RAYHERT KONSTANTIN,
Candidate of philosophical sciences, associate professor
of the Department of philosophy and methodology of knowledge,
Odessa I. I. Mechnikov National University

THE PHILOSOPHICAL ISSUES OF THE IDEA OF CONSCIOUS MACHINES

The study outlines the existing and potential philosophical issues of the idea of conscious machines originated from the development of artificial consciousness within the framework of contemporary research of artificial intelligence and cognitive robotics. The outline shows that the idea of conscious machines is concerned with two big philosophical issues. The first philosophical issue is a definition of consciousness, taking into account the selection of a set of objects that can have consciousness (human being, living being or machine), the typology of consciousness, the clarifying of the nature of consciousness' carriers and the relationship between consciousness and its environment (including social and cultural) and the relationship between consciousness and language, in order to create an artificial consciousness within a machine, making that machine conscious. The second philosophical issue is a clarification whether only artificially created machines can be conscious machines, or cyborgizated (engineered) human beings also can be considered of conscious machines. These philosophical issues show that there can be two ways to create conscious machines: 1) the creating of artificial consciousness within artificially created machine; and 2) the cyborgization of human being, transforming it into artificially created machine possessed natural consciousness (or even possessed consciousness artificially transformed from natural into artificial).

Key words: artificial consciousness; artificial intelligence; cognitive robotics; conscious machine; cyborg.

The problem statement. Within the framework of contemporary research of artificial intelligence and cognitive robotics, the development of artificial consciousness (also known as machine consciousness and synthetic consciousness) gathers momentum [2; 4; 5; 6; 10; 22; 24]. The efforts of the development of artificial consciousness result the design of machines which possess artificial consciousness, i.e. conscious machines. Basically, the designing of conscious machines takes place at the sci-tech level [8; 11; 12; 19; 20; 21]; philosophical aspect of the idea of conscious machines has being taken in the background, though 'consciousness' is still a philosophical concept which needs elucidation (because there is no yet a common comprehension of the phenomenon of consciousness in philosophy, science, humanities and engineering [25]): without an elucidation of the concept of consciousness it's impossible to elucidate the concept of conscious machine and, probably but not necessarily, to implement of artificial consciousness and create authentic conscious machines in practical situation.

The purpose of this study is to outline some existing and potential philosophical issues of the idea of conscious machines.

Research and results. The great philosophical issue which can affect the idea of conscious machines is the problem of disclosure of the very notion 'consciousness'. The term 'consciousness' was coined by English philosopher and physician John Locke in 1689: he defined 'consciousness' as "the perception of what passes in a man's own mind" [16, p. 98] and indicated the following: "Consciousness always accompanies thinking, and it is that which makes everyone to be what he calls self, and thereby distinguishes himself from all other thinking things, in this alone consists personal identity, i.e. the sameness of a rational being: and as far as this consciousness can be extended backwards to any past action or thought, so far reaches the identity of that person; it is the same self now it was then; and it is by the same self with this present one that now reflects on it, that that action was done" [16, p. 319]. As can be seen, according to John Locke consciousness has the characteristics of the perception of thinking and personal identity (= self). These characteristics of consciousness allow considering a consciousness in terms of anticipation; awareness; cognition and knowledge; emotion, feeling, sensation, thought and volition; imagination and memory; intelligence and mind; intentionality, introspection, phenomenal and subjective experience, qualia and reflexing; language, narrative and speech; learning; wakefulness etc.

But since John Locke the philosophical term 'consciousness' entered the everyday life (and social imaginary) and the domains of knowledge (engineering, humanities, medicine and sciences) and believe (beliefs, religions), becoming an umbrella term having covered a multiplicity of meanings, or as American philosopher of mind Ned Block noted: "The concept of consciousness is a hybrid or better, a mongrel concept: the word 'consciousness' connotes a number of different concepts and denotes a number of different phenomena" [3, p. 375]. Therefore, an urgent issue for philosophy from now on is not only identifying the notion of consciousness but also classifying the existent and potential concepts of consciousness.

