ethical frameworks and public adoption.

Table 4. Disjunctions between ethical frameworks and public adoption realities (RQ4).

Dimension of Disjunction	Ethics Literature Focus	Public Adoption Perspective	Integration Strategy
Ethical Principle Articulation	Abstract, decontextualized (e.g., autonomy, beneficence)	Emotionally grounded, socially situated trust and acceptance norms	Translate abstract principles into lived experiences and symbolic interpretations
Normative Evaluation Criteria	Compliance with algorithmic transparency and logic	Perceived respect, relational trust, and cultural fit	Co-develop evaluative indicators with user communities
Design Priorities	Formal correctness, explainability, and safety	Empathy, adaptability, and emotional intelligibility	Embed socio-emotional design into early algorithm development
Methodological Approach	Top-down, expert-driven	Bottom-up, user-centered, and participatory	Institutionalize participatory ethics in design and deployment cycles
Disciplinary Silos	Ethics vs. HRI vs. Sociology vs. Engineering	Fragmented perspectives and limited shared vocabulary	Foster cross-disciplinary research consortia and convergence frameworks

4.5 Structural Omissions That Hinder Ethical and Inclusive Human–Robot Integration

Findings related to RQ5 identify major blind spots in both the theoretical and methodological approaches to modeling ethical behavior in humanoid robotics. Currently, expert-driven models often marginalize the perspectives of end users—especially those from non-Western or underrepresented backgrounds—limiting the diversity of values embedded in algorithms and increasing the risk of social misalignment during deployment. Most frameworks overlook the diverse range of emotional norms, cultural narratives, and social expectations that influence human responses to robots across various contexts. Methodologically, the field is constrained by an overreliance on laboratory simulations, which cannot fully capture how trust and ethical acceptance evolve in real-world, dynamic environments. The lack of ecologically valid, field-based, and longitudinal studies hinders the assessment of how ethical behaviors persist or adapt over time and misses emerging concerns that only arise in practice. A further limitation is the insufficient use of participatory design practices. Few studies actively involve diverse user groups in the co-design and evaluation of ethical algorithms, limiting opportunities to identify and address normative misalignments before deployment. Addressing these gaps requires a shift toward inclusive, context-sensitive, and participatory research methodologies that reflect real-world diversity and complexity. Such approaches are crucial for developing robotic systems that are not only technically proficient but also ethically credible and socially acceptable. Key theoretical and methodological challenges are summarized in Table 5.

Table 5. Key theoretical and methodological gaps in ethical humanoid robot design (RQ5).

Gap Type	Specific Deficiencies	Implications	Proposed Remedy
Theoretical	Overreliance on universalist ethical principles	Ignores moral pluralism; limits adaptability in cross-cultural deployments	Integrate moral anthropology and contextual ethics into framework development.
Participatory Design	Lack of user-in-the-loop methods; marginalization of affected groups	Excludes community values; weakens legitimacy and inclusivity	Employ co-design strategies; democratize the ethical modeling process
Cross-Cultural Validity	Western-centric focus on case selection and evaluation	Undermines generalizability; risks cultural mismatch in non-Western contexts	Incorporate comparative studies; apply localized value-sensitive calibration
Methodological Rigor	Dependence on controlled lab studies; lack of real-world complexity	Fails to capture emergent behaviors and long-term social effects	Conduct longitudinal, field-based HRI evaluations
Evaluation Metrics	Absence of emotional, symbolic, and community-based trust indicators	Incomplete assessment of "successful" ethical behavior	Develop trust resonance metrics tied to cultural and affective dynamics

