Frontiers in Arts, Humanities
& Social Sciences

A Study on Human-Robot Interaction Design and Ethical Issues in Educational Robots

Wang Xueting,Gao Qingwen

Wenzhou University of Technology, Wenzhou, 325000, China

Abstract. Educational robots are increasingly used in homes and classrooms for teaching assistance and emotional companionship, but the convenience they bring conceals underlying ethical risks. This paper investigates how to define the ethical boundaries of human-robot interaction in educational robots across three dimensions: privacy security, human agency, and creativity impact. A systematic analysis is conducted, integrating existing empirical research and theoretical frameworks to structurally examine ethical issues from these three dimensions. The findings reveal that educational robots risk privacy breaches by collecting sensitive student data for commercial purposes; their "perfect interaction" may lead students to avoid real relationships, weakening empathy and social skills; and overreliance on standardized answers can stifle independent thinking, homogenizing creativity. Therefore, ethical boundaries should be based on data minimization, local processing, and transparency; robots should be positioned as "transitional objects" guiding students toward authentic relationships rather than replacing them; and independent thinking must be preserved to prevent cognitive outsourcing. Future education should pursue human-machine collaboration, where ethical boundaries guide technological development with respect for human agency at the core.

Keywords: educational robots； human agency； ethical boundaries； human-robot interaction；autonomy

Received: April 20, 2026
Revised: May 8, 2026
Accepted: May 12, 2026
Published: May 18, 2026

1. Introduction

The history of educational robots dates back to the mid-20th century. With advances in computer technology and the development of artificial intelligence, educational robots have gradually become an integral part of the educational field. Their development has broadly undergone four stages: the initial exploration stage, the interactive enhancement stage, the intelligent expansion stage, and the system integration stage.

During the initial exploration stage, educational robots were primarily used for basic learning assistance and had relatively limited functions. In 1970, Waseda University in Japan created the world's first bipedal robot, WABOT-1, equipped with dialogue, walking, and visual control systems, marking the birth of bipedal robots. In 1986, Honda Motor Co. of Japan began researching bipedal robots and launched ASIMO in 2000. This robot could perform complex tasks in specific environments, signaling the formal entry of bipedal educational robots into the education sector.

During the interactive enhancement phase, driven by technological advancements, bipedal educational robots began to develop stronger interactive capabilities, enabling them to integrate into classroom instruction. In 2008, France's Aldebaran Robotics launched the NAO robot, which not only performed physical movements but also interacted with students through voice and facial expressions, becoming a valuable assistant to teachers. In 2014, Japan's SoftBank Group and Aldebaran Robotics jointly developed the Pepper robot, further enhancing the interactivity and emotional expression capabilities of educational robots.

In the intelligent expansion phase, as technologies such as artificial intelligence, cloud computing, and machine vision matured, the applications of bipedal educational robots became more widespread and intelligent. In 2016, China's UBTech launched bipedal educational robots such as Walk X and Alpha Ebot. These robots can not only autonomously identify learners' progress but also provide personalized instructional support. Classic bipedal educational robots such as ASIMO and NAO were also updated during this phase, further improving their adaptability and interactivity in educational settings. Bipedal educational robots are no longer limited to the classroom but have expanded into home education and self-directed learning environments, offering a more personalized learning experience [1].

With the advancement of digital transformation in education, the application prospects for bipedal educational robots in the education sector are vast. In December 2021, China's Ministry of Industry and Information Technology, along with 14 other departments, jointly issued the "14th Five-Year Plan for the Development of the Robot Industry," explicitly calling for the development of robot application scenarios and the demonstration and promotion of products in the education sector. In February 2023, the Ministry of Education and 16 other departments jointly issued the "Implementation Plan for the 'Robot+' Application Initiative," further emphasizing the new form of "Robot+" campus services [1]. To date, teaching assistant robots have been gradually introduced in many regions across China. In Qinghai Province, this technology has been implemented in over 3,000 classrooms; across Guangxi, more than 200 schools have adopted the "teacher + teaching assistant robot" human-robot co-teaching model. Policy support is driving this educational transformation, helping to reduce the workload for both teachers and students. Against this backdrop, the importance of human-robot interaction is self-evident.

