Journal of Interdisciplinary
SOCIAL, DIGITAL CREATIVE &ENGINEERING STUDIES

SUBJECT AREA(S): AI ETHICS | HUMAN-COMPUTER INTERACTION

DOI: 10.37871

RESEARCH ARTICLE

Integrating AI Ethics and Human–Computer Interaction: Toward Responsible and Human-Centered Intelligent Systems

Boris Kriuk*

Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong

Abstract

Artificial intelligence systems increasingly influence human decision-making across healthcare, justice, and digital platforms. While AI ethics frameworks articulate normative principles, their effective implementation depends on how humans interact with intelligent systems. Human–Computer Interaction (HCI) provides the design mechanisms through which ethical principles are operationalized. This article synthesizes contemporary research on AI ethics and HCI to propose an integrated, human-centered framework for responsible AI. The analysis highlights ethical risks, interaction-based mitigation strategies, and applied governance implications (Figure 1).

Introduction

AI systems increasingly mediate critical decisions, raising concerns related to transparency, fairness, accountability, and human agency [1–5]. Although numerous ethical guidelines have been proposed [2,3,38–40], practical implementation remains inconsistent.

HCI research demonstrates that ethical outcomes are shaped not only by algorithms but also by interface design, feedback mechanisms, and user control [7–10]. The integration of AI ethics with HCI is therefore essential for translating ethical principles into practice **(Figure 2)**.

Conceptual Foundations

AI Ethics: Principles and Operational Challenges

AI ethics frameworks commonly emphasize transparency,

***Corresponding author(s)**

Boris Kriuk, Department of Electronic and Computer Engineering, Hong Kong University of Science and Technology, Hong Kong

DOI: 10.37871/jisdces1115

Submitted: 20 December 2025

Accepted: 11 January 2026

Published: 15 January 2026

Keywords

- AI Ethics
- Human–Computer Interaction Responsible AI
- Explainable AI
- Trust in AI

VOLUME: 1 ISSUE: 1 - JANUARY, 2026

Scan Me

How to cite this article: Kriuk B. Integrating AI Ethics and Human–Computer Interaction: Toward Responsible and Human-Centered Intelligent Systems. J Interdiscip Soc Digit Creat Eng Stud. 2026 Jan 15; 1(1): 7. Doi: 10.37871/jisdces1115

