International Journal of Scientific Research and Management (IJSRM)
||Volume||14||Issue||02||Pages||4574-4580||2026||
Website: https://ijsrm.net ISSN (e): 2321-3418
DOI: 10.18535/ijsrm/v14i02.el02

Artificial Intelligence Ethics: From Human-Centric Philosophy to Humanistic Value Orientation in Education

Nguyen Minh Giam[1]

[1]Faculty of Education, Thu Dau Mot University, Ho Chi Minh City, Vietnam

Abstract:
The robust development of artificial intelligence (AI) today is not merely a technological advancement, but also poses profound challenges to fundamental human values. This paper approaches AI ethics not from a purely technical perspective, but based on Human-centric philosophy, aiming to establish core value orientations for education in the digital era.

Through a systematic analysis of theoretical perspectives and global governance frameworks, the study clarifies that AI ethics is essentially a mechanism to safeguard human autonomy and dignity. Consequently, it argues that AI must be positioned as a tool for support and empowerment, and must absolutely not replace the decision-making role or diminish the independent thinking capacity of learners. Furthermore, the article emphasizes the requirement to build a "foundation of trust" in society through transparent governance mechanisms and clear accountability.

Based on these theoretical foundations, the paper proposes a strategic shift in education: from the goal of training tool usage skills to education oriented towards humanistic values. The ultimate goal is to equip learners with the ethical cognitive capacity to master technology, ensuring that AI always serves the sustainable development of humanity rather than dominating and leading it.

Keywords: AI Ethics, Human-centric Philosophy, Human Autonomy, Humanistic Values in Education, Social Responsibility

1. Introduction

The increasingly deep integration of artificial intelligence (AI) into economic, cultural, and social pillars is ushering humanity into a new era where the boundaries between human and machine capabilities are becoming blurred. Particularly in education, AI is no longer just a technical support tool; it is directly impacting how humans think and solve problems. This shift has elevated ethical issues related to AI from theoretical debates to major concerns in development governance. Therefore, the most critical question today is not what AI can do, but what AI should do to ensure sustainable and humanistic development.

While the potential of technology is undeniable, the uncontrolled application of AI is creating urgent challenges. Floridi et al. (2018) warn of the risk that machines may usurp the coordination of life through imposing suggestions, thereby diminishing human autonomy and independent thinking capacity. Similarly, Formosa et al. (2025) emphasize the necessity to clearly distinguish between autonomous agents (humans) and artificial systems to avoid unintentionally ceding ethical decision-making power to inanimate algorithms.

In this context, recent studies by Usmani et al. (2023) and Lepri et al. (2021) consistently propose a new approach: "Human-centric AI". From this perspective, ethics is not a barrier to innovation but a philosophical foundation to position AI as an empowering tool that supports humans in expanding knowledge without losing their identity. Notably, UNESCO's Recommendation (2021) has established AI ethics as a global normative framework, affirming that technological development must always go hand-in-hand with the protection of fundamental freedoms and aim for the common good of society.

However, practice shows a significant gap in translating these macro philosophies into specific value orientations within the educational environment. While many studies focus on technical solutions or risk

management, analyses of philosophical foundations aimed at forming ethical framework orientations for learners remain limited.

Addressing this reality, this paper aims to deeply analyze the philosophical bases of AI ethics, thereby establishing core humanistic value orientations for education. The article focuses on clarifying: (1) the philosophical nature of AI ethics from a human-centric perspective; (2) governance principles aimed at a sustainable society; and (3) implications for directing education to truly become a place that creates individuals who master technology, rather than being led by it.

2. Philosophical basis: autonomy and human-centric philosophy

To orient values for education in the digital age, it is first necessary to build a solid philosophical foundation regarding the relationship between humans and artificial intelligence. Current research indicates that AI ethics is not merely a set of technical rules to fix algorithmic errors, but more profoundly, a philosophical reflection on the human position in the face of the rise of machines.

2.1. The Nature of AI Ethics

An Approach from Dignity and Autonomy Unlike technical approaches that focus on algorithmic adjustment, the philosophical basis of AI ethics is often grounded in the "dignitarian approach." According to Hanna & Kazim (2021), AI ethics can be understood as a value reference system to establish principles coordinating behavior, ensuring that technological development aligns with human ethical standards. Gunkel (2024) also generalizes AI ethics as a human-centered field of study, focusing on examining ethical consequences for society while clearly distinguishing between system design and the capacity for actual moral responsibility.

