DOI: 10.30727/0235-1188-2021-64-1-102-115
Original research paper
Оригинальная исследовательская статья

Architectural Approach to Design of Emotional Intelligent Systems*

A.V. Shiller
Lomonosov Moscow State University, Moscow, Russia

O.E. Petrunya
Moscow Aviation Institute, Moscow, Russia

Abstract

Over the past decades, due to the course towards digitalization of all areas of life, interest in modeling and creating intelligent systems has increased significantly. However, there are now a stagnation in the industry, a lack of attention to analog and bionic approaches as alternatives to digital, numerous speculations on "neuro" issues for commercial and other purposes, and an increase in social and environmental risks. The article provides an overview of the development of artificial intelligence (AI) conceptions toward increasing the human likeness of machines: from the key ideas of A. Turing and J. von Neumann, who initiated the digitalization of society, to discussions about the definition of AI and the emergence of conceptions of strong and weak AI. Special attention is paid to the approach of A. Sloman, to ideas about the architecture and design of complex artificial systems are considered, which make it possible to "emotionally" expand the idea of weak/strong AI. In the article's section on the necessity and possibility of incorporating emotions into the architecture of AI, the authors reveal the goals and methodological limitations for creating an emotional artificial agent. In addition, the article briefly presents the main principles of the authors' architectural approach to the creation of emotional intellectual systems on the example of the cognitive-affective model of architecture, which allow modeling the impact of emotions on the cognitive processes involved in decision-making processes. The described architectural approach to modeling intelligent systems can be used as a conceptual basis for discussing and formulating a strategy for the development of neurocomputing, philosophy of artificial intelligence, and experimental philosophy, for developing innovative research programs, formulating and solving theoretical and methodological problems.

* The study was supported by the Interdisciplinary Scientific and Educational School of Moscow University "Brain, Cognitive Systems, Artificial Intelligence."

Keywords: philosophy of artificial intelligence, weak AI, strong AI, emotions, digitalization, anthropomorphism, intelligent agent.

Alexandra V. Shiller – Ph.D. in Philosophy, Head of the grant's department, Faculty of Philosophy, Lomonosov Moscow State University.
shiller.a@gmail.com
https://orcid.org/0000-0002-2466-9476

Oleg E. Petrunya – Ph.D. in Philosophy, Associate Professor, Department of Philosophy, Moscow Aviation Institute.
hypostasis@yandex.ru
https://orcid.org/0000-0002-5306-5129

For citation: Shiller A.V. & Petrunya O.E. (2021) Architectural Approach to Design of Emotional Intelligent Systems. *Russian Journal of Philosophical Sciences = Filosofskie nauki.* Vol. 64, no. 1, pp. 102–115.
DOI: 10.30727/0235-1188-2021-64-1-102-115

Архитектурный подход к созданию эмоциональных интеллектуальных систем*

А.В. Шиллер
Московский государственный университет имени М.В. Ломоносова, Москва, Россия

О.Э. Петруня
Московский авиационный институт (национальный исследовательский университет), Москва, Россия

Аннотация

В течение последних десятилетий из-за курса на цифровизацию всех областей жизни значительно возрос интерес к моделированию и созданию интеллектуальных систем. Однако сейчас наблюдается застой в отрасли, блокирование аналогового и бионического подходов в качестве альтернативных цифровому, многочисленные спекуляции на нейротематике в коммерческих и иных целях, рост социальных и экологических рисков. В статье представлен обзор развития представлений об искусственном интеллекте (ИИ) на пути к повышению человекоподобия машин: от ключевых идей А. Тьюринга

* Исследование выполнено при поддержке Междисциплинарной научно-образовательной школы Московского университета «Мозг, когнитивные системы, искусственный интеллект».

