DOI: 10.26565/2226-0994-2023-68-3
УДК 004.8;130.3

**Volodymyr Kuklin**

# ON THE QUESTION OF THE APPEARANCE OF CONSCIOUSNESS IN NEURAL NETWORKS

Revolutionary changes in the intellectual development of mankind are discussed. This is the emergence of a second human signaling system, awareness of the nature of the formation of cognitive abilities of the human brain and the formation of highly intelligent neural networks. The nature of the formation of the second signaling system was determined not only by an increase in the number of neurons and synapses of the cerebral cortex, but also by the emerging possibility of unprecedented interactive communication within the growing number of human communities, based on the development of speech. The painful refusal of neurophysiologists and their colleagues from religious mythical ideas about the nature of consciousness is considered. It became clear that all human intellectual abilities are provided by the work of the brain and other parts of the nervous system. This departure from mystical ideas about consciousness was influenced by the development of neural networks, programming languages, and the work of neurophysiologists together with neurosurgeons. The development of artificial neural networks was facilitated not only by the efforts of neurophysiologists and their colleagues from related fields of science, but, above all, by unprecedented technical progress in the creation of high-speed computing tools with incredible amounts of memory. Analogies of the recent revolution in the development of artificial intelligence systems and the nature of the emergence of a second signaling system in humans are discussed. The development of revolutionary trends in the process of creating multiparameter neural networks made it possible to expect that a detailed study of this process will make it possible to understand the nature of the appearance of the second signaling system in higher animals. The work shows analogies between these two phenomena. The possibilities of the neural network assessing its internal state and the ability to realize its position in the external environment, which are key characteristics of self-identity and consciousness, are discussed.

***Key words:*** *human second signaling system, cognitive revolutions, self-identity and consciousness of neural networks.*

## 1. INTRODUCTION

The problem of consciousness and the organization of human thinking processes has worried scientists since ancient times. Since most scientists were believers, it was not strange for them that the human brain and his nervous system provide control of all organs, but thoughts, as they imagined, were connected with the spiritual, non-material sphere. Therefore, until the middle of the 20th century, despite the enormous amount of information about the human body and, in particular, about the brain, even serious neurobiologists and neurophysiologists did not part with this idea. Since the spirit was responsible for the thought process, even the nature of the appearance of intelligent beings on the planet (the emergence of the second signal system) was not discussed; perhaps the widespread idea of the origin of the world, instilled in childhood from childhood, was to blame. It took a detailed study of the brain, involvement of mathematicians in thinking on this topic, and most importantly, the first attempts to create connectivist models of the brain from artificial neurons in order to dispel the ethereal fog of spiritual thinking. Indeed, a small number of neurons in the model turned out to be capable of finding answers, that is, generating a thought. There was no longer room for the disembodied spirit. It made an impression.

People sobered up from the religious dope began the process of rapidly creating and mastering brain models - neural networks. As usual, from a complete denial of the possibility of describing intelligence by systems with real elements (here these are neurons with synapses) they quickly moved on to intentions to create a thinking machine that would be on par with the brain of a thinking person. It took less than a century for neural networks to appear, the intellectual capabilities of which were equal to the capabilities of human intelligence, and in some important cases even exceeded these human capabilities. Now, for human society, worried about the appearance of an unexpected competitor, other problems have arisen. To what extent are the emerging neural networks able to compete with humans and what advantages will humans retain? These possible remaining priority qualities and abilities for people included the presence of consciousness. But opinion of humans now denies neural networks the possession of this consciousness. If we believe that consciousness in its simplest form comes down to understanding one's internal state and the ability to be aware of one's position in the external environment, then it seems unclear whether neural networks are capable of coping with the task of forming one's identity. But something tells us that if neural networks, with the help of people, have come such a long way in their intellectual development, surpassing humans in many cases, then they are unlikely to stop there and will not develop their consciousness on the same scale as people. The purpose of the work is to consider changes in the nature of the intellectual development of civilization in the last two centuries and discuss the question of whether neural networks will be able to assess their identity in the near future and whether we can expect them to display manifestations of consciousness similar to human ones.