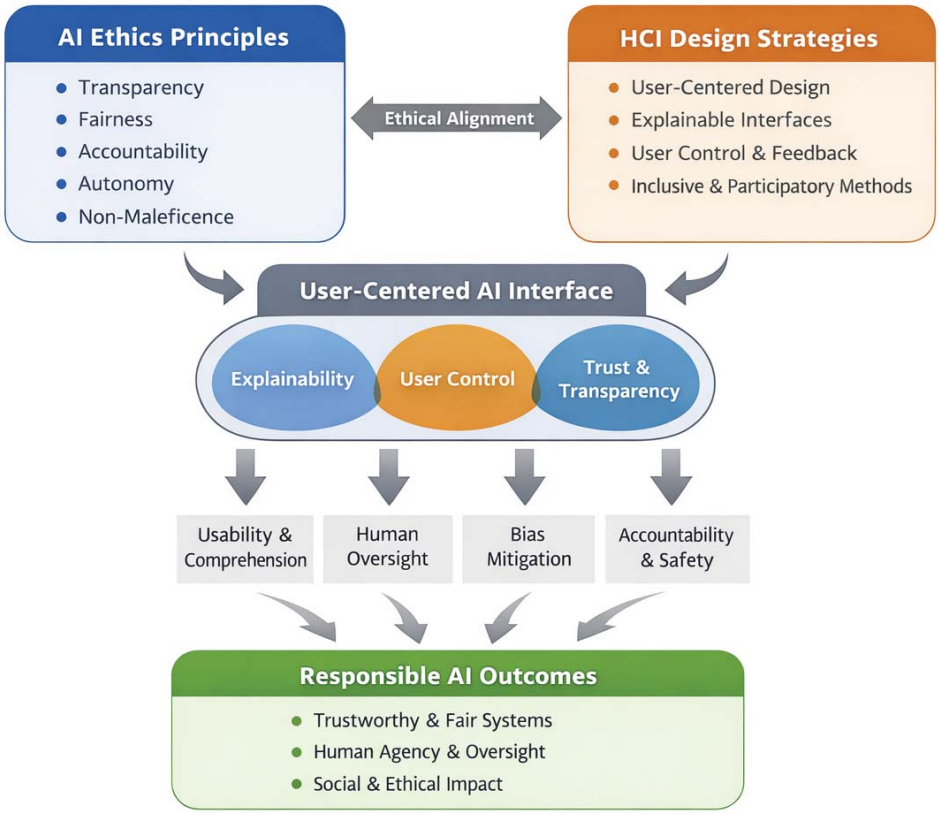

Figure 1 **Conceptual framework integrating AI ethics and human–computer interaction.**
This figure illustrates the relationship between core AI ethical principles, human–computer interaction design strategies, and responsible AI outcomes, emphasizing the role of user-centered interfaces in mediating ethical implementation.

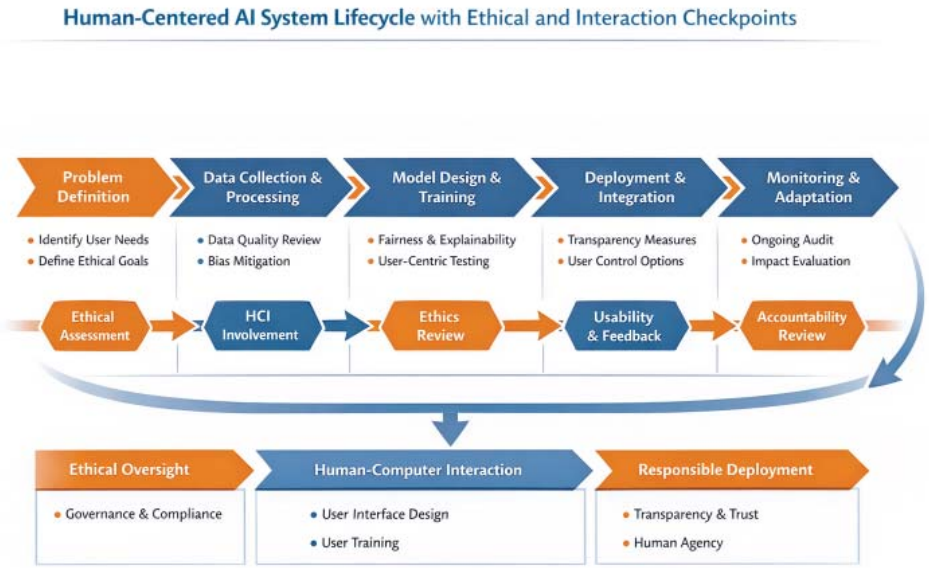

Figure 2 **Human-centered AI system lifecycle with ethical and interaction checkpoints.**
The figure presents key stages of the AI system lifecycle, highlighting ethical review points and HCI-based interventions that support transparency, accountability, and human agency throughout development and deployment.

fairness, accountability, and non-maleficence [1,3,6]. However, these principles often lack concrete operational pathways. Studies have shown that ethical failures frequently arise from poor system interaction rather than algorithmic intent [13,20].

The relationship between AI ethics principles and interaction requirements is summarized in **Table 1**, illustrating how ethical goals depend on specific HCI design features.

Human–Computer Interaction and Intelligent Systems

HCI theory provides models for understanding how users interpret and rely on AI systems [15–19]. Poor interaction design can amplify automation bias and over-reliance [17,18].

The mapping of ethical risks to HCI-based mitigation strategies is presented in **Table 2**, highlighting design interventions that support responsible use.

Integrating AI Ethics and HCI

Explainability and User Understanding

Explainable AI is widely recognized as an ethical requirement [10–12]. However, explanations must be cognitively aligned with user expertise and context. Layered explanations and interactive visualizations improve user comprehension and calibrated trust **(Figure 3)**.

Human Agency, Control, and Responsibility

Maintaining meaningful human control is central to ethical AI deployment [21–24]. Interfaces that allow users to challenge or override AI outputs reduce moral disengagement and reinforce accountability.

The interaction pathways influencing trust and reliance are illustrated in **(Figure 4)**.

Applied Domains of AI Ethics and HCI

The integration of AI ethics and HCI is particularly critical in high-stakes domains. In healthcare, explainable interfaces support shared decision-making [21–24]. In criminal justice, transparency and contestability reduce ethical and legal risks [25–28].

