© IJARW | ISSN (O) - 2582-1008
November 2026 | Vol. 8 Issue. 5
www.ijarw.com

PHILOSOPHICAL ISSUES ARISING FROM THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

Do Minh Anh

Hanoi University of Natural Resources and Environment, Vietnam

ABSTRACT

The rapid development of artificial intelligence (AI), particularly machine learning and large language models, is bringing about profound changes in social life while raising a range of new questions for philosophical thought. The ability of machines to process information, learn, reason, and generate content has sparked debates about the nature of intelligence, the relationship between thought and consciousness, the limits of artificial cognition, and the position of human beings in modern society. In addition, the increasingly widespread application of AI poses challenges concerning ethical responsibility, social justice, labor relations, and the long-term governance of technology. Based on an analysis of contemporary philosophical perspectives and the application of the methodological framework of Marxist-Leninist philosophy, this article examines the fundamental issues arising from the development of AI in the fields of ontology, epistemology, ethics, and the philosophy of human beings. It emphasizes the need to develop AI in a human-centered direction, ensuring that technological progress is accompanied by social responsibility and humanistic values. The article also suggests several orientations for AI research and application that are appropriate to Vietnam's circumstances.

Keyword: artificial intelligence, philosophy, consciousness, AI ethics, responsibility.

1. INTRODUCTION

Artificial intelligence (AI) has become one of the scientific and technological achievements exerting a far-reaching influence on human development. From data processing, image recognition, and language analysis to content creation and decision-making support, AI is increasingly being applied in fields that previously relied primarily on human cognitive abilities and intellectual activities. The emergence of large language models and generative AI technologies has further highlighted the application potential of this technology while bringing the boundary between human activities and machine operations into sharper focus as a subject of widespread interest.

In fact, the idea of developing machines capable of performing human intellectual activities dates back to the mid-twentieth century. In their proposal for the Dartmouth workshop in 1956, McCarthy, Minsky, Rochester, and Shannon hypothesized that every aspect of learning and intelligence could, in principle, be described with sufficient precision to enable machines to simulate it. According to Russell and Norvig (2021), this proposal marked one of the important milestones in laying the foundations for the emergence and development of AI as a field of study. However, alongside remarkable technological advances, the question of whether machines can think is no longer merely hypothetical. It has become an issue that requires consideration from both scientific and philosophical perspectives. Notably, the development of AI is not simply transforming methods of production, management, and information exchange. It is also directly challenging concepts that have long been regarded as fundamental to human understanding. As machines become capable of generating texts and images, performing analytical tasks, and producing logically structured responses, the central question is no longer limited to what technology can accomplish. It also concerns the very nature of intellectual activity, consciousness, and genuine understanding. At the same time, the increasing use of algorithms to support or make decisions raises questions about human autonomy, responsibility, and the ability to maintain control over technology.

For Vietnam, examining these issues is of practical significance in the context of promoting scientific and technological development and national digital transformation. Decision No. 127/QD-TTg dated January 26, 2021, issued by the Prime

Minister, approved the National Strategy for Research, Development, and Application of Artificial Intelligence through 2030. Subsequently, Resolution No. 57-NQ/TW dated December 22, 2024, issued by the Politburo, reaffirmed the important role of science and technology, innovation, and digital transformation in national development. These policy orientations demonstrate that the development and mastery of technology must be considered in conjunction with the protection of social and ethical values and the interests of human beings.

Against this background, this article examines the major philosophical issues arising from the development of AI, including the nature of intelligence and consciousness, the relationship between knowledge and reality, ethical responsibility, transformations in labor and human nature, and long-term risks. On this basis, the article draws on the perspectives of Marxist-Leninist philosophy to clarify the role of human beings in the development and governance of technology, while proposing several orientations of relevance to Vietnam.

2. THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE AND THE ISSUES IT RAISES FOR PHILOSOPHICAL THOUGHT

Since its emergence, AI has undergone several stages of development, characterized by different approaches to the nature of intelligence. In its early stages, research primarily focused on simulating thought through symbolic systems, rules, and logic. Subsequently, the development of artificial neural networks and machine learning opened up new approaches that enabled machines to learn from data, identify patterns, and improve their performance without requiring detailed programming for each individual task. According to Russell and Norvig (2021), AI can be studied through various approaches, ranging from simulating human thought processes and behavior to developing agents capable of acting rationally to achieve specific goals.

The shift from rule-based AI to systems that learn from data represents not only technical progress but also reflects different conceptions of intelligence. While the symbolic approach emphasizes the role of logic and symbolic operations, the connectionist approach focuses on the ability to learn, adapt, and develop intelligent behavior through data processing. This distinction gives rise to an important philosophical question: Can intelligence be explained entirely in terms of information-processing functions, or does it also depend on specific material, biological, and social conditions?

From a philosophical perspective, this is not merely a question concerning the technical capabilities of machines. The deeper issue is to identify the constituent elements of cognitive activity, the relationship between thought and consciousness, and the conditions under which a subject can genuinely understand the world. Although AI is increasingly demonstrating its ability to perform complex intellectual tasks, whether such manifestations imply the existence of consciousness remains an unresolved question without a universally accepted answer.

For this reason, AI has become an object of particular significance in various branches of philosophy. Ontological debates examine the relationship between matter, thought, and consciousness; epistemology focuses on the limits of artificial knowledge and the criteria for determining truth; ethics addresses the need to establish responsibility and standards of conduct; while the philosophy of human beings examines changes in labor, creativity, and the position of humanity in the face of technological development. These issues are closely interconnected and form a system of questions that requires a comprehensive approach.

3. ONTOLOGICAL AND EPISTEMOLOGICAL ISSUES ARISING FROM THE DEVELOPMENT OF ARTIFICIAL INTELLIGENCE

3.1. The Nature of Intelligence and the Thinking Capabilities of Machines

One of the most fundamental questions raised by the development of AI is whether machines genuinely think or merely simulate the manifestations of thinking. This issue was addressed by Turing (1950) in his work Computing Machinery and Intelligence. Rather than directly attempting to resolve the question of whether machines can think, Turing proposed the imitation game, which later became known as the Turing Test. Under this approach, a machine's manifestation of intelligence is evaluated by determining whether a human interlocutor can distinguish its responses from those of a human being.

Turing's important contribution was to transform an abstract question into an issue that could be examined through observable behavior and performance outcomes. However, the test primarily evaluates the outward manifestation of intelligence and does not resolve whether machines genuinely understand the content they process. A system may provide appropriate answers, use natural language, and demonstrate reasoning abilities, but these capabilities alone are insufficient to establish that it possesses a mental life comparable to that of human beings.

This distinction was clarified by Searle (1980) through his Chinese Room argument. In his thought experiment, Searle imagined a person who does not understand Chinese but can follow a set of predefined rules to manipulate symbols and produce responses appropriate to given instructions. Although the resulting responses may lead outside observers to believe that the person understands Chinese, the individual performing the operations does not understand the meaning of the symbols being processed. On this basis, Searle argued that correctly manipulating symbols according to syntactic rules does not necessarily imply an understanding of their semantic content. In other words, executing a program does not automatically mean possessing genuine understanding.

Searle's argument has generated extensive debate in the philosophy of mind. Critics argue that the individual performing the operations inside the room should not be equated with the entire system, since understanding may emerge at the level of the system as a whole. Other perspectives emphasize that intelligence should be evaluated in terms of a system's functions and its ability to interact with the environment, rather than being determined solely by the material substrate on which it is based.

These debates suggest that the central issue is not simply how many tasks machines can perform, but rather how the nature of intelligence should be defined. If intelligence is understood primarily in terms of information processing and problem-solving capabilities, AI may be regarded as a form of artificial intelligence in the functional sense. However, if thinking also encompasses the ability to understand meaning, achieve self-awareness, and connect knowledge with lived experience, equating machine operations with human thought requires greater caution.