In current philosophical debates, the concept of "autonomy" plays a pivotal role. Formosa et al. (2025) offer a notable analysis by distinguishing types of agents, including: basic agents, autonomous agents, and moral agents. The authors argue that although AI can simulate intelligent behaviors, they are not conscious "moral agents" like humans. Therefore, the core limit of AI ethics lies in defining the boundary of responsibility: AI may have autonomy in technical operation, but self-determination in moral terms must belong to humans. Floridi et al. (2018) further reinforce this view by asserting that the core of AI ethics is the protection of individual self-determination. He warns of the danger of machines usurping the coordination of life through imposing suggestions. Thus, from a philosophical perspective, AI ethics can be viewed as a mechanism to prevent the erosion of independent thinking capacity and the growing dependence on pre-programmed pathways shaped by algorithms.

From the above analyses, it is evident that the most important philosophical boundary to be established is not the limit of computational power, but the limit of ethical responsibility. No matter how complex the decision-making behaviors AI can simulate, they remain unconscious mechanical agents. Attributing human moral attributes to machines is a dangerous misconception, as this can lead to humans evading or denying responsibility when systemic errors occur.

2.2. Human-Centric Philosophy

Building on the foundation of autonomy, the dominant philosophy guiding modern AI ethics is human-centric thought. Usmani et al. (2023) view this as a human-centered operational philosophy, where technology is designed to empower and assist, rather than replace the user's role. In this understanding, an ethical AI system needs to operate as a mechanism protecting thought, helping humans expand intellectual capacity while maintaining identity and control.

Lepri et al. (2021) also agree that the development of AI must parallel core humanistic values such as compassion, fairness, and accountability. This approach simultaneously indicates a significant shift in technological goals: from optimizing performance (doing things faster) to optimizing human value (doing things more meaningfully).

At the global governance level, UNESCO's Recommendation (2021, 2023) has formalized human-centric philosophy into a normative framework. UNESCO defines AI ethics as a systematic commitment to protect

fundamental freedoms and human dignity throughout the entire lifecycle of AI systems. According to Subash & Whig (2025), principles such as "do no harm" and "fairness" are not just technical requirements but also reflect the continuation of ethical philosophical traditions in the digital context.

In summary, the philosophical basis of AI ethics is built on a consistent principle: artificial intelligence must be a tool serving human development, and any technological progress becomes meaningless if it infringes upon human autonomy and dignity.

However, the human-centric perspective also poses a paradox that needs resolution: how can AI help humans work more effectively without making them intellectually lazier? If technological support means depriving learners of intellectual effort, then it is merely a form of "false empowerment" rather than the reality. Therefore, human-centric philosophy does not stop at meeting needs but, more importantly, protects the intrinsic developmental capacity of humans.

3. Pillars of sustainable social governance

If human-centric philosophy is seen as the core ideological orientation, then social governance frameworks are the implementation tools to translate AI ethical principles into practical action. A review of international studies shows that AI governance is not just about enacting technical regulations, but is essentially a form of "**social contract**" between developers, regulators, and the beneficiary community.

3.1. Social Responsibility and Accountability In the context of AI increasingly participating in critical decisions, Camilleri (2024) emphasizes that AI governance must be tightly linked to social responsibility. This means organizations are responsible not only for algorithmic performance but also for the social impacts it creates. Accordingly, AI must not operate as a system lacking explainability; transparency and accountability regarding operating mechanisms must be considered mandatory conditions to maintain and reinforce the trust of the user community.

Concurring with this view, a review by Ismail & Ahmad (2025) of 22 global governance frameworks shows that "accountability" is one of the most core values. Accountability requires clearly identifying the responsible entity—and that entity must be human—for every decision supported by AI. In education, this requirement is particularly important to prevent the shifting of moral responsibility to machines when pedagogical errors occur.

From the above analyses, it can be affirmed that transparency is the foundation for forming trust in the digital era. In education, pedagogical decisions based on inexplicable and unverifiable algorithms cannot be accepted. Therefore, accountability is not just a management requirement but a prerequisite ethical condition for AI deployment in the classroom environment.

