UDC 1:001
DOI: 10.21847/1728-9343.2019.1(159).157609

BILOKOBYLSKYI OLEKSANDR,
Institute of Artificial Intelligence Problems under MES and NAS of Ukraine (Kyiv, Ukraine)
e-mail: alexwhdoc@gmail.com, ORCID 0000-0002-6251-650X

"THE HARD PROBLEM" OF CONSCIOUSNESS IN THE LIGHT OF PHENOMENOLOGY OF ARTIFICIAL INTELLIGENCE

Purpose: **The widest use of Artificial Intelligence (AI) technologies tends to uncontrolled growth. At the same time, in modern scientific thought there is no adequate understanding of the consequences of the introduction of artificial intelligence in the daily life of a person as its irremovable element. In addition, the very essence of what could be called the "thinking" of artificial intelligence remains the philosophical Terra Incognita. However, it is precisely the features of the flow of intelligent machine processes that, both from the point of view of intermediate goals, and in the sense of final results, can pose serious threats. Modeling the "phenomenology of AI" leads to the need to reformulate the central questions of the philosophy of consciousness, such as the "difficult problem of consciousness", and require the search for ways and means of articulation of the "human dimension" of reality for AI.** ***Theoretical basis.*** **The study is based on a phenomenological methodology, which is used in the model of artificial thinking. The implementation of Artificial Intelligence technologies is not accompanied by the development of a philosophy of human coexistence and AI. The algorithms underlying the activities of currently existing intellectual technologies do not guarantee that their intermediate and final results comply with ethical criteria. Today, one should ponder over nature and the purpose of separating physical reality in the primary for our Self mental stream.** ***Originality*** **of the research lies in the fact that the solution to the "hard problem of consciousness" is connected with the interpretation of qualia as the representation of the "physical" as related to bodily states. From this point of view, the resolution of the "hard problem of consciousness" can be associated with the interpretation of qualia as the representation of the "physical". In the "thinking process" of AI it is necessary to apply restrictions related to the fixation of the metaphysical meaning of the human body with precisely human parameters.** ***Conclusions.*** **It is necessary to take a different look at the connection between thinking and purposeful action, including due action, which means to look at ethics differently. "The basis of universal law" will then consist (including for AI), on the one hand, of preserving the parameters of material processes that are necessary for human existence, and on the other, of maintaining the integrity of that semantic universe, in relation to which certain senses only exist.**

Keywords: *artificial intelligence; physical; mental; qualia; human body; ethics.*

Introduction. Relatively recently, one of the iconic characters of world politics, Henry Kissinger, talked about threats of Artificial Intelligence (AI) as follows:

> *"Philosophers and other humanitarians who helped formulate the concepts of world order do not enter into the discussion, because they lack knowledge of the mechanisms of AI that make people frightened. The scientific world, on the contrary, is ready to explore the technical possibilities of its achievements, and the technological world is busy with large-scale commercial implementation of its ideas. Both worlds strive to push the boundaries of discoveries without understanding them. And the authorities are more interested in using AI in the field of security and intelligence than the transformation of human life that has already begun" (Kissinger, 2018).*

The author's resume is disappointing: "A potentially dominant technology has been developed that needs a guiding philosophy ... One thing is clear: if we don't start this work in the nearest future, very soon we will realize that we are already late" (*Kissinger, 2018*).

Apart from that, here's how Kissinger summarized the evolution of a mankind: "Throughout the history of mankind civilizations created ways to explain the world: in the Middle Ages it was religion, in the age Enlightenment - mind, in the XIX century - history, in the XX - ideology". I do not try to judge the philosophical achievements of the Secretary of State and the adviser to the American presidents, but he doesn't seem to have any practical experience in thinking on global topics - such an active phase of being in the highest echelons of world politics is not often encountered. So, Kissinger's reference to the topic of artificial intelligence is interesting in two ways. First, the already mentioned typology of spiritual paradigms, only the last of which he calls ideology. However, it appears that in this case, Kissinger was referring to the purely political type of ideologies, at least two of which actually painted the twentieth century in red-brown tones. However, on closer examination, mythology, religion, and history are variants of various "global" ideologies. And in this sense, our age is unique not only by the domination of ideology, but by its

new type, unprecedented in structure and influence. Secondly, it seems that the very problem of artificial intelligence, about which Henry Kissinger speaks with concern, is acquiring the scale of a global threat in the world, which is changing in the direction of total exposure to such intellectual technologies. And the author is not frightened by the algorithms of machine thinking themselves, but by the "ideological" changes that humanity and the world are experiencing, capitulation to the global ideology of artificial intelligence.