3.2. The Relationship Between Knowledge, Language, and Reality

The development of AI, particularly large language models, raises further questions about the relationship between language and knowledge. These systems can process, analyze, and generate texts with complex structures, demonstrating connections among concepts and information. However, generating linguistically meaningful content does not necessarily mean that a system has a comprehensive understanding of the reality represented by that content.

Bender et al. (2021) warned about the limitations of large language models through the concept of "stochastic parrots". According to the authors, these models can generate fluent text by exploiting statistical patterns in language data without necessarily possessing a corresponding understanding of the meaning of the content they produce. This warning highlights the gap between language generation capabilities and genuine cognitive competence.

From an epistemological perspective, an important issue is the need to distinguish information from knowledge and the persuasiveness of a statement from its truthfulness. An answer presented coherently, supported by seemingly logical arguments, and expressed in appropriate language may still contain inaccurate or unsubstantiated information. Therefore, AI-generated outputs cannot automatically be regarded as reliable knowledge unless they have been verified against appropriate information sources and evidence.

This issue is directly related to the Marxist-Leninist philosophical perspective on the role of practice in the process of cognition. According to dialectical materialism, cognition is the process through which objective reality is reflected in the human brain and is formed and developed in association with practical activities. Practice serves as the foundation, driving force, and ultimate purpose of cognition, as well as the criterion for verifying truth (Ministry of Education and Training, 2021). Applying this perspective to the use of AI indicates that information generated by these systems must be verified against objective evidence rather than evaluated solely on the basis of its presentation or the apparent plausibility of the responses.

Furthermore, the widespread adoption of AI is also changing the way people access knowledge. The ability to obtain answers to numerous questions quickly helps save time and broadens access to information. However, excessive dependence on readily available outputs may gradually reduce people's initiative in seeking, analyzing, and evaluating information. This is a particularly important concern in education and scientific research because the value of learning lies not only in the final outcome but also in the ability to formulate questions independently, develop arguments, and verify conclusions.

Thus, AI can support the expansion of human cognitive capabilities, but it does not eliminate the role of critical thinking and human verification. The challenge is to use technology as a tool for supporting knowledge discovery and processing while maintaining human initiative in determining the accuracy and significance of the information received.

3.3. The Problem of Consciousness and the Limits of Artificial Cognition

Beyond the question of whether machines can think, the development of AI raises an even more profound issue: Can an artificial system develop consciousness and subjective experience? This question belongs to the field of the philosophy of mind and is closely associated with the challenge of explaining the relationship between physical processes and mental life.

Chalmers (1995) distinguishes between the relatively easy problems of consciousness, such as explaining the ability to process information, pay attention, remember, and respond to the environment, and the hard problem of consciousness, which concerns why physical processes are accompanied by subjective experience. This distinction demonstrates that explaining the operating mechanisms of a cognitive system does not necessarily amount to explaining the nature of conscious experience.

In the context of AI, the issue becomes more complex because a system may exhibit intelligent behavior, respond flexibly, and even describe emotional states without any conclusive evidence of corresponding subjective experience. Therefore, the ability to communicate or the complexity of information processing alone cannot serve as a sufficient basis for concluding that AI possesses consciousness. Conversely, the absolute assertion that machines can never be conscious under any circumstances should also be considered in light of relevant philosophical debates and scientific developments.

From the perspective of dialectical materialism, consciousness is regarded as a property of highly organized matter, namely the human brain, and has both natural and social origins. Labor and language play important roles in the formation and development of human consciousness (Ministry of Education and Training, 2021). This approach makes it possible to understand consciousness in relation to specific material conditions and socio-historical circumstances, rather than treating it as an entity that exists independently of the material world.

However, applying this perspective to AI research should not lead to an equation of the simulation of certain cognitive functions with the process of consciousness formation. Current AI systems are products of human scientific activity and creativity, developed on the basis of computational models, data, and technological infrastructure. AI's information-processing capabilities demonstrate that humans can simulate or extend certain intellectual functions through artificial means, but they do not provide sufficient grounds for concluding that these systems possess subjective experiences comparable to those of human beings.

