## COGNITIVE SPACE

## Philosophy of Consciousness

# Do Chatbots Dream of Androids? Prospects for the Technological Development of Artificial Intelligence and Robotics

*A.R. Efimov*
*Robotics Laboratory of Sberbank, Moscow, Russia*

DOI: 10.30727/0235-1188-2019-62-7-73-95
Original research paper

**Abstract**

The article discusses the main trends in the development of artificial intelligence systems and robotics (AI&R). The main question that is considered in this context is whether artificial systems are going to become more and more anthropomorphic, both intellectually and physically. In the article, the author analyzes the current state and prospects of technological development of artificial intelligence and robotics and also determines the main aspects of the impact of these technologies on society and economy, indicating the geopolitical strategic nature of this influence. The author considers various approaches to the definition of artificial intelligence and robotics, focusing on the subject-oriented and functional ones. In the article, AI&R abilities and human abilities are compared in such areas as categorization, pattern recognition, planning and decision making, etc. Based on this comparison, it is concluded in which areas AI&R's performance is inferior to a human and in which cases it is superior to one. The modern achievements in the field of robotics and artificial intelligence create the necessary basis for further discussion of the applicability of goal setting in engineering, in the form of a Turing test. It is shown that development of AI&R is associated with certain contradictions that impede the application of Turing's methodology in its usual format. The basic contradictions in the development of AI&R technologies imply that there is to be a transition to a post-Turing methodology for assessing engineering implementations of AI&R. In

such implementations, on the one hand, the "Turing wall" is removed, and on the other hand, artificial intelligence gets its physical implementation.

**Keywords:** robotics, artificial intelligence, Turing, Turing test, philosophy of artificial intelligence.

**Albert R. Efimov** – Head of the Sberbank Robotics Laboratory, postgraduate student at the Faculty of Philosophy, Lomonosov Moscow State University.

makkawity@gmail.com,
https://orcid.org/0000-0001-6857-8659

**For citation:** Efimov A.R. (2019) Do Chatbots Dream of Androids? Prospects for the Technological Development of Artificial Intelligence and Robotics. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. Vol. 62, no. 7, pp. 73–95. DOI: 10.30727/0235-1188-2019-62-7-73-95

## Снятся ли чат-ботам андроиды? Перспективы технологического развития искусственного интеллекта и робототехники

*А.Р. Ефимов*
*Лаборатория робототехники Сбербанка, Москва, Россия*

DOI: 10.30727/0235-1188-2019-62-7-73-95
Оригинальная исследовательская статья

**Аннотация**

Статья посвящена обобщению основных трендов развития систем искусственного интеллекта и робототехники (ИИиР). Основной вопрос, который рассматривается в этом контексте: будут ли искусственные системы становиться все более антропоморфными как в интеллектуальном, так и в физическом отношении? В статье автор не только проводит анализ современного состояния и перспектив технологического развития искусственного интеллекта и робототехники, но и определяет основные аспекты влияния этих технологий на общество и экономику, указывая на геополитический стратегический характер данного влияния. Автор рассматривает различные подходы к определению искусственного интеллекта и робототехники, выделяя предметно-ориентированный

и функциональный. Также производится сопоставление способностей ИИиР и человеческих способностей в таких областях, как категоризация, распознавание образов, планирование и принятие решений и др. На основе этого сопоставления сделаны выводы о том, когда ИИиР уступают человеку, а в каких случаях превосходят его. Проанализированные автором современные достижения в области робототехники и искусственного интеллекта создают необходимый базис для дальнейшего рассуждения о применимости инженерного целеполагания в виде теста Тьюринга. Показано, что развитие ИИиР связано с определенными противоречиями, затрудняющими применение методологии Тьюринга в привычном формате. Рассмотренные автором базовые противоречия в развитии технологий ИИиР позволяют сделать основной вывод о переходе к посттьюринговой методологии оценки инженерных реализаций искусственного интеллекта и робототехники, в которых, с одной стороны, снимается «стена Тьюринга», а с другой стороны, искусственный интеллект получает свое физическое воплощение.

