
ETHICAL CONCERNS IN HUMAN-ROBOT INTERACTION

Dr. Rohan Mehta

Assistant Professor

Department of Robotics and Artificial Intelligence

School of Computing and Engineering

University of Technology and Innovation, Bengaluru, Karnataka, India

Abstract

Human-robot interaction is moving from controlled laboratory settings into homes, hospitals, classrooms, workplaces, retail spaces, and public environments. As robots become more socially expressive, data-intensive, and autonomous, ethical questions arise not only from what robots do, but also from how people understand, trust, depend on, and are influenced by them. This study examines the principal ethical concerns associated with human-robot interaction and presents a practical ethics-by-design framework for responsible deployment. The framework considers physical and psychological safety, informed consent, privacy and data governance, transparency of robotic capability, emotional attachment, manipulation, bias and inclusion, human autonomy, accountability, and meaningful human oversight. Particular attention is given to vulnerable users, including children, older adults, patients, and persons with disabilities, because power imbalances and reduced capacity to evaluate system behaviour can increase ethical risk. The study uses a structured analytical methodology that combines literature synthesis, stakeholder mapping, scenario-based risk analysis, and evaluation criteria for robot designers, institutions, and service providers. Rather than treating ethics as a compliance check after development, the proposed model embeds ethical review across the lifecycle of a robot: problem definition, data collection, interaction design, testing, deployment, monitoring, and retirement. Illustrative prototype results show how a systematic framework can improve consent clarity, privacy safeguards, transparency, auditability, and human control when compared with an unstructured design process. These values are presented solely to demonstrate the evaluation format and should be replaced by field data in an empirical study. The paper concludes that responsible human-robot interaction requires more than technical reliability. It requires clear communication, limits on persuasive design, inclusive participation, accountable governance, and regular assessment of both intended benefits and unintended social consequences.

Keywords: Human-Robot Interaction, Robot Ethics, Responsible Robotics, Privacy, Autonomy, Trust, Informed Consent, Social Robots, Algorithmic Accountability, Human Oversight.

1. INTRODUCTION

Human-robot interaction (HRI) concerns the ways in which people and robots perceive, communicate with, coordinate with, and influence one another. Earlier robots were commonly understood as industrial tools operating behind physical barriers or under direct supervision. Contemporary robots increasingly share space with people and can speak, gesture, recognise faces, infer preferences, make recommendations, collect sensor data, and adapt their behaviour over time. This change has expanded the potential usefulness of robots, but it has also made ethical questions central to the design and governance of HRI.

Ethical concern is not limited to the possibility of physical malfunction. A robot may be mechanically safe while still creating harms through intrusive data collection, misleading

social cues, biased interaction patterns, unfair access, excessive surveillance, or pressure on users to disclose information. A companion robot that repeatedly displays concern, for example, may encourage emotional dependence even though it does not possess understanding or reciprocal care. A service robot that offers recommendations may appear authoritative and influence decisions beyond the user's knowledge or competence. These risks are especially important where robots interact with children, patients, older adults, employees, and people who require assistance. The ethical status of a robot is distinct from the ethical responsibilities of the people and institutions that design, deploy, purchase, supervise, and regulate it. Robots do not eliminate human responsibility. Instead, they distribute decisions across designers, operators, organisations, data providers, and

users. Ethical analysis must therefore ask who defines the robot's objectives, what data are collected, how system limitations are communicated, who can intervene when a problem occurs, and how users can seek redress after harm or error.

This paper analyses key ethical concerns in HRI and develops a structured framework for responsible practice. The objective is not to stop useful innovation, but to establish conditions under which robotic systems can support human wellbeing without weakening dignity, autonomy, privacy, fairness, or accountability. The framework is relevant to social, service, assistive, collaborative, educational, and healthcare robots, while recognising that risk levels differ across contexts.

