

BASIC METHODS AND APPROACHES TO ARTIFICIAL CONSCIOUSNESS MODELING

I.A. PSHENOKOVA

Institute of Computer Science and Problems of Regional Management –
branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences
360000, Russia, Nalchik, 37-a I. Armand street

Annotation. The article provides a brief analysis of some methods and approaches to the modeling of artificial consciousness. Some approaches to the definition of consciousness in philosophy, psychology and neurobiology are presented. In particular, functional and neurobiological models of consciousness are considered. Some approaches to artificial consciousness modeling are presented. The considered models fully satisfy the purpose for which they were created, however, none of them yet shows the real creation of personal preferences acquired and processed through the agent's body and emotions, which are considered the basis for the creation of a potential artificial consciousness. This area of general artificial intelligence is actively developing and there is still no unified theory of the fundamental principles and methods for creating intelligent systems that are conscious, capable of understanding their actions and goals, as well as self-awareness.

Keywords: artificial intelligence, consciousness, artificial consciousness, robotics, simulation

Information about the author

Pshenokova Inna Auesovna, Candidate of Physics and Mathematics sciences, Head of Laboratory «Intellectual Habitats», Institute of Computer Science and Regional Management Problems – branch of Kabardino-Balkarian Scientific Center of the Russian Academy of Sciences;

360000, Russia, Nalchik, 37-a I. Armand street;

pshenokova_inna@mail.ru, ORCID: <https://orcid.org/0000-0003-3394-7682>

СПИСОК ЛИТЕРАТУРЫ / REFERENCES

1. Petukhov V.V. *Lektsii po obshchey psikhologii* [Lectures on General Psychology]. Moscow: Izdatelstvo Moskovskogo universiteta [Moscow University Publishing House], 1997. 597 p. ([In Russian](#))
2. Carruthers P. Higher-order theories of consciousness, in *The Stanford Encyclopedia of Philosophy*, ed. Zalta E. N. Metaphysics Research Lab; Stanford University, CA. 2016.
3. Van Gulick R. Consciousness: in *The Stanford Encyclopedia of Philosophy*, ed. E.N. Zalta (Metaphysics Research Lab, Stanford University). 2018.
4. Anokhin P.K. *Uzlovyye voprosy teorii funktsional'nykh sistem* [Key questions of the theory of functional systems]. Moscow: Nauka, 1980. 203 p. ([In Russian](#))
5. Luria A.R. *Yazyk i soznaniye* [Language and consciousness]. Ed. E.D. Khomskaya. Moscow: Izdatelstvo Moskovskogo universiteta [Moscow University Publishing House], 1979. 320 p. ([In Russian](#))
6. Baars B.J. *In the Theater of Consciousness. The Workspace of the Mind*. Oxford: Oxford University Press. 1997. 88 p.
7. Shanahan M.P. A cognitive architecture that combines internal simulation with a global workspace. *Conscious. Cognit.* 15. 2006. Pp. 433–449. DOI: 10.1016/j.concog.2005.11.005.
8. Franklin S., Madl T., D'Mello S. et al. LIDA: a systems-level architecture for cognition, emotion, and learning. *IEEE Trans. Auton. Ment.* 2014. Vol. 6. No. 1. Pp. 19–41. DOI: 10.1109/TAMD.2013.2277589.
9. Grossberg S. Towards solving the hard problem of consciousness: The varieties of brain resonances and the conscious experiences that they support. *Neural Networks*. 2017. Vol. 87. Pp. 38–95. <https://doi.org/10.1016/j.neunet.2016.11.003>
10. Crick F., Koch C. The problem of consciousness. *Scientific American*. Special edition. 2002. Vol. 12. No. 1. Pp. 11–17.
11. Rosenthal D. *Consciousness and Mind*. Oxford, Clarendon. 2005. 400 p.
12. Seth A.K. Explanatory correlates of consciousness: Theoretical and computational challenges. *Cognitive Computation*. No. 1. 2009. Pp. 50–63.
13. Chalmers D. Facing up to the problem of consciousness. *Journal of Consciousness Studies*. 1995. Vol. 2. No. 3. Pp. 200–219.
14. Dehaene S., Lau H. and Kouider S. What is consciousness, and could machines have it? *Science*. Vol. 358. 2017. Pp. 486–492. DOI: 10.1126/science.aan8871.
15. Tononi G. Consciousness as integrated information: a provisional manifesto. *Biol. Bull.* Vol. 215. 2008. Pp. 216–242. DOI: 10.2307/25470707.
16. Floridi L. Consciousness, agents and the knowledge game. *Mind Mach.* No. 15. 2005. Pp. 415–444. DOI: 10.1007/s11023-005-9005-z.
17. Morin A. Possible links between self-awareness and inner speech. *Journal of Consciousness Studies*. Vol. 12. No. 4-5. 2005. Pp. 115–134.
18. Holland O. A strongly embodied approach to machine consciousness. *Journal of Consciousness Studies*. Vol. 14. No. 7. 2007. Pp. 97–110.
19. Hesslow G. Conscious thought as simulation of behaviour and perception. *Trends Cogn. Sci.* Vol. 1. No. 6. 2002. Pp. 242–247. DOI: 10.1016/S1364-6613(02)01913-7.
20. O'Regan J.K., Noë A. A sensorimotor account of vision visual consciousness. *Behav. Brain Sci.* Vol. 24. No. 5. 2001. Pp. 939–973. DOI: 10.1017/S0140525X01000115.