## 2. COGNITIVE REVOLUTIONS

The first cognitive revolution is undoubtedly the appearance of a second signaling system on the planet. The reasons for the appearance of a second signaling system in humans have always interested humanity. There have been many attempts to explain this phenomenon. It is clear that it was the growth in the volume of the cerebral cortex that allowed the emergence of intelligence in a human being. They also note the possibility of influence on the formation of the second signal system by the manifestation of percolation (passage) of signals in the general array of the cortex.

In addition, new communication opportunities due to the formation of articulate speech provided an unprecedented volume of contacts between individuals. The number of human communities began to far exceed the number of the absolute majority of packs of other higher animals. In addition, communication made it possible to better educate children and disseminate information. Extensive connections between numerous community members dramatically accelerated the spread of information.

The nature of the appearance of the mind in humans until the middle of the 20th century was explained by the existence of little-understood supernatural forces, a certain spirit operating outside the brain. Moreover, the role of the brain and nervous system in controlling the body's systems was already recognized by this time. It is not necessary to even say that this state of affairs completely suited the churchmen, who already agreed with the descriptions of all the intricacies of the activity of the nervous system, but left open the possibility of the existence of spiritual forces in the manifestations of the mind. It must be said that the abundance of varying degrees of beliefs among representatives of the scientific community also created a split in their ranks and a painful breaking of the prevailing ideas was required that nervous activity and the process of thinking are processes of different natures.

Cognitive revolution of neuropsychology in the mid-20th century. The pre-revolutionary state of mind was characterized by a struggle against pragmatism and behaviorism. Pragmatism was based on experimental research and supported the "relationship of stimuli and response" scheme. These approaches treated the brain as a black box.

The desire to look inside this box arose among cognitive psychologists. This field of psychology originated in the 1950s through the efforts of Allen Newell, Herbert Simon, Noam Chomsky, Donald Hebb and George Miller (who defected from the behaviorist camp) and a number of other neuropsychologists who refused to perceive the brain as a black box. Back in 1949, D. Hebb's book *The organization of behavior, Neuropsychological Theory* [Hebb, 1949] was published. D. Hebb drew attention to the ability of neurons to cooperate with the formation of a single process "neurons that fire together are connected." He also drew attention to the fact that the brain is always excited. Experiments helped to study the brain, in particular the Canadians W. Penfield and G. Jasper (see [Posner, 1978]), who stimulated the brain with electrodes, studied the reactions of patients and, based on this, removed parts of the cerebral cortex to stop seizures. Psychologist D. Hebb also worked in this group. He was trying to figure out the consequences of brain injuries. It was believed that the emergence of languages like IPL and the first developments of artificial intelligence stimulated the process of revolution in brain science, but so far this did not concern answers to questions about what consciousness is. It is amazing that the defector J. Miller from Harvard became the personification of this cognitive revolution and, at his own suggestion, many froze research to find answers to the questions of what consciousness is [Miller]. Nevertheless, scientists continued to work on this problem. In 1964, the school "Brain and Conscious Experience" under the auspices of the Pontifical Academy of Sciences and under the leadership of J. Eccles, who was famous for discovering the chemical side of information transmission in synapses. But at the same time, he and many of his colleagues continued to see some other supernatural nature in the structures of the mind and memory, although they recognized that experience is formed in the patterns of neurons and through the activation of synapses. The meeting was never able to answer J. Eccles' question about how exactly neural activity in the cerebral cortex provokes sensory experience. Reaction of G.-L. Teubera was more categorical: "When we tried to outline probable systems and mechanisms without which there would be no consciousness, the opinions expressed were very different and contradictory. There was not even confidence... regarding understanding why consciousness is However, having freed himself from the influence of religiosity and having gone through the realization of his mistakes, R. Sperry, later after this meeting, formulated the conclusions that formed the basis of the cognitive revolution: the process of consciousness is impossible without the brain, and mental factors are associated with it, which are consistent with the phenomena of subjective experience, but do not imply disembodied supernatural forces operating outside the brain mechanism [Sperry, 1970]. Useful were the considerations of J. Eckle and the philosopher K. Popper, that thought can influence the state of the brain, and T. Nagel, that this state can correspond to external influence [Nagel, 2000]. For neuropsychologists, the opinion of J. von Neumann that the nervous system functions in conditions of parallelism was also important, especially since the structure of the brain, as it turned out, consists of neural clusters oriented to different functions with strong internal short connections and weak connections among themselves [Gazzaniga, 2018; Cline, Mouret, Lipson, 2013]. Thus, all nervous processes, including mental ones, were finally attributed to brain activity, although exactly how this was implemented in an array of neurons remained unclear.