4.6 Integrative Framework Proposal: Operationalizing Ethical-Social Alignment through HR-IEM

Drawing upon the synthesis of findings across all five research questions, this study proposes the HR-IEM as a strategic tool to bridge the gap between abstract ethical principles and the concrete realities of public trust and adoption. The HR-IEM is designed to systematically align algorithmic objectives—including transparency, fairness, emotional sensitivity, and cultural adaptability—with empirically derived social trust indicators such as perceived inclusiveness, emotional resonance, and symbolic legitimacy. This alignment is operationalized through a matrix that can be flexibly applied in varied contexts, ranging from healthcare and education to public services, providing a practical roadmap for developers, regulators, and policymakers. To demonstrate how the framework is applied, consider a healthcare scenario: A hospital introduces a humanoid robot assistant designed to support nursing staff and patient care. Applying the HR-IEM, developers first map ethical goals (e.g., ensuring fairness in patient prioritization, transparent communication of medical information, and emotional sensitivity in end-of-life conversations) to corresponding trust indicators, such as patient comprehension, perceived respect, and emotional comfort. The matrix then guides iterative user testing and participatory feedback sessions, where real patients and caregivers assess the robot's behavior for cultural fit and ethical alignment. This process enables adaptive refinements, ensuring the robot's actions remain attuned to local norms and emotional needs while still adhering to universal ethical guidelines. The same approach can be adapted to educational contexts, where transparency and inclusiveness in classroom assistance are calibrated to specific cultural and demographic settings.

Crucially, the HR-IEM is structured to accommodate cross-cultural and context-specific adaptation. For example, in Japan, where social harmony and relational trust are highly valued, the matrix prioritizes emotional sensitivity and non-disruptive communication as key indicators of trust. In contrast, Western settings may emphasize transparency, agency, and user autonomy. In resource-constrained or emerging economies, inclusiveness, accessibility, and participatory feedback are foregrounded, with scenario-specific adaptation of algorithmic parameters. This cultural responsiveness is not just a theoretical add-on; it is embedded into the HR-IEM's iterative use, making the tool viable for global deployment.

As a practical roadmap, HR-IEM guides interdisciplinary teams through a sequence of steps: (1) identify context-relevant ethical objectives; (2) map these objectives to local social trust indicators; (3) implement participatory evaluation cycles with target users; (4) adapt algorithmic behavior through continuous stakeholder feedback; and (5) measure and report on trust, inclusion, and emotional alignment over time. The matrix also supports longitudinal assessment, allowing for the tracking of how trust and acceptance evolve as robotic systems are iteratively deployed and refined. Presented in Table 6, the HR-IEM transforms ethical design from a static, compliance-based exercise into an ongoing, co-creative process that integrates empirical research, cultural sensitivity, and participatory governance. By grounding ethical algorithmic development in real-world social trust dynamics, the framework not only addresses persistent gaps in current practice but also charts a practical and inclusive path forward for human–robot coexistence.

Table 6. Human–robot integration ethics matrix.

Algorithmic Design Objective	Corresponding Social Trust Indicator	Evaluation Context
Transparency	Perceived clarity and user understanding of decisions	Healthcare and education scenarios
Fairness	Perceived inclusiveness and absence of bias	Multi-user and demographic-sensitive environments
Predictability	Confidence in consistent and reliable robot behavior	Domestic and public service tasks
Emotional Sensitivity	Emotional resonance and avoidance of manipulation	Eldercare and emotionally sensitive interactions
Cultural Responsiveness	Cultural alignment and symbolic appropriateness	Cross-cultural deployments and localization
Accountability	Perceived oversight, recourse, and ethical governance	Regulated domains (e.g., public institutions, finance)

Figure 2 illustrates the HR-IEM as an integrative framework that systematically maps core algorithmic objectives, such as transparency, fairness, emotional sensitivity, and cultural responsiveness, to their corresponding social trust indicators and application domains. This diagram visualizes the dynamic interplay between ethical design imperatives and real-world contextual demands, serving as a practical guide for aligning technical development with public expectations. By providing a structured visualization of these relationships, Figure 3 complements the thematic synthesis presented in the accompanying table and supports the operationalization of socio-ethical alignment in humanoid robotics.

