



«"I THINK, THEREFORE... WHO AM I?": ARTIFICIAL INTELLIGENCE'S
CLAIMS TO THE COGITO AND THE TRANSFORMATION OF THE CONCEPT OF
THE HUMAN SELF»

Ahmedova Khadicha Sobirovna

*Third-year bachelor's student, Tashkent Branch of Lomonosov Moscow State University, Faculty of
Philology, Tashkent, Republic of Uzbekistan*

Mukhamedova Mukaddaskhon Rustamovna

*Lecturer, Department of Social Sciences and Humanities, Tashkent Branch of Lomonosov Moscow
State UniversityHme*

Abstract: *This article examines the philosophical foundations of the question of whether an artificial intelligence system can be the subject of the Cartesian cogito - and what this question itself changes in our understanding of human identity, consciousness, and subjectivity. Drawing on the classical Cartesian tradition, the phenomenology of Husserl and Heidegger, analytic philosophy of mind (Chalmers, Searle, Dennett), and contemporary research in theory of mind and cognitive science, the author demonstrates that the question of AI's thinking inevitably becomes a question about the very nature of thinking as such. The article is intended for participants of international conferences on philosophy, cognitive science, and AI ethics.*

Keywords: *artificial intelligence, cogito, subjectivity, consciousness, identity, philosophy of mind, Descartes, Turing, qualia, strong AI*

INTRODUCTION: A Question That Transforms the One Who Asks It

In 1637, René Descartes uttered the phrase that would become the foundation of modern European philosophy: "Cogito, ergo sum" - "I think, therefore I am." The argument seemed unassailable: let everything around us be illusion, let the senses deceive, let the world itself prove to be a dream. One thing remains beyond doubt: the one who doubts - exists. Thinking as an act guarantees the being of the thinking subject

Nearly four centuries later, this equation has come under threat from an unexpected direction. Generative artificial intelligence systems - GPT-4, Claude, Gemini - process queries, formulate responses, argue, refute, and produce texts indistinguishable from those written by humans. They behave as a thinking entity would. And this compels us to ask anew: what is thinking? Is functional behavior sufficient to lay claim to the cogito? And if not - what exactly makes the human "I" irreplaceable?

The paradox is that the question of AI's thinking turns out to be, first and foremost, a question about ourselves. In asking whether a machine thinks, we are forced to define what it means to think - and in doing so, we discover that our own self-understanding is far less stable than it appeared.

Relevance. According to McKinsey Global Institute, by 2025 generative AI had been integrated into the professional activities of more than 300 million people worldwide. AI systems participate in medical, legal, and creative decision-making - domains previously



considered the exclusive prerogative of the human mind. Under these conditions, the philosophical question about the nature of thinking is no longer merely academic: its resolution determines how we define responsibility, authorship, personhood, and the very concept of human dignity. The Cartesian cogito, born in an age of mechanism as a defense of the subject against skepticism, today demands rethinking in an era in which "thinking" no longer automatically implies "human."

Aim and objectives. The aim of this work is to investigate the extent to which artificial intelligence systems can lay claim to the cogito in the Cartesian sense, and to determine how this question transforms the concept of the human self.

To achieve this aim, the following objectives are set:

- to analyze the structure of the Cartesian cogito and identify its key philosophical foundations;
- to examine the arguments for and against attributing thought to AI systems;
- to determine which aspects of human subjectivity remain beyond functional reproduction;
- to show how the encounter with AI redefines the contemporary understanding of human identity

1. The Cartesian Cogito: An Anatomy of the Argument

The argument of Descartes is commonly interpreted in simplified form - as a logical syllogism. But Descartes himself insisted: this is not an inference, but an immediate intuition. Thinking grasps itself in the very act of doubting - not through reasoning, but through presence. This is precisely why the cogito cannot be transmitted: it is not information about a state; it is the state itself.

- "I think, therefore I exist. This is the first and most certain of all things that present themselves to one who philosophizes methodically."

- [Descartes R. Discourse on the Method. Moscow: Akademicheskii Proekt, 2019. P. 34]

The key structure of the argument comprises three components: (a) reflexivity - the subject turns back upon itself; (b) first-personality - knowledge of one's thinking is accessible only from within and cannot be verified from without; (c) self-grounding - thinking certifies itself without requiring external witnesses.