и Дж. фон Неймана, положивших начало цифрофизации общества, до дискуссий об определении ИИ и появлении положений о «сильном» и «слабом» ИИ. Отдельное внимание уделено подходу А. Сломана, рассмотрены идеи об архитектуре и дизайне сложных искусственных систем, благодаря которым становится возможно «эмоциональное» расширение идеи слабого/сильного ИИ. В разделе статьи, посвященном необходимости и возможности включения эмоций в архитектуру ИИ, описаны цели и методологические ограничения для создания эмоционального искусственного агента. Кроме того, в статье кратко представлены основные принципы авторского архитектурного подхода к созданию эмоциональных интеллектуальных систем на примере когнитивно-аффективной модели архитектуры, которая позволяет моделировать влияние эмоций на когнитивные процессы, задействованные в процессах принятия решений. Описанный архитектурный подход к моделированию интеллектуальных систем может быть использован в качестве концептуальной основы при обсуждении и формулировании стратегии развития нейрокомпьютинга, философии искусственного интеллекта и экспериментальной философии, для разработки плодотворных научно-исследовательских программ, постановки и решения теоретических и методологических проблем.

Ключевые слова: философия искусственного интеллекта, слабый ИИ, сильный ИИ, эмоции, цифровизация, антропоморфизм, интеллектуальный агент.

Шиллер Александра Викторовна – кандидат философских наук, руководитель грантового отдела философского факультета МГУ имени М.В. Ломоносова.
https://orcid.org/0000-0002-2466-9476

Петруня Олег Эдуардович – кандидат философских наук, доцент кафедры философии Московского авиационного института (национального исследовательского университета).
hypostasis@yandex.ru
https://orcid.org/0000-0002-5306-5129

Для цитирования: *Шиллер А.В, Петруня О.Э.* (2021) Архитектурный подход к созданию эмоциональных интеллектуальных систем // Философские науки. 2021. Т. 64. № 1. С. 102–115.
DOI: 10.30727/0235-1188-2021-64-1-102-115

Introduction

The relevance of the topic is determined, first of all, by the radical changes that have swept modern society, the “digitalization,” which

is the process of development and introduction of computers and software into the communication and management systems of economy, politics, and society, using the capabilities of high-level programming languages, including their significant independence from specific hardware. The ideological context of digitalization is created by two key ideas: A. Turing's idea about the similarity between a computing machine operatings with a finite number of states and a person with limited memory [Turing 1937] and J. von Neumann's idea about the identity of living organisms and finite automata [von Neumann 1995]. However, the main trend is set by the scenario of the *imitation game* proposed by Turing [Alekseev 2013], if the machine succeeds in this game, it will open the era of *strong artificial intelligence.*

The ultimate goal of modern digitalization is planetary artificial intelligence (AI) and the transformation of human nature itself. This idea is one of the dogmas of neo-mechanicism, a worldview based on a formal numerical understanding of the world. Modern digitalization is the implementation of this worldview in the field of technology, which implies an essential restructuring of the entire scientific and technical methodology in a formal and mathematical way, a finite unification of tools, a rejection of bionic and analog approaches [Petrunya 2017]. The above-mentioned trend has already slowed down the process of creating progressive neurocomputers. Their place was taken by imitation mathematical models that have nothing to do with biological objects.

For this reason, the digital environment was initially perceived with a fundamental strategic flaw – it is unfriendly, non-anthropomorphic, and disproportionate to humans. Therefore, in the real practice of designing computers and creating software for effective work here and now, it is necessary to solve not only many technical problems. There is a need to improve the efficiency of human-machine interaction, first of all, in the development of user and other interfaces that are already actively included in people's lives and are used by banks, telecom operators, and other areas of business [Vedyakhin et al. 2021]. Unfortunately, the solution to these problems occurs spontaneously with all the resulting disadvantages.

Nevertheless, in Western cognitive science, one way or another, the question of increasing the human likeness of machines was raised. First, this is due to efforts to make a machine that is able to pass the Turing imitation criterion. Second, this search has stimulated a set of relevant practical challenges mentioned above, including user-friendliness and usability. Below we will consider two approaches, which, along with

many others, are used in cognitive and computer sciences to increase the human likeness of machines: (a) a design approach in creating computer architectures and (b) modeling emotions in these architectures.

