

Research on User Trust Modeling and Interface Design Paradigm of AI Systems under Ethical Constraints

Woxing Ren, Long Fu, Chunjing Liu

Liaocheng University Liaocheng City, Shandong Province 252000

Abstract: The rapid development of artificial intelligence technology has sparked ethical controversies, and the issue of user trust in AI systems has gradually become a key obstacle to its application and promotion. In the context of ethical constraints, the mechanism for establishing trust urgently needs to be clarified. To solve the problem of fuzzy trust construction path for AI system users, this article summarizes the core principles of AI ethics, analyzes the impact of ethical mechanisms on user trust, constructs a multidimensional structure of trust, and explores the main methods of trust modeling. In terms of interface design, emphasis is placed on enhancing user trust through improving transparency, controllability, and emotional connection, and a design paradigm that integrates ethical perception is proposed. Research provides theoretical support for building trustworthy and efficient human-computer interaction systems, promoting the sustainable development of AI systems on the basis of safeguarding user rights.

Keywords: AI ethics; User trust modeling; Interface Design

Fund Project:

This paper is the phased research result of the 2024 undergraduate teaching reform research project of Shandong Province, “Research and Practice on the Teaching Reform of Design Majors in Colleges and Universities Empowered by AIGC Technology”, project number: M2024175.

Introduction

Artificial intelligence technology is profoundly changing social operations and individual lifestyles, and is widely applied in key fields such as healthcare, finance, and education. With the increasing participation of AI systems, public attention to their transparency, fairness, and security continues to rise. User trust has become a core element affecting the promotion and continuous operation of AI systems, especially in scenarios involving sensitive data and high-risk decisions, where its importance is even more prominent. Lack of trust will directly lead to barriers to technological acceptance, and even trigger ethical and social risks. Therefore, exploring the trust construction mechanism in AI systems and responding to human expectations for technological controllability and value consistency has important theoretical value and practical significance for promoting the healthy and sustainable development of artificial intelligence.

1. Analysis of trust issues in AI systems under ethical constraints

1.1 Overview of Core Principles of AI Ethics

The widespread application of artificial intelligence has raised high concerns about ethical issues. Transparency, fairness, privacy protection, and accountability constitute the core principles of AI ethics. Transparency enhances the interpretability of the system, which helps to strengthen user understanding and trust; Fairness requires algorithms to avoid bias and ensure fair results; Privacy protection emphasizes the standardized use of personal data to alleviate concerns about user information leakage; Clarifying the responsibilities of all parties in the accountability mechanism helps to enhance the reliability of the system. These ethical principles run through the entire process of AI system design and application, forming the foundational framework for user trust. Embedding ethics into AI systems is an important guarantee for sustainable technological development.

1.2 The Impact Mechanism of Ethical Constraints on AI User Trust

Ethical constraints profoundly influence the generation and maintenance of trust by regulating users' cognitive and emotional responses to artificial intelligence systems. In the context of increasingly complex technological systems, users often lack a comprehensive understanding of the internal operational logic of AI, which can lead to a sense of uncertainty and insecurity. The intervention of ethical mechanisms provides users with a reference for understanding system operation and judging behavioral boundaries, which helps to construct psychological safety expectations. AI systems with ethical self-discipline often exhibit higher information transparency, clear responsibility, and fairness, thereby reducing users' anxiety and doubts and enhancing the formation of trust.

Ethical constraints are also embedded in system interactions through institutionalization and design, exerting a deep influence on user trust construction. For example, adding informed consent mechanisms, feedback transparency mechanisms, and error accountability mechanisms during system use not only enhances users' sense of participation and control, but also establishes a perceptible foundation of trust. When users feel that their personal interests and rights are fully respected, and the system has clear value boundaries, it is easier to generate trust and continue trust behavior in interactions.

2. AI system user trust modeling

2.1 Definition and Dimensions of User Trust

User trust refers to an individual's positive expectations and psychological acceptance of the behavior of an artificial intelligence system when faced with uncertainty and risk. This trust is not only based on cognitive judgments of system performance, but also deeply influenced by user emotional reactions, interactive experiences, and socio-cultural backgrounds. In the process of human-computer interaction, trust is a key psychological mechanism that promotes users' sustained use, active participation, and risk tolerance.

User trust has multidimensional attributes, among which the cognitive dimension emphasizes the trust formed by users based on rational analysis, mainly including the evaluation of system reliability, accuracy, and professional ability. The emotional dimension is reflected in the emotional connection and psychological comfort of users during the interaction process, such as the familiarity and resonance with the system language and behavioral style. The behavioral dimension involves the degree to which users rely on system functionality in actual operations, such as whether they are willing to entrust critical tasks to AI processing and whether they are willing to accept system suggestions.

2.2 Key factors affecting user trust

The trust of users in AI systems is influenced by multiple factors such as technological performance, interactive experience, social cognition, and institutional safeguards. The stability, accuracy, and interpretability of the system form the foundation of trust, and frequent errors or lack of transparency will weaken user trust. A good interactive experience, clear information presentation, and moderate user control can help enhance trust. The individual background of users, such as technical literacy, cultural differences, and past experience, can also affect their trust attitude. The reputation, ethical standards, and institutional safeguards of the institutions to which the system belongs provide external support for trust, collectively forming the key foundation of user trust in AI systems.