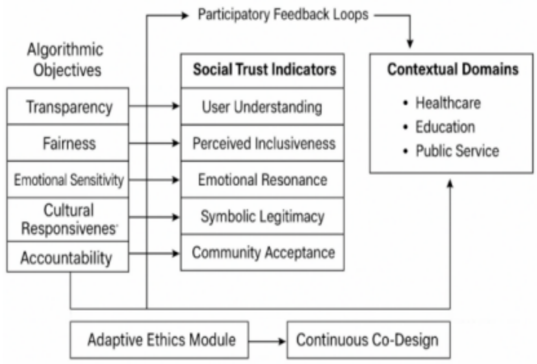


Figure 2. Human–robot integration ethics matrix.

5. Conclusion

5.1 Thematic Synthesis

This study examined the intricate relationship between public acceptance of humanoid robots and the ethical design of behavioral algorithms, revealing that these dimensions are deeply intertwined. While technological performance and aesthetic design remain relevant to adoption, the literature emphasizes that emotional resonance, normative alignment, and cultural intelligibility are equally critical to fostering public trust. Therefore, ethical behavior in robots must be conceptualized not merely as compliance with abstract principles, but as a dynamic, emotionally and socially responsive process shaped by the contextual expectations of humans. A key contribution of this study is the development of the HR-IEM, which systematically links algorithmic goals (e.g., transparency, fairness, responsiveness) to social trust indicators, offering a structured foundation for designing, evaluating, and governing ethical behavior in humanoid systems. This framework bridges two traditionally fragmented literatures—algorithmic ethics and HRI—and shifts the ethical conversation beyond technocratic and laboratory-centric paradigms toward one that is participatory, emotionally attuned, and culturally grounded. The key thematic insights and their practical implications are synthesized in Table 7. This table presents a concise overview of how each primary research focus—public trust, emotional response, media and culture, ethical algorithm design, disciplinary disjunctions, participatory design, empirical scope, and framework proposal—translates into concrete design and policy recommendations for advancing ethically robust and socially accepted humanoid robotics.

Table 7. Synthesis of key thematic findings and research implications.

Thematic Focus	Key Insights	Implications
Public Trust and Acceptance (RQ1 & RQ2)	Public adoption is shaped by emotional, cultural, and social factors, not just technical utility or design appeal.	Robot design must consider social roles, anthropomorphism, and culturally sensitive interaction norms to ensure effective and respectful communication.
Emotional and Symbolic Responses	Feelings of empathy or alienation are triggered by how robots represent or disrupt human identity and roles.	Include socio-emotional awareness in robot behavior and communication design.
Media and Cultural Influence	Media representations and local cultural beliefs significantly shape public expectations and fears.	Communication strategies and deployment plans should be tailored to the local context and informed by cultural considerations.
Ethical Algorithm Design (RQ2 & RQ3)	Current ethics frameworks focus on explainability and logic but neglect contextual, relational, and emotional aspects.	Develop adaptive, value-sensitive algorithms with embedded societal feedback loops.
Disconnection Between Ethics and Adoption (RQ4)	Ethics and public adoption research are treated separately, missing their reciprocal influence.	Foster interdisciplinary collaboration to bridge technical ethics and social acceptance research.
Lack of Participatory Design (RQ5)	Most ethical design processes exclude users and stakeholders from the decision-making process.	Encourage co-design frameworks that incorporate diverse user input from early development stages.
Overreliance on Lab Studies	Many findings are derived from lab-based experiments, lacking real-world contextual validity.	Increase investment in field-tested, longitudinal HRI studies in public and domestic environments.
Proposed Framework: HR-IEM	Aligns algorithmic behavioral goals with social trust indicators to evaluate ethical-social fit.	Use as a tool for interdisciplinary teams to assess deployment readiness and ethical alignment.