***The nature of the cognitive revolution of artificial intelligence.*** Created in the image and likeness of the cerebral cortex, the neural networks that emerged at the end of the last century were small (insignificant number of layers, number of neurons, number of synapses-connections). During this period, the creators of such devices were mainly neurophysiologists and partly mathematicians. In the early 2000s, developers realized that they needed to move away from narrow application areas of artificial intelligence (narrow AI) and begin to use knowledge and skills across different contexts and disciplines. New technologies and new architectural solutions have appeared. The technical progress that accompanied or even provoked this was supported by 1) the development of the so-called matrix notation, as well as 2) the formation of a third level of representation of software solutions using technology libraries. We must also take into account that all these events took place in the context of the rapid development of the information revolution, a significant reduction in the cost of equipment and software.

The massive emergence of graphic processors, a significant increase in computing speed, explosive growth in memory capacity, multiple multiplication of the number of neurons, connections and layers in neural networks against the backdrop of growing instrumental power, ensured the success of this direction - the creation of artificial intelligence, which could significantly expand the scope of perceived knowledge. The volumes of information mastered by these giant neural networks have become so significant that they are talking about their encyclopedicity (multimodality).

But the main result of revolutionary transformations in the artificial intelligence industry was the discovery of obvious cognitive abilities of new neural networks - Large language Models of the GPT series [OpenAI, 2023]. These LLM cognitive abilities have become closer to human abilities compared to the capabilities of previous narrow or low-parameter AI models of the same and other series. When considering the abilities of the latest multi-parametric LLM models, we had to return more than once to attempts to determine intelligence (see, for example, [Linda S Gottfredson, 1997] In addition, now the comparison of the latest neural network models, starting with GPT-4, was no longer focused on the achievements of previous models, but on comparison with human intelligence.

Although growing pains manifested themselves in the weakness of new LLM models. It was not easy for them to learn from their own experience, additional input of information often led to disruptions in their work, and there were difficulties in adapting to a changing environment. The models did not expand the memory capacity well, did not demonstrate confidence, and did not insist on the resulting solution. The model made up ideas if it did not find a direct answer in the training set; hints were often not perceived. The desire of the developers to see the traits of a genius in LLM was in vain; tasks requiring insight - "eureka" - were not solved. The neural network models demonstrated in their decisions errors that existed in the training data (the irrationality of thinking was based on artifacts of human culture). Although new generation LLMs [Bubeckar, 2023] were already able to explain their decision a posteriori. Once again in the history of the development of neural networks, the excessive demands of customers were in conflict with the obvious fact that all these shortcomings were also characteristic of an educated person with average mental abilities.

Useful analogies. One can hope that the phenomena accompanying this AI cognitive revolution, which we are now fortunately able to study in detail, can clarify the nature of the emergence of a second signaling system in humanity. One of the reasons for the emergence of the second signaling system in evolution is most likely the excess of a certain threshold in the volume of the human cerebral cortex. This phenomenon is similar

an increase in the number of neurons and connections between neurons in the network above a threshold value, which was observed in the context of the cognitive revolution of artificial intelligence. And with the formation of the second signaling system and with the emergence of giant neural networks, this growth in the volume of the neural system increased the ability to process large amounts of information and qualitatively changed the intellectual capabilities of these systems. A concomitant and supporting process of development of cognitive abilities with an increase in the number of neurons in humans at one time, and in neural networks in modern times, was immersion in a vast information environment; means of communication and visualization were formed. Noting these analogies, we can with a high degree of confidence assume that the processes of formation of the second signaling system in humans and the phenomena accompanying the revolutionary transition from low-parameter to multi-parameter neural networks are quite similar. Further study and development of neural networks will make it possible to understand not only the features of the cognitive revolution of artificial intelligence, but also the nature of the evolution of the intelligence of civilization as a whole.

## 3. HAVE NEURAL NETWORKS REACHED THE LEVEL OF HUMAN CONSCIOUSNESS?

Many scientists and experts in the field of artificial intelligence recognize the significant progress in the intellectual capabilities of modern giant neural networks, but they doubt the formation of consciousness close to human in such networks. Therefore, below we will dwell on the problem of consciousness from a formal number of position. Note that it is a comparison of the description of human mental activity and the processes that occur deep in neural networks that helps to understand the nature of the intellectual activity of such large systems as the human brain and giant neural networks.