**Ключевые слова:** робототехника, искусственный интеллект, Тьюринг, тест Тьюринга, философия искусственного интеллекта.

**Ефимов Альберт Рувимович –** руководитель Лаборатории робототехники Сбербанка, аспирант философского факультета МГУ им. М.В. Ломоносова.

makkawity@gmail.com

https://orcid.org/0000-0001-6857-8659

**Для цитирования:** Ефимов А.Р. (2019) Снятся ли чат-ботам андроиды? Перспективы технологического развития искусственного интеллекта и робототехники // Философские науки. 2019. Т. 62. № 7. С. 73–95. DOI: 10.30727/0235-1188-2019-62-7-73-95

## Introduction

Artificial intelligence and robotics are among the newest areas of science and technology. Despite the long traditions of technological anthropomorphism and organ-projection, dating back to the ancient Greek philosophers Protagoras and Socrates [Dreyfus 1978], the true history of machines capable of at least remotely reasoning or acting like a human being has lasted only a few decades. After all, seventy

years ago, even calculators were office job positions of "specialists in computing," rather than equipment units in many organizations.

Over the brief history of computer technology, cybernetics, computer science, and robotics, there has been no shortage of advertising slogans praising the advantages and effectiveness of such technologies over previous management tools. In an effort to attract the attention of as many consumers as possible, their names change dramatically – from cybernetics to artificial intelligence. However, we cannot overlook the apt remark of the Nobel laureate Robert Solow, which he made more than 30 years ago, "You can see the *computer age everywhere but* in the productivity *statistics*" [Solow 1987, 36].

Over the past three decades, technology has taken a huge step forward, and right now, most economists agree that artificial intelligence and robotics (hereinafter referred to as AI&R) will have a significant impact on our civilization. However, not only economy determines increasing attention of states and governments to the opportunities that AI&R has to offer. These technologies become a full-fledged tool of geopolitics, with which the winners stand to obtain everything and the losers will be forced to abide by the rules that the winners establish [Shaw 2017].

Science fiction provides a huge list of topics for both filmmakers and philosophers. Even the title of this article is a paraphrase of the title of P.K. Dick's famous novel *Do Androids Dream of Electric Sheep?* (1), which is entirely devoted to the complex problem of the emergence of consciousness on a non-biological substrate. However, the current paper is devoted to some aspects of interaction of artificial intelligence (2) (or software, the "psyche") and robots (or generally, hardware as media for artificial intelligence, the "body," "motility") (3).

## Definitions of Artificial Intelligence and Robotics

Due to the fact that the demand for research in AI&R has so dramatically increased, it is extremely challenging to give an accurate definition of AI. Representatives of academic science and business

do not have a common point of view on what AI&R is. However, lack of a unified definition does not prevent the growth of investment and business activity, which, in turn, makes it even more urgent to offer businesses more adequate ways to define the AI&R subject area, using a specific set of concepts.

The model proposed by V.K. Finn seems to be the most accurate one. Finn suggests that AI should be interpreted through a definition of natural intelligence, as having has the following characteristics, only implemented on a non-biological substrate (as software, algorithms, and hardware implementation):

1. Ability to highlight the essential in existing knowledge, i.e. to organize the latter (this is a necessary aspect of intuition).

2. Ability to set goals and to plan activities –generation of multiple sequences phrased as 'goal → plan → action.'

3. Ability to select knowledge (logical premises and conclusions, relevant goals of reasoning).

4. Ability to make conclusions from available knowledge, i.e. the ability to reason, which may contain both plausible conclusions, to be further used to formulate hypotheses, and reliable conclusions…;

5. Ability to perform well-founded decision-making, using ordered knowledge (presentation of knowledge) and results of reasoning suitable for the goal.

6. Ability to reflect, i.e. to assess knowledge and actions.

7. Presence of cognitive inquisitiveness: the cognizing subject should be able to ask the question, 'What is it?' and to look for an answer to it.

8. Ability and need to find an explanation (not necessarily deductive!) as an answer to the question, 'why?'

9. Ability to synthesize cognitive procedures that form the heuristic of problem solving and addressing issues, for example, via combination of induction, analogy and abduction (the latter is due to falsification of some hypotheses, through searching for counterexamples), with subsequent deduction.

10. Ability to learn and to use memory.

11. Ability to rationalize ideas, planning to clarify them as concepts.

12. Ability to create a holistic picture regarding the subject of thinking, combining knowledge relevant to the goal (i.e. the formation of at least an approximate theory of the subject area).

13. Ability to adapt to changing life circumstance and available data, which means correction of 'theories' and activities.