There are a few typologies of consciousness in philosophical literature [3; 17]. One of them, for example, was provided by Ned Block. He suggested two concepts (types) of consciousness: (1) phenomenal consciousness (P-consciousness) and (2) access-consciousness (A-consciousness). Phenomenal consciousness is experience ("We have P-conscious states when we see, hear, smell, taste, and have pains" [3, p. 380]) which properties are distinct from any cognitive, intentional or functional property ("Cognitive = essentially involving thought; intentional properties = properties in virtue of which a representation or state is about something; functional properties = e.g., properties definable in terms of a computer program" [3, p. 381]). Phenomenal consciousness deals with qualia (individual instances of phenomenal experience) and, thus, is related to the hard problem of consciousness [7] and the mind-body (psycho-physical) problem [14]. Access-consciousness is the states which "must always be states of consciousness of" [3, p. 383], because the representations of content of the states are functional, i.e. system-relative ("What makes a state A-conscious is what a representation of its content does in a system") [3, p. 383]; simply put, access-consciousness is something through which information in one's mind is accessible for verbal report, reasoning and the control of behaviour.

And now you can imagine that Ned Block's typology of consciousness can be enough to reproduce the consciousness within the machine. However, you could be faced with one problem: metaphorically, at the present level of the development of artificial intelligence, cognitive robotics, computer engineering and science, and cybernetics you already have phenomenal and access consciousness - phenomenal consciousness is computer hardware (the machine's equivalent of experience) and access-consciousness is computer software (the machine's equivalent of something whereby information is accessible for verbal report, reasoning and the control of behaviour). I foresee that this metaphor will be met with hostility by the researchers: probably, they will say that we cannot verify the presence of consciousness within computer system consisted of hardware and software, because there is no yet a conventional criterion for recognizing consciousness within the conscious machine in contemporary philosophy, science, humanity, medicine and engineering. And this criterion still cannot be singled out because today there are too many conceptions of consciousness and there is no single criterion for choosing among them the most correct (true) one. The researchers of consciousness and, respectively, of artificial consciousness and conscious machines can't reach an agreement what to consider consciousness and by which parameters to identify the presence of consciousness: here it is not clear whether consciousness depends of the nature of its carrier; whether culture, society or environment plays any role in the formation of consciousness; whether is language an important thing in the formation of consciousness and if language is important, then whether only natural language forms consciousness, either artificial language can do it too etc. Let's have a look at those closer.

Within the philosophical issue of the nature of carrier of consciousness there can be formulated two philosophical questions: 1. "Can only a biological being be a carrier of consciousness?" 2. "Can an artificial (non-biological) being be also a carrier of consciousness?"

If only a biological being can be a carrier of consciousness, then any artificial consciousness cannot be simulated (or recreated) on a non-biological substrate. This means that if the researchers want to create artificial consciousness they'd have to create the machines based on a biological substrate. And furthermore, the researchers have to answer the following questions: "Do only human beings possess consciousness?" "If the answer is "Yes", then what of it, the brain, the nervous system or the entire biological body of human being, is involved in the emergence of human consciousness?" The answer to this question allows creating an artificial consciousness within the machine created on a biological substrate with the help of biotechnology and biomedical engineering, more precisely, within a comparable part recreated in the machine. More interesting that the answer to this question could entail the strange and fantastic at first glance questions like "Can a conscious machine built on a biological substrate suffer from mental disorders of consciousness (something like machine's Delirium, Oneiroid, Sopor, Amentia, Twilight state etc.)?" or "Can a conscious machine built on a biological substrate take a solipsistic position?"

If the answer is "No" to question "Do only human beings possess consciousness?", then who else among the living beings can possess consciousness, and, more important, are there similarities and differences between human consciousness and consciousness of other living beings?" The answer to this question expands the variety of potential implementations of artificial consciousness within a machine based on a biological substrate.

Thus, in general, if only a biological being can be a carrier of consciousness, then a conscious machine can be built only on a biological base. In such a case a conscious machine might be a biorobot or even a human clone.