2. Literature Review

This literature review aims to systematically summarize existing research on human-robot interaction design and its ethical implications in the field of educational robotics, with a focus on three core issues: privacy and security, autonomy, and the impact on creativity. Through the synthesis and critical analysis of relevant studies, this review provides a theoretical foundation and research directions for future work.

2.1 Research on Human-Robot Interaction in Educational Robotics

Human-Robot Interaction (HRI) is a core area of research in educational robotics. Existing studies have shown that educational robots can effectively enhance learners' engagement and motivation through various interaction modes, including voice, emotions, and behavior [2]. Tanaka and Matsuzoe [3] found that children exhibited longer attention spans and higher task completion rates when interacting with educational robots. Kanda et al. [4] demonstrated through long-term experiments that the presence of robots in the classroom can promote collaborative learning among students.

Some studies also point out that current educational robots still face numerous limitations in their interaction design. Robots have limited ability to understand nonverbal cues (such as facial expressions and body posture), making it difficult to achieve truly natural interaction [5]. Furthermore, robots' "pseudo-emotional" expressions may lead students to overestimate their "comprehension abilities," thereby fostering unwarranted trust [6]. Yuan [7] notes that human-robot intimate relationships involve inherent contradictions; the misalignment between technological logic and human emotional logic is a deep-seated cause of incomplete interactive experiences.

2.2 Research on Privacy in Educational Robots

As educational robots become increasingly prevalent in classrooms and homes, data privacy issues are becoming increasingly prominent. Drobniak et al. [8] point out that the data collected by educational robots includes not only students' basic information but also highly sensitive content such as learning behaviors, emotional states, and voice recordings. If not properly handled, this data is highly susceptible to misuse or leakage. Research from Georgia Southern University [9] further reveals that the current educational robot industry exhibits a trend toward "natural monopoly," with a handful of large technology companies controlling vast amounts of student data, thereby forming a "surveillance capitalism"-style business model.

Regarding informed consent, Lee et al. [10] found that the vast majority of parents and students are unaware of how the data collected by robots will be used. Even when privacy policies exist, their complex language and verbose length render them virtually ineffective. Consequently, scholars have called for the establishment of an ethical framework specifically tailored to data processing in educational robots, emphasizing the principles of "data minimization" and "local processing" [11]. Chinese scholars Gong Xue et al. [1], while examining the design logic of bipedal educational robots, also pointed out that data security design should be a fundamental requirement in the development of educational robots, rather than an after-the-fact remedy.

2.3 Research on Emotional Dependency in Educational Robots

Emotional dependency is an emerging topic in the ethical study of educational robots. In her seminal work Alone Together[12], Turkle noted that the "perfect interaction" provided by robots may lead users to avoid real human relationships, thereby fostering emotional dependency. A longitudinal study by Nomura et al. [13] further confirmed that children who interacted with companion robots over the long term exhibited significantly lower empathy than the control group. Researchers speculate that this may be because children concentrate a large amount of emotional interaction on non-human entities, lacking opportunities to practice understanding others' emotions.

On the other hand, some studies suggest that educational robots can serve as "transitional objects," helping students establish emotional security and thereby better transition into real-life interpersonal relationships [14,12]. The key lies in the robot's role: whether it serves as a substitute for real relationships or as a bridge. Research by Pizarro et al. [15] indicates that, with appropriate guidance, social robots can assist children on the autism spectrum with emotional regulation training; however, this effect depends on the robot's "instrumental" role rather than its "substitutive" role.

2.4 Research on the Impact of Educational Robots on Creativity

Regarding the impact of educational robots on creativity, existing research shows some divergence. An experiment by Usher et al. [16] found that students using robot-assisted learning performed better in terms of originality and richness of detail; the researchers attributed this to the embodied interactive experience provided by the robots. A longitudinal study by the Sydney Business School at Shanghai University revealed that students who used generative AI over the long term exhibited a significant decline in divergent thinking and innovative performance; the researchers termed this the "self-defeating" effect [17]. Furthermore, the "automation paradox" proposed by Qadir and Mumtaz [18] suggests that overreliance on technology to perform cognitive tasks may lead to the degradation of human cognitive abilities. Mao Weihua and Wu Yaqian [19], in their exploration of ChatGPT's application in higher education and continuing education, also noted that while large language models can provide rapid answers, they may undermine learners' critical thinking skills.