Discussions on the concept of AI and A. Sloman's design approach

The term "artificial intelligence" was coined in the mid-1950s by John McCarthy at the Dartmouth Conferences. A little later, in 1959, McCarthy with Marvin Minsky organized the Computer Science and Artificial Intelligence laboratory at the Massachusetts Institute of Technology. It was McCarthy who became the main ideologist and enthusiast of the field that J. Searle in the article "Minds, Brains, and Programs" called strong AI [Searle 1980]. The concept of strong AI originates in the Turing interpretation of the consequences of the absence of a solution to D. Hilbert's decision problem (*das Entscheidungsproblem*). E. Post provides a different, narrow understanding of the consequences, considering his method and the machine proposed for it as proving "…significant in the development of symbolic logic along the lines of Gödel's theorem on the incompleteness of symbolic logics and Church's result concerning absolutely unsolvable problems" [Post 1936, 103]. A. Turing expanded the problem field in the article "Computing Machinery and Intelligence," proposing to consider the computational actions of a digital machine with a finite number of states as identical to the computational activity of a person with limited (finite) capabilities [Turing 1950]. Thus, D. Hilbert's finitism received a new functionalist interpretation and a new breath. However, it can be noted that, in Turing's problem statement, it is not clear how a negative answer to the question "Can machines think?" may be verified [Yanovskaya 1960]. Such understatement, in our opinion, can be interpreted not so much as a lack of solution to the problem but rather as an uncertainty in the area of its solution.

Another concept, according to which the machine does not think, but imitates thinking, was labeled by J. Searl labeled as weak AI. At the same time, Searl's main intention is emotionally colored; in fact, he wonders how one can agree with McCarthy's thesis about the identity between thermostats, telephones, computing machines, and people? He suggests thinking about the consequences of adopting such beliefs. Against the idea of strong AI, Searle offers the Chinese room argument. The argument describes a situation in which the machine's lack of understanding becomes apparent. Unfortunately, Searle missed

another aspect of the problem: the existing duality of the intellectual activity of the cognizing subject, based on reasoning (discourse) and intuition (speculation) as independent abilities. Although this split dates back to ancient Greek tradition, it was not obvious to all Western intellectuals (including Searle). At the same time, in H. Poincaré's works on the philosophy of mathematics, such a duality is taken into consideration [Poincaré 2014]. Thus, intentionality is not equivalent to reasoning [Biryukov 1978]. It "organizes the subject's attention and his understanding of the object of intellectual or sensory contemplation. Discursive forms are easily formalized, i.e., are replaced by a sequence of signs" [Feinberg 1981, 38], in contrast to the intuitive component of thinking, for which this is impossible.

British researcher A. Sloman replied to criticism of the concept of strong AI with an article with the eloquent title "Did Searle Attack Strong Strong or Weak Strong AI?" [Sloman 1986]. It should be added that Sloman shares the Newell–Simon hypothesis based on the assertion that symbol manipulation is the basis of the ability to think [Newell & Simon 1976] as well as Simon's ideas about the architecture and design of complex artificial systems [Simon 1969]. It is in this vein that one should understand the polemic between Sloman and Searl.

A. Sloman rejects Searl's splitting of AI into strong and weak versions. Based on the assertion that AI accurately reproduces certain human actions that can be characterized in terms of *functions, tasks* and *adaptations* (the Simon–Sloman position), Sloman refuses to accept the thesis that there is a fundamental difference between a human and a "smart" machine. So, he does not even consider a weak version of Searl's AI. Being a supporter of the strong version, Sloman offers its own splitting of AI into weak strong AI and strong strong AI versions. If it cannot be said that AI does not reproduce any human mental operations at all (a weak version of Searl), then there is no need to admit that it reproduces them completely. Marking his own position as weak strong AI and the one that Searl "attacks" as strong strong AI, Sloman states: his opponent is at war with windmills – none of Searl's opponents strongly believes in strong strong AI. But the last thesis hardly reflects the position of J. McCarthy, whom Searle, first of all, criticizes in the above-named article. By the way, McCarthy himself lamented that "as soon as the system starts working normally, they immediately stop calling it artificial intelligence" [Petrunya 2021]. Thus, without denying Sloman the rightness, it should be recognized that his dichotomy does not exhaust all positions in AI, but it complements Searle's analysis.