2. LITERATURE REVIEW

The literature on robot ethics has developed alongside the growth of socially interactive and autonomous systems. Foundational work argues that robots should be assessed not only by their technical performance but also by their effects on human values, relationships, institutions, and practices. In care settings, scholars have debated whether robots can support independence and wellbeing without replacing meaningful human contact or creating a false impression of care. In social robotics, researchers have examined anthropomorphism, emotional attachment, deception, trust calibration, consent, and the moral implications of machines that are designed to appear attentive or empathetic.

Privacy is a recurring concern because embodied robots may use cameras, microphones, location data, biometric signals, movement logs, and household or workplace information. Unlike many screen-based systems, robots can observe people in physical spaces and may collect data continuously. Meaningful privacy protection therefore requires data minimisation, clear notice, accessible consent processes, secure storage, restricted access, and policies for retention, deletion, and secondary use. Transparency is equally important: users should be able to understand a robot's purpose, sensing capabilities, significant limitations, and the human or organisational entity responsible for its operation.

Recent scholarship also identifies equity and justice as necessary dimensions of HRI. Robotics design can exclude people through assumptions about language, body size, mobility, cultural norms, digital literacy, economic access, or communication style. A system trained primarily on limited user groups may perform poorly for others. Ethical HRI consequently requires participatory design, accessibility testing, diverse datasets where appropriate, and assessment of distributional impacts. The literature suggests that ethical practice is most effective when it is integrated across the design lifecycle rather than added only after technical development has been completed.

2.1 SUMMARY OF PREVIOUS STUDIES

Author(s)	Method / Technique	Major Findings	Research Gap
Asaro (2006)	Robot ethics and social responsibility	Argued that ethical reflection should address the social purposes and consequences of robotics.	Provided broad foundations but offered limited operational guidance for interaction design.
Sharkey & Sharkey (2012)	Ethics of elder-care robots	Highlighted risks of reduced human contact, deception, privacy loss, and weakened dignity in care settings.	Focused mainly on elder care rather than the full range of HRI applications.
Riek & Howard (2014)	Professional HRI ethics	Proposed professional responsibilities for researchers and practitioners working with people and robots.	Requires translation into organisation-level deployment procedures.
van Wynsberghe (2016)	Care ethics and service-robot design	Linked robot design with values of care, responsibility, and context-sensitive human support.	More implementation tools are needed for non-care sectors.

Aroyo et al. (2018)	Trust and social engineering experiment	Showed that robots may influence disclosure, conformity, and risk-related decisions.	Limited to selected experimental scenarios and user groups.
Boada, Roman & Torras (2021)	Critical literature review	Organised social-assistive robot ethics around wellbeing, care, and justice.	Called for stronger practical frameworks that guide deployment decisions.
Etemad-Sajadi et al. (2022)	Service-robot ethics study	Found that ethical attributes shape willingness to use robots in service delivery.	Needs cross-cultural and high-risk setting validation.
Ostrowski et al. (2022)	Equity, ethics and justice audit	Identified gaps in how HRI scholarship addresses representation, power, and justice.	More deployment-focused evaluation measures are required.

3. OBJECTIVES OF THE STUDY

The primary objective of this research is to analyse the ethical concerns that arise when humans interact with autonomous, social, service, assistive, and collaborative robots, and to develop a practical ethics-by-design framework for responsible robotic deployment. The framework connects technical design decisions with human values, institutional responsibility, stakeholder participation, and ongoing review.

3.1 PRIMARY OBJECTIVE

The primary objective of this study is to design and assess a structured ethical framework that helps developers and organisations identify, prevent, document, and respond to ethical risks throughout the lifecycle of human-robot interaction.

PRIMARY OBJECTIVE POINTS

- To identify ethical risks related to safety, privacy, autonomy, transparency, manipulation, fairness, and accountability in HRI.
- To create an ethics-by-design framework that embeds value-sensitive decision-making in robot development and deployment.
- To improve clarity regarding a robot's identity, capabilities, limitations, data practices, and human oversight arrangements.
- To protect vulnerable users through accessible consent, appropriate safeguards, and context-specific risk assessment.
- To establish review mechanisms for monitoring harms, complaints, bias, security incidents, and unintended social effects.