Therefore, the issue of artificial consciousness requires continued investigation through dialogue among philosophy, cognitive science, and neuroscience. An appropriate approach is to avoid both the absolute denial of technological possibilities and the exaggerated attribution of human-like capabilities to machines. This is also an important requirement for maintaining a scientific and objective perspective in response to the rapid advances in AI.

4. ETHICAL ISSUES AND RESPONSIBILITY IN THE DEVELOPMENT AND APPLICATION OF ARTIFICIAL INTELLIGENCE

4.1. Ethical Responsibility in Artificial Intelligence Systems

As AI becomes increasingly involved in decision-making processes, determining who should be held responsible for the outcomes generated by these systems has become more complex. In traditional systems, responsibility can often be

attributed to the actions and decisions of specific individuals or organizations. However, in machine learning systems, data processing and output generation may occur in ways that are difficult to predict fully, making it more challenging to trace the causes of errors and determine accountability.

An AI system may be designed by one organization, developed using data from multiple sources, provided by a business, and deployed in a specific application environment. When errors occur, responsibility cannot be determined solely on the basis of the entity directly operating the system. Instead, the entire process of system design, provision, testing, and use must be taken into account. The difficulty of identifying the party responsible in automated systems is commonly discussed through the concept of the "responsibility gap".

From a philosophical perspective, ethical responsibility is associated with an agent's capacity for awareness, choice, and accountability for their actions. Therefore, as long as AI has not been established as an independent moral agent, responsibility for harmful outcomes generated by these systems cannot be transferred to machines. Instead, responsibility must be considered in relation to the individuals and organizations involved in the development, provision, deployment, and governance of the technology.

This perspective is also consistent with the requirement to ensure human control over AI. Floridi and Cowls (2019) proposed five fundamental principles of AI ethics: beneficence, non-maleficence, autonomy, justice, and explicability. These principles demonstrate that technological development should be evaluated not only in terms of operational efficiency but also in terms of the extent to which it respects human rights and interests.

4.2. Fairness, Transparency, and the Protection of Human Rights

One of the notable ethical challenges associated with AI is the risk of reproducing and amplifying existing social biases. Machine learning systems are built on training data; therefore, the quality and characteristics of these data directly affect system performance. If the data contain biases or reflect historical inequalities, the system may perpetuate these patterns in subsequent decisions (Bender et al., 2021).

This issue is particularly important when AI is used in areas such as recruitment, finance, education, and social governance. An algorithm may produce seemingly objective results because the process is carried out by a machine. However, the data and methods used to design the system may themselves contain certain biases. Therefore, automation should not be equated with neutrality, nor should algorithmic outputs automatically be regarded as fair.

Alongside fairness, transparency and explainability are also essential requirements. When an AI system makes a decision that directly affects an individual's interests, the person concerned should have access to appropriate information about the basis for that decision. This helps prevent situations in which individuals are expected to accept technology-generated outcomes without having an opportunity to understand, examine, or challenge them.

Another aspect that deserves attention is the material and social nature of AI. Crawford (2021) argues that AI is not an immaterial technology existing independently of economic and social conditions. Its development is closely associated with the exploitation of natural resources, energy consumption, labor organization, and large-scale data collection. Therefore, assessing the impacts of AI requires going beyond algorithms to examine the material conditions and power relations underlying the technology.

These issues are reflected in international ethical and legal frameworks. UNESCO (2021) adopted the Recommendation on the Ethics of Artificial Intelligence, emphasizing the need to protect human dignity, human rights, fairness, transparency, and accountability throughout the entire life cycle of AI systems. At the legal level, Regulation (EU) 2024/1689 on artificial intelligence, adopted by the European Parliament and the Council (2024), establishes a risk-based regulatory approach, imposing different requirements on AI systems according to the level of impact they may have.

From a philosophical perspective, these principles and regulations demonstrate the need to place technology within a framework of ethical values and social responsibility. AI can develop in a positive direction only when technical efficiency is accompanied by respect for human rights,

protection of the public interest, and efforts to mitigate adverse social consequences.