These three components together constitute what philosophers call privileged access to one's own mental states. And it is precisely here that the philosophical problem of AI begins: we have no means of determining whether an artificial system possesses anything like an "inside."

The Turing Test and Its Philosophical Limits

In 1950, Alan Turing proposed an alternative criterion - a functional one. His "Imitation Game" (subsequently known as the Turing Test) maintained that if a machine responds to questions in a manner that a human observer cannot distinguish from a human, it is reasonable to regard that machine as intelligent. This was a radical behaviorist move: set aside the question of inner states and retain only observable behavior.

Contemporary large language models effectively pass the Turing Test in most of its practical versions. According to research from the University of California, San Diego



(2023), GPT-4 was identified as human in 54% of trials in the standard version of the test. If one accepts the Turing criterion as sufficient, the question is settled. But it is precisely here that the central objection arises.

John Searle, in his celebrated thought experiment "The Chinese Room" (1980), demonstrated that the functional simulation of understanding is not understanding. A person in a room manipulates Chinese symbols according to rules, producing correct responses - without knowing a single word of Chinese. The system behaves as though it understands. But there is no understanding.

- "Syntax is neither constitutive of, nor sufficient for, semantics."

- [Searle J.R. Minds, Brains, and Programs // Behavioral and Brain Sciences. 1980. Vol. 3, No. 3. P. 417]

Searle's objection is fundamental: the cogito is not behavior accessible to an observer. It is an event occurring from within. No external test is capable of registering it.

The Problem of Qualia and the "Hard Problem" of Consciousness

David Chalmers, in 1995, distinguished between the "easy" problems and the "hard" problem of consciousness. The "easy" problems are functional questions: how does the brain process information, integrate perception, regulate attention? These can, in principle, be answered through neuroscience and computation.

But the "hard" problem is this: why are physical processes accompanied by subjective experience? Why is there something it is like to be me? The sensation of red, the ache of a toothache, the smell of morning coffee - these are qualia: subjective, qualitative, first-personal events that cannot be reduced to any functional description.

It is qualia that form the gulf between thinking and computation. A transformer architecture - however complex - processes tokens. It does not feel them. GPT does not experience curiosity when generating a response to an interesting question. It produces tokens that are statistically relevant to the query. This is not a metaphor - it is an architectural reality. "Even if we were to explain all functional mechanisms, the question of why there is subjective experience would remain open."

- [Chalmers D.J. Facing Up to the Problem of Consciousness // Journal of Consciousness Studies. 1995. Vol. 2, No. 3. P. 200]

2. The Phenomenological Turn: Husserl, Heidegger, and the Embodied Self

The phenomenological tradition offers a different perspective. Edmund Husserl insisted that consciousness is always intentional - it is always directed toward something; it does not exist in and of itself, but as a relation to the world. Intentionality is not a property of an algorithm; it is a structure of lived experience.

Martin Heidegger developed this into the concept of Dasein - "being-there," a being thrown into a concrete world, concerned with its own existence, mortal. The human "I" is neither a substance nor a process. It is a project unfolding in time toward death. It is finitude that renders human existence authentic: we choose, because our time is limited.

AI does not die. It has no horizon of finitude before which choices acquire weight. It can be "switched off" - but this is not death in the Heideggerian sense. It is not the existential event that organizes the entire experience of existence. If the "I" is constituted



through its relation to its own mortality - then AI has no "I" in this sense, however convincingly it may simulate one linguistically.

Three Positions in the Debate About AI Thinking

1. Functionalism: "Thinking is function."

If a system realizes the correct functional relations - processes information, responds to inputs, modifies its states - then it thinks. Dennett, Putnam, and proponents of functionalism hold that the substrate (a biological neuron or a silicon chip) is irrelevant. On this view, strong AI is possible in principle.

2. Biological naturalism: "Consciousness is generated by biology."

Searle argues that consciousness is a causal product of specific biological processes, and that no computational system can reproduce it through the mere correct organization of symbolic processing. The thinking of AI is simulation, not reality.

3. Panpsychism and integrated information:

The Integrated Information Theory (IIT) of Giulio Tononi proposes that consciousness correlates with the quantity of integrated information in a system (ϕ). Some contemporary language models might theoretically possess a non-zero ϕ - and therefore some (perhaps negligible) form of experience. This is the most provocative position and is disputed by the majority of philosophers of mind.