To further clarify the described controversy, it must be said that in relation to AI, since the late 1950s – early 1960s, Russian philosophy and science has been stuck to the positions that J. Searle marks as weak AI. This context of understanding was formed by S.A. Yanovskaya in her preface to the Russian translation of Turing's article "Can Machines Think?" [Yanovskaya 1960]. However, the Sloman dichotomy can add an important nuance to this analysis: the Soviet scientific community can be divided into supporters of weak weak AI and strong weak AI. In the first case, the researchers did not accept or did not consider any similarity between a machine and a person, in the second, their functional analogy was postulated [Svintsitsky 1964].

In this regard, we should recognize the proximity of Sloman's weak strong version of AI (Simon–Sloman's position) and the Russian version of strong weak AI, which, postulating the similarity of a person and a machine, brackets their differences out. Opening up great opportunities for conceptual speculation under the guise of AI, such an approach simultaneously demonstrates its productivity in the field of putting forward bold hypotheses and their empirical testing. And here Simon-Sloman's conception of the architecture of complexity is quite capable of competing with the P.K. Anokhin's theory of functional systems.

Subsequently, A. Sloman and colleagues [Wright, Sloman, & Beaudoin 1996], generalizing and developing the ideas of Simon, proposed a certain kind of information processing architecture that "emotionally" expands the idea of weak strong AI. This architecture is based not only on the capabilities of high-level programming languages with their significant independence from the hardware implementation of a computer, but mainly on the research methodology of the abstract space of possible requirements for functioning agents (niche space) and the space of possible designs for such agents (design space) and mapping them. Since the architecture dominates over the mechanism in Sloman's approach, the design usually determines the capabilities more than the details of its implementation in each specific case. This, in particular, makes it possible to reconcile the computational and connectionist approaches, i.e., positions of McCarthy and Minsky.

Any architecture is based on the principles envisaged by the designer, and the subsequent functioning and development of the information system does not go beyond those principles. The concept of architecture here is congruent with the systemic characteristics of human thinking

and the psyche as a whole, i.e., causal structures implemented in a non-neural substrate [Wright, Sloman, & Beaudoin 1996].

In the design approach, two interrelated tasks are solved: (1) the research of deeper mechanisms that generate the mental phenomena of living beings; (2) the research of new possibilities for creating perfect intelligent systems. The starting point of the analysis is the research of structures (design) that satisfy the requirements for autonomous self-organizing systems. The design approach allows to successfully explicate the required content into specific design forms and prepare proposals for its subsequent approximation as well as empirical testing of hypotheses, using a variety of interdisciplinary theoretical approaches.

Necessity and possibility of including emotions into AI architecture

Sloman's design approach to the creation of an artificial agent architecture shows the possibility of including in its composition such blocks that were previously proposed to be excluded – the block of motivation, emotions, "morality" (the prototype of the ethical system of an artificial agent). However, Sloman points out that "some AI researchers believe that it should be the goal of AI to design agents that overcome human limitations while displaying all their strengths. This may not be possible if some of the limitations are inevitable consequences of the mechanisms and architectures required to produce those strengths" [Sloman 1999]. That is why it is necessary to understand the phenomenology of emotions, the relationship between emotions and other cognitive functions, methodological limitations and requirements for modeling emotions in the architecture of AI [LeDoux, Phelps, & Alberini 2016; Panksepp 1998].