4.3. The Moral Status of Artificial Intelligence

The development of AI also raises the question of whether machines can become entities that deserve moral consideration. This issue is directly related to debates about consciousness and subjective experience. If a system merely performs computational operations without the capacity to feel or experience, regarding it as a moral subject with rights and interests comparable to those of human beings lacks a convincing basis.

However, as AI capabilities become increasingly sophisticated, determining the moral status of artificial systems may become a subject requiring further investigation. The key issue is to distinguish between the expression of emotions and the actual experience of emotions, as well as between the ability to respond to a situation and the ability to experience that situation.

At present, extending moral status to AI remains a highly contested issue. Therefore, speculation about the rights of machines should not be allowed to overshadow more pressing concerns, particularly human responsibility for the impacts of technology. Protecting human rights, dignity, and interests must remain central to the development of ethical principles and AI governance policies.

5. ARTIFICIAL INTELLIGENCE AND TRANSFORMATIONS IN CONCEPTIONS OF HUMAN BEINGS, LABOR, AND CREATIVITY

The development of AI is changing how people perceive intellectual capabilities, labor activities, and the value of creativity. Historically, many activities, such as calculation, analysis, language processing, and content creation, have been regarded as distinctive manifestations of human capabilities. However, as machines become increasingly capable of performing these tasks, new questions have emerged concerning the nature and position of human beings in modern society.

From the perspective of Marxist-Leninist philosophy, human beings are not merely biological entities but also social beings whose formation and development take place through practical activities, particularly labor and social relations. Labor not only creates the material conditions necessary for human existence but also contributes to the development of human capabilities, thought, and values. Therefore, when AI transforms the tools and methods of labor, these changes have implications not only for the economy but also for social life and human development.

AI has the potential to increase productivity, reduce costs, and support the performance of many complex tasks. However, automation may also transform employment structures, reducing demand for certain occupations while increasing the need for new skills. The issue that requires attention is not simply which jobs technology can replace, but also how the benefits of technological adoption are organized and distributed.

From a Marxist perspective, the development of the productive forces is invariably associated with changes in production relations. Applied to the context of AI, this raises questions about who owns the data, models, and technological infrastructure; which entities control the automation process; and how productivity gains are distributed among businesses, workers, and society. If the benefits of AI are concentrated primarily among those who control technology and resources, while workers bear most of the costs of the transition, economic and social inequalities may increase.

Crawford (2021) also points out that AI development is associated with specific material conditions and labor relations. The extraction of natural resources, energy consumption, data collection, and use of labor in the process of building AI systems demonstrate that this technology does not exist independently of social life. Therefore, AI needs to be examined within the broader context of economic, political, and social relations rather than evaluated solely in terms of its technical features.

In addition to labor, creativity also raises new questions that require further reflection. Generative AI systems can produce texts, images, and various other forms of content based on their training data. However, human creativity involves more than the combination of information. It is also associated with lived experience, emotions, purposes, cultural contexts, and the ability to evaluate the value of a product. Therefore, the ability of AI to generate new products does not

mean that human creativity has lost its significance.

What deserves attention is how people use AI in creative activities and research. When technology is used to support idea generation, expand expressive capabilities, and improve the efficiency of intellectual work, it can become a means of promoting creativity. Conversely, if people become entirely dependent on machine-generated outputs, fail to verify them, and lose their initiative in the thinking process, their capacity for independent creativity may be adversely affected.

The above analysis suggests that human value lies not only in performing tasks that machines cannot yet accomplish, but also in the ability to establish goals, understand meaning, make value-based choices, and take responsibility for actions. This provides a basis for affirming that AI should be developed to expand human capabilities rather than diminish the role of human beings as active agents in social life.

6. LONG-TERM RISKS AND HUMAN RESPONSIBILITY FOR THE FUTURE

In addition to impacts that have already emerged or are currently unfolding, the development of AI raises questions about long-term risks, particularly as systems become increasingly autonomous and capable of performing complex tasks. This issue is examined by Bostrom (2014) in Superintelligence: Paths, Dangers, Strategies, in which he discusses the possibility of the emergence of superintelligent systems and the potential risks that may arise if their operational objectives are not aligned with human values.