[Finn 2018, 37]

According to Finn, the abilities indicated in paragraphs 1, 3, 4, 5, 6, 8, and 9 can be implemented by machines without human intervention, granted sufficient software and hardware, but the abilities presented in listed in Paragraphs. 2, 7, 12, and 13 cannot be realized without partial participation of a person interacting with the machine.

The very concept of "robotics" is difficult to define, although there is already a global standard in industrial robotics that defines the concept of a robot: "actuated mechanism programmable in two or more axes with a degree of autonomy, moving within its environment, to perform intended tasks" [ISO 8373: 2012].

The drawback of this definition becomes obvious when you think about the fact that it refers to quite different objects, such as unmanned aerial vehicles, self-driving cars and industrial manipulators, featuring various degrees of freedom. Moreover, this definition has no correlation with a number of cutting-edge studies that indicate that robots of the future are going to be, first of all, tools that are capable of group interaction with one another, of self-organization and of communication with humans in a natural language.

This "lag" of AI&R definitions from actual technological progress provides an impetus for simplified engineering definitions in these areas of study. In AI&R, this continues the example of A. Turing, who approached the definition of AI (asking the question, "Can a machine think?") through an anti-essentialist refusal to deeply investigate the "essence" of AI, formulating this as an engineering problem instead. In particular, abandoning ineffective ontological disputes about AI, Sberbank experts regard AI as algorithms, soft-

ware and hardware that solve the following applied problems at the human level (or higher):

– computer vision;

– recommender systems and intelligent decision support systems;

– natural language processing;

– speech recognition and synthesis;

– promising methods and technologies of AI.

The author proposed an *applied* functional definition of robotics, which is based on the adaptation of the "duck test" (4). According to this definition, a robot can be any device that is capable of simultaneously performing the following:

1. *Sense:* a robot is able to perceive the world using its sensors. Such sensors can be microphones (sonars), cameras (for all ranges of the electromagnetic spectrum), various electromechanical sensors (accelerometer), and others.

2. *Think:* a robot is able to interpret (understand) the signals that it receives from sensors observing the physical world, to build and adapt behaviors, and make decisions depending on the models selected. This ability can be realized in different ways: the on-board computer of the robot, an "intelligent" cloud, or a person who controls the robot using teleoperation or an interface with feedback support.

3. *Act:* a robot is able to affect the physical world in an effective way [Efimov et al. 2019].

The approaches of Russian Sberbank's specialists, including those proposed by the author, are, in fact, a case of application of the philosophical method of M. Heidegger, who moved from a Platonic rational abstraction to an "emotional-practical" attitude towards the world in general, and to the use of AI&R technology in particular [Dreyfus 1991, 8].

As J. McCarthy put it, the phrase "artificial intelligence" is a "suitcase word," i.e. a word that has an incredibly large number of meanings. Currently, the number of possible meanings may already have equaled the number of all possible approaches to research or to

commercialization of technologies, which renders senseless all static or formal logical definitions of "artificial intelligence" and makes researchers and engineers look for more relevant ways to describe their specific areas of activity. One of such approaches is adopted at Sberbank, describing AI as a set of application areas, or robotics as a series of critical functions. Researchers and engineers, acting in their own interests and driven by awareness of the possibility of realizing themselves through technical solutions or scientific discoveries that such solutions offer, can rely on such practical definitions and pursue a stable emotional connection with a clearly described subject or functional area.

## Current and future areas of application of artificial intelligence and robotics

Since the beginning of the industrial revolution, those sectors of human activity in which automation was applicable have been under scrutiny of entrepreneurs and engineers. They tried to automate any mechanical work that could be automated at the current level of technological development, provided that this could bring significant savings in resources. Moreover, industries where human activities were based on rules or knowledge were largely excluded from the "dehumanizing" influence of technology. If we look at Rasmussen's proposed classification of activity types: (a) skill-based, (b) rule-based, and (c) knowledge-based, we can see that automation concerned mainly the first and second types of activities [Rasmussen 1979].

AI has the unique capacity of transforming activity areas that were not previously influenced by automation and were deemed as areas for "creative" people (i.e., the (c) type, according to Rasmussen) into technical industries, just like those where automation has actively been used for centuries. And this happens due to the possibility of accumulating data in digital form. Jurisprudence, education, finance, medicine, and economics are gradually transforming into sets of algorithms and data, and the activities of people in these spheres are becoming closer to crafts (rather than skills), or to what the

ancient Greeks called βαναυσία, where a skill of Type C turns into commonplace work of Type A.

