Отримана 31.10.2025
Прорецензована: 14.11.2025
Прийнята до друку: 24.11.2025
Електронна адреса: i.kostrov@khai.edu
DOI: http://doi.org/10.25264/2312-7112-2025-29-67-72

Kostrov I. Philosophical studies of H. Dreyfus on artificial intelligence. *Наукові записки Національного університету «Острозька академія»: серія «Філософія»*: *науковий журнал.* Острог : Вид-во НаУОА, 2025. № 29. С. 67–72.

*УДК 1:004.8*

***Illia Kostrov***

## PHILOSOPHICAL STUDIES OF H. DREYFUS ON ARTIFICIAL INTELLIGENCE

*The analysis conducted in the article shows that the question of the ability of artificial intellectual systems to reproduce thinking, outlined by A. Turing in the middle of the 20th century, is not limited only by practical implementation, but rather exacerbates fundamental methodological, epistemological and ontological problems that concern the very understanding of the nature and structure of human intelligence and knowledge. This necessitates raising the question of the definition of the very concept of «intelligence». For this purpose, the author of the article appeals to the ancient heritage, namely, to the multi-valued term «Logos», introduced by Heraclitus. Heraclitus distinguishes two definitions of Logos: «Logos-word» and «Logos-law». Logos as a «word» is a sign, the name of an object that can be formalized and introduced as an element of a formal system, which will find its embodiment in the creation of formal symbolic systems (Aristotle's logic). Such an understanding of Logos focuses on the analytically discrete properties of thinking. In Dreyfus's terminology, «Logos-word» is designated as «Knowledge-That». At the same time, the understanding of Logos as a «law», «fate» of things, which Dreyfus calls «Knowledge-How», emphasizes the ability of human thinking to operate with holistic images, that is, to its associative and analogical properties. «Knowledge-How» is manifested in the ability to operate with «images» – holistic representations of what constitutes the essence of an object or phenomenon for us in a certain situation, turns to its conceptual, «background» understanding. H. Dreyfus's critical analysis demonstrates the failure of reducing human thinking to formal systems in which knowledge is understood exclusively as the manipulation of symbols. His distinction between the concepts of «Knowledge-That» and «Knowledge-How» allows us to more deeply understand aspects of human mental activity and to comprehend the possibilities of modern intellectual systems, which, although capable of depicting analytical thinking, are unable to reproduce the contextual meanings of human thinking.*

***Keywords:*** *artificial intelligent systems», «Logos-word», «Logos-law», Hubert Dreyfus, «Knowledge-That», «Knowledge-How».*

***Ілля Костров***

## ФІЛОСОФСЬКІ ДОСЛІДЖЕННЯ Г. ДРЕЙФУСА ЩОДО ШТУЧНОГО ІНТЕЛЕКТУ

*Проведений у статті аналіз свідчить, що питання про здатність штучних інтелектуальних систем до відтворення мислення, окреслене ще А. Тюрінгом у середині XX ст., не зводиться лише до практичної реалізації, а радше загострює фундаментальні методологічні, епістемологічні та онтологічні проблеми, які стосуються самого розуміння природи та структури людського інтелекту та знання. Саме це спонукає звернутися до проблеми визначення поняття «інтелект». З цією метою автор статті звертається до античної спадщини, а саме, до багатозначного терміну «Логос», що впроваджується Гераклітом. Геракліт відокремлює два визначення Логосу: «Логос-слово» та «Логос-закон». Логос як «слово» – це позначка, назва предмета, яке можна формалізувати та ввести як елемент формальної системи, що знайде своє втілення у створенні формальних символічних систем (логіка Арістотеля). Таке розуміння Логосу зосереджує на аналітично дискретних властивостях мислення. У термінології Дрейфуса «Логос-слово» позначено як «Знання-Що». Водночас розуміння Логосу як «закону», «долі» речей, що Дрейфус називає «Знання-Щодо», акцентує увагу на здатності людського мислення оперувати цілісними образами, тобто, до його асоціативних та аналогових властивостей. «Знання-Щодо» виявляється у вмінні оперувати «образами» – цілісними уявленнями про те, що становить сутність предмету або явища для нас у певній ситуації, обертає до його концептуального, «фонового» розуміння. Критичний аналіз Г. Дрейфуса демонструє неспроможність редукції людського мислення до формальних систем, у яких знання розуміється виключно як маніпуляція символами. Розмежування ним понять «Знання-Що» та «Знання-Щодо», дозволяють глибше зрозуміти аспекти людської розумової діяльності та осмислити можливості сучасних інтелектуальних систем, які хоча й здатні зображати аналітичне мислення, однак не спроможні відтворити контекстуальні сенси людського мислення.*

***Ключові слова:*** *«штучні інтелектуальні системи», «Логос-слово», «Логос-закон», Г'юберт Дрейфус, «Знання-Що», «Знання-Щодо».*

The problem of artificial intelligence has become a subject of discussion among specialists from various fields of philosophy, science and culture. The specificity of the philosophical approach to the study of artificial intelligence is the understanding of its essential nature, possibilities and limits of use in view of the

integrity, value and harmonious development of the individual, society and the world as a whole. One of the main philosophical questions is the task of finding out to what extent it can be likened to human (natural) intelligence. A feature of the philosophical comprehension of reality is the desire to determine such essential features of the object of reflection that determine the solution of problematic issues. The beginning of the study of artificial intelligence should be preceded by the question of what «intelligence» is in general, that is, we must investigate our own thinking and determine what it is, what its structure is, in order to imply this on the capabilities of computer systems.

In the modern scientific space, the issues of artificial intelligence and intelligent systems attract significant attention from researchers from various fields of knowledge, being actively discussed both at the theoretical and methodological and empirical and practical levels. The issues of the capabilities of computer systems were considered in their works by A. Turing, H. Dreyfus, D. Chalmers, D. Dennett, J. Searle, M. Minsky, A. Newell, S. Garnad, Paul and Patricia Churchland, T. Nagel and others. Among Ukrainian researchers studying the philosophical foundations of artificial intelligence, it is worth noting V. Shtanko, A. Synytsia, K. Reichert, O. Dobrovolska, Yu. Melnyk, O. Polishchuk, etc. At the same time, the issues of philosophical understanding of the capabilities and limits of artificial intelligent systems, raised in the works of H. Dreyfus, remained without sufficient attention from scientists.

The aim of the article is to study the understanding of «intellect« in ancient philosophy, in particular, the concept of «Logos» in Heraclitus and the involvement of H. Dreyfus' ideas on the analysis of hypothetical assumptions of the theoretical foundations of artificial intelligence and the concepts of «Knowledge-That» and «Knowledge-How», which he developed based on ancient philosophical definitions of the concept of «intelligence», is noteworthy.

Philosophers of antiquity – from Heraclitus to Aristotle – tried in various ways to comprehend the nature and uniqueness of human (natural) thinking and cognition, explored the possibilities of reproducing their structures and forms (in particular Aristotle), which closely echoes the modern problems of formalizing language (Wittgenstein) and developing artificial (digitized) intellectual systems. However, despite the significant contribution of thinkers of the past to this field, M. Heidegger in the middle of the twentieth century noted that no era was so ignorant in fundamental issues of human essence, being and thinking as the modern one.

The starting point for considering the issues of modern digital artificial intelligence was the works of Alan Turing, among which the most famous is his concept of the Turing Test. The main conclusion of this concept is that if a system is able to demonstrate behavior that an external observer cannot distinguish from the behavior of a natural intelligent system, then such a system has the properties of intelligence. The Turing Test became a key stage in the formation of the theoretical foundations of artificial intelligence and a significant milestone in the history of its development. Although the model proposed by Turing was later subjected to significant criticism from the scientific community, it continues to retain significant theoretical and practical influence nowadays. Objections to the Turing Test contributed to the formation of new approaches to understanding the nature of intelligence and helped outline the boundaries of the use of machine imitation of human behavior within artificial intelligence. Despite the fact that the Turing Test is over seventy-five years old, its critical understanding remains relevant today. Turing's problem of creating a digital system capable of reproducing human intellectual activity has found practical application both in numerous modifications of the test itself (for example, Stephen Garnad's «Turing Test Hierarchy») and in the development of new programs such as Joseph Weizenbaum's «ELIZA» (1966) and Kenneth Kolb's «PARRY» (1972). In addition, since 1990, the annual Loebner Prize has been established, awarded for the creation of a program that best imitates human thinking.

The first to initiate the discussion around the Turing Test was Alan Turing himself in his work «Computing Machinery and Intelligence». In it, he outlined a number of critical remarks and counterarguments that differed from his own position, and also offered his own, sometimes ironic, responses to them. A generalized presentation of these considerations is given in the table [6, p. 52]:

The critical remarks made by Turing in his article initiated a wide range of discussions – from purely technical ones, concerning the development of computer machines and programs, to philosophical and anthropological ones, within which the question of the limitations of our understanding of the essence, nature, principles and structures of our own, human thinking is repeatedly raised.

| Remarks | Turing's answers |
|---|---|
| Theological objection – a question about the metaphysical essence of man: thinking is a function of the soul, how to deal with it? | We must appeal to factual knowledge |
| The objection of «heads in the sand»: machine thinking is a source of danger and risks, and therefore it is better not to even start studying it | We must be careful, but scientific and technological progress can hardly be stopped |
| Mathematical objection about the fundamental incompleteness and limitedness of formalization (Gödel's theorems confirm this) | However, the Cartesian slogan about the limitlessness of the human mind has not been proven either |
| Argument from consciousness: machines are unlikely to feel, experience | Over time, this issue will also be resolved |
| Argument from various inabilities: a computer cannot be kind, inventive, with a sense of humor, etc. | Such characteristics are our interpretations. By functionally developing machines, everything will be possible to do as in the case of humans |
| Lady Lovelace's objection: machines will only do what they are told, not something new | One day they will be told to think for themselves |
| Argument from informal behavior: people are not programmed in advance | Complex algorithms of machine behavior will make their behavior unpredictable and incomprehensible to us |
| Argument from a continuous nervous system: the nervous system is continuous, and the computer is discrete | At very small values, this will not even be noticeable |
| Argument from extrasensory perception: what to do with telepathy, clairvoyance, foresight, telekinesis? | Perhaps, by generating answers, it will be possible to approach the frequency of guesses by psychics. It is also necessary to better investigate the nature of psychokinetic energy |

One of the first critics of the idea that it was possible to draw clear analogies between the functioning of computers and the activity of the human brain was the American philosopher Hubert Dreyfus, a graduate of Harvard University, where he received his Ph.D. While working at the Massachusetts Institute of Technology, Dreyfus gained recognition as one of the leading researchers of the phenomenology of E. Husserl, M. Heidegger, as well as the philosophical ideas of M. Foucault and other thinkers.

Dreyfus conducts a detailed analysis of Turing's arguments regarding the possibilities and limits of artificial intelligent systems. In his research, he draws on philosophical concepts of cognition and thinking developed by G. Leibniz, D. Hume, I. Kant, and other thinkers, who, in turn, were based on philosophical ideas of antiquity.

It is worth noting that etymologically the term «intellect« has a Latin origin, which was reflected in most European languages, whose peoples primarily used Latin in the scientific community. Even more interesting results are obtained from the analysis of the search for equivalents of the Latin word «intellectus» in different languages. As Alain de Libera notes in the «European Dictionary of Philosophy», in the Ukrainian language this term has such semantic equivalents as «ability to understand», «meaning», «sense».

The term «intellect» itself was formed and became widespread in the medieval period as an attempt to find a counterpart to Greek concepts, among which the term «Logos» is of particular importance.

The doctrine of Logos occupies a central place in the philosophy of Heraclitus. Having created a dialectical model of the world and emphasizing its variability («you cannot step into the same river twice»), constant formation («everything flows, everything changes») and the uncertainty of the existence of things, he simultaneously outlines a certain «limit« of this variability – Logos. Heraclitus interprets Logos as «word», «law», «teaching» or «fate». It is the Logos, that is, the word (title, name), that determines the inner essence of things, which ensures their identity with themselves despite external changes and the fluidity of being.

Heraclitus distinguishes two opposite meanings of Logos: first, as a general rational law that determines the rhythm of continuous changes and mutual transitions of things, ensuring the stability and harmony of the world (principle, reason, basis); second, as Logos in the world of people – a word that they use according to their understanding and will. Because of this, according to the philosopher, most people do not comprehend the true meaning of the «eternal being» Logos, which governs the world. Later, Logos as an ontological,

form-creating essence of being becomes a central concept in Stoic philosophy, where it is interpreted as «creative fire», divine reason and universal expediency, which orders both the cosmos and man. In a similar sense, the concept of Logos is used by Philo of Alexandria, who, using the method of exegesis to interpret sacred Jewish texts (which later became part of the European-Christian Old Testament of the Bible), reveals their content through the system of concepts of ancient Greek philosophy. God in his concept appears as a transcendent Monad, which in the process of creating the world requires a Mediator – the Logos, which is understood as God's word, mind, wisdom and the law of the existence of the universe.

Thus, Logos, which is one of the Greek etymological equivalents of the term «intellect», is understood as «formed» thought. In Aristotle's logic, Logos takes the form of a «concept» – a primary form of thinking that unites objects into classes and distinguishes their essential features. At the same time, it can be assumed that Logos as a «law» or «fate» of things and phenomena appears as a holistic image of an object, thanks to which we recognize it. This idea finds its embodiment in the ancient tradition, in particular in Plato's teaching about ideas that appear as mental (thought) patterns. In the subsequent philosophical understanding of the nature and abilities of the human mind, this property is called «background knowledge», which plays an important role in modern philosophical research on the problems of artificial intelligence.

Over time, the term «intellectus» acquires a wide range of meanings, among which it is worth highlighting such as «ratio», «conception», «Erkenntnisvermögen» german for «the ability to know», «comprehendere» german for «comprehension», «notion», «Bedeutung» german for «meaning», «Sinn» german for «meaning», etc. In the process of developing philosophical thought, these semantic shades expanded, intertwined and interacted in different contexts. Over time, the terms of one language became part of another, and attempts to find linguistic equivalents only complicated the already complex system of concepts.

In his book «What Computers Can't Do» (1972) and the article «Alchemy and AI» (1965), H. Dreyfus raises the question: what exactly in human mental activity cannot be imitated? Thus, he addresses the problem of determining the essence of human mental abilities, that is, the question of the nature of intelligence. Dreyfus emphasizes the role of unconscious (intuitive) thought processes, which, precisely because of their «unconsciousness», are not amenable to formalization, and therefore cannot be reproduced or modeled.

H. Dreyfus identifies four main scientific and philosophical assumptions – biological, psychological, epistemological and ontological – which, in his opinion, have become the basis for researchers' confidence that human intelligence can be reduced to the manipulation of symbols. He emphasizes that each of these assumptions should be considered not as a proven truth, but as a hypothesis, which within the framework of artificial intelligence research is mistakenly perceived as an axiom – that is, an a priori truth that supposedly guarantees the success of creating real intelligent systems. Let us consider the biological and psychological assumptions in which H. Dreyfus addresses the features of human thinking, namely, the combination of discrete and analog processes in it.

According to the biological hypothesis [3, p. 71–74], it is believed that the human brain processes information through discrete operations, similar to the biological analogue of the «on/off» mechanism. Hubert Dreyfus questions this assumption, noting that brain activity is not exclusively discrete. Based on the results of neurophysiological studies, in particular by Professor Walter A. Rosenblith of the Massachusetts Institute of Technology, he draws attention to the existence of analog components in the temporal and dynamic parameters of neural activity. Thus, according to the conclusion of H. Dreyfus, the presence of not only discrete, but also analog processes in the brain refutes the axiomatic nature of the biological assumption.

H. Dreyfus begins his consideration of the psychological hypothesis with a question-precaution: «Whether the brain works as a digital computer is a purely empirical question that neurophysiology must resolve; however, it is impossible to give an unambiguous answer to a related but completely different question: whether the mind *functions* (as we have highlighted) as a digital computer, that is, whether the use of a computer model in psychology is justified» [3, p. 75]. The philosopher emphasizes that a significant part of our knowledge about the world is not reduced to mechanical operation with already available information, but is based on complex attitudes, intentions, and inclinations that determine the ways of interpreting situations.

Even Plato in his dialogues, in particular in «Euthyphro» («On Piety»), raised the issue of analyzing the structures of thinking, formulating it as follows: despite the fact that human thinking and behavior have a rational structure based on rules and «pragmatic layers», is it necessary to be aware of these rules in order to think or act intelligently? H. Dreyfus develops this problem, noting that even when we use familiar, everyday concepts and judgments, they function within an epistemological context that is not always realized by the subject. He defines this context as the sphere of so-called «background knowledge» – knowledge that is not

formalized in specific, clearly defined images, but is a necessary condition for thinking. In subsequent works, Dreyfus designates these two types of mental activity by the terms «*Knowledge*-That» and «*Knowledge*-How».

H. Dreyfus emphasized that human problem-solving and professional competence are not based on a formal analysis of all possible combinations to find the right solution, but on a *contextual* understanding of the situation – that is, on an intuitive feeling of what is essential or appropriate in specific conditions. He explained this difference through the distinction between «Knowledge-That» and «Knowledge-How».

«*Knowledge*-That» is knowledge about the essence of an object or phenomenon, its properties, signs and qualities. According to I. Kant, such knowledge forms the basis of analytical judgments; they reflect our conscious, consistent ability to solve problems by decomposing the complex into simple elements or determining the characteristics inherent in the object itself. We use this type of thinking when we are faced with complex tasks that require stopping, analyzing and gradually reviewing ideas. According to H. Dreyfus, this is a simple, formalized form of mental activity, similar to combining information according to a certain algorithm. Such knowledge acquires a clear structure in the form of judgments, symbols and formulas devoid of context, which we operate with the help of logic and language. This is the same Logos-word in Heraclitus, which designates things, gives them a name and functions in the human world as a tool of cognition without necessarily understanding the deep meaning. This is also the Logos that Aristotle transforms into the basic element of a formal logical system – the «concept». Dreyfus recognizes that artificial intellectual systems are able to successfully imitate «Knowledge-That», because it reproduces the discrete, formally structured aspect of human thinking.

Instead, «*Knowledge*-How» is, first of all, an associative form of mental activity, a way of direct human interaction with the world around us. Such knowledge does not require clear verbalization or contextualization and often does not reduce to a system of judgments. «*Knowledge*-How» is manifested in familiar, everyday «images» – holistic representations of what an object or phenomenon is for us in a certain situation. We act without resorting to conscious symbolic reasoning, when, for example, we recognize an object, perform familiar actions, make intuitive decisions. In such cases, we seem to «move» directly to an adequate reaction, without analyzing all possible alternatives – we «grasp» the object in its contextual, background affiliation. This, according to H. Dreyfus, constitutes the essence and uniqueness of human perception: the ability to understand what an object is before it is consciously perceived, that is, to perceive meaning without relying on previously formulated knowledge or rules.

Thus, according to H. Dreyfus, human thinking is based not only on discrete «Knowledge-That», but also on analog «Knowledge-How», on holistic images, intuitions, attitudes, the formation of which is influenced by many factors that cannot be unambiguously determined, and therefore formalized. This «context» or «background knowledge» is, in the terminology of I. Kant, the «transcendental» conditions of any knowledge. This echoes the ambiguous term of ancient philosophy for the intellect, namely, with Logos.

In the philosophy of Heraclitus, two different understandings of it were already introduced. Logos is a word, a sign, a name of an object, which can be formalized and introduced as an element in a formal system (Aristotle's formal logic). At the same time, the Logos-word contains within itself a set of properties that make up a given object, which can be abstracted (separated) from it and synthesized in a new object. This is exactly what Dreyfus designates as Knowledge-That. In return, the Logos-law, «fate» or «limit of variability» reflects the holistic image of the object, its ideal mental essence, which ensures the identity of the object with itself. It is thanks to this holistic figurative Logos that the human mind «recognizes» previously unfamiliar objects, classifies them, identifies them with others. Dreyfus defines this ability of our thinking as Knowledge-How, that is, «background», «contextual».

In conclusion, Dreyfus emphasizes that Knowledge-That, as analytically discrete, can be reproduced in computer systems. This is the operation (manipulation) of information. In contrast, Knowledge-How, which operates on holistic images of objects, is analog, associative thinking, an appeal to the meaning and inner essence of an object, which cannot be formalized, and therefore reproduced in artificial intellectual systems.

**Література:**

*1. Антологія сучасної філософії науки, або усмішка ASIMO* / за наук. ред. В. Мельника та А. Синиці. Львів: ЛНУ імені Івана Франка, 2007. 568 с.

2. Більчук Н. Л., Костров І. А. *Межі імітації: штучний інтелект між «знанням-що» та «знанням-як»*. Вісник ХНУ імені В.Н. Каразіна. Серія «Філософія. Філософські перипетії». Вип. 72, 2025, С. 128-138

3. Dreyfus H. L. *What Computers Can't Do: A Critique of Artificial Reason*. New York : Harper & Row, 1972. 231 p.

4. Dreyfus H. L., Dreyfus S. E. *Mind over Machine: The Power of Human Intuition and Expertise in the Era of the Computer*. New York : Free Press, 1986. 240 p.

5. Райхерт К. Г'юберт Л. Дрейфус проти Деніела К. Деннета у фільмі «Робокоп». *Філософія та політологія в контексті сучасної культури*. 2017. № 3. С. 56–64.

6. Синиця А. *Вступ до філософії штучного інтелекту: підручник*. Львів: ЛНУ імені Івана Франка, 2023. 292 с.

7. Turing A. M. *Computing Machinery and Intelligence*. *Mind*. 1950. 49. P. 433–460. https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf

8. Searle J. R. Is the brain's mind a computer program? *Scientific American. ф*1990. Nr 262. P. 26-31. https://www.cs.princeton.edu/courses/archive/spr06/cos116/Is_The_Brains_Mind_A_Computer_Program.pdf