5.2 Implications for Research and Practice

These findings have important implications for multiple stakeholders. Developers and engineers are encouraged to incorporate ethical objectives, including emotional intelligence, cultural sensitivity, and continuous user feedback, into the design and deployment lifecycle. The HR-IEM serves as a practical compass for navigating this complexity and ensuring that robotic systems are responsive not only to technical standards but also to the values and expectations of diverse user groups.

For policymakers and regulators, the study suggests a shift from generic ethical codes toward trust-based, context-specific evaluation metrics, including the institutionalization of public trust as a formal design outcome and the requirement of social impact assessments in real-world deployments. Researchers are encouraged to prioritize interdisciplinary methodologies that integrate technical performance metrics with rich qualitative insights, thereby making ethical–social alignment a central objective in both HRI and AI ethics scholarship. The HR-IEM framework provides a platform for dialogue among engineers, ethicists, social scientists, and end-users, fostering adaptive algorithmic governance and supporting the social legitimacy of next-generation robotic technologies.

5.3 Limitations

Despite its integrative and conceptual strengths, this study is subject to several important limitations. First, the reliance on a narrative literature review rather than a systematic, PRISMA-guided approach introduces constraints in terms of transparency, replicability, and the breadth of source inclusion. The review was guided by thematic and conceptual relevance, which may have omitted some pertinent perspectives, particularly from adjacent fields such as anthropology, human factors, or law.

Second, the HR-IEM framework, while theoretically robust, remains untested in empirical settings. Its practical efficacy must be validated through field applications and experimental studies. Third, much of the literature synthesized centers on Western, technologically advanced societies, which limits the cross-cultural generalizability of the findings. Broader global engagement is required to capture the full spectrum of values, norms, and expectations relevant to HRI. Lastly, the rapid pace of technological change in robotics and AI necessitates continuous updating of the HR-IEM and ongoing reassessment of ethical standards to maintain relevance.

5.4 Future Research Directions

Building on the limitations and open questions identified, several promising avenues for future research are proposed. There is a pressing need for contextual, cross-cultural studies that investigate how ethical behavior in humanoid robots is interpreted and valued in diverse real-world settings, such as households, care facilities, or public institutions, across different cultures and demographics. The development and systematic evaluation of participatory co-design frameworks are likewise essential for ensuring that marginalized voices and user communities have a meaningful role in shaping robotic ethics.

Research should prioritize the development of adaptive ethics modules: AI components that can interpret emotional cues, adapt to dynamic social contexts, and learn from ongoing user feedback. Most importantly, future studies must empirically validate the HR-IEM framework, applying it in longitudinal field studies and experimental deployments to assess its diagnostic and predictive power for ethical–social alignment. Such work will not only deepen our understanding of the complex interplay between technology and society but also provide a robust foundation for the responsible, legitimate, and sustainable integration of humanoid robots into everyday human life.

Generative AI Statement

The authors declare that no Gen AI was used in the creation of this manuscript.

References

[1] Gao Y, Chang Y, He Y, Yu Z. Under what conditions will I let social robots into my home? A configuration theory perspective. Journal of Research in Interactive Marketing, 2025. DOI: 10.1108/JRIM-07-2024-0363

[2] Vozna A, Costantini S. Ethical, legal, and societal dimensions of ai-driven social robots in elderly healthcare. Intelligenza Artificiale, 2025, 19(1), 30-40. DOI: 10.1177/17248035241310192

[3] Maj K, Modliński A, Małaszniak W, Kamerys J. When the robot says “no”: The impact of robot agreement and disagreement on user's trust, perceived expertise and compliance. PsyArXiv Preprints, 2025. DOI: 10.31234/osf.io/2m89h

[4] Gao W, Shen S, Ji Y, Tian Y. Human perception of the emotional expressions of humanoid robot body movements: evidence from survey and eye-tracking measurements. Biomimetics, 2024, 9(11), 684. DOI: 10.3390/biomimetics9110684

[5] Donati D. Trust, trustworthiness and the moral dimension in human-AI interactions. Ethics, Politics & Society, 2024, 7(2). DOI: 10.21814/eps.7.2.6181