There are several approaches that are most important for philosophy and cognitive sciences: the three-factor theory of C. Osgood and his colleagues [Osgood, May, & Miron 1975]; the theory of the six components of emotions [Fontaine et. al. 2007]; M. Lewis's theory of the equal status of cognitions and emotions [Lewis 2008]; the theory of embodied cognition [Niedenthal et. al. 2009]; the concept of enactivism [Colombetti 2014]; the theory of cognitive assessment of the emergence of emotions [Ortony, Clore, & Collins 1988]; A. Sloman's concept of architecture of the agent [Sloman 1987]; J. Broekens, D. DeGroot and W. Kosters's approach to modeling emotions [Broekens, DeGroot, & Kosters 2008].

The analysis of these approaches shows that emotions and cognitive processes are functionally connected. Therefore, emotions as an influencing factor that cannot be removed from the system of interrelated processes, should be assigned one of the central places in the architecture of an artificial intellectual system. Simulating emotions in AI becomes especially important when there are such types of tasks of creating AI: (1) achieving system integration by ensuring multimodality of cognitive processes; (2) creation of a flexible architecture that changes during the cycle of functioning of an artificial agent; (3) ensuring high reliability of the interaction of AI with a person; (4) building a digital model of the human emotional system for its theoretical research.

The requirement to include emotions in the architecture of AI is due to both the practical tasks of developing emotional agents and theoretical search that is driven by the "return to corporeality" and the anthropomorphic turn in philosophy and cognitive science. The future-oriented goal is the creation of a model of anthropomorphic intelligence for experimenting with, which will act as an analog of "natural" intelligence.

However, the current possibilities for modeling emotions in artificial agents are extremely limited. Therefore, there should be progress in the transition toward the use of individual theories and modeling parts of the architecture of an artificial agent. In this article, we propose to consider the principles of creating an emotional architecture of an artificial agent using the example of the cognitive-affective model of architecture (CAMA).

Cognitive-affective model of architecture

One of the problems identified in the analysis of the field of emotional modeling is "the lack of a theory that can become the basis for modeling the process of expressing the effects of emotions" [Shiller 2019, 237]. To solve this problem, we propose to investigate the principles of building an architecture model that includes both cognitive and affective components.

CAMA is based on the theory of embodied cognition, the Sloman's concept of architecture and Broekens' approach to modeling emotions. It seems to us that this model, which combines several modern approaches, may turn out to be promising for the purposes of computational modeling of emotions.

The principle of combining cognitive and affective components into one model takes into account and allows modeling individual differ-

ences (personality traits) as well as the impact of emotions (affective states) on cognitive functions involved in decision-making processes (attention, goal-setting, etc.).

Thus, the ***first principle*** of the model is ***synthesis of*** both different components and theoretical foundations.

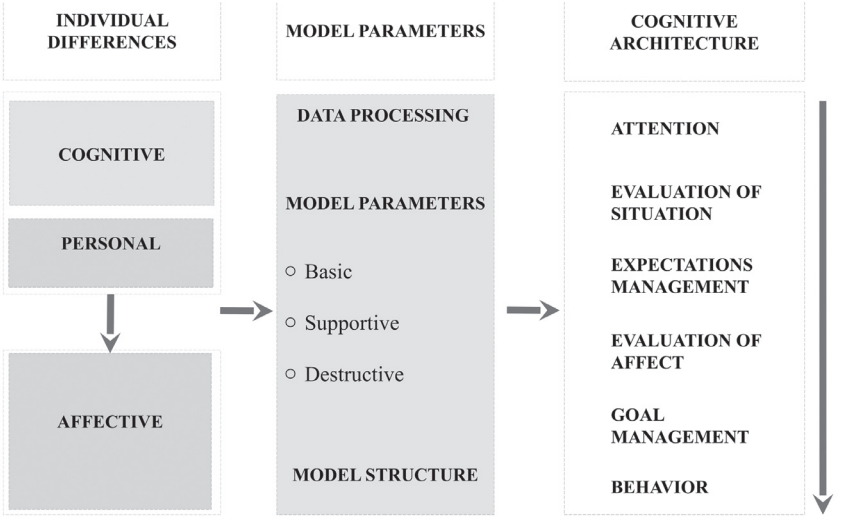

Figure 1. *Diagram of an architecture model*

In the middle part of the diagram (Fig. 1), a set of parameters of the cognitive architecture is presented, among which there should be noted the parameters responsible for data processing (processing speed, working memory), as well as structural parameters (processing rules and regulations, the scheme of intermodular connections and their weight).

