ARTIFICIAL INTELLIGENCE AND ITS EPISTEMOLOGICAL STATUS

Kozhokaru N.I. (Russian Federation)

Kozhokaru Natalia Igorevna — Senior Lecturer
FACULTY OF PHILOSOPHY
STATE ACADEMIC UNIVERSITY FOR THE HUMANITIES,
MOSCOW

Abstract: *The rise of artificial intelligence (AI) as a technological and cognitive phenomenon challenges the foundations of classical epistemology. This article examines AI not merely as a computational artifact, but as a potential epistemic agent whose operations intersect with, mimic, or disrupt established models of knowledge. By analyzing the compatibility and tensions between AI systems and the major epistemological paradigms—rationalism, empiricism, and constructivism—we attempt to clarify the ambiguous status of AI within the framework of contemporary theory of knowledge. Special attention is given to the ontological and phenomenological constraints of machine cognition and the implications for our understanding of truth, justification, and belief.*
Keywords*: artificial intelligence, epistemology, knowledge, rationalism, empiricism, constructivism, machine learning, cognitive agency, epistemic subject, understanding.*

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ЕГО ЭПИСТЕМОЛОГИЧЕСКИЙ СТАТУС

Кожокару Н.И. (Российская Федерация)

Кожокару Наталия Игоревна — старший преподаватель
философский факультет
Государственный академический университет гуманитарных наук,
г. Москва

Аннотация: *рост искусственного интеллекта (ИИ) как технологического и когнитивного феномена бросает вызов основам классической эпистемологии. В этой статье ИИ рассматривается не просто как вычислительный артефакт, но как потенциальный эпистемический агент, операции которого пересекаются, имитируют или нарушают устоявшиеся модели знания. Анализируя совместимость и противоречия между системами ИИ и основными эпистемологическими парадигмами — рационализмом, эмпиризмом и конструктивизмом — мы пытаемся прояснить неоднозначный статус ИИ в рамках современной теории познания. Особое внимание уделяется онтологическим и феноменологическим ограничениям машинного познания и последствиям для нашего понимания истины, обоснования и веры.*
Ключевые слова: *искусственный интеллект, эпистемология, знание, рационализм, эмпиризм, конструктивизм, машинное обучение, когнитивное агентство, эпистемический субъект, понимание.*

Introduction: From Cognitive Simulation to Epistemic Agency

Contemporary artificial intelligence has evolved beyond algorithmic automatism and now demonstrates complex capacities such as pattern recognition, adaptive learning, and even autonomous decision-making. These abilities invite a reconsideration of AI's role in the broader epistemological landscape. Is AI merely an extension of human cognitive tools, or does it possess features that elevate it to the status of an epistemic agent?

Within classical epistemology, knowledge has traditionally been defined as *justified true belief*, following Plato's seminal formulation, and further systematized by thinkers such as Descartes, Locke, and Kant. Yet AI complicates this triad: it produces outcomes that function as knowledge without possessing beliefs or engaging in justification in the human sense. As Floridi has argued, AI should be considered an "informational agent" rather than a knowing subject. This distinction foregrounds the need to interrogate the epistemological implications of non-conscious cognitive systems.

Weak and Strong AI: Cognitive Capabilities without Consciousness Philosophical discourse on AI often distinguishes between "weak AI"—systems designed for narrow tasks like language processing or visual recognition—and "strong AI", a hypothetical construct characterized by general intelligence and self-awareness. While strong AI remains speculative, weak AI is already embedded in knowledge-generating processes.

Examples abound: large language models (e.g., GPT, BERT) generalize data in ways reminiscent of inductive reasoning; AlphaFold predicts protein folding patterns with superhuman accuracy; AlphaGo demonstrated unprecedented strategic behavior in the game of Go, outmaneuvering human champions using novel moves unanticipated even by experts.

These performances raise pressing questions: Can such outputs be called *knowledge* if their origins are not understood by either the machine or its human users? Does the absence of intentionality disqualify machine-generated results from epistemic consideration?

AI and Classical Epistemological Paradigms AI's behavior resonates with several foundational theories of knowledge, though with important qualifications:

Rationalism. Descartes and Leibniz located the source of knowledge in reason and innate structures. Many expert systems and rule-based AI models are structured around deductive reasoning, employing formal logic to derive conclusions from given axioms. In this sense, AI echoes rationalist methodologies, albeit without internal consciousness or intuition.

Empiricism. Empiricism, articulated by Locke and Hume, grounds knowledge in sensory experience. AI trained via deep learning operates on similar principles, extracting patterns from large datasets—a process akin to the acquisition of knowledge through repeated exposure and statistical association.

Constructivism. Constructivist epistemologies (e.g., Kant, Piaget, Vygotsky) emphasize the active construction of knowledge through interaction with the environment. Reinforcement learning in AI—a paradigm wherein agents learn by trial and error—can be viewed as a mechanistic analog to constructivist processes[7]. However, this construction lacks the self-reflexivity and meaning-making capacity central to human cognition.

Epistemic Limitations: Knowledge without Understanding. Despite these parallels, key epistemological limitations persist. AI lacks beliefs, intentionality, and self-understanding—attributes traditionally considered necessary for genuine knowledge acquisition. As John Searle's Chinese Room argument illustrates, syntactic manipulation of symbols does not equate to semantic understanding.

Moreover, AI systems do not construct hypotheses or critically evaluate their outputs. They operate within probabilistic frameworks, optimizing for outcomes rather than truth. This raises epistemic concerns: if AI-generated conclusions cannot be justified within an interpretive framework accessible to human rationality, can they be epistemologically valid?

Conclusion: The Ambiguous Epistemological Status of AI. AI technologies challenge the anthropocentric boundaries of epistemology. While current AI systems perform cognitive functions with unprecedented efficiency, their lack of intentional states, normative reasoning, and self-awareness restricts their epistemic status. They may generate data and even produce truths, but without belief or understanding, they do not fulfill traditional criteria for knowledge.

Thus, AI occupies an epistemologically ambiguous position: it is a powerful tool of cognitive extension but not yet a cognitive *subject*. The ongoing task of philosophy is to refine our conceptual frameworks to adequately address this new form of intelligence—neither wholly alien nor entirely familiar.

References / Список литературы

1. Plato. (1992). Theaetetus (M. J. Levett, Trans.; M. Burnyeat, Rev. ed.). Hackett.
2. *Floridi L.* (2011). The philosophy of information. Oxford University Press.
3. *Russell S., & Norvig P.* (2021). Artificial intelligence: A modern approach (4th ed.). Pearson.
4. *Silver D., Huang A., Maddison C.J., Guez A., Sifre L., Van Den Driessche G. & Hassabis D.* (2016). Mastering the game of Go with deep neural networks and tree search. Nature, 529(7587), 484–489. https://doi.org/10.1038/nature16961
5. *Nilsson N.J.* (1998). Artificial intelligence: A new synthesis. Morgan Kaufmann.
6. *Lake B.M., Ullman T.D., Tenenbaum J.B. & Gershman S.J.* (2017). Building machines that learn and think like people. Behavioral and Brain Sciences, 40, e253. https://doi.org/10.1017/S0140525X16001837
7. *Sutton R.S. & Barto A.G.* (2018). Reinforcement learning: An introduction (2nd ed.). MIT Press.
8. *Bender E.M. & Koller A.* (2020). Climbing towards NLU: On meaning, form, and understanding in the age of data. Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics, 5185–5198.
9. *Searle J.R.* (1980). Minds, brains, and programs. Behavioral and Brain Sciences, 3(3), 417–457. https://doi.org/10.1017/S0140525X00005756