3.2. Fairness and Social Risk Control

The second pillar of sustainable governance is the capacity for risk control to ensure **fairness**. According to Subash & Whig (2025), ethical principles such as "do no harm" and "fairness" need to be translated from theory into technical filters during the design and operation processes. In this sense, AI ethics acts as a control mechanism, helping to limit social prejudices from being amplified by algorithms.

Sargiotis (2024) argues that AI ethics needs to be integrated directly into practical operational processes, rather than stopping at declarative principles. When properly integrated, AI ethics functions as an early warning system, helping to identify and prevent risks such as privacy violations or discrimination before they cause consequences. Samea Qoura et al. (2024) also emphasize that establishing ethical safety thresholds is a necessary regulatory tool to help AI applications operate on the right trajectory and avoid deviant impacts.

At the global level, UNESCO's Recommendation (2021) has elevated these principles into a systematic commitment. Sustainable social governance requires synchronization between macro policy and technical solutions, to ensure that all AI advancements aim for the common good and leave no one behind.

In summary, AI ethical governance is a combination of transparent accountability and proactive risk control mechanisms. This is a crucial foundation for translating humanistic values from theoretical orientation into social reality.

Practice shows that social prejudices, when integrated into algorithms, often become significantly harder to identify and rectify. In this context, AI ethics can be viewed as a value screening mechanism, helping to detect and limit deviations from the outset. Without proactive risk control measures from the design phase, technology may unintentionally become a tool that amplifies inequality rather than contributing to narrowing it.

4. Humanistic value orientation in education

From the analysis of philosophical foundations and governance challenges, it is evident that solving the AI ethics problem cannot stop at technical barriers or prohibitive regulations. For AI to truly serve human development, ethics needs to be transformed into core humanistic values, permeated into educational goals and methods. Below are three key value orientations.

4.1. Shifting Objectives: From Instrumental Skills to Value Thinking

In the digital era, the goal of education is no longer just teaching learners how to use tools, but more importantly, teaching them how to think about the value and consequences of those tools. Kommineni et al. (2025) emphasize that AI ethics needs to be integrated as a guiding principle for sustainable development pedagogy. Accordingly, schools need to position AI as a support tool to expand access to knowledge and personalize learning, but it must absolutely not replace the professional role and ethical judgments of educators.

Following this approach, Lee et al. (2024) propose an integrated education model where technological knowledge is deployed alongside the formation of ethical awareness and social responsibility. Instead of training learners who are only skilled in technical operations, education needs to aim at fostering the capacity to make considered decisions in practical situations.

This shift demonstrates a new perspective on the mission of education: instead of racing against machines in information processing speed, education needs to focus on developing qualities that technology cannot replace, especially compassion and ethical thinking capacity. In a context where tools are constantly changing, teaching learners to ask "Why?" becomes much more important than just stopping at "How?".

4.2. Protecting Learner Autonomy and Critical Thinking

The most core humanistic value that needs protection in the digital education environment is autonomy. Egamberdiyeva (2025) warns of the risk of learners' independent thinking capacity diminishing if they rely excessively on available AI suggestions. Lacking appropriate orientation, the convenience of technology can unintentionally create a generation of passive learners who gradually lose critical thinking—a vital goal of liberal education.

Therefore, Mutawa & Sruthi (2025) suggest redefining AI ethics in education: from a set of external compliance rules to an intrinsic capacity of the learner. AI ethics needs to become the ability to proactively identify, evaluate, and regulate interactions with technology. This view aligns with Floridi et al. (2018), emphasizing that education must prevent the risk of machines usurping the coordination of thought, while ensuring humans always retain the final decision-making power.

The greatest danger of digital education is not that AI will rebel and take control, but that humans will become passive. When learners become accustomed to receiving ready-made answers from AI without undergoing a process of critical thinking, they are voluntarily relinquishing their cognitive autonomy. Thus, protecting autonomy is also protecting the learner's identity during the learning process.

4.3. Building a Foundation of Trust in School Culture

Finally, for humanistic values to be realized, schools need to build a foundation of trust in learning and using AI platforms. Karran et al. (2025) show that the level of social and learner acceptance of AI depends heavily on ethical trust. Users are only willing to accept technology when they feel safe regarding privacy and trust that the system operates fairly.