2.5 Research Gaps and the Positioning of This Study

In summary, although existing research has explored human-robot interaction and its ethical issues in educational robots from multiple perspectives, the following shortcomings remain: First, there is a lack of systematic analysis of the intrinsic connections between privacy, emotions, and creativity; second, most studies focus on the technical level, and the construction of ethical boundaries remains unclear; third, empirical research targeting university students is relatively scarce. This study aims to address these gaps by constructing a systematic ethical framework for human-robot interaction in educational robots.

3. Human-Robot Interaction Modes in Educational Robots

3.1 Classification of Human-Robot Interaction Modes

In educational robots, human-robot interaction generally encompasses three primary modes: voice interaction, emotional interaction, and learning assistance.

In the voice interaction mode, users typically wake the robot with specific commands, pose questions, and receive answers from the robot. Most robots currently on the market can utilize large language models to provide students with "heuristic education," such as ChatGPT and Gemini. However, these large language models have several drawbacks: poor reliability, lack of moral judgment and ideological risks, rigid thinking, and limited creativity [19]. The advantage of voice interaction lies in its naturalness and efficiency, but the risk is that students may become overly reliant on the robot's answers, thereby abandoning independent thinking.

In the emotional interaction mode, the robot extracts emotional states from user-inputted text, voice, or visual information, and provides positive feedback or guidance to establish an emotional connection with the user. For example, Joey, a social robot from Hong Kong, China, interacts with those around it using vivid expressions and gestures; QTrobot, a humanoid desktop robot from the UK and Luxembourg, has been validated by research from the University of Cambridge and others; it uses structured dialogue to help students (including those on the autism spectrum) with emotional regulation training [15]. The risk of emotional interaction lies in the potential for students to become overly dependent on the robot, thereby weakening their ability to navigate real-life interpersonal relationships.

In learning assistance mode, robots can observe student behavior to develop personalized learning plans for both students and teachers, and monitor student progress. In 2017, the picture book reading robot Luka won international design awards such as the German iF Design Award and the Red Dot Award upon its launch, with cumulative sales exceeding 1 million units across 12 countries; Since establishing its Education Division in 1980, LEGO has developed multiple educational robots, such as WeDo 2.0 (2016), the SPIKE™ Prime Innovation Set (2019), and the SPIKE™ Innovation Starter Set (2021). The risk of the learning assistance mode lies in the potential for students to "outsource" their thinking, thereby weakening their ability to solve problems independently.

3.2 Case Analysis of Educational Robots

Although ChatGPT itself is not a physical robot, its conversational interaction logic has been widely applied in educational robots. Therefore, this section takes ChatGPT as an example to analyze its ethical risks. As a teaching aid, ChatGPT is an AI system based on a large language model—not a physical robot—but its conversational interaction logic has been widely adopted in educational robots. ChatGPT's strengths include broad information coverage, rapid response times, and the ability to engage in multi-turn conversations. However, it has significant ethical shortcomings: it lacks moral judgment and may generate content unsuitable for minors; its training data contain biases, potentially leading to answers with ideological leanings; and overuse may result in students' thinking becoming rigid, causing them to lose their innovative capacity [19]. Therefore, ChatGPT can only serve as a teaching aid and cannot replace the teacher's role in stimulating student interest or providing in-depth guidance.

NAO and Pepper robots. The NAO robot, launched by the French company Aldebaran Robotics in 2008, features 25 degrees of freedom, multiple sensors, and voice interaction capabilities, and is widely used in universities and research institutions worldwide. The Pepper robot builds upon this foundation by adding emotion recognition capabilities, enabling it to assess a user's emotional state through facial expressions and voice intonation. The application of these two robots in the education sector demonstrates that embodied robots are more effective than pure software systems at capturing students' attention, but they are also more likely to induce emotional dependency [2].

Among China's domestic educational robots, Walk X and Alpha Ebot, launched by UBTech, are typical bipedal educational robots capable of autonomously identifying learners' progress and providing personalized instructional support. The picture book reading robot Luka offers reading companionship services to children through image recognition and voice interaction. The widespread adoption of these products reflects China's technological advancements in the field of educational robotics, but it also exposes issues such as insufficient data privacy protection and a lack of informed consent from users [1].