Consciousness and unconsciousness. It is rational to imagine the activity of our cerebral cortex as a set of various intellectual operations, the definition of which we do not yet provide. They occur at different times and many of them are executed simultaneously. Those of them that we are aware of and represent with our inner gaze in a language that is mastered in the process of life (this could be the language of our everyday communication, a language that we learned later, it could be the language of algebra, mathematics, music, etc.), occur as we imagine in our minds. But, as neurosurgeon W. Pensfield noted (see, for example, [Miller, 1962], this becomes possible with a sufficiently high concentration of our attention on these processes.

But we are not aware about of many operations, especially those responsible for the functioning of various life support systems. The body does not consider it necessary to provide us with access to influence them and even to observe them (although persistent people, yogis, for example, try and often unsuccessfully gain control over a number of vital processes, for example, changing the pulse rate, blood pressure, the degree of heating of different parts of the body, relieve pain, etc.).

**Unconscious.** The unconscious, as A. Schopenhauer understood it [Shopenhauer, 1996], that is, very little-conscious impulses and aspirations, is a not entirely true term that could define the internal actions of our brain and nervous system as a whole, hidden from our attention. Practice shows that these internal actions of our brain can include assessments and calculations, structuring information, even drawing conclusions and decisions, that is, everything that a person does consciously, under the control of his consciousness, using hearing and vision, discussing with colleagues, reading and writing down...There are many examples of such intellectual activity hidden from humans in the scientific practice of many scientists. This is similar to how hidden states are formed in neural networks, and how intermediate and final solutions are formed from them.

Therefore, there are different levels of description of consciousness. What is controlled by a person, the fact that he sees images and ideas with his inner gaze, is formulated in the language familiar to him in the internal thought process, this is the external side of consciousness, a kind of interface that allows us to access the deep hidden layers of consciousness and memory. That is, consciousness is a huge "iceberg", where the thoughts and ideas we are aware of roam at its top. We are not aware of what is happening inside this "iceberg".

Therefore, the unknown here is rather unconscious. The concept of 'inaccessible to awareness" is a term that better defines that huge layer of uncontrollable activity of the cerebral cortex, coupled with other sections of the nervous system. Obviously, neural networks also hide many processes from us if we have not bothered to create the necessary interface for observing them. Here, in this aspect of consciousness, there is apparently no fundamental difference between natural and artificial intelligence.

***Self-identity.*** Undoubtedly, awareness of one's Self is due to the presence of memory, which tells us not only different episodes of our life, but also all sorts of interactions with the people, objects, and phenomena around us.

All these numerous contacts and observations that took place created and are creating a web of relationships between our Self and the surrounding community, environment and world. This web records the position of our Self in these environments, allows us to realize who we are and how we are perceived. Therefore, zoologist W. Thorpe believed that consciousness includes "awareness of one's own essence." That is, awareness of our Self is impossible without the environment formed around us. Just as our intellect and global knowledge base were unable to develop outside of society, so our Self could not be formed without the environment we are aware of and its reaction to our existence.

There is the problem of searching for consciousness in modern neural networks that are exceptionally educated and capable of performing intellectual feats. To form their consciousness, their position, relationships with members of the intellectual community must first be formed, then through repeated interactions they and this intellectual community will come to understand their meaning and role.

Sometimes you can hear that artificial intelligence systems solve only those tasks that are assigned to them, do not show initiative and do not have far-reaching intentions in the future, unlike humans. But in these manifestations, artificial intelligence systems are unlikely to differ much from the mass of people who do not show initiative and are inclined to carry out tasks that others and circumstances set before them. Equip a neural network with systems for contact with the outside world in full and give them the means to influence this world and you will be amazed at how quickly they will become like a person.

## REFERENCES

Hebb, D.O. (1949). *The organization of behavior, Neuropsychological Theory*. Wiley, New York.

Posner, M.I. (1978). *Chronometric Exploration of Mind.* / Hillsdail N.j.: Lawrence Erlbaum Associated.

Miller, G.A. (1962). *Psychology: The science of Mental Life*. New York: Harped and Row.