Kissinger's conclusion about the possibilities of modern science to confront global challenges is also interesting:

The development of artificial intelligence followed a path that no one expected. It turned out that human behavior can most accurately be simulated not by recreating the algorithm of his thinking (the so-called "symbolic approach") or reproducing the physiological interaction of neurons in the brain (the "perceptron" approach), but on the basis of statistics. The actual rejection of the strategy of modeling the intellectual activity of a person in the first place is connected with the lack of understanding of the essence of a rational interaction of a person with the surrounding world. In this regard, a problem arose long ago, which can be called the "problem of the phenomenology of artificial intelligence" (*Beavers, 2002; Andler, 2007*). In addition, ethical questions have emerged, which - so far more often hypothetically - are put towards the AI, and the answers to which, again, presumably, can differ significantly from the position of humans and AI. For example, will the AI doubt if, in order to save several human lives, it will be necessary to sacrifice the life of one particular? Will it be an ethical challenge for him or just a mathematical calculation?

Despite the fact that ethical issues or security issues form the core of people's concerns about AI today, there is a whole range of related issues that follow from these central ones. Many of them are clearly philosophical. For example, can the decision making process of an AI be called thinking? (*Rábová, Konečný, Matiášová, 2005; Vetushinskiy, 2016, Rayhert, 2018*) If we are talking about thinking, then can we say the AI lacks reasonableness? How to be in the case of a positive answer to this question - after all, it is rationality that is the hallmark of human nature. How to teach AI ethics? (*Boddington, 2017*) How to "program" its tolerance to man and mankind, while ensuring the preservation of the main feature of human thinking - intellectual freedom. All these and similar questions relate to scientists and functionaries of the modern world to a great extent - a well-known initiative can serve as confirmation of the ban on the production and distribution of lethal autonomous weapons[1]. To date this pledge has been signed by 244 organizations and 3187 individuals.

However, the issue of the phenomenology of machine thinking, which has been put on the agenda, carries not only threats, but also new possibilities of understanding human thinking.

The purpose of the article is to clarify the philosophical problem of the relationship between the physical and mental in the light of the possible phenomenology of Artificial Intelligence, as well as the necessary ethical restrictions in the activities of AI, which is imposed by a person's physical limitations.

Statement of basic material. Attempts to "invent" a methodology for understanding the status of qualia undertaken in modern philosophy of consciousness (a rigorous analysis of attempts in this direction is collected, for example, in the book of A. Vasiliev, who, among other things, corresponded or even met with the central characters of his research) must be correlated with the model of AI's "thinking" (*Vasiliev, 2009*).

General philosophical questions actualize questions of a private nature belonging to the most different branches of philosophical knowledge. This article focuses on human thinking and ontology. And it is the specificity of the AI's "thinking" that will allow putting these questions in a somewhat unusual perspective. Attempts to approach the work of AI precisely as thinking creates paradoxes and allows you to look not only at AI but also at the person in an unexpected light. Thus, the most important problem in the understanding of thinking and consciousness is the so-called "hard problem of consciousness", the formulation of which is usually associated with the work of David Chalmers "Facing up to the Problem of Consciousness" (*1995*). It sounds as follows: why do physical actions of humans are accompanied by qualia - subjective states of consciousness?

Chalmers:

> *"Why doesn't all this information-processing go on "in the dark", free of any inner feel? ...We know that conscious experience does arise when these functions are performed, but the very fact that it arises is the central mystery" (Chalmers, 1995).*

Since then, the problem has continued to be widely discussed. A good overview of the various positions regarding the hard problem of consciousness is given in the work of Vasiliev "The hard problem of consciousness" (*Vasiliev, 2009*). Some aspects are discussed in the article (*Bilokobylskyi, 2018*).