4. Ethical Issues in Human-Machine Interaction

4.1 Identification of Ethical Issues

In this age of information overload, the leakage of personal information has become a widespread concern. Users cannot be certain whether the collection of private data—such as every conversation, saved video content, and personal information—is stored solely on cloud platforms or whether it is resold or leaked. In 2019, Amazon was exposed for allowing unprotected access by human reviewers to recordings from its Echo Dot for Kids, with some recordings remaining undeleted for extended periods, revealing systemic vulnerabilities in smart voice devices regarding children's data protection [9]. Research from Georgia Southern University indicates that the resource demands of training generative AI have led to the formation of a "natural monopoly" in the industry, with a handful of large technology companies controlling the flow of educational data [9].

Drobniak et al. [8] further reveal that the privacy risks posed by educational robots are unique: in addition to raw sensor data, the "derived inferences" they generate (such as learning ability assessments and emotional state labels) may be obtained by third parties through "inference leakage," while existing legal frameworks (such as the GDPR and CCPA) can only provide limited protection in such human-robot interaction scenarios. A reflective diary study of Chinese university students found that learners are generally concerned about data breaches, "unethical data utilization" based on big data analysis, and algorithmic biases that may exacerbate educational inequality.

As educational robots' emotional interaction capabilities improve, the phenomenon of students developing human agency on robots is drawing increasing attention. A longitudinal study by Nomura et al. [13] found that children who interacted with companion robots over the long term scored significantly lower on empathy scales than the control group. The researchers speculated that this is related to children focusing their emotional interactions excessively on non-human entities. Attachment theory suggests that individuals tend to form attachments with objects that provide a sense of security and consistent responses. Educational robots possess precisely these characteristics: they are emotionally stable, never tire, respond predictably, and can continuously provide positive feedback.

In terms of creativity, a six-month follow-up experiment conducted by the Sydney Business School at Shanghai University found that 63% of students using generative AI felt their ideas were inferior to those of others; without AI assistance, these students' innovative performance declined by an average of 7%, and their divergent thinking also weakened [17]. Furthermore, a study in Shanghai, China, indicated that after prolonged AI use, the similarity of product ideas increased by 11%, suggesting that AI may lead to homogenization of thinking. The "automation paradox" proposed by Qadir and Mumtaz [18] suggests that tools intended to save us effort can, with prolonged use, cause us to lose our ability to think critically. If students assume that the answers provided by robots are correct, they will no longer consider "whether there are other ways," and creativity will naturally be lost.

4.2 Summary of Ethical Issues

By analyzing the issues discussed above, the ethical concerns in human-robot interaction within educational robotics can be summarized into three core dimensions:

The core issue in the privacy risk dimension lies in the lack of transparency and user control over the collection, storage, use, and transfer of data. Sensitive information such as students' facial features, voices, learning behaviors, and emotional states is controlled by a handful of large technology companies, forcing users to "consent to surveillance" as the price for accessing educational services. The existing legal framework fails to provide adequate protection, and corporate self-regulatory mechanisms have yet to be established.

The core issue in the emotional dependency dimension lies in the fact that the "perfect interaction" provided by robots may cause students to avoid real interpersonal relationships, thereby weakening their empathy and social skills. The value of robots as "transitional objects" is distorted into that of "substitute objects," leading students to immerse themselves in virtual relationships while avoiding the practice of dealing with uncertainty in real-world interactions.

The core issue in the creativity dimension lies in the fact that overreliance on robots' standardized answers may lead to the "outsourcing" of students' thinking, stifling independent thought and innovative capabilities. The homogenization of thinking, the "self-deprecation" effect, and the "automation paradox" collectively pose a threat to creativity.

These three dimensions are not isolated from one another but are interrelated: privacy breaches undermine students' autonomy, emotional dependency weakens their social skills, and the suppression of creativity diminishes their agency. Together, they point to a fundamental question: how can human agency be respected and defended in interactions with machines?

5. Research Findings

Based on the above analysis, this study proposes three ethical boundaries for human-robot interaction in educational robots and provides corresponding design principles and usage recommendations.