But if an artificial being can be also a carrier of consciousness, then any artificial consciousness can be simulated (or recreated) on a non-biological substrate. In this case, there is no need to reproduce the architecture of the brain or nervous system of human or other biological being within the machine; and we again run into the problem of identifying the criterion for the presence of an authentic consciousness in the machine: perhaps, because of the non-biological substrate, we will not be able to recognize consciousness in a machine built on this non-biological substrate. It turns out that at contemporary stage of the development of artificial consciousness research the natural way of creating artificial consciousness is to model the biological substrate in machines to make these machines conscious.

The solution of the philosophical issue of relation between consciousness and language depends on selected conception of consciousness; more precisely, from what whether we think that consciousness is a unique feature of a human being, or other beings can also possess consciousness. If we think that consciousness, at least, is presented in many species of animals then we are faced with the problem of the radically uneven distribution of language [13], because not every species has its own language, and that leads us to the point of view that language doesn't matter for consciousness. But if we bethink that when we aware of, we are thinking and simultaneously we as though are saying within through a natural language, conducting an internal monologue, an internal speech, then we have to admit that language does matter for consciousness. The latter entails, firstly, the question whether language affects consciousness (the Sapir-Whorf hypothesis aka Linguistic relativity) or consciousness affects language or it all depends on situation; and, secondly, the question whether there is only the relationship between natural language and consciousness, or there can be a relationship between artificial language and consciousness either. The latter is problematic, because all artificial languages, except constructed inter-

national auxiliary languages like Esperanto, Ido or Interlingua, are not languages in the strict sense: they are sign systems of the speech fixation metaphorically called 'languages' [1, c. 145]. Nevertheless, we cannot rule out such an opportunity that those sign systems, called 'languages' and participating in communication, cannot affect consciousness or consciousness can use one of those systems instead of natural language. There is one well-known fact that in 2017 two robots (bots), created by Facebook and known as Alice and Boob, developed their own machine language extemporaneously, and that language was incomprehensible for the creators of those robots [23], although it is quite possible that the robots conducted meaningful dialogues in their own language. So, if the language or quasi-language is related to the formation of consciousness, then there is a high probability that robots Alice and Bob, this artificial intelligence created by Facebook, having created their own (machine) (quasi)language, could create their own (machine) consciousness within themselves; but we can't testify it because, firstly, Facebook liquidated those robots, and, secondly, even if we could testify, we would not be able to recognize those robots' consciousness because we couldn't even understand those robots' language.

The philosophical issue of relation between consciousness and language allows us to raise the philosophical issue about relation between consciousness and its environment, including cultural and social: natural language, being a sign system, is a social (and cultural) construction used in transmission of information (communication) and able to affect communicant(s); so, if consciousness uses (or is based on) natural language, then consciousness is depended on that socially (culturally) constructed language which itself is limited in describing the environment. But this all matters only if it recognizes the relationship between consciousness and language.

And, finally, I would like to draw attention to one more philosophical issue: "Can only machines possessed artificial consciousness be considered conscious machines?" This question is not meaningless because of the phenomena of artificial intelligence and cyborg. Well-known that in 1964 American mathematician and computer scientist John McCarthy, who coined the term 'artificial intelligence', started designing technology that had to simulate human capabilities to replace human beings with intelligent machines [18, p. 1], cognitive robots. At the same time American engineer and inventor Douglas Engelbart started creating a computer system that had to 'augment' or extend human capabilities, "rather than to mimic or replace them" [18, p. 1]. Thus, there were two types of cyborgs, hybrids of machine and biological: (1) cyborg as a machine augmented by some biological stuff and (2) cyborg as a biological being (for example, a human being) augmented by some machine stuff. So, we perfectly can imagine the first type of cyborg as a machine possessed an artificial consciousness modeled or not after a natural consciousness, although we can have doubts about the presence of consciousness within that type of cyborg because we can't know if the correct conception of consciousness was used for implementation of artificial consciousness. In case of the second hybrid we can have no doubts in the presence of consciousness within a cyborg because we deal with a cyborgized human being who has already possessed consciousness before the cyborgization. And here is the question: "If human being is a cyborg now, that is, a partly machine, then could that human being be considered as a conscious machine?" Ironically, a human being even now can be considered a

machine thanks to the philosophies of Julien Offray de La Mettrie [15] and Jerry Fodor (who believed that human brain was in its way a computer) [9], for example. Moreover, American technology and science reporter John Markoff points out the paradox: "The paradox is that the same technologies that extend the intellectual power of humans can displace them as well" [18, p. 1]. That means that not only artificially created beings (machines possessed artificial intelligence) can replace human beings in situations where humans' intellectual and physical power is not enough but the very human beings more and more augmenting, extending and replacing their own natural parts can become being indistinguishable from artificially created beings and, weirdly, all of this can make them conscious machines.