3. The moderating effect of interface design on user trust

3.1 Basic Theory of User Interface (UI) and User Experience (UX)

The user interface (UI) and user experience (UX), as core components of human-computer interaction, have a profound impact on the establishment and regulation of user trust in AI systems. The interface not only carries the basic functions of information display and functional operation, but also conveys the reliability, transparency, and affinity of the system through visual structure, interaction process, and emotional expression, thereby affecting users' psychological expectations and trust level.

UI design focuses on the visual presentation and element layout of the interface, and its rationality directly affects users' perception of the system's professionalism and rigor. Clear icons, a unified color system, and a concise layout style help reduce users' cognitive load and enhance their sense of control over the system. Interaction consistency and predictability can also enhance users' trust perception of system response at the behavioral level. When interface design can guide users to smoothly complete tasks and receive feedback in a timely manner, the positive user experience it brings will further strengthen trust at the emotional level.

3.2 Design principles and patterns to promote trust

Interface design not only carries information transmission and interaction functions in AI systems, but also plays a key role in trust building. Effective interface design should be based on user psychology and cognitive mechanisms, combined with ethical requirements, to build an interactive experience that is transparent, controllable, and emotionally connected, thereby enhancing user

trust in the system. In terms of design principles, information transparency is the primary condition for promoting trust. The interface should clearly present the system's behavioral logic, data usage methods, and decision-making basis, enabling users to understand the AI's judgment process and reduce their doubts about "black box" operations.

Controllability is another core dimension in building trust. In design, users should be given sufficient operational freedom and choice to perceive control over the system during interaction, rather than passively accepting the results. For example, allowing users to manually adjust recommended content, select algorithm preferences, or provide "recall" and "correction" functions can effectively enhance their subjective control over the system. Personalized feedback mechanism is also an important strategy to enhance trust. Through user behavior data and preference analysis, the interface can generate targeted and adaptive feedback content, making users feel understood and respected in interaction, thus forming an emotional bond of trust.

Ethical perception design should be integrated throughout the interface logic. The system should actively demonstrate its commitment to privacy protection, fairness, and accountability through visualization, allowing users to intuitively perceive the system's value position at the visual and interactive levels. For example, privacy instructions, risk warning labels, data usage statements, and other modules can be embedded in the interface to make ethics a part of the interface rather than additional information. This design can enhance users' psychological security, making trust built not only on technical abilities, but also on value consensus.

4. Conclusion

Against the backdrop of the continuous integration of artificial intelligence into various fields of society, building user trust in AI systems has become a key issue for sustainable technological development. This article focuses on the construction of trust and interface design under ethical constraints, emphasizing the fundamental role of ethical principles in enhancing user trust, and pointing out that design should have characteristics such as information transparency, controllable operation, and emotional connection. User trust not only depends on system performance, but is also deeply influenced by interaction experience and value recognition. The design of future AI systems should further integrate ethical norms and human-machine interaction logic, providing systematic support from the perspective of trust modeling and design paradigms, in order to achieve a balance between technological progress and social acceptance, and promote the evolution of intelligent systems towards a more trustworthy and human-oriented direction.

References:

- [1] Wan Liyong, Du Jing, Xiong Ruoxin Human Computer Co Creation: A New Paradigm for Digital Education Resource Development Based on AIGC [J]. *Research on Modern Distance Education*, 2023, 35(5):12-21. DOI:10.3969/j.issn.1009-5195.2023.05.002.
- [2] Yu Guoming, Teng Wenqiang, Ye Jingyi Process Thinking: Building a Positive Interaction between Technological Development and Technological Ethics - Reflections on the Ecological Governance of Generative AI Communication from the Perspective of Responsible Innovation [J]. *Media Observation*, 2023(12):37-44.
- [3] Wang Pengtao, Xu Runjie Research on the Reconstruction of Academic Publishing Trust Mechanism under AIGC Intervention in Knowledge Production [J]. *Library and Information Knowledge*, 2023, 40(5):87-96. DOI:10.13366/j.dik.2023.05.087.
- [4] Wang Pengtao, Xu Runjie Research on the Reconstruction of Academic Publishing Trust Mechanism under AIGC Intervention in Knowledge Production [J]. *Library and Information Knowledge*, 2023, 40(5):87-96. DOI:10.13366/j.dik.2023.05.087.

About the author:

Woxing Ren (1999-), female, Han, from Huzhou, Zhejiang Province, is a master's student in the School of Fine Arts and Design, Liaocheng University, Class of 2023, with a research focus on art and design

Long Fu (2000 -), male, Han, born in Qingdao, Shandong Province, master's student in the School of Fine Arts and Design, Liaocheng University, Class of 2023, with a research focus on art and design

Chunjing Liu (1979-), female, Han, from Liaocheng, Shandong, holds a PhD and is an associate professor at Liaocheng University. Her research focuses on art and design