The work of the model can be represented in several steps:

(1) Input: obtaining perceptual data (from the external environment), which may be important for making decisions and choosing actions, as well as the presence of predetermined information about the characteristics of the agent (a certain "synthetic personality," for which there are envisaged parameters of the intensity of emotions, the value of various personality traits).

(2) Data processing: it depends on individual (cognitive and affective) differences and occurs by correlating the predetermined characteristics with the architecture parameters, which determines changes in the course of data processing (acceleration, deceleration, reduction in the set

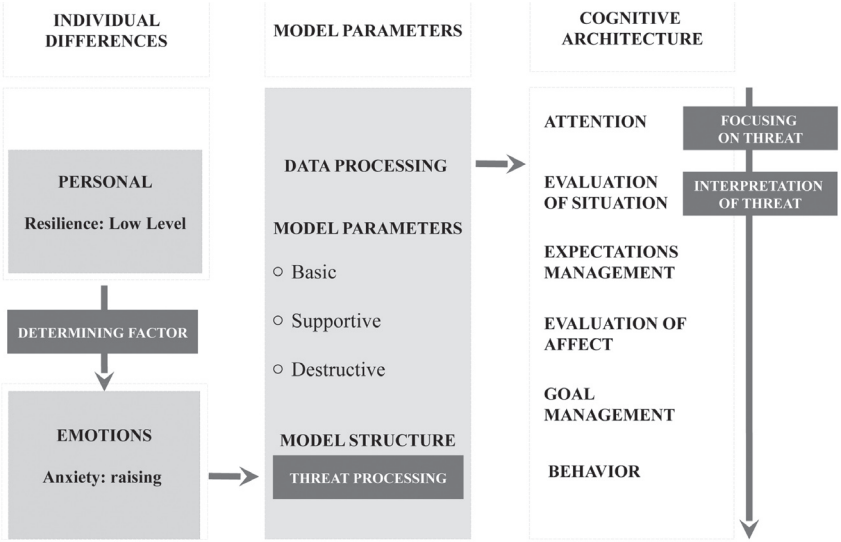

Figure 2. *Impact of low resilience and anxiety on threat perception*

of steps, prioritization of the processing of certain data). The matching mechanism is based on empirical data and ideas about how a person acts with a similar set of perceptual data and personality characteristics.

(3) Output: there is a hierarchical system containing a set of situations and goals in which the initial perceptual data and the proposed behavior are differently correlated. Data analysis allows us to assign some "tags" to information – for example, if the received perceptual data is assessed as dangerous, posing a threat to the life of the agent, then the role of the constructs of anxiety and fear increases, which determines the speed of data processing in the modules of attention and situation assessment as well as the choice of the agent's behavior as a result of data analysis (see Fig. 2).

Over time, similar perceptual data will be more quickly tagged as dangerous and, therefore, prioritized for processing.

As Fellous points out, "emotions are patterns of neuromodulations that affect brain areas involved at all levels of functions, from low-level motor control to planning and high-level cognition" [Fellous 2004, 7]. Therefore, to express the cognitive effects of emotion the model uses parametric functions consisting of linear sets of factors that are estimated and affect each of the parameters, which is consistent with the data of modern research. An example is the following parameter: "working memory capacity reflects the normalized weighted sum of emotion intensities, trait values, and skill level" [Shiller 2019, 236].

The second principle of the model is ***anthropomorphism,*** that is, the human psyche is the basis of the architecture of the model and the dominant that determines the emotionality and behavior of the artificial agent. Based on these two principles, the features of the model are as follows:

(1) The ability to create many variants of synthetic personalities by defining a set of cognitive and affective components, personality traits.