21. Chella A., Manzotti R. *Artificial Consciousness*. Andrews UK Limited. 2013. 281 p.
22. Esser S., Lustig C., Haider H. What triggers explicit awareness in implicit sequence learning? Implications from theories of consciousness. *Psychological Research*. 2021. <https://doi.org/10.1007/s00426-021-01594-3>
23. Graziano MSA. The Attention Schema Theory: A Foundation for Engineering Artificial Consciousness image. *Front. Robot. AI*. Vol. 4. No. 60. 2017. <https://doi.org/10.3389/frobt.2017.00060>
24. Graziano M.S., Guterstam A., Bio B.J. et al. Toward a standard model of consciousness: Reconciling the attention schema, global workspace, higher-order thought, and illusionist theories. *Cogn. Neuropsychol.* Vol. 37. No. 3-4. 2020. Pp. 155–172. DOI: 10.1080/02643294.2019.1670630.
25. Reggia J.A., Katz G.E., Davis G.P. Humanoid Cognitive Robots That Learn by Imitating: Implications for Consciousness Studies. *Consciousness in Humanoid Robots*. 2018. DOI: 10.3389/frobt.2018.00001.
26. Manzotti R., Chella A. Conscious Machines: A Possibility? If So, How? *Journal of Artificial Intelligence and Consciousness*. 2020. Vol. 07. No. 02. Pp. 183–198. <https://doi.org/10.1142/S2705078520710022>
27. Linson A., Clark A., Ramamoorthy S. et al. The Active Inference Approach to Ecological Perception: General Information Dynamics for Natural and Artificial Embodied Cognition. *Consciousness in Humanoid Robots*. 2018. <https://doi.org/10.3389/frobt.2018.00021>
28. Chang A., Biehl M., Yu Y. et al. Information closure theory of consciousness. *Frontiers in Psychology*. 2020. Vol. 11. <https://doi.org/10.3389/fpsyg.2020.01504>
29. Winfield A. «Why Did You Just Do That?». Explainability and Artificial Theory of Mind for Social Robots. *Frontiers in Artificial Intelligence and Applications*. Vol. 335. 2020. Pp. 8–10. DOI:10.3233/FAIA200892.
30. Cominelli L., Mazzei D., De Rossi D.E. SEAI: Social Emotional Artificial Intelligence Based on Damasio’s Theory of Mind. *Front. Robot. AI*. 2018. DOI: 10.3389/frobt.2018.00006.
31. Damasio A. *Feeling & Knowing: Making Minds Conscious*. Pantheon. 2021. 256 p.
32. Bosse T., Heyselaar E.S. Linking Theory of Mind in human-agent interactions to validated evaluations: Can explicit questionnaires measure implicit behaviour? : In *Proceedings of the 21th ACM International Conference on Intelligent Virtual Agents*. 2021. Pp. 120–127. New York, NY: Association for Computing Machinery. DOI: 10.1145/3472306.3478343.
33. Pei Wang, Patrick Hammer, Hongzheng Wang. An Architecture for Real-Time Reasoning and Learning. *AGI 2020: Artificial General Intelligence*. 2020. Pp. 347–356. DOI: 10.1007/978-3-030-52152-3_37.
34. Kinouchi Y., Mackin K.J. A Basic Architecture of an Autonomous Adaptive System With Conscious-Like Function for a Humanoid Robot. *Front. Robot. AI*. 2018. DOI: 10.3389/frobt.2018.00030.
35. Anderson J. The Ersatz Brain Project: A brain-like computer architecture for cognition. *IEEE 11th International Conference on Cognitive Informatics and Cognitive Computing*. 2012. DOI: 10.1109/ICCI-CC.2012.6311125.
36. Golden M. *Statistical Machine Learning*. CRC Press. 2020. 506 p.
37. Stuart G., Spruston N. Dendritic integration 60 years of progress. *Nat. Neurosci.* 2015. 18. Pp. 1713–1721. DOI:10.1038/nn.4157.
38. Frank van der Velde. In Situ Representations and Access Consciousness in Neural Blackboard or Workspace Architectures. *Front. Robot. AI*. 2018. DOI: 10.3389/frobt.2018.00032.

39. Balkenius C., Trond A. Tjøstheim, B. Johansson et al. The missing link between memory and reinforcement learning. *Frontiers in Psychology*. 2020. Vol. 11. Pp. 34-46. <https://doi.org/10.3389/fpsyg.2020.560080>

40. Raja Chatila et al. Toward Self-Aware Robots. *Front. Robot. AI*. 2018. DOI: 10.3389/frobt.2018.00088