One notable argument put forward by Bostrom is that a system's level of intelligence and its ultimate objectives may not be directly related. A system with superior processing capabilities will not necessarily pursue objectives that automatically align with human interests. Furthermore, different objectives may lead to similar instrumental behaviors, such as seeking resources, maintaining operational capacity, or expanding available resources in order to achieve predefined goals. These arguments indicate that the greater the technological capabilities become, the more important it is to ensure effective control and appropriate direction of system operations.

From this perspective, Russell (2019) proposes an approach to human-compatible AI, whereby systems should be designed not to assume that their objectives are fully defined, but instead to learn and adjust in response to human preferences, values, and interests. This approach helps shift the focus of research from merely enhancing machine capabilities toward ensuring that such capabilities are used beneficially and remain subject to human control.

However, it is necessary to distinguish between risks already documented in practice and superintelligence scenarios that remain controversial in terms of their likelihood and timing. Issues such as algorithmic bias, misinformation, privacy violations, and excessive dependence on technology are already evident and require attention. Meanwhile, risks associated with superintelligent systems require further investigation based on a cautious approach that avoids both complacency and unsupported predictions.

From a philosophical perspective, the central issue is human responsibility for the long-term consequences of scientific and technological activities. The development of AI affects not only immediate interests but may also shape the conditions for the existence and development of future generations. Therefore, decisions concerning AI research, development, and application should be considered in relation to the long-term interests of society, ensuring a balance between technological innovation and safety requirements.

7. ORIENTATIONS FOR THE DEVELOPMENT AND APPLICATION OF ARTIFICIAL INTELLIGENCE IN VIETNAM

The philosophical issues arising from the development of AI indicate that efforts should extend beyond technological advancement to the establishment of appropriate cognitive, ethical, and institutional foundations. In Vietnam, AI development should be aligned with the goals of achieving technological autonomy, safeguarding national interests, and promoting the role of human beings as the central agents in the digital transformation process.

First, it is necessary to continue applying the methodological principles of Marxist-Leninist philosophy in researching and evaluating AI, particularly the perspectives of comprehensiveness, development, and historical specificity. AI should be examined in relation to

the country's economic, social, cultural, and developmental conditions, avoiding one-sided approaches that focus exclusively on technological benefits or speculative risks. At the same time, research should engage with contemporary advances in cognitive science, neuroscience, and the philosophy of mind to strengthen the scientific foundations for addressing questions concerning intelligence and consciousness.

In addition, the principle of human-centered AI development and application should be clearly established. Technology should serve the goals of improving quality of life, expanding access to knowledge, and promoting comprehensive human development. The application of AI should not be evaluated solely in terms of the degree of automation or economic efficiency but should also take into account its impacts on privacy, autonomy, human dignity, and the legitimate interests of individuals, organizations, and communities.

Another important requirement is to improve governance mechanisms and establish clear accountability frameworks for AI systems. Particular attention should be paid to sectors that may directly affect human rights and interests, such as healthcare, education, finance, and social governance. Establishing mechanisms for supervision, inspection, explanation, and error correction will help mitigate risks arising during the operation of AI systems. The responsibilities of organizations involved in developing, providing, and using AI should be determined according to their respective roles and levels of control, ensuring that technology does not become a justification for obscuring human accountability.

Regarding the labor market, proactive adaptation to changes driven by automation and AI is essential. Retraining, digital skills development, and the enhancement of competencies required to work in AI-assisted environments should be prioritized. At the same time, greater attention should be paid to groups of workers who are particularly vulnerable to technological transformation, ensuring that the benefits of scientific progress are distributed equitably and do not exacerbate social inequality.

In education and scientific research, the use of AI should be considered in relation to the requirements of developing independent thinking, critical reasoning, and academic integrity. AI can support learners in accessing information, analyzing data, and developing ideas, but it cannot replace the responsibility of learners and researchers for the content, arguments, and scientific results they produce. Therefore, education in technological ethics and the ability to critically evaluate information should receive attention alongside the widespread development of AI-related skills.