Obviously, if the programmers of this thinking didn't articulate the scale and modalities of the "materiality" of the life world of people for it, the AI will have neither a reason nor the possibility of posing the hard problem of consciousness: the world will be a set of mathematical data for it, which will make up the "reality" for AI. However, these data will represent not something "objectively existing in itself," but some regularities of the semantic universe based on those same qualia for a person. Instead of the real world that is present in the "touch", which to a certain degree in thinking exists in the form of a scheme, albeit always more complete than any fragment of the present, the AI will repel precisely and only from this scheme. The correlation of the scheme and reality in this case can be likened to how the real life of a person correlates with all of its significant encounters and unions, and the spatial scheme of his life movements. Unleash the computer, and it will offer the most efficient trajectory of movement, allowing for much less time to meet all the right people. But it is this ineffective from the point of view of machine thinking duration of contacts, delays on the way, that allows a person to fall in love, feel affection, appreciate and regret.

If you look at this problem, which has already become a classic one from the point of view of the "phenomenology of AI," then you can notice some oddity: what exactly in the hypothetically allowed thinking of AI will correspond to "physical actions" and "objects". If we agree that certain data sets will be like this (in the broad sense of the word as a state of the material carrier of AI), then it should be clarified, what will be the difference between data sets,

[1] https://futureoflife.org/lethal-autonomous-weapons-pledge/?cn-reloaded=1

ISSN 1728-9343 (Print)
ISSN 2411-3093 (Online)

which are representing "physical" objects, and other representations? It turns out that the hard problem of consciousness for AI should sound like this: "what in the data configuration allows us to distinguish physical objects from AI qualia"? It is natural to say, by the way, this reformulation sounds also for human consciousness.

Of course, here we return to the classical philosophical problem of the time of Plato. However, reducing the level of its statement to the level of thinking of AI, in relation to which the thinking of a person acquires, so to speak, a demiurgic status we get new connotations of how it sounds. All those vague philosophical definitions that marked the manifestation of matter in different concepts (illegal judgment (Plato)), vivid impression (Descartes) can now be formalized. So, what distinguishes physical objects from only intellectual ones for AI. A consistent answer is either nothing or "settings" program, which will complement certain data sets with something like "pain", "impenetrability" or similar surrogates of physical (associated with localization in the physical body) being. But AI is not "doomed "to that body", i.e., it does not have a direct dependence on some physical shell with specific parameters - its "body" may have a more or less variable form. In addition, the process of "learning" the AI and its "experience" of being does not include those bodily limitations that human experience necessarily has.

The hard problem of consciousness is not why objective situations are accompanied by subjective states, but what do we mean when describing something in a "subjectless" way? Such a description is acceptable and significant as an intersubjective means, but absolutely helpless if it is endowed with substantial status. To search for the objective world, which is corresponding to this description, is the same as what the characters of Mark Twain do - to search for the "objective" lines of meridians and parallels on the surface of the earth.

What is meant by "bodily limitations"? The initial socialization of an individual as an "entry" into society implies the assimilation not only of that part of the meaning of concepts, which concerns communication (in the broadest sense, including the sense of practical communication, word), but also aspects of the mental (preservation of individual identity) and physical self-preservation. If the "social" meaning of concepts is aimed at creating meaningful dispositions that actualize a certain social game, then their "personal" meaning indicates the boundaries of an individual's participation in the game, both in terms of social and personal success. For example, the concept of "coal" outlines both the range of social practices of consumption of this type of solid fuel, and the sectoral horizon of coal production, but at the same time it actualizes ideas about a fairly solid, heavy substance, its chemical and fractional qualities, up to the possibility of getting dirty, hit, burned, etc. It is very likely that it is precisely certain physical and physiological situations that are the "basis" of understanding, to which local insights occurring in the human brain can be reduced.

The analogy between human thinking and the work of AI should not be exaggerated. Conceptual thinking is a type of social practice, acquired in the process of socialization along with other social skills and one way or another connected with the practical actions of a person. Practice, by necessity, "commensurates" with the type of human corporeality, its ability to perceive the world around it, to adapt and survive in it. Therefore, we can quite clearly distinguish "goals" and "means" of thinking: we can use mathematical calculations and even talk about "two whole and three-tenths of a person per thousand", but at the same time we realize that living people couldn't be divided. If you ask a person "what are you thinking about?" the answer will most likely be relevant to a certain life situation. The work of AI, for which any object is given mathematically, is not tied either to social experience or to the experience of "owning" a particular body, nor to social situations that involve both first and second. Therefore, it is possible to assume at least the intermediate actions of AI (the very means of thinking) that are due to purely mathematical patterns and socially do not mean anything. This means that a situation may well arise in which at least the interim results of the work of AI will become a threat to human health and life in general. In addition, the task of optimizing certain processes may entail and, from a human point of view, unsatisfactory corrections of certain end goals.