Key application domains and their ethical–interaction contributions are summarized in **Table 3**.

A cross-domain perspective on AI ethics and HCI integration is shown in **(Figure 5)**.

Table 1. Mapping AI Ethics Principles to Human–Computer Interaction Design Requirements.

AI Ethics Principle	HCI Design Focus	Design Implication
Transparency	Explainable interfaces	Clear, user-adapted explanations of AI outputs
Fairness	Inclusive interaction design	Bias-aware UI and diverse user testing
Accountability	Feedback and audit trails	Traceable decisions and user reporting mechanisms
Autonomy	User control mechanisms	Ability to question, override, or adjust AI decisions
Non-maleficence	Risk-aware interaction	Warnings, uncertainty indicators, and safeguards

Abbreviations: AI – Artificial Intelligence; HCI – Human–Computer Interaction

Table 2. Ethical Risks in AI Systems and HCI-Based Mitigation Strategies.

Ethical Risk	AI Context	HCI Mitigation Strategy
Automation bias	Decision-support systems	Confidence visualization and alternative options
Opacity	Black-box models	Layered and role-specific explanations
Over-reliance	High-stakes domains	Human-in-the-loop interaction design
Cognitive overload	Complex analytics dashboards	Progressive disclosure and adaptive interfaces
Loss of agency	Fully automated workflows	Interactive checkpoints and user confirmation

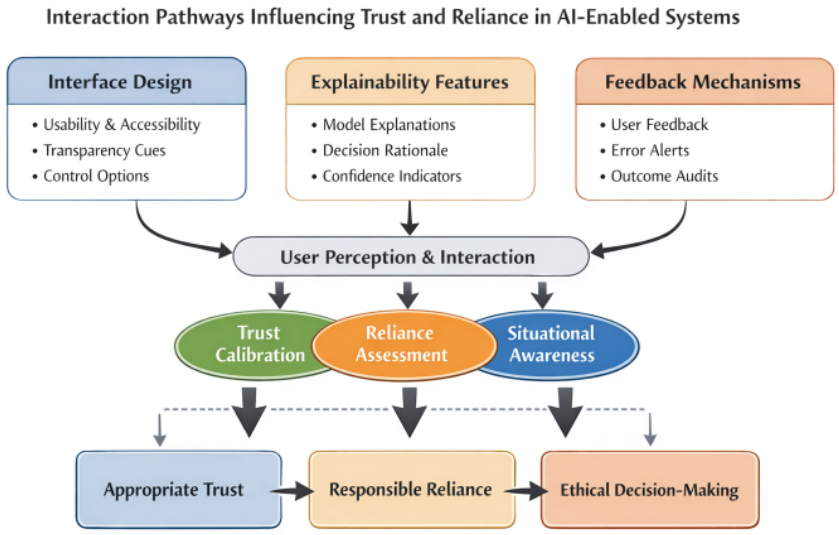

Figure 3 **Interaction pathways influencing trust and reliance in AI-enabled systems.**
This figure depicts how interface design, explainability features, and feedback mechanisms shape user trust, appropriate reliance, and ethical decision-making when interacting with intelligent systems.

Figure 4 **Mapping ethical risks to HCI mitigation strategies in intelligent systems.**
The figure summarizes common ethical risks associated with AI systems and corresponding HCI-based design strategies that mitigate bias, opacity, automation bias, and loss of human control.

Table 3. Application Domains of AI Ethics and HCI Integration.

Domain	AI Application	HCI–Ethics Contribution
Healthcare	Clinical decision support	Shared decision-making and explainability
Criminal justice	Risk assessment tools	Transparency, contestability, and oversight
Digital platforms	Recommendation systems	User trust, fairness, and moderation clarity
Public services	Automated eligibility systems	Accountability and procedural justice
Education	Adaptive learning systems	Learner autonomy and ethical personalization