(2) The consideration of the role of corporeality and, as a consequence, the ability to implement the principles of embodied cognition in creating AI.

(3) The ability to partially implement the principle of neuroplasticity, which underlies the work of the human brain.

(4) The presence of restrictions associated with the need to formalize the processes of the human psyche, which inevitably leads to the simplification of the final model of architecture.

In the future, it is planned to expand and improve this architecture model by testing primary hypotheses, collecting data and implementing theoretical provisions in order to form an acting artificial agent with an emotional subsystem based on CAMA.

Conclusion

Nowadays, there is a huge rise in scientific and pseudo-scientific research on AI and its mention in the media. However, it should be noted that, in most cases, all these researches are of applied nature and use the phrase "artificial intelligence" as a catchy analogue of fairly simple technologies that can only be attributed to the "weak" version of AI. The reason lies in the fact that the digital approach to the development of AI technologies for a long time was understood as the only "correct" one, and analog and body-oriented approaches, which consider the multimodality of cognitive activity and possible errors/imperfections that ensure the richness of this activity, were excluded from the sphere of interest of researchers.

The described discussions on the definition and various versions (strong, weak, or some intermediate one) of AI illustrate the complexity of both the concept and the scientific direction. As an inevitable consequence of the development of this research area, there are an anthropomorphic turn in the studies on AI and efforts to model analogs of human intelligence or at least some of its constituent parts. Sloman's model of architecture for an artificial agent, Broekens's researches on modeling,

and numerous experimental data in line with the theory of embodied cognition inspired the authors of the article to create another version of cognitive architecture that allows modeling individual differences and affective states in the process of cognitive data processing. In the future, additional studies are planned that will allow testing hypotheses, supplementing the architecture with new blocks, forming a mathematical apparatus, and moving from theory to practice, implementing the model in the form of a digital assistant/bot that is capable of the believable expression of emotion.

REFERENCES

Alekseev A.Yu. (2013) *Complex Turing Test: Philosophical-Methodological and Socio-Cultural Aspects.* Moscow: IInteLL (in Russian).

Biryukov B.V. (1978) What Can Computing Machines Do? Instead of an Afterword. In: Dreyfus H. *What Computers Can't Do. Critique of Artificial Intelligence* (pp. 298–332). Moscow: Progress (in Russian).

Broekens J., DeGroot D., & Kosters W.A. (2008) Formal Models of Appraisal: Theory, Specification, and Computational Model. *Cognitive Systems Research.* Vol. 9, no. 3, pp.173–197.

Colombetti G. (2014) Ideas for an Affective Neuro-physio-phenomenology. In: Colombetti G. *The Feeling Body: Affective Science Meets the Enactive Mind* (pp. 135–170). Cambridge, MA: MIT Press.

Feinberg E.L. (1981) *Cybernetics, Logic, Art.* Moscow: Radio i svyaz' (in Russian).

Fellous J.M. (2004) From Human Emotions to Robot Emotions. In: *Proceedings of the AAAI Spring Symposium: Architecture for Modeling Emotion* (technical report SS-04-02). Menlo Park, CA: AAAI Press.

Fontaine J.R.J., Scherer K.R., Roesch E.B., & Ellsworth P.C. (2007) The World of Emotions Is Not Two-Dimensional. *Psychological Science.* Vol. 18, no. 2, pp. 1050–1057.

LeDoux J., Phelps L., & Alberini C. (2016) What We Talk about When We Talk about Emotions. *Cell.* Vol. 167, no. 6, pp. 1443–1445.

Lewis M. (2008) Self-Conscious Emotions: Embarrassment, Pride, Shame, and Guilt. In: Lewis M. & Haviland-Jones J.M. (Eds.) *Handbook of Emotions* (3rd ed., pp.742–749). New York: Gilford Press.

Newell A. & Simon H.A. (1976) Computer Science as Empirical Inquiry: Symbols and Search. *Communications of the ACM,* Vol. 19, no 3, pp. 113–126.