Many economists see AI&R as an opportunity for growth in the total factor productivity of labor, which has been steadily declining since the 1980s, despite the transition of our civilization to the information age (cf.: [Brignolfson et al. 2019]). The current economic estimates discuss a 40% increase in labor productivity across a whole economy, due to the use of AI&R. In the case of the full-scale penetration of technologies based on AI&R into an economy, the economic growth due to their introduction will be very significant, comparable in effect to the appearance of the steam engine during the Industrial Revolution in the 18th century. In economically developed countries, additional AI&R-related growth can range from 1.8% to 4.6% of GDP annually [Purdy et al. 2016].

In fact, AI&R is a typical cross-cutting technology used in any area of national economy and public administration. In this regard, it is worth indicating only the most significant opportunities for application of AI&R. For robotics, these areas include: image processing, sensor data processing, predictive analysis and modeling, voice assistants, chatbots, logistics, retail, agriculture, and elderly care. As for artificial intelligence, in addition to all of the above, these technologies are used in education, synthesis of images and virtual worlds, analysis of sensor data, and analysis of medical data.

## Some significant achievements of artificial intelligence and robotics over the past 20 years

The founder of quantum computing, David Deutsch, rightly observes that any engineering research in the field of general AI that is not supported by serious epistemology and methodological foundations will be fruitless [Deutsch 2012]. However, if philosophy and philosophers themselves are not open to the latest technological advances and fail to understand their current and future opportunities, as well as limitations, there will be underestimation of the prospects of AI&R or, vice versa, unjustified high expectations may appear. This section attempts to bridge the gap between the

engineering achievements of recent years and their philosophical conceptualization.

To begin with, we answer the question of where artificial intelligence and robots perform better than humans. At the very beginning of the information technology era, Alan Turing laid a foundation for a comparison between man and computer, as well as the basics of machine functionalism [Putnam 1975]. Continuing this tradition, J. Fodor [Fodor 1990] defined computerism, to be further developed into test functionalism [Alekseev 2013]. In many ways, the subsequent history of AI&R became a field of contest where some researchers sought to prove that a machine would never rise to human thinking [Dreyfus 1978], while others convincingly showed that this was not only possible, but already achieved [Krol 1999]. However, the overall comparison of the achievements of AI&R as related to human capabilities so far is not in favor of machines.

Let us consider some of the areas in which AI&R and humans can be compared. Here, we go beyond the traditional juxtaposition of natural and artificial logical-linguistic abilities. We are also involved in the psychomotor functions and the limits of their implementation in artificial systems. This is due to the fact that intellectual activity itself has a psychomotor basis. So, it is well known that the development of motor functions in children is one of the key factors for further development of intelligence in general. This is confirmed by our knowledge of general evolution, which demonstrates a close relationship between motor and mental functions. That is why the emergence and development of the psyche occurred precisely in those complex organisms that actively moved in their ambient environment [Dubrovsky 2018].

Comparing the capabilities of AI&R and a human, we come to the conclusion that man still surpasses artificial systems in most respects. For example, a human person beats AI&R in contextualization of relations between things, in forecasting that is necessary for understanding of cause-effect relationships, planning and decision making, in fine motor skills, in dynamics and movement, in weightlifting, and issues of energy efficiency. Here, a human is inferior in

two aspects: perception and categorization of the surrounding world and accuracy (recurrence) of movements. According to the data presented in the AI Index Report [Shoham Y. et al. 2018], the accuracy of pattern category recognition using systems based on deep learning neural networks statistically exceeded the level of a human person in year 2015. Further steady progress has been achieved in the quality of pattern recognition by machines. The comparative accuracy (repeatability) of the movements of humans and robots is measured in biomechanics and in ergonomics, but in general we can say that the error in movements of modern robots in the same point from the same direction and under the same conditions does not exceed 0.1 mm, considering several thousand recurrences. A human person cannot achieve such accuracy, because s/he will need breaks for rest, sleep, food, etc.