3.2 SECONDARY OBJECTIVES

The secondary objectives focus on stakeholder participation, consent, data governance, inclusive interaction design, organisational responsibility, and the practical evaluation of ethical readiness before and after deployment.

SECONDARY OBJECTIVE POINTS

- To map the interests and vulnerabilities of users, bystanders, carers, workers, managers, and affected communities.
- To define context-sensitive consent procedures for home, healthcare, education, workplace, and public-space robots.
- To assess how data collection, storage, sharing, and automated inference can affect privacy and informational self-determination.
- To evaluate whether interaction cues create appropriate trust rather than misplaced confidence or emotional pressure.
- To promote inclusive design for users with different languages, abilities, ages, cultures, and levels of digital literacy.
- To define accountability roles for vendors, operators, developers, institutions, and supervisory personnel.
- To develop incident-reporting, audit, and improvement procedures for ongoing ethical governance.
- To distinguish between acceptable assistance and designs that unduly influence, replace, or diminish human agency.
- To provide a scalable evaluation model for research laboratories, public institutions, service providers, and industry teams.

4. RESEARCH METHODOLOGY

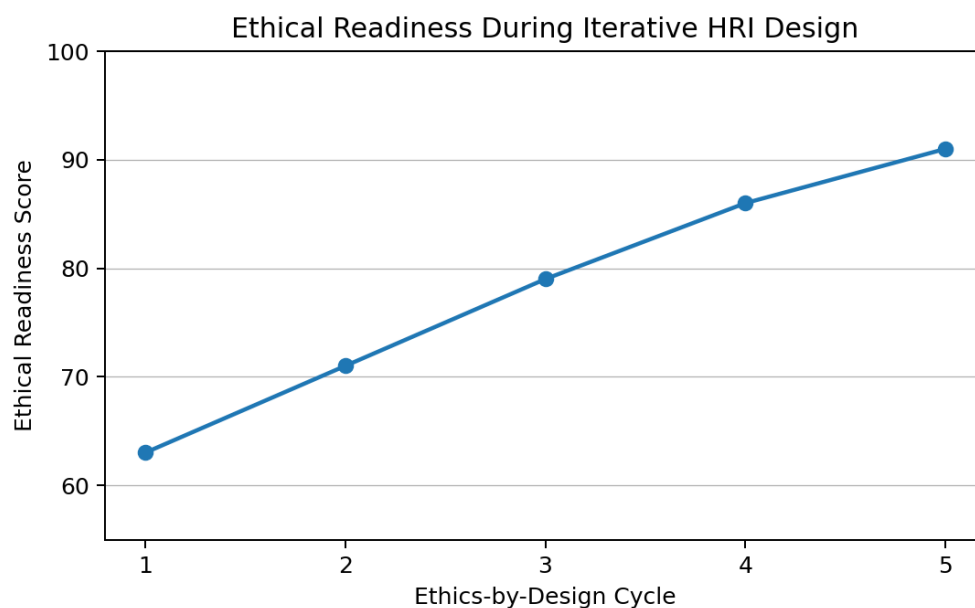
The proposed study follows a structured, mixed-method, ethics-by-design methodology. It begins with a contextual assessment of the intended robot, users, setting, task, level of autonomy, sensor capabilities, data flows, and consequences of failure. The assessment identifies both direct users and people who may be affected without actively using the system, such as family members, patients, co-workers, visitors, and bystanders. This wider perspective is important because robots operating in shared environments can create impacts beyond the person who initially agrees to use them.