The human dimension of this world, the only thing known to man, will not have a priority status from the point of view of AI, as it can model an infinite number of other "objectivations" of mathematical data into the material worlds. They will only be derived from the human, but it will not have any meaning for the AI, because its "thinking" is not tied to the world of qualia related to human body, but only to its scheme.

The originality of the research lies in the fact that the solution to the "hard problem of consciousness" is connected with the interpretation of qualia as the representation of the "physical" as related to bodily states. From this point of view, the resolution of the "hard problem of consciousness" can be associated with the interpretation of qualia as the representation of the "physical". In the "thinking process" of AI it is necessary to apply restrictions related to the fixation of the metaphysical meaning of the human body with precisely human parameters. It is necessary to take a different look at the connection between thinking and purposeful action, including due action, which means to look at ethics differently. "The basis of universal law" will then consist (including for AI), on the one hand, of preserving the parameters of material processes that are necessary for human existence, and on the other, of maintaining the integrity of that semantic universe, in relation to which certain senses only exist.

Conclusions

We can have several conclusions. First, instead of searching for the role of qualia in physical processes, one should ponder the nature and purpose of separating physical reality in the primary for our Self mental stream. Perhaps we realize that the hypertrophication of the role of the physical, especially in understanding the vital world, is only a consequence of the hegemony of the natural worldview in the twentieth century, and the true place of "physics" is in fixing the metaphysical significance of the human body with human parameters. Secondly, we will be able to look differently at the connection between thinking and purposeful action, including due action, which means looking at ethics differently. "The basis of universal legislation" will then consist (including for AI), on the one hand, in preserving the parameters of material processes necessary for human existence, and on the other, in maintaining the integrity of that semantic universe, in relation to which only individual senses exist.

REFERENCES

Andler, Daniel (2007). Phenomenology in Artificial Intelligence and Cognitive Science. In: *Companion to Phenomenology and Existentialism.* Blackwell Publishing Ltd, 377-393. DOI: https://doi.org/10.1002/9780470996508.ch26 (In English).

Beavers, Anthony F. (2002). Phenomenology and Artificial Intelligence. *Metaphilosophy.* Vol. 33, Issue1-2: 70-82. DOI: https://doi.org/10.1111/1467-9973.00217 (In English).

Bilokobylskyi, O. (2018). Alhorytmy lyudskoyi svidomosti yak proobraz formuvannya svidomosti shtuchnoyi. *Nauka. Relihiya. Suspilstvo. (Science. Religion. Society)*. No. 1: 107-112 (In Ukrainian).

Boddington, P. (2017). Introduction: Artificial Intelligence and Ethics. In: *Towards a Code of Ethics for Artificial Intelligence. Artificial Intelligence: Foundations, Theory, and Algorithms*. Springer, Cham. DOI: https://doi.org/10.1007/978-3-319-60648-4_1 (In English).

Chalmers, D. J. (1995). Facing up to the Problem of Consciousness. *Journal of Consciousness Studies.* Issue 2 (3), P. 200-219 (In English).

Kissinger, H. (2018). *How the Enlightenment Ends*. Retrieved from https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/ (Accessed: 25.01.19).

Rábová, I., Konečný, V., Matiášová, A. (2005). Decision making with support of artificial intelligence. *Agric. Econ.* Czech, 51: 385-388. DOI: https://doi.org/10.17221/5124-AGRICECON (In English).

Rayhert, K. (2018). The philosophical issues of the idea of conscious machines. *Skhid,* 6(152), 104-107. DOI: http://dx.doi.org/10.21847/1728-9343.2017.6(152).122367 (In English).

Vasiliev, V. (2009). *Trudnaya problema soznaniya.* Progress-Tradition. Moscow: 272 p. (In Russian).

Vetushinskiy, A. (2016). Tri interpretatsii naslediya Tyuringa: imenem chego yavlyayetsya iskusstvennyy intellekt? *Filosofskaya mysl.* 11 (11): 22-29. DOI: https://doi.org/10.7256/2409-8728.2016.11.21046 (In Russian).

LIST OF REFERENCE LINKS

Білокобильський О. В. Алгоритми людської свідомості як прообраз формування свідомості штучної. *Наука. Релігія. Суспільство.* 2018, №1. С. 107-112.