5.1 Privacy Protection Must Be Grounded in Respect for Student Agency

Data collection and processing by educational robots must be fundamentally grounded in respect for student agency. Student data must not be commodified as capital, and students must not be forced to relinquish their privacy rights as a result of using educational services. Specifically, the following principles should be established:

The Principle of Data Minimization: Only the minimum data necessary to fulfill educational functions should be collected; private information unrelated to teaching (such as home addresses or non-essential biometric data) should not be collected. The Principle of Local Processing Priority: Data processing that can be completed on local devices should not be uploaded to the cloud, thereby reducing the data's exposure. The Principle of Process Transparency: Users must be clearly informed of the scope of data collection, the purpose of use, the storage period, and the data flow path, ensuring genuine informed consent rather than a mere formal click-through confirmation. The Principle of User Control: Users should have the right to access, correct, and delete their own data at any time, as well as the right to refuse the collection of non-essential data.

Enterprises should establish enforceable data governance mechanisms, clearly define the boundaries of responsibilities between schools and suppliers, implement retention period controls, and establish age-appropriate informed consent procedures. Educational robots should not treat privacy as a "ticket" for use.

5.2 Clarifying the Role of Educational Robots as "Transitional Objects"

Psychologist Winnicott [14] introduced the concept of "transitional objects," referring to intermediary items (such as stuffed animals) that help individuals transition from dependence to independence. Turkle [12] applied this concept to the technological realm, noting that the value of robots and AI lies in helping individuals build a sense of security, thereby giving them the courage to engage with the real world, rather than serving as the endpoint of real relationships.

Accordingly, the design and use of educational robots should adhere to the following principles: A tool-oriented approach: Robots are tools for teaching and emotional practice, not substitutes for emotional relationships. Encouraging real-world interaction: The interactive design of robots should guide students toward genuine interpersonal communication. For example, students should be encouraged to discuss topics raised by the robot with classmates or parents, rather than confiding all their emotions to the robot. Limiting usage duration and contexts: Avoid allowing robots to become students' sole emotional outlet. Reasonable usage boundaries should be established, particularly in after-school and home settings.

For college students, the university itself is a vital setting for learning social skills and building relationships. If students pour out all their emotions to AI instead of communicating with roommates, friends, or family, then AI is not helping them but isolating them. Therefore, students should constantly remind themselves that AI can be used to practice social expression but should not be used to replace real-life communication.

5.3 Preserve Space for Independent Thinking and Avoid "Outsourcing" Thought

In the learning process, educational robots should serve as facilitators and assistants, not as the ultimate providers of answers. Specific principles include being heuristic rather than didactic. When answering questions, robots should use Socratic questioning to guide students to draw their own conclusions, rather than directly providing standard answers. Preserving Core Creative Processes: Robots can assist with organizing materials, checking for grammatical errors, and providing background information, but core creative tasks—such as conceptualization, problem definition, and solution selection—must be completed independently by students. Fostering Critical Thinking: Students should be encouraged to question the robot's responses and compare information from different sources, rather than blindly trusting the robot.

Just as navigation tools have gradually caused us to lose our ability to find our way, overreliance on AI can also cause us to lose our ability to think independently. Faced with the overwhelming flood of information, only by developing our own thinking skills can we avoid being swept along by the tide.

5.4 The Intrinsic Connection Among the Three Ethical Boundaries

The three dimensions of privacy, emotional well-being, and creativity may appear distinct, but they all ultimately point to respect for human agency. Privacy protection safeguards students' autonomy—their right to decide who uses their information and within what scope; emotional boundaries uphold students' social well-being, enabling them to form and maintain authentic interpersonal relationships; and creativity safeguards students' agency, ensuring they possess both the willingness and ability to engage in independent thinking and creation. Together, these three ethical dimensions form the pillars of the ethical framework for educational robots.

6. Conclusion

Defining the ethical boundaries of human-robot interaction in educational settings essentially answers a fundamental question: what kind of relationship do we want the next generation to have with machines? If the answer is "human-robot collaboration with a human-centered approach," we must not compromise on our bottom line. As Turkle [12] noted, the value of robots lies in their role as "transitional objects" that help individuals build a sense of security, rather than becoming the endpoint of genuine relationships. This should be the fundamental consensus among all educational robot designers, users, and policymakers.