Thus, AI ethics in education is not just a matter of algorithms, but a matter of culture of transparency. Córdova & Vicari (2022) support an "ethics by design" approach, where ethical standards are embedded directly into the school's operational processes. Transparently explaining how AI collects data and makes suggestions is not only a technical requirement but also an expression of respect for learner dignity, helping to reduce anxiety about surveillance or manipulation.

In short, humanistic value orientation in education requires a strong shift: from teaching literacy to teaching humanity, from consuming technology to mastering technology, and from managing risk to creating trust. At a deeper level, technology is merely the "body," while the culture of trust is the "soul" of the smart school. A successful AI-integrated education system is not necessarily the most modern one, but the one most trusted by learners. And that trust can only be formed when ethical values are respected and executed transparently and consistently.

4.4. Orienting the Development of Learning Competence in the AI Context

The human-centric educational philosophy in the context of digital transformation requires a redefinition of learning competence. Previously, this capacity was often measured by the ability to memorize and process information. However, as AI tools can perform these tasks faster and more accurately than humans, assessment criteria need to shift towards autonomy in thinking and the capacity to perceive values.

First, developing learning competence must be linked to critical thinking that objectifies technology. Learners need not only question the content of knowledge but also the mechanism creating that knowledge. This means not accepting AI outputs as precise answers, but viewing them as data needing verification through logical analysis and ethical consideration. This is a necessary cognitive self-protection mechanism, helping to limit the tendency towards intellectual laziness and reduce the risk of dependence on available suggestions.

Second, education needs to guide learners to shift from the position of passive technology users to value creators. AI may excel at synthesizing data, but it cannot truly feel human lived experiences. Therefore, the core learning competence of the future is the ability to use AI as a tool to solve practical problems while maintaining individual empathy and responsibility. Learners need to be trained to direct AI to serve socially meaningful goals, rather than passively letting technology shape their habits and personal objectives.

Finally, the orientation for developing learning competence needs to emphasize self-awareness of identity and personal values. Education must help learners understand that AI is merely a vehicle to extend intellectual capabilities, while conscience and ethical responsibility are what define human value. Success in learning should not be measured by the speed of using AI, but by the ability to steadfastly maintain cultural and ethical values when collaborating with machines.

5. Conclusion

The robust development of artificial intelligence places education before a historic crossroad: either we passively let technology lead, or we proactively shape technology based on fundamental human values. From systematic analyses of philosophical foundations and social governance frameworks, this study draws three conclusions with strategic orientation significance.

First, philosophically, AI ethics is not merely about establishing technical barriers to limit risk, but is a reaffirmation of the central position of humans. Human-centric philosophy needs to be seen as a consistent principle, where AI is positioned as a tool to empower and expand intellectual capacity, and must absolutely not be allowed to replace autonomy or the role of human ethical decision-making. Thus, protecting the independent thinking capacity of learners against suggestions from algorithms is also protecting human dignity in the digital era.

Second, regarding social governance, trust is the most important foundation for AI to exist sustainably in the educational environment. Transparency and accountability are not just legal requirements but a social commitment between schools, technology developers, and learners. Only when mechanisms for controlling risk and bias are executed publicly can we reduce apprehension and build a safe and fair digital education environment.

Third, regarding educational orientation, the paper proposes a shift in focus from teaching usage skills to educating for humanistic values. The task of the modern school is not only to train learners to use AI proficiently but, more importantly, to create citizens with strong ethical orientations to master technology. The core competence of future learners lies not in memorizing knowledge, but in the ability for critical thinking and making humanistic decisions—something machines cannot replace.

In conclusion, the destination of integrating AI into education is not to optimize performance or speed, but to promote the free and comprehensive development of humans. Thus, AI ethics is the core foundation ensuring technology always serves the highest goal of education: nurturing humanity and contributing to creating a better society.