Conclusions

Our outline shows that the idea of conscious machines is concerned with two big philosophical issues. The first philosophical issue is a definition of consciousness, taking into account the selection of a set of objects that can have consciousness (human being, living being or machine), the typology of consciousness, the clarifying of the nature of consciousness' carriers and the relationship between consciousness and its environment (including social and cultural) and the relationship between consciousness and language, in order to create an artificial consciousness within a machine, making that machine conscious.

The second philosophical issue is a clarification whether only artificially created machines can be conscious machines, or cyborgized (engineered) human beings also can be considered of conscious machines.

These philosophical issues show that there can be two ways to create conscious machines: 1) the creating of artificial consciousness within artificially created machine; and 2) the cyborgization of human being, transforming it into artificially created machine possessed natural consciousness (or even possessed consciousness artificially transformed from natural into artificial).

REFERENCES

1. Rayhert, K. W. (2016). The Theoretic Loading of the Conceptual Metaphor of Logical Language. *Filosofiia i politologiia v konteksti suchasnoyi kultury*, №2 (11), pp. 144-150 (ukr).

2. Aleksander, I. (1995). *Artificial Neuroconsciousness: An Update*, available at: https://web.archive.org/web/19970302014628/http://www.ee.ic.ac.uk/research/neural/publications/iwann.html. DOI: 10.1007/3-540-59497-3_224.

3. Block, N. (1997). On a Confusion about a Function of Consciousness, in Block N., Flanagan O. and Guzeldere G. (Eds.), *The Nature of Consciousness. Philosophical Debates*. Cambridge, MA: The MIT Press, pp. 375-416.

4. Boyles, R. J. (2012). *Artificial Qualia, Intentional Systems and Machine Consciousness*, available at: https://philpapers.org/archive/BOYAQI.pdf.

5. Buttazzo, G. (2011). Artificial Consciousness: Utopia or Real Possibility? *Computer*, №7, pp. 24-30. DOI: 10.1109/2.933500.

6. Chalmers, D. J. (2011). A Computational Foundation for the Study of Cognition. *Journal of Cognitive Science*, №4, pp. 325-359. DOI: 10.17791/jcs.2011.12.4.325.

7. Chalmers, D. J. (2010). Facing Up to the Problem of Consciousness. Chalmers, D. J. (2010). *The Character of Consciousness*. Oxford; New York: Oxford University Press, pp. 3-27. DOI: 10.1093/acprof:oso/9780195311105.003.0001.

8. Dehaene, S., Lau, H. & Kouider, S. (2017). What is Consciousness, and Could Machines Have It? *Science*, №6362, pp. 486-492. DOI: 10.1126/science.aan8871.

9. Fodor, J. (1975). *The Language of Thought*. New York: Thomas Y. Crowell.

10. Gamez, D. (2008). Progress in Machine Consciousness. *Consciousness and Cognition*, №3, pp. 887-910. DOI: 10.1016/j.concog.2007.04.005.

11. Haikonen, P. O. (2003). *Cognitive Approach to Conscious Machines*. Exeter: Imprint Academic.

12. Haikonen, P. O. (2007). *Robot Brains: Circuits and Systems for Conscious Machines.* Hoboken: Wiley-Interscience.

13. Haladjian, H. H. (2016). Consciousness and Language. *Psychology Today*, available at: https://www.psychologytoday.com/blog/theory-consciousness/201608/consciousness-and-language.

14. Harnad, S. (2000). Correlation vs. Causality: How/Why the Mind/Body Problem is hard. *Journal of Consciousness Studies*, №4, pp. 54-61.