[6] Isop WA. A conceptual ethical framework to preserve natural human presence in the use of AI systems in education. Frontiers in Artificial Intelligence, 2025, 7, 1377938. DOI: 10.3389/frai.2024.1377938

[7] Ferhataj A, Memaj F, Sahatcija R, Ora A, Koka E. Ethical concerns in AI development: Analyzing students’ perspectives on robotics and society. Journal of Information, Communication and Ethics in Society, 2025, 23(2), 165-87. DOI: 10.1108/JICES-08-2024-0111

[8] Zafeirakopoulos D, Stefaneas P. Toward responsible AI: A framework for ethical design utilizing deontic logic. International Journal on Artificial Intelligence Tools, 2024, 33(08), 2550003. DOI: 10.1142/S0218213025500034

[9] Díez-Calvo S, Lidon I, Rebollar R, Gil-Perez I. Problems of participatory processes in policymaking: A service design approach. Journal of Services Marketing, 2025, 39(3), 180-201. DOI: 10.1108/JSM-03-2024-0137

[10] Giger JC, Piçarra N, Pochwatko G, Almeida N, Almeida AS. Intention to work with social robots: The role of perceived robot use self-efficacy, attitudes towards robots, and beliefs in human nature uniqueness. Multimodal Technologies and Interaction, 2025, 9(2), 9. DOI: 10.3390/mti9020009

[11] Nertinger S, Herzog O, Mühlbauer A, Naceri A, Haddadin S. Designing humanoids: How robot posture influences users’ perceived safety in HRI. In 2024 IEEE-RAS 23rd International Conference on Humanoid Robots (Humanoids) 2024, pp. 45-52. DOI: 10.1109/Humanoids58906.2024.10769912