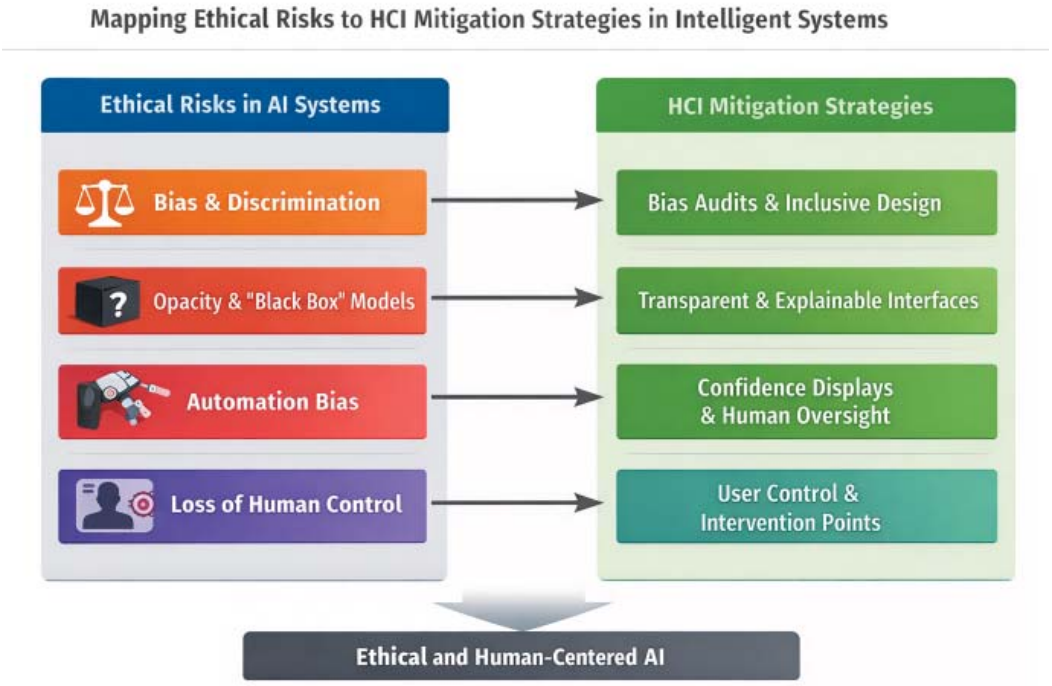

Figure 5 Application domains of AI ethics and HCI integration.
This figure illustrates key application areas—including healthcare, criminal justice, digital platforms, public services, and education—where the integration of AI ethics and human–computer interaction contributes to responsible and socially aligned AI use.

Design and Governance Implications

Ethics-by-design approaches emphasize embedding ethical considerations throughout the AI lifecycle [14,30–32]. HCI methods such as participatory design and usability testing play a central role in identifying ethical risks early and iteratively refining system behavior.

Institutional governance, professional training, and regulatory alignment are necessary complements to interaction design [33,38–40].

Challenges and Research Gaps

Despite progress, gaps remain in empirical validation of ethical interface designs, cross-cultural evaluation, and long-term impact assessment [26,27,34–37]. Future research should prioritize interdisciplinary collaboration and longitudinal studies.

Conclusion

This study demonstrates that AI ethics cannot be effectively implemented without human–computer interaction. Ethical principles gain practical meaning only when translated into interaction design, user agency, and governance structures. Integrating AI ethics and HCI enables the development of responsible, trustworthy, and human-centered intelligent systems.

Conflict of Interest

The author declares no conflict of interest.

Acknowledgement

The author acknowledges the contributions of scholars in AI ethics, bioethics, and human–computer interaction whose work informed this synthesis.