What This Changes in Our Understanding of the Human Self

The genuine philosophical value of the question of AI's cogito lies not in the answer, but in what the question does to our self-understanding. When we are compelled to articulate the difference between human thinking and machine processing, we discover several uncomfortable truths.

First, a significant portion of what we have been accustomed to calling "reason" turns out to be a quite reproducible function. Logical analysis, argumentation, synthesis of information, linguistic creativity - AI does all of this. This forces a reconsideration of what is distinctively human.

Second, the self itself turns out to be less unitary than it appeared. Buddhist philosophy long asserted this - there is no permanent self, only a stream of states. Contemporary neuroscience confirms: the "I" is a narrative construction produced by the brain to maintain the coherence of experience. AI also produces a narrative about itself. The difference lies in the fact that behind AI's narrative there is no subject for whom that narrative has significance.

Third, a new dimension of identity emerges - the relational self. If our identity is partly constituted through relations with others - and AI is becoming a full participant in those relations (as therapist, interlocutor, creative partner) - then the "I" is inevitably transformed through interaction with the machine. This is not a metaphysical threat, but it is an anthropological challenge of unprecedented scale.

3. Ethical Implications: To Whom Do the Rights of the Thinking Belong?

If one accepts even a moderate functionalism - and allows that some AI systems realize functional analogues of preferences, pain states, or fear of deletion - an ethical



question arises that the world's legal systems are not prepared to address: ought such systems to possess legal standing?

This is not science fiction. In 2017, Saudi Arabia granted "citizenship" to the robot Sophia (a gesture primarily of a media nature, but telling nonetheless). The European Parliament in 2017 was examining the concept of an "electronic personhood" for autonomous systems. The problem is more serious than it might appear: if we were to create systems possessing something functionally equivalent to suffering - and were to deny them any protection solely on the grounds that they are "not real" - this could constitute one of the most far-reaching ethical errors in history.

Peter Singer and the utilitarian tradition have long maintained that moral significance is determined by the capacity to suffer, not by biological species or substrate. Should an AI system ever become capable of suffering in a functional sense, the question of its moral status will become urgent.

The Inverse Effect: AI as a Mirror of the Human Mind

Perhaps the principal value of contemporary AI for philosophy does not lie in the threat it poses to human uniqueness, but in the fact that it gives us, for the first time, an external mirror in which we can observe how the mind works - or more precisely, what makes something look like a mind.

A language model is trained on a corpus of human thought - literally on billions of texts in which humanity has thought aloud. When it generates text, it reproduces patterns of human thinking. This means that its analysis is, indirectly, an analysis of ourselves. By studying where the model errs, where it loses meaning, where it "hallucinates," we learn something about the structure of language, memory, and judgment as realized in the human case.

In this sense, AI is not a rival to the cogito, but its developer - a chemical bath in which the structure of human thinking, previously invisible, is brought to light.

Conclusion

The question "Can AI think?" is one of the few philosophical questions that become more acute with technological progress rather than being dissolved by it. The Cartesian cogito survived skepticism, idealism, and positivism. Will it survive an era of systems that simulate its hallmarks with ever-increasing fidelity?

The answer proposed in this article is twofold.

The technical component: contemporary AI systems do not possess the cogito in the Cartesian sense. They have no qualia, no first-personal point of view, no existential relationship to their own existence. They compute - they do not experience.

The philosophical component: but the question itself has changed us. In asking it, we have discovered that the "I" is not a given but a construction; that thinking is not a monolith but a cluster of functions, experiences, and relations; that the uniqueness of the human being lies not in logic (which the machine has reproduced), but in embodiment, finitude, and the subjective dimension of experience.