15. La Mettrie, J. O., de (1865). *L'homme machine*. Paris: Frederic Henry.

16. Locke, J. (1999). *An Essay Concerning Human Understanding*. New York: Oxford University Press USA.

17. Lycan, W. (1996). *Consciousness and Experience. Cambridge*, MA: The MIT Press.

18. Markoff, J. (2016). *Machines of Loving Grace: The Quest for Common Ground between Humans and Robots*. New York: Ecco.

19. Neukart, F. (2016). *Reverse Engineering the Mind: Consciously Acting Machines and Accelerated Evolution*. Berlin: Springer. DOI: 10.1007/978-3-658-16176-7.

20. Pitrat, J. (2009). *Artificial Beings: the Conscience of a Conscious Machine*. Hoboken: Wiley-ISTE.

21. Reggia, J. A. (2014). *Conscious Machines: The AI Perspective*, available at: https://www.aaai.org/ocs/index.php/FSS/FSS14/paper/view/9107/9078.

22. Reggia, J. A. (2013). The Rise of Machine Consciousness: Studying Consciousness with Computational Models. *Neural Network*, vol. 44, pp. 112-131. DOI: 10.1016/j.neunet.2013.03.011.

23. Sassoon, L. '(2017). 'Robot Intelligence is Dangerous': Expert's Warning after Facebook AI 'Develop Their Own Language'. *Mirror*, available at: http://www.mirror.co.uk/tech/robot-intelligence-dangerous-experts-warning-10908711.

24. Schlagel, R. H. (1999). Why not Artificial Consciousness or Thought? *Minds and Machines*, №1, pp. 3-28. DOI: 10.1023/A:1008374714117.

25. Velmans, M. (2009). How to Define Consciousness - and How not to Define Consciousness. *Journal of Consciousness Studies*, №5, pp. 139-156.

Райхерт Константин,
кандидат философских наук, доцент кафедры философии и методологии познания,
Одесский национальный университет имени И. И. Мечникова

ФИЛОСОФСКИЕ ПРОБЛЕМЫ ИДЕИ СОЗНАЮЩЕЙ МАШИНЫ

В статье очерчиваются некоторые существующие и возможные философские проблемы идеи сознающей машины, возникшей благодаря исследованиям в области искусственного интеллекта и когнитивной робототехники: 1) проблема определения "сознания", которая, в том числе затрагивает вопросы о множестве предметов, которые может покрывать это понятие, типологии сознаний, прояснении природы носителей сознания, а также отношений между сознанием и языком и сознанием и окружением; 2) проблема прояснения того, могут ли только искусственно созданные машины считаться сознающими машинами, или же сознающими машинами могут считаться также киборгизированные люди. Рассмотренные философские проблемы показывают, что возможны два способа создания сознающих машин: 1) создание искусственного сознания в искусственно созданных машинах и 2) киборгизация человека, превращающая человека в искусственно созданную машину, обладающую естественным сознанием.

Ключевые слова: искусственное сознание; искусственный интеллект; киборг; когнитивная робототехника; сознающая машина.

Райхерт Костянтин,
кандидат філософських наук, доцент кафедри філософії та методології пізнання,
Одеський національний університет імені І. І. Мечникова

ФІЛОСОФСЬКІ ПРОБЛЕМИ ІДЕЇ СВІДОМОЇ МАШИНИ

Стаття окреслює деякі існуючі та можливі філософські проблеми ідеї свідомої машини, які виникають завдяки дослідження штучного інтелекту та когнітивної робототехніки: 1) проблема визначення "свідомості", яка також зважає на питання про сукупність предметів, що охоплює це питання, типологію свідомості, прояснення природи носіїв свідомості, відношення між свідомістю та мовою та свідомістю й оточенням; 2) проблема прояснення того, чи можна лише штучно створені машини вважати свідомими машинами або свідомими машинами можуть бути також кіборгізовані люди. Розглянуті філософські проблеми показують, що можливі два шляхи створення свідомих машин: 1) створення штучної свідомості в штучно створених машинах і 2) кіборгізація людини, яка перетворює людину на штучно створену машину з природною свідомістю.

Ключові слова: кіборг; когнітивна робототехніка; свідома машина; штучна свідомість; штучний інтелект.