[12] Kartika D, Bahri S. Social humanoid robot dalam perspektif syariat dan etika. Reslaj: Religion Education Social Laa Roiba Journal, 2024, 6(12), 5171. DOI: 10.47467/reslaj.v6i12.3750
[13] Kim T, Lee OK, Kang J. Why people trust AI software robots: A mediated moderation perspective on the interaction between their intelligence and appearance. Industrial Management & Data Systems, 2025. DOI: 10.1108/IMDS-04-2024-0329
[14] Sasaki K, Yonemitsu F, Ariga A. The uncanny valley phenomenon can be explained by categorization failure rather than categorization difficulty. Visual Cognition, 2024, 32(5), 388-399. DOI: 10.1080/13506285.2024.2448697
[15] Hannig M, Stock-Homburg R, Knof M. The privacy paradox in interactions with service robots in the workplace. The International Journal of Human Resource Management, 2025, 36(3), 385-413. DOI: 10.1080/09585192.2024.2445135
[16] Adebiyi AA, Adejumobi HO, Aliyu ME, Adeosun KG. Fake news and national security threats in Nigeria: The role of media in shaping public perception and policy responses. African Journal of Social Sciences and Humanities Research, 2025, 8(1), 46-60. [Online]. Available: https://abjournals.org/ajsshr/papers/volume-8/issue-1/fake-news-and-national-security-threats-in-nigeria-the-role-of-media-in-shaping-public-perception-and-policy-responses/ (accessed on 24 May 2025)
[17] Sarwar B. The politics of representation: arundhati roy's re-framing of media narratives in the ministry of utmost happiness. Spectrum, 2023, 18, 10-21. DOI: 10.3329/spectrum.v18i1.76354
[18] Zafar MB, Khalid R, Iqbal M, Hussain, M. Cultural contours of justice: Law and crimes narratives in popular imagination. Journal for Social Science Archives, 2025, 3(1), 118-127. DOI: 10.59075/jssa.v3i1.102
[19] Tsai G, Choo LC. Debts of gratitude in cross-cultural perspective: Confucian and western ethics. Journal of Applied Philosophy, 2025, 42(1), 131-154. DOI: 10.1111/japp.12767
[20] Gerlich M. Societal perceptions and acceptance of virtual humans: Trust and ethics across different contexts. Social Sciences, 2024, 13(10), 516. DOI: 10.3390/socsci13100516
[21] Sagar K, Shende T, Jaiswal A. The ethical algorithm: Human values in autonomous AI. International Journal For Multidisciplinary Research, 2024, 6(6). DOI: 10.36948/ijfmr.2024.v06i06.32275
[22] Grace O, Drummond J. Investigating public attitudes toward ai-driven medical technologies: A review of literature on trust and ethical considerations. OSF Preprints, 2025. DOI: 10.31219/osf.io/a9sxw
[23] Pasipamire N, Muroyiwa A. Navigating algorithm bias in AI: Ensuring fairness and trust in Africa. Frontiers in Research Metrics and Analytics, 2024, 9, 1486600. DOI: 10.3389/frma.2024.1486600
[24] Nadim MA, Di Fuccio R. Unveiling the potential: Artificial intelligence's negative impact on teaching and research considering ethics in higher education. European Journal of Education, 2025, 60(1), e12929. DOI: 10.1111/ejed.12929
[25] Abbo GA, Belpaeme T, Spitale M. Concerns and values in human-robot interactions: A focus on social robotics. arXiv preprint arXiv:2501.05628, 2025. DOI: 10.48550/arXiv.2501.05628
[26] Poulsen A, Burmeister OK, Greig J, Ulhaq A, Tien D. Value sensitive design of social robots: Enhancing the lives of LGBT+ older adults. International Journal of Social Robotics, 2025, 17(1), 147-162. DOI: 10.1007/s12369-024-01201-6
[27] Mårell-Olsson E, Bensch S, Hellström T, Alm H, Hyllbrant A, Leonardson M, Wet al. Navigating the human-robot interface-exploring human interactions and perceptions with social and telepresence robots. Applied Sciences, 2025, 15(3), 1127. DOI: 10.3390/app15031127
[28] Willem T, Fritzsche MC, Zimmermann BM, Sierawska A, Breuer S, Braun M, et al. Embedded ethics in practice: A toolbox for integrating the analysis of ethical and social issues into healthcare AI research. Science and Engineering Ethics, 2024, 31(1), 3. DOI: 10.1007/s11948-024-00523-y
[29] Gutsche K, Genovese J, Serstjuk P, Altindis S. User-centered design of professional social service robots. In Intelligent Human Systems Integration (IHSI 2025): Integrating People and Intelligent Systems, AHFE Open Acces, 2025, 160, 10-20. DOI: 10.54941/ahfe1005806
[30] Tang Y. UX and ethics. In The Routledge Handbook of Ethics in Technical and Professional Communication, 1st ed.; Routledge, 2025.
[31] Chien SY, Wang YF, Cheng KT, Chen YC. Comparative study of XAI perception between eastern and western cultures. International Journal of Human-Computer Interaction, 2024, 1-7. DOI: 10.1080/10447318.2024.2441015
[32] Wang G, Zhang Z, Xie S, Guo Y. Province of origin, decision-making bias, and responses to bureaucratic versus algorithmic decision-making. Public Administration Review, 2025. DOI: 10.1111/puar.13928
[33] Chatterjee T, Guzzo A, Tlaie A, Kaddour A, Papaxanthis C, Gaveau J, et al. Following the robot's lead: Predicting human and robot movement from EEG in a motor learning HRI task. iScience, 2025. DOI: 10.1016/j.isci.2025.112914
[34] Chojnowski O, Eberhard A, Schiffmann M, Müller A, Richert A. Human-like nonverbal behavior with metahumans in real-world interaction studies: An architecture using generative methods and motion capture. In 2025 20th ACM/IEEE International Conference on Human-Robot Interaction (HRI), 2025, pp. 1279-1283. DOI: 10.1109/HRI61500.2025.10973925