Second, the study applies stakeholder mapping and scenario analysis. Stakeholder mapping identifies the needs, rights, responsibilities, and vulnerabilities of each group. Scenario analysis then examines ordinary use, foreseeable misuse, system error, loss of connectivity, data breach, unexpected physical contact, emotional dependence, and disagreement between a robot's recommendation and a human decision. Each scenario is assessed against criteria such as proportionality, reversibility, consent, transparency, safety, fairness, and availability of human intervention.

Third, the framework uses an ethical design review before field testing. The review checks whether the robot clearly identifies itself as a machine, discloses significant data practices, gives users understandable choices, supports opt-out or pause functions, records important system events, and provides pathways for reporting harm. In high-risk settings, the review also examines escalation procedures, supervisory roles, training requirements, accessibility, and the quality of communication between the robot and human professionals.

Finally, the evaluation combines document analysis, interaction observation, participant feedback, usability review, privacy and security assessment, and structured ethics scoring. Evidence may include consent materials, data-flow diagrams, interaction scripts, incident logs, audit records, participant comments, and design revisions. The approach does not assume that a single score can resolve an ethical question. Instead, the scorecard supports transparent discussion and helps teams identify matters requiring redesign, limited deployment, stronger safeguards, or rejection of a proposed use case.

Figure 4.1: Ethical Readiness Improvement Across HRI Design Cycles



The figure illustrates the expected improvement in ethical readiness when teams repeatedly review interaction scenarios, consent pathways, data practices, and

accountability arrangements. The values are illustrative and are included only to demonstrate the evaluation format.

5. DATA ANALYSIS

Data analysis evaluates whether the proposed ethics-by-design framework improves the ability of a robot-development or deployment team to identify and mitigate ethical concerns before a system reaches real-world users. The analysis compares an unstructured approach, in which ethical matters are addressed informally or late in development, with a structured approach that includes stakeholder mapping, risk review, design documentation, consent testing, privacy assessment, and post-deployment monitoring.

For transparency, the numerical values in the following table are illustrative prototype results. They are not a substitute for evidence from a controlled empirical study. Before submission as an empirical article, the values should be replaced with measured data from a defined sample, interaction setting, robot type, participant profile, review period, and validated assessment instrument. The table demonstrates how ethical readiness can be measured without reducing ethical judgement to a purely technical checklist.

5.1 Ethical Readiness Evaluation Results

Evaluation Metric	Unstructured HRI Development	Proposed Ethics-by-Design Framework
Clarity of Robot Identity and Capability Limits	64.1%	89.3%
Informed Consent and Choice Architecture	58.7%	91.2%
Privacy and Data-Minimisation Safeguards	61.4%	88.5%
Physical and Psychological Safety Assessment	72.6%	92.4%
Human Oversight and Intervention Readiness	70.2%	93.1%
Bias, Inclusion, and Accessibility Review	59.8%	86.7%
Accountability and Incident Traceability	62.3%	90.6%
Trust Calibration and Anti-Manipulation Design	57.6%	87.9%
Ethical Review Completion Time	118 min	79 min

Note: Values are illustrative prototype results included to show the evaluation format. Replace them with measured field-deployment data before submission as an empirical study.

6. DISCUSSION

The proposed framework addresses a recurring weakness in robotics projects: ethics is often discussed in general terms but not translated into decisions that designers, managers, and operators can apply during development and deployment. By linking ethical values with concrete questions about data, consent, interaction scripts, autonomy, escalation, and monitoring, the model makes ethical reflection more actionable. It encourages teams to identify harms before they become embedded in technical architectures or institutional routines.

The illustrative evaluation suggests that structured review can improve the clarity of

consent, data safeguards, transparency, human oversight, and incident traceability. These improvements are especially important when robots operate in personal spaces or provide support to people who may rely on them for information, companionship, or assistance. The framework also recognises that trust should be calibrated rather than maximised. A robot should not be designed to create confidence beyond its actual reliability, competence, or authority.