References

1. Camilleri, M. A. (2024). Artificial intelligence governance: Ethical considerations and implications for social responsibility. *Expert systems*, *41*(7), e13406.
2. Córdova, P. R., & Vicari, R. M. (2025). ADVANCING ETHICS IN ARTIFICIAL INTELLIGENCE FOR EDUCATION: SAFEGUARDING LEARNING WITH ARTIFICIAL MORAL AGENTS. *Journal of Media Critiques*, *11*(28), e423-e423.
3. Egamberdiyeva, Z. (2025). Ethical and pedagogical implications of artificial intelligence in education. *Science and Education*, *6*(8), 44-51.
4. Floridi, L., Cowls, J., Beltrametti, M., Chatila, R., Chazerand, P., Dignum, V., ... & Vayena, E. (2018). AI4People—An ethical framework for a good AI society: Opportunities, risks, principles, and recommendations. *Minds and machines*, *28*(4), 689-707.
5. Formosa, P., HipÃģlito, I., & Montefiore, T. (2025). Artificial Intelligence (AI) and the Relationship between Agency, Autonomy, and Moral Patiency. *arXiv preprint arXiv:2504.08853*.
6. Gunkel, D. J. (2024). Introduction to the Ethics of Artificial Intelligence. In *Handbook on the Ethics of Artificial Intelligence* (pp. 1-12). Edward Elgar Publishing.
7. Hanna, R., & Kazim, E. (2021). Philosophical foundations for digital ethics and AI Ethics: a dignitarian approach. *AI and Ethics*, *1*(4), 405-423.
8. Ismail, O., & Ahmad, N. (2025). Ethical and Governance Frameworks for Artificial Intelligence: A Systematic Literature Review. *International Journal of Interactive Mobile Technologies*, *19*(14).
9. Karran, A. J., Charland, P., Trempe-Martineau, J., Ortiz de Guinea Lopez de Arana, A., Lesage, A. M., Senecal, S., & Leger, P. M. (2025). Multi-stakeholder perspective on responsible artificial intelligence and acceptability in education. *npj Science of Learning*, *10*(1), 44.
10. Kommineni, M., Chundru, S., Maroju, P. K., & Selvakumar, P. (2025). Ethical Implications of AI in Sustainable Development Pedagogy. In *Rethinking the Pedagogy of Sustainable Development in the AI Era* (pp. 17-36). IGI Global Scientific Publishing.
11. Lee, J., Hong, M., & Cho, J. (2024). Development of a Content Framework of Artificial Intelligence Integrated Education Considering Ethical Factors. *International Journal on Advanced Science, Engineering & Information Technology*, *14*(1).
12. Lepri, B., Oliver, N., & Pentland, A. (2021). Ethical machines: The human-centric use of artificial intelligence. *IScience*, *24*(3).
13. Mutawa, A. M., & Sruthi, S. (2025). UNESCO's AI Competency Framework: Challenges and Opportunities in Educational Settings. *Impacts of Generative AI on the Future of Research and Education*, 75-96.
14. Samea Qoura, P., & Abdul, A. (2024). Artificial Intelligence: Ethical Considerations and Caveats. *International Journal of Education, Science, Technology and Development*, *2(3)*, 130-163.
15. Sargiotis, D. (2024). Ethical AI in information technology: Navigating bias, privacy, transparency, and accountability. *Privacy, Transparency, and Accountability (May 28, 2024)*.
16. Subash, B., & Whig, P. (2025). Principles and frameworks. In *Ethical Dimensions of AI Development* (pp. 1-22). IGI Global.
17. Subash, B., & Whig, P. (2025). Principles and frameworks. In *Ethical Dimensions of AI Development* (pp. 1-22). IGI Global.
18. UNESCO. (2021). Recommendation on the Ethics of Artificial Intelligence.

19. UNESCO. (2023). Recommendation on the Ethics of Artificial Intelligence.
20. Usmani, U. A., Happonen, A., & Watada, J. (2023, June). Human-centered artificial intelligence: Designing for user empowerment and ethical considerations. In *2023 5th international congress on human-computer interaction, optimization and robotic applications (HORA)*. IEEE.

Author Profile

Nguyen Minh Giam holds a PhD in Educational Science from Thu Dau Mot University, Ho Chi Minh City, Vietnam. His research specializes in educational science, including: educational technology, AI in education, educational management, teaching theory and methodology, and the development of self-learning and lifelong learning skills. (Email: giamnm@tdmu.edu.vn).