**Engineering implementations of the Turing test (2014–2019)**

The transition from abstract dreams to engineering solutions has more than once helped mankind to overcome major limitations. For example, at the dawn of aviation development, it was record flights across the English Channel, then across the Atlantic and the Arctic Ocean that made transport (and, indirectly, our entire modern civilization) develop in the direction that is now familiar to us. Alan Turing may not have drawn such a parallel, when he proposed a "simulation game" to determine the thinking ability of machines, but transposition of his idea into development of technology was inevitable. A number of technological contests are held in the world, which are aimed at checking TT implementations on the base of a simulation game. Such contests can be divided into two types: classical (they implement TT of type 2 [Alekseev 2013; Warwick & Shah 2016]) and non-classical ones (these implement various further modifications of TT).

An example of the classical approach to TT is the experiment conducted by K. Warwick and H. Shah in the Royal Society in London in 2014 and described in various works (e.g., in [Warwick & Shah 2016]). There is also criticism of this approach to conducting the

test, even earlier presented in a number of works, for example, in the book *The Turing Test: The Elusive Standard of Artificial Intelligence* [Moor 2003]. Despite known criticism of this approach, the test conducted in 2014 exactly reproduced the conditions presented by A. Turing in his article "Computing Machinery and Intelligence" [Turing 1950]. To put things in a nutshell, these conditions are as follows (the terms follow [Alekseev 2013]): a judge (J.) must incorrectly identify a man (M.) or a computer (C.) in more than 30% of cases, following a dialogue that lasts five minutes. Over 30 researchers took part in this test, including the author himself (as J.). As a result, the (C.) chatbot Eugene Goostman, simulating a teenager from Odessa, managed to deceive 33% of J.'s, as compared to M. Despite the fact that this achievement drew largely fair criticism, it should be recognized as significant, since it indirectly testifies to the irrelevance of the classical approach and the literal behaviorist interpretation of the Turing test. After all, passing the test does not lead to new breakthroughs in the field of development of artificial intelligence, whereas, for example, in the case of transatlantic flights at the beginning of the 20$^{th}$ century, or space flights of the middle of the same century, solving key industry problems led to a breakthrough in new areas of science and technology.

Non-classical interpretations of the Turing test, such as tests carried out in Skolkovo in 2015 and in 2019 at the OpenTalks.AI conference, were rather more promotion activities than an attempt to determine the capabilities of the machine for dialogue with a person. In particular, the authors of the test in 2019 emphasize that during the dialogues, M. often deliberately imitated the style of the chatbot dialogue in order to mislead J. In fact, this does not prove an increase in the intelligence of the machine but a conscious decrease in the intelligence of the Human (M.) in the eyes of the Judge [Nanosemantics 2019]. In addition, instead of five minutes of communication, the organizers limited J.'s conversation with C. or M. to only twelve cues, which is clearly not enough to recognize C. or M. through the dialogue. It is precisely by these two circumstances that we can account for the threshold of deception in this test, which

was 58% (i.e., in 58% of all these cases the C. managed to pass off his identity as M. in the eyes of J.). In other words, if M. talking to AI&R ceases to communicate in his usual way and begins to speak in a command-order modality, AI&R will paradoxically begin to understand the man better, and the man, realizing that he is talking with a chatbot, will experience much less frustration than in conversation with the chatbot (robot) acting as a human. Intuitively, we can assume that we are at the very beginning of a special simplified language of communication between man and machine, which is deeply combined, in its linguistic (words and their meanings) and non-linguistic features (sounds, light signals). This assumption may be the subject of a very interesting study, if carefully planned.

## If you are afraid of a robotic rebellion, then just close the door in front of them (DARPA Robotics Challenge)

The behaviorist approach to assessing the achievements of modern robotics remains an everlasting intellectual paradigm, as there is always a desire to answer the question of whether the robot can perform actions at the same level or better than a human. This is justified, because one of the main reasons for the introduction of robotics is preservation of human lives during elimination of natural disasters, such as nuclear explosions. The accident and subsequent tragic events at the Fukushima nuclear power plant in Japan in 2011 served as a prerequisite for launching the Robotics Challenge technology contest, held by the American Advanced Defense Research Agency DARPA [Krotkov et al. 2016]. The essence of the task was that robots must perform actions similar to those performed by human rescuers in disaster management: drive a vehicle on your own, get out of it, open a door, enter a room, find a tool, perform actions with it, go through a blockage, and climb stairs. For any adult, these actions do not require any effort and the obstacle course designed by the organizers can be completed by a person without any preparation in a few minutes. The final stage of the contest was held in 2015 in the California and showed that robots needed much longer than 40 minutes (the best result was 44 minutes), in order

to successfully complete all the nine tasks. The main conclusions about the status of robotics based on the results of this engineering contest were as follows:

Firstly, fine motor skills and locomotion of robots do not exceed the level of a one-year-old child. The most difficult tasks for all robots were to open the door and to climb the stairs. These tasks do not cause problems for any fit people, but they seem incredibly complex to robots. That is why one of the journalists advises: if one of robots "goes berserk, here's a useful tactic: Shut the door behind you out" [Hernandez 2017].