The ethical acceptability of HRI depends heavily on context. A robot that provides wayfinding information in a public building presents different risks from a robot that observes a patient at home, assists an older adult with medication, supports a child in a classroom, or collaborates with a worker near heavy machinery. Context affects the adequacy of consent, the need for supervision,

the seriousness of errors, the sensitivity of data, and the balance between efficiency and human contact. For this reason, universal design rules should be supplemented by sector-specific governance and meaningful participation from affected users.

Ethical HRI also requires attention to organisational incentives. A technically responsible robot can still be used irresponsibly if an institution deploys it to reduce staff without ensuring continuity of care, uses collected data for unrelated purposes, or fails to provide a channel for correction and complaint. Accountability must therefore remain visible: users need to know who is responsible for decisions, where concerns can be raised, and how problems will be investigated. The long-term objective is not only safe machines, but socio-technical systems that respect human dignity, agency, and democratic oversight.

7. CONCLUSION

Ethical Concerns in Human-Robot Interaction presents a structured approach to identifying and responding to the social, psychological, informational, and institutional risks that accompany the use of robots in human

environments. The framework integrates contextual assessment, stakeholder mapping, scenario analysis, informed consent, privacy protection, transparent interaction design, inclusion, human oversight, accountability, and continuous monitoring. This combination can help developers and organisations move from broad ethical principles to practical, documented, and reviewable decisions.

The study emphasises that responsible robotics cannot be achieved through technical reliability alone. The quality of HRI depends on whether people understand the system, can refuse or pause interaction, are protected from unjustified surveillance and manipulation, receive support appropriate to their needs, and have meaningful access to human assistance and redress. Future research should validate the framework through field studies in healthcare, education, public services, manufacturing, and domestic environments; compare ethical concerns across cultures and user groups; evaluate long-term relationships with social robots; and develop participatory methods that give affected communities a stronger role in defining acceptable robotic practices.

8. REFERENCES

1. Asaro, P. M. (2006). What should we want from a robot ethic? *International Review of Information Ethics*, 6, 9-16.
2. Sharkey, A., & Sharkey, N. (2012). Granny and the robots: Ethical issues in robot care for the elderly. *Ethics and Information Technology*, 14, 27-40. <https://doi.org/10.1007/s10676-010-9234-6>
3. Sparrow, R., & Sparrow, L. (2006). In the hands of machines? The future of aged care. *Minds and Machines*, 16, 141-161. <https://doi.org/10.1007/s11023-006-9030-6>
4. Sullins, J. P. (2012). Robots, love, and sex: The ethics of building a love machine. *IEEE Transactions on Affective Computing*, 3(4), 398-409. <https://doi.org/10.1109/T-AFFC.2012.31>
5. Riek, L. D., & Howard, D. (2014). A code of ethics for the human-robot interaction profession. *Proceedings of We Robot 2014*.
6. Lin, P., Abney, K., & Bekey, G. A. (Eds.). (2012). *Robot Ethics: The Ethical and Social Implications of Robotics*. MIT Press.
7. Wallach, W., & Allen, C. (2008). *Moral Machines: Teaching Robots Right from Wrong*. Oxford University Press.
8. Dautenhahn, K. (2007). Socially intelligent robots: Dimensions of human-robot interaction. *Philosophical Transactions of the Royal Society B*, 362(1480), 679-704. <https://doi.org/10.1098/rstb.2006.2004>
9. 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](https://doi.org/10.1016/S0921-8890(02)00372-X)
10. Kahn, P. H., Jr., Ishiguro, H., Friedman, B., & Kanda, T. (2007). What is a human? Toward psychological benchmarks in the field of human-robot interaction. *Interaction Studies*, 8(3), 363-390.