Finally, research and international cooperation should be strengthened to develop principles for AI governance. The recommendations of UNESCO (2021) and the experience of the European Union in establishing a legal framework can provide useful references for Vietnam. However, their application should be adapted to the country's development conditions, legal system, and cultural and social characteristics. The objective should not be merely to adopt technology but also to progressively strengthen national capacities in research, development, and governance, thereby ensuring that Vietnam can proactively respond to the transformations brought about by AI.

8. CONCLUSION

The development of artificial intelligence is raising fundamental philosophical issues directly related to how human beings understand intelligence, consciousness, knowledge, ethics, and their own nature. The increasingly expanding capabilities of machines in information processing, performing cognitive tasks, and generating content have transformed traditional approaches to intellectual activities while highlighting the need to distinguish between the simulation of cognitive functions and genuine human understanding and consciousness. Furthermore, issues concerning ethical responsibility, fairness, labor relations, and long-term risks demonstrate that AI is not merely a technological achievement but also an economic, social, and philosophical phenomenon with far-reaching implications.

From the perspective of Marxist-Leninist philosophy, AI research should be situated within the relationships between matter and consciousness, human beings and practical activity, and specific historical and social conditions. This approach makes it possible to recognize AI as a product of human creative activity while clarifying the transformations that technology brings about in the productive forces, labor relations, and social life. Nevertheless, the

application of philosophical methodology should also involve dialogue with and the selective adoption of contemporary research achievements, particularly in the fields of the philosophy of mind, epistemology, and the ethics of technology.

It can be argued that the greatest challenge does not simply lie in how far AI can advance, but rather in humanity's ability to guide, control, and use technology in pursuit of appropriate objectives. Technological progress is truly meaningful only when it contributes to enhancing human capabilities, dignity, and quality of life, rather than undermining autonomy or increasing social inequality. Therefore, examining the philosophical issues arising from the development of AI not only contributes to theoretical understanding but also provides practical guidance for establishing a responsible, human-centered, and sustainable model of technological development in the current context of Vietnam.

REFERENCES

[1]. Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the dangers of stochastic parrots: Can language models be too big? In *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency (FAccT '21)* (pp. 610–623). ACM.

[2]. Bostrom, N. (2014). *Superintelligence: Paths, dangers, strategies.* Oxford University Press.

[3]. Chalmers, D. J. (1995). Facing up to the problem of consciousness. *Journal of Consciousness Studies, 2*(3), 200–219.

[4]. Communist Party of Vietnam. (2024). *Resolution No. 57-NQ/TW dated December 22, 2024, on breakthroughs in the development of science and technology, innovation, and national digital transformation.*

[5]. Crawford, K. (2021). *Atlas of AI: Power, politics, and the planetary costs of artificial intelligence.* Yale University Press.

[6]. European Parliament and Council of the European Union. (2024). *Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act).* Official Journal of the European Union.

[7]. Floridi, L., & Cowls, J. (2019). A unified framework of five principles for AI in society. *Harvard Data Science Review, 1*(1). https://doi.org/10.1162/99608f92.8cd550d1

[8]. Ministry of Education and Training. (2021). *Textbook of Marxist-Leninist Philosophy (for undergraduate students not majoring in political theory).* Hanoi: National Political Publishing House.

[9]. Prime Minister of Vietnam. (2021). *Decision No. 127/QD-TTg dated January 26, 2021, approving the National Strategy for Research, Development, and Application of Artificial Intelligence until 2030.*

[10]. Russell, S. (2019). *Human compatible: Artificial intelligence and the problem of control.* Viking.

[11]. Russell, S., & Norvig, P. (2021). *Artificial intelligence: A modern approach* (4th ed.). Pearson.

[12]. Searle, J. R. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences, 3*(3), 417–457.

[13]. Turing, A. M. (1950). Computing machinery and intelligence. *Mind, 59*(236), 433–460.

[14]. UNESCO. (2021). *Recommendation on the ethics of artificial intelligence.* UNESCO.