The future of education is not a confrontation between humans and machines, but a collaboration between them. In this collaboration, ethical boundaries are not restrictions on technology, but rather guidelines for it. They tell us what kind of technological development is truly "for the greater good" and what kind of applications can truly serve human growth.

Ethical issues surrounding educational robots will not automatically disappear with technological progress; on the contrary, the more powerful the technology becomes, the more prominent the ethical risks will be. Therefore, as technology advances, we must continue to question, reflect, and adjust in a timely manner. Only in this way can we ensure that educational robots truly become a force for promoting student growth, rather than a hidden threat that erodes their agency.

References

[1] Gong, X., Liu, B. X., Zhang, R. F., & Xie, Z. Q. (2023). The historical development and design logic of bipedal educational robots. Academy, 16(5), 45-52.

[2] Belpaeme, T., Kennedy, J., Ramachandran, A., Scassellati, B., & Tanaka, F. (2018). Social robots for education: A review. Science Robotics, 3(21), eaat5954. https://doi.org/10.1126/scirobotics.aat5954

[3] Tanaka, F., & Matsuzoe, S. (2012). Children teach a care-receiving robot to promote their learning. IEEE Transactions on Affective Computing, 3(4), 468-479.

[4] Kanda, T., Hirano, T., Eaton, D., & Ishiguro, H. (2004). Interactive robots as social partners and peer tutors for children: A field trial. Human-Computer Interaction, 19(1-2), 61-84.

[5] Saerbeck, M., Schut, T., Bartneck, C., & Janse, M. D. (2010). Expressive robots in education: Varying the degree of social supportive behavior of a robotic tutor. In Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (pp. 1613-1622). ACM. https://doi.org/10.1145/1753326.1753567

[6] Salem, M., Lakatos, G., Amirabdollahian, F., & Dautenhahn, K. (2015). Towards safe and trustworthy social robots: Ethical challenges and practical guidelines. International Journal of Social Robotics, 7(3), 331-345.

[7] Yuan, G. F. (2025). Freedom and security: The certainty and inner conflict of human-robot intimate relationships. Media Observer, 41(2), 88-95.

[8] Drobniak, A., Terzic, I., Kirin, B., & Boticki, I. (2023). Privacy in the age of robotics: Protecting personal data in classrooms. ICCE Proceedings.

[9] Georgia Southern University. (2024). Cyberethics, data surveillance, and the classroom: Addressing corporate control of generative AI to protect student learning environments [Poster presentation]. SoTL Commons Conference, Savannah, GA, United States.

[10] Lee, J., Kim, S., & Lee, H. (2020). Privacy concerns of parents and children in smart classroom environments. Journal of Educational Technology & Society, 23(3), 45-57.

[11] Sharkey, N., & Sharkey, A. (2012). The crying shame of robot nannies: An ethical appraisal. Interaction Studies, 13(1), 9-29.

[12] Turkle, S. (2011). Alone together: Why we expect more from technology and less from each other. Basic Books.

[13] Nomura, T., Suzuki, T., Sato, M., Kanda, T., Yamada, S., & Ishiguro, H. (2019). Long-term effects of interaction with a social robot on children's empathy development. International Journal of Social Robotics, 11(4), 613-625.

[14] Winnicott, D. W. (1971). Playing and reality. Tavistock Publications.

[15] Pizarro, C., García, J., López, M., Fernández, A., & Ruiz, R. (2025). Effectiveness of robot-mediated learning in fostering children's social and cognitive development. Applied Sciences, 15(7), 3567.

[16] Usher, M., Reznik, N., Bronshtein, G., Ben-Abu, R., & Hershkovitz, A. (2025). Bridging virtual and physical: Exploring students' computational thinking and creativity in robot-guided vs. simulation-based learning. Journal of Learning Analytics, 12(1), 45-63.

[17] Li, P., Mou, L., & Li, Y. H. (2024). Research on the path of ChatGPT empowering the construction of digital resources in vocational education. Innovation Education, 12(4), 78-84.

[18] Qadir, J., & Mumtaz, S. (2020). Automation paradox: The unintended consequences of AI in education. IEEE Transactions on Technology and Society, 1(4), 198-206.

[19] Mao, W. H., & Wu, Y. Q. (2024). Application and risk challenges of ChatGPT in higher continuing education. Computer Products & Circulation, 22(7), 112-117.