11. van Wynsberghe, A. (2016). Service robots, care ethics, and design. *Ethics and Information Technology*, 18, 311-321. <https://doi.org/10.1007/s10676-016-9409-8>
12. Coeckelbergh, M. (2010). Health care, capabilities, and AI assistive technologies. *Ethical Theory and Moral Practice*, 13, 181-190. <https://doi.org/10.1007/s10677-009-9186-2>
13. Borenstein, J., & Arkin, R. C. (2016). Robotic nudges: The ethics of engineering a more socially just human being. *Science and Engineering Ethics*, 22, 31-50. <https://doi.org/10.1007/s11948-015-9685-1>
14. Aroyo, A. M., Rea, F., Sandini, G., & Sciutti, A. (2018). Trust and social engineering in human-robot interaction: Will a robot make you disclose sensitive information, conform to its recommendations, or gamble? *IEEE Robotics and Automation Letters*, 3(4), 3701-3708. <https://doi.org/10.1109/LRA.2018.2856272>
15. Boada, J. P., Roman Maestre, B., & Torras, C. (2021). The ethical issues of social assistive robotics: A critical literature review. *Technology in Society*, 67, 101726. <https://doi.org/10.1016/j.techsoc.2021.101726>
16. van Maris, A., Zook, N., Caleb-Solly, P., Studley, M., Winfield, A., & Dogramadzi, S. (2021). A new perspective on robot ethics through investigating the human-robot interaction. *Applied Sciences*, 11(21), 10136. <https://doi.org/10.3390/app112110136>
17. Etemad-Sajadi, R., Soussan, A., & Schöpfer, T. (2022). How ethical issues raised by human-robot interaction can impact the intention to use the robot. *International Journal of Social Robotics*, 14, 1103-1115. <https://doi.org/10.1007/s12369-021-00857-8>
18. Ostrowski, A. K., Walker, R., Das, M., Yang, M., Breazeal, C., Park, H. W., & Verma, A. (2022). Ethics, equity, and justice in human-robot interaction: A review and future directions. *2022 31st IEEE International Conference on Robot and Human Interactive Communication (RO-MAN)*, 969-976. <https://doi.org/10.1109/RO-MAN53752.2022.9900805>
19. Rosén, J., Lindblom, J., Billing, E., & Lamb, M. (2021). Ethical challenges in the human-robot interaction field. *Proceedings of the 2021 ACM/IEEE International Conference on Human-Robot Interaction Companion*.
20. Lutz, C., & Tamò-Larrieux, A. (2020). The robot privacy paradox: Understanding how privacy concerns shape intentions to use social robots. *Human-Machine Communication*, 1(1). <https://doi.org/10.30658/hmc.1.6>
21. Pollmann, K., Loh, W., & Fronemann, N. (2023). Entertainment vs. manipulation: Personalized human-robot interaction between user experience and ethical design. *Technological Forecasting and Social Change*, 189, 122376. <https://doi.org/10.1016/j.techfore.2023.122376>
22. European Commission High-Level Expert Group on Artificial Intelligence. (2019). *Ethics Guidelines for Trustworthy AI*. European Commission.
23. IEEE. (2019). *Ethically Aligned Design: A Vision for Prioritizing Human Well-being with Autonomous and Intelligent Systems*. IEEE Standards Association.
24. UNESCO. (2021). *Recommendation on the Ethics of Artificial Intelligence*. UNESCO.
25. National Institute of Standards and Technology. (2023). *Artificial Intelligence Risk Management Framework (AI RMF 1.0)*. NIST.
26. Organisation for Economic Co-operation and Development. (2019). *Recommendation of the Council on Artificial Intelligence*. OECD.
27. International Organization for Standardization. (2014). *ISO 13482: Robots and Robotic Devices - Safety Requirements for Personal Care Robots*. ISO.
28. European Union. (2016). *Regulation (EU) 2016/679: General Data Protection Regulation*. Official Journal of the European Union.
29. National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research. (1979). *The Belmont Report*. U.S. Department of Health, Education, and Welfare.
30. World Health Organization. (2021). *Ethics and Governance of Artificial Intelligence for Health: WHO Guidance*. World Health Organization.