Secondly, the robots that participated in the competition had very limited autonomy, and most of the time they were in a remote control mode, i.e., in fact, they did not differ from radio-controlled toys or more complex machines. On their own, without help of humans, the robots were unable to complete any tasks. This raises the question of the need to increase the degree of autonomy and intelligence of robots when conducting their tests in real conditions, and not during studio shooting (where everything is usually fine).

**Basic contradictions in the development of AI&R technologies**

In its development, all technology moves along the lines of basic contradictions, or so-called "bearing contradictions" [Beskaravayny 2018]. In the development of AI&R, one can also see several areas of technological contradictions. The author does not offer an exhaustive list, yet each of the directions can be represented as a vector of development and a possible opposition between competing technologies. Sometimes, such basic contradictions are called the competition between *armor* and *projectile* (5) [Beskaravayny 2018]. Developing this metaphor, we will name a number of technological contradictions in the AI&R area, presenting them in the form of "armor and projectile."

*Mind or motor skill.* Here, the *psyche* can be considered as the *armor.* AI&R constantly has to improve its sensors and software, in order to better understand the world around. This allows AI&R to build adequate models and to take faster and more effective de-

cisions. The *projectile* here is a motor skill. Robots must learn to move around in their environment in the same way as humans and animals do, in order to learn about the objects of the world and to study them, constantly improving perception skills. Without fine motor skills for manipulation, robots cannot sense the world, like humans do. The confrontation between the *psyche* and the *motor skill* of robots is not new in robotics; it is known as the Moravec paradox [Moravec 1988]. The essence of the paradox is that it is relatively easy to reach the level of an adult in tasks such as an intelligence test or in a game of checkers, but it is difficult or even impossible to achieve the skills of a one-year-old child in tasks of perception or mobility.

*Computer power or logic rules.* Computer power can be considered as *armor* because it is necessary to constantly increase the amount of computer power available for processing deep neural networks. According to researchers of the OpenAI project, the growth in processor power used to train neural networks exceeds the growth rate of processor power according to Moore's law by about five times, and the generation of systems in 2019 exceeds the generation of systems in 2012 by 300K times. The *projectile* here is the logical rules that can be used to create programs that simulate intelligence, since the main problem of using neural networks and other statistical methods is a rational explanation of the results. Many studies indicate that the most effective way to achieve Artificial General Intelligence (AGI) is to create combined solutions that correlate logical approaches and deep learning neural networks [Brooks 2019]. It should be noted that Moore's law, in the form where it governed the entire computer industry for almost 60 years, has lost its relevance. The creation of AI&R technologies certainly relies on microprocessor technologies. However, where further breakthrough research will be carried out it will not be due to excessive computer capacities, but to various other ways that increase the efficiency of calculations outside hardware solutions.

*Digital tsunami vs. digital filter.* In this case, the digital tsunami plays the role of *armor*. The bottom line is that there is a widespread

misconception about the exponential increase in the amount of information accumulated by mankind. If it does grow, it remains linear and relates to the growth of mankind's population itself. However, the ability of mankind to transform its impressions of the surrounding world into a digital format is growing exponentially, thereby creating a digital cloud around the physical world (but the latter generally does not change much). In the case of a *digital filter*, the *projectile* is the ordering of information (filtering), which can be either a human function or a company's mission (Google). The problem of filtering the necessary information can also be reformulated as a problem of learning proceeding from small data, similar to the way young children (who need to be shown or told only once), and then they are able to repeat the example without large computational and, therefore, energetic costs.

*The symbiosis of man and machine.* Increasing complexity of technology and acceleration of life cycles (LCs) are observed in all areas of technology development, including AI&R. *The armor* here is the training process – the main way to adapt a person to changes. Practically in all occupations, people are made to undergo continuous training. Human education has not changed since the Stone Age or even since the appearance of speech. *Cyborgs,* people