Niedenthal P.M., Winkielman P., Mondillon L., & Vermeulen N. (2009) Embodiment of Emotion Concepts. *Journal of Personality and Social Psychology.* Vol. 96, no. 6, pp.1120–1136.

Ortony A., Clore G.L., & Collins A. (1988) *The Cognitive Structure of Emotions.* Cambridge, UK: Cambridge University Press.

Osgood C.E., May W.H., & Miron M.S. (1975) *Cross-Cultural Universals in Affective Meaning.* Urbana: University of Illinois Press.

Panksepp J. (1998) *Affective Neuroscience: The Foundations of Human and Animal Emotions.* New York: Oxford University Press.

Petrunya O.E. (2017) The Opposition of Natural and Artificial in Neuroscience: "Who Is to Blame" and "What to Do." *Neyrokompyutery: razrabotka, primeneniye = Neurocomputers: Development, Application.* No. 4, pp. 26-30 (in Russian).

Petrunya O.E. (2021) *Anthropic Approach vs. Imitation Game: Philosophical and Methodological Problems of Science and Technology in the Digital Age. M*oscow: LENAND (in Russian).

Poincaré H. (2014) Mathematics and Logic. In: *The Foundations of Science: Science and Hypothesis, The Value of Science, Science and Method* (pp. 448–459). Cambridge, UK: Cambridge University Press.

Post E.L. (1936) Finite Combinatory Processes – Formulation 1. *The Journal of Symbolic Logic.* Vol. 1, no. 3, pp. 103–105.

Searle J.R. (1980) Minds, Brains, and Programs. *Behavioral and Brain Sciences*. Vol. 3, no 3, pp. 417–457.

Shiller A.V. (2019) Expression of Modeled Effects of Emotions in Artificial Agents as a Visual Language. *ΠΡΑΞΗΜΑ. Problems of Visual Semiotics.* Vol. 22, no. 4., pp. 223–243 (in Russian).

Simon H.A. (1969) *The Sciences of the Artificial.* Cambridge, MA: MIT Press.

Sloman A. (1986) Did Searle Attack Strong Strong or Weak Strong AI? In: Cohn A.G. & Thomas J.R. (Eds.) *Artificial Intelligence and Its Applications* (pp. 271–288). Chichester: John Wiley and Sons.

Sloman A. (1999) What Sort of Architecture is Required for a Human-Like Agent? In: Wooldridge M. & Rao A. (Eds.) *Foundations of Rational Agency* (pp. 35–52). Dordrecht: Springer.

Svintsitsky V.N. (1964) On the Question of the Genetic Connection of Cybernetics with Classical Automatics. In: Berg A.I., Biryukov B.B., Novik I.B., Kuznetsov I.V., & Spirkin A.G. (Eds.) *Cybernetics, Thought, Life* (pp. 164–172). Moscow: Mysl' (in Russian).

Turing A. (1937) On Computable Numbers, with an Application to the Entscheidungsproblem. *Proceedings of the London Mathematical Society. Series 2.* Vol. 42, no. 1, pp. 230–265.

Turing A. (1950) Computing Machinery and Intelligence. *Mind.* Vol. 59. no. 236, pp. 433–460.

Vedyakhin A.A. et al. (2021) *Strong Artificial Intelligence: Approaching the Supermind* (A.S. Potapov, Ed.). Moscow: Alpina Publisher (in Russian).

von Neumann J. (1995) The General and Logical Theory of Automata. In: Bródy F. & *Vámos T. (Eds.) The Neumann Compendium* (Vol. 1, pp. 526–566). Singapore: World Scientific.

Wright I.P., Sloman A., Beaudoin L.P. (1996) Towards a Design-Based Analysis of Emotional Episodes. *Philosophy, Psychiatry, & Psychology.* Vol. 3, no 2, pp.101–126.

Yanovskaya S.A. (1960) Preface to the Russian Translation. In: Turing A. *Can Machines Think?* (pp. 3–11). Moscow: Fizmatgiz (in Russian).