Eccles, J.C. (1964). Brain and Conscious Experience. *Study Week* September 28 to October 4, 19, of the Pontificia Academia Scientiarum.

Sperry, J.C. (1970). Perception in the Absense of theNeocortical Commissures. *Perception and Its Disorders*, 48, pp.123-128.

Nagel, T. (2000). The Psychophysical Nexus. *New Essays on the A Priori*. Oxford: Oxford University press, pp. 432-471.

Gazzaniga, M.S. (2018). *The consciousness instinct: Unraveling the mystery of how the brain makes the mind.* Farrar: Straus and Giroux.

Cline, J., Mouret, J-B., Lipson, H. (2013). *The Evolutionary Origins of Modularity* / Proc. Of the Royal Sosciety of London B.: Biological Sciences 280.

OpenAI. (2023). GPT-4 technical report, 2023. *arXiv preprint arXiv*:2303.08774 [cs.CL] https://doi.org/10.48550/arXiv.2303.08774.

Linda S Gottfredson. (1997). Mainstream science on intelligence: An editorial with 52 signatories, history, and bibliography, 1997.

Bubeckar, S. (2023). Sparks of Artificial General Intelligence: *Early experiments with GPT-4/ arXiv:2303.12712v2* [cs.CL] 24 Mar.

Shopenhauer, A. (1996). *The World as Will and Representation* (1818). New York: Dover, 2t.

**Kuklin Volodymir M.**
Doctor of Physical and Mathematical Sciences, Professor
Head of Department of Artificial Intelligence and Software
4, Maidan Svobody, Kharkiv, Ukraine
E-mail: v.m.kuklin@karazin.ua
ORCID: http://orcid.org/0000-0002-0310-1582

Article arrived: 21.03.2023 Accepted: 10.05.2023

## ДО ПИТАННЯ ПРО ПОЯВУ СВІДОМОСТІ НЕЙРОННИХ МЕРЕЖ

**Куклін Володимир Михайлович**
доктор фізико-математичних наук, професор,
завідувач кафедри штучного інтелекту та програмного забезпечення
Харківський національний університет імені В. Н. Каразіна
майдан Свободи, 4, Харків, 61022, Україна
E-mail: v.m.kuklin@karazin.ua
ORCID: http://orcid.org/0000-0002-0310-1582

## АНОТАЦІЯ

Обговорюються революційні зміни в інтелектуальному розвитку людства. Це поява другої сигнальної системи людини, усвідомлення природи формування когнітивних здібностей людського мозку та формування високоінтелектуальних нейронних мереж. Природу формування другої сигнальної системи було зумовлено як збільшенням числа нейронів і синапів кори мозку, так і можливістю безпрецедентного інтерактивного спілкування всередині зростаючої чисельності людських громад, з урахуванням розвитку промови. Розглянуто болючу відмову нейрофізіологів та їх колег від релігійних міфічних уявлень про природу свідомості. Стало зрозумілим, що всі інтелектуальні здібності людини забезпечені роботою мозку та інших розділів нервової системи. На цей відхід від містичних уявлень про свідомість вплинув розвиток нейронних мереж, мов програмування й діяльність нейрофізіологів спільно з нейрохірургами. Розвитку штучних нейронних мереж сприяли зусилля нейрофізіологів та їх колег із суміжних галузей науки, але передусім небачений технічний прогрес у створенні швидкодіючих обчислювальних засобів, які мають неймовірні обсяги пам'яті. Обговорюються аналогії революції, що виникла останнім часом у розвитку систем штучного інтелекту і природи появи другої сигнальної системи у людини. Розвиток революційних тенденцій у процесі створення багатопараметричних нейронних мереж дав змогу очікувати, що детальне вивчення цього процесу дозволить розібратися в природі появи другої сигнальної системи у вищих тварин. У роботі показані аналогії між цими двома явищами. Обговорюються можливості оцінки нейронною мережею свого внутрішнього стану та здатність усвідомлювати своє становище у зовнішньому оточенні, що є ключовими характеристиками самоідентичності та свідомості.

***Ключові слова:*** *друга сигнальна система людини, когнітивні революції, самоідентичність та свідомість нейронних мереж.*

Стаття надійшла до редакції: 21.03.2023 Схвалено до друку: 10.05.2023