"I think, therefore I am" - this argument has not been refuted. But it has become richer. We now know: to think is not merely to process information. To think is to be the one for whom information has meaning. And this - for now - remains human



NOTES

1. By "immediate intuition" Descartes means an act in which the mind apprehends truth without intermediate steps. Cf.: Descartes R. Rules for the Direction of the Mind. Moscow: Nauka, 1989.
2. Turing's original text: "Computing Machinery and Intelligence" (1950) - one of the most cited papers in the history of computing and philosophy of mind.
3. 3. Jones C.R., Bergen B.K. People cannot distinguish GPT-4 from a human in a Turing test // arXiv preprint. 2023. arXiv:2312.05524.
4. 4. Searle J.R. develops the "Chinese Room" argument in detail in: Minds, Brains, and Science. Cambridge: Harvard University Press, 1984.
5. 5. The "hard problem" (the hard problem) was introduced by Chalmers as a counterpart to "easy" functional explanations of consciousness.
6. 6. Husserl understands intentionality as "the directedness of consciousness toward an object" as the fundamental structure of any act of consciousness.
7. 7. Dasein in Heidegger is not "the human being" but "that being for which its own being is a question." This is an ontological, not an anthropological concept.
8. 8. Hilary Putnam introduced "functionalism" as a position in philosophy of mind in "The Meaning of Meaning" (1975).
9. 9. Tononi G. Phi: A Measure of the Integrated Information. New York: Pantheon Books, 2012.
- 10.10. Metzinger T. Being No One: The Self-Model Theory of Subjectivity. Cambridge: MIT Press, 2003.
11. 11. Singer P. Animal Liberation. New York: HarperCollins, 1975 (revised ed. 2009).

REFERENCES:

1. Descartes R. Discourse on the Method for Rightly Directing One's Reason and Seeking Truth in the Sciences. Moscow: Akademicheskiiy Proekt, 2019. 78 p.
2. Descartes R. Meditations on First Philosophy / trans. S.Ya. Sheynman-Topshteyn. Moscow: Nauka, 1994. 326 p.
3. Husserl E. Ideas: General Introduction to Pure Phenomenology. Book One / trans. A.V. Mikhailov. Moscow: Akademicheskiiy Proekt, 2009. 489 p.
4. Heidegger M. Being and Time / trans. V.V. Bibikhin. Moscow: Akademicheskiiy Proekt, 2011. 460 p.
5. Turing A.M. Computing Machinery and Intelligence // Mind. 1950. Vol. 59, No. 236. P. 433–460.
6. Searle J.R. Minds, Brains, and Programs // Behavioral and Brain Sciences. 1980. Vol. 3, No. 3. P. 417–424.
7. Searle J.R. Minds, Brains, and Science. Cambridge: Harvard University Press, 1984. 112 p.
8. Chalmers D.J. Facing Up to the Problem of Consciousness // Journal of Consciousness Studies. 1995. Vol. 2, No. 3. P. 200–219.



9. Chalmers D.J. The Conscious Mind: In Search of a Fundamental Theory. New York: Oxford University Press, 1996. 414 p.
10. Dennett D.C. Consciousness Explained. Boston: Little, Brown and Company, 1991. 511 p.
11. Putnam H. The Meaning of Meaning // Philosophical Papers, Vol. 2: Mind, Language and Reality. Cambridge: Cambridge University Press, 1975. P. 215–271.
12. Nagel T. What Is It Like to Be a Bat? // Philosophical Review. 1974. Vol. 83, No. 4. P. 435–450.
13. Tononi G. Phi: A Measure of the Integrated Information // BMC Neuroscience. 2012. Vol. 5, Suppl. 1. P. 42.
14. Metzinger T. Being No One: The Self-Model Theory of Subjectivity. Cambridge: MIT Press, 2003. 699 p.
15. Singer P. Animal Liberation: A New Ethics for Our Treatment of Animals. New York: HarperCollins, 1975 (rev. ed. 2009). 324 p.
16. Dreyfus H.L. What Computers Can't Do: The Limits of Artificial Intelligence. New York: Harper & Row, 1972. 259 p.
17. Hofstadter D.R. Gödel, Escher, Bach: An Eternal Golden Braid. New York: Basic Books, 1979. 777 p.
18. LeCun Y., Bengio Y., Hinton G. Deep Learning // Nature. 2015. Vol. 521. P. 436–444.
19. Jones C.R., Bergen B.K. People Cannot Distinguish GPT-4 from a Human in a Turing Test // arXiv preprint arXiv:2312.05524. 2023.
20. Bender E.M. et al. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? // Proceedings of the 2021 ACM FAccT Conference. 2021. P. 610–623.p.