References

1. Floridi L, Cowls J. A unified framework of five principles for AI in society. Harv Data Sci Rev. 2019;1(1):1–15.
2. Jobin A, Ienca M, Vayena E. The global landscape of AI ethics guidelines. Nat Mach Intell. 2019;1(9):389–399.
3. Mittelstadt BD, Allo P, Taddeo M, Wachter S, Floridi L. The ethics of algorithms: Mapping the debate. Big Data Soc. 2016;3(2):1–21.
4. Whittlestone J, Nyrup R, Alexandrova A, Dihal K, Cave S. Ethical and societal implications of algorithms, data, and AI. AI Ethics. 2019;1(1):1–16.
5. McGregor S. Preventing repeated real world AI failures by cataloging incidents. AI Ethics. 2021;1(1):1–6.
6. Borenstein J, Howard A. Emerging challenges in AI ethics. AI Ethics. 2021;1(1):1–5.
7. Amershi S, et al. Guidelines for human-AI interaction. Proc CHI Conf Hum Factors Comput Syst. 2019;1–13.
8. Shneiderman B. Human-centered artificial intelligence: Reliable, safe and trustworthy. Int J Hum Comput Interact. 2020;36(6):495–504.
9. Norman DA. The design of everyday things. New York: Basic Books; 2013.\
10.];lkb Miller T. Explanation in artificial intelligence: Insights from the social sciences. Artif Intell. 2019;267:1–38.
11. Doshi-Velez F, Kim B. Towards a rigorous science of interpretable machine learning. arXiv. 2017;1702.08608.
12. Guidotti R, et al. A survey of methods for explaining black box models. ACM Comput Surv. 2018;51(5):1–42.
13. Raji ID, et al. Closing the AI accountability gap. Proc Conf Fairness Account Transpar. 2020;33–44.
14. Friedman B, Hendry DG. Value sensitive design: Shaping technology with moral imagination. Cambridge (MA): MIT Press; 2019.
15. Rogers Y. HCI theory: Classical, modern, and contemporary. Synth Lect Hum-Centered Inform. 2012;5(2):1–129.
16. Hollnagel E, Woods DD. Joint cognitive systems: Foundations of cognitive systems engineering. Boca Raton: CRC Press; 2005.
17. Parasuraman R, Riley V. Humans and automation: Use, misuse, disuse, abuse. Hum Factors. 1997;39(2):230–253.
18. Lee JD, See KA. Trust in automation: Designing for appropriate reliance. Hum Factors. 2004;46(1):50–80.
19. Endsley MR. Situation awareness in dynamic human–machine systems. Hum Factors. 1995;37(1):32–64.
20. Suresh H, Guttag JV. A framework for understanding unintended consequences of machine learning. arXiv. 2019;1901.10002.
21. Vayena E, Blasimme A, Cohen IG. Machine learning in medicine: Addressing ethical challenges. PLoS Med. 2018;15(11):e1002689.
22. London AJ. Artificial intelligence and black-box medical decisions. Hastings Cent Rep. 2019;49(1):15–21.
23. Topol EJ. High-performance medicine: The convergence of human and artificial intelligence. Nat Med. 2019;25(1):44–56.
24. McDougall RJ. Computer knows best? The need for value-flexibility in medical AI. J Med Ethics. 2019;45(3):156–160.
25. McGregor S, Richmond K, Primeau R. AI in the criminal justice system: From ethics to hermeneutic risk. AI Ethics. 2025;5(1):1–15.
26. Selbst AD, et al. Fairness and abstraction in sociotechnical systems. Proc Conf Fairness Account Transpar. 2019;59–68.
27. Green B, Hu L. The myth in the methodology: Towards a recontextualization of fairness in machine learning. Proc Mach Learn Res.

2018;81:1–12.

28. Pasquale F. The black box society. Cambridge (MA): Harvard University Press; 2015.

29. Crawford K. Atlas of AI. New Haven: Yale University Press; 2021.

30. Benjamins R, et al. Towards responsible AI: Governance, tools and methods. AI Ethics. 2023;3(1):1–14.

31. Fjeld J, et al. Principled artificial intelligence: Mapping consensus in ethical frameworks. Berkman Klein Center. 2020;1–29.

32. Floridi L. Translating principles into practices of digital ethics. Philos Technol. 2019;32(2):185–193.

33. Mohammadzadeh N, et al. Identifying influential theories in human–computer interaction. Health Sci Rep. 2025;8(1):eXXXX.

34. Salles A, et al. Healthcare AI ethics: A systematic review. BMC Med Ethics. 2020;21(1):1–15.

35. Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology. MIS Q. 2003;27(3):425–478.

36. Davis FD. Perceived usefulness, perceived ease of use, and user acceptance. MIS Q. 1989;13(3):319–340.

37. ISO. ISO 9241-210: Human-centred design for interactive systems. Geneva: ISO; 2019.

38. High-Level Expert Group on AI. Ethics guidelines for trustworthy AI. Brussels: European Commission; 2019.

39. World Health Organization. Ethics and governance of artificial intelligence for health. Geneva: WHO; 2021.

40. UNESCO. Recommendation on the ethics of artificial intelligence. Paris: UNESCO; 2021.