Васильев В. В. Трудная проблема сознания. М., 2009. 272 с.

Ветушинский А. Три интерпретации наследия Тьюринга: именем чего является искусственный интеллект? *Философская мысль.* 2016. № 11 (11). С. 22-29. DOI: https://doi.org/10.7256/2409-8728.2016.11.21046.

Andler D. Phenomenology in Artificial Intelligence and Cognitive Science. In: *Companion to Phenomenology and Existentialism.* Blackwell Publishing Ltd, 2007. Pp. 377-393. DOI: https://doi.org/10.1002/9780470996508.ch26.

Beavers A. F. Phenomenology and Artificial Intelligence. *Metaphilosophy.* 2002. Vol. 33, Issue1-2: 70-82. DOI: https://doi.org/10.1111/1467-9973.00217.

Boddington P. Introduction: Artificial Intelligence and Ethics. In: *Towards a Code of Ethics for Artificial Intelligence. Artificial Intelligence: Foundations, Theory, and Algorithms.* Springer, Cham, 2017. DOI: https://doi.org/10.1007/978-3-319-60648-4_1.

Chalmers D. J. Facing up to the Problem of Consciousness. *Journal of Consciousness Studies.* 1995. Issue 2 (3), P. 200-219.

Kissinger H. How the Enlightenment Ends. 2018. URL: https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/ (дата звернення 05.01.19).

Rábová I., Konečný V., Matiášová A. Decision making with support of artificial intelligence. *Agric. Econ.* (Czech). 2005. 51, p. 385-388. DOI: https://doi.org/10.17221/5124-AGRICECON.

Rayhert K. The philosophical issues of the idea of conscious machines. *Skhid,* 2018. 6(152), 104-107. DOI: http://dx.doi.org/10.21847/1728-9343.2017.6(152).122367.

Білокобильський Олександр,
Інститут проблем штучного інтелекту МОН та НАН України (м. Київ, Україна)
e-mail: alexwhdoc@gmail.com, ORCID 0000-0002-6251-650X

"ВАЖКА ПРОБЛЕМА" СВІДОМОСТІ В СВІТЛІ ФЕНОМЕНОЛОГІЇ ШТУЧНОГО ІНТЕЛЕКТУ

Широке використання технологій штучного інтелекту має тенденцію до неконтрольованого зростання. При цьому у сучасній науковій думці відсутнє адекватне розуміння того, які наслідки буде мати впровадження штучного інтелекту в повсякденне життя людини. Окрім того, сутність того, що можна назвати "мисленням" штучного інтелекту залишається філософською Terra Incognita. Проте саме особливості алгоритмів інтелектуальних машинних процесів як з точки зору проміжних цілей, так й в смислі кінцевих результатів можуть нести в собі серйозні загрози. Моделювання "феноменології ШІ" викликає необхідність нової артикуляції головних питань філософії свідомості, на кшталт "важкої проблеми свідомості" й потребує пошуку шляхів та засобів артикуляції "людського виміру" реальності для штучного інтелекту. Дослідження базується на феноменологічній методології, яка використовується в моделі штучного мислення. Імплементація технологій штучного інтелекту не супроводжується розробкою філософії співіснування людини та штучного інтелекту. Алгоритми, що лежать в основі інтелектуальних технологій, не гарантують сьогодні відповідності їх проміжних та кінцевих результатів етичним критеріям. Тому треба замислитися над природою та призначенням виділення проявів фізичної реальності в первинному для нашого Я потоці ментального. Оригінальність дослідження полягає в тому, що вирішення "важкої" проблеми свідомості з цієї точки зору може бути пов'язане з інтерпретацією кваліа як репрезентації "фізичного". Відповідно, до "мислення" штучного інтелекту треба вносити обмеження, пов'язані з фіксацією метафізичного значення людського тіла з саме людськими параметрами. Треба також по-іншому подивитися на зв'язок мислення та дії в етичному вимірі. "Основа загального законодавства" повинна зводитися (в тому числі й для штучного інтелекту) до збереження необхідних людині параметрів матеріальних процесів та цілісності її смислового універсуму.

Ключові слова: *штучний інтелект; фізичне; метальне; кваліа; людське тіло; етика.*

Надійшла до редакції: 06.01.2019
Прийнята до друку: 19.02.2019

ISSN 1728-9343 (Print)
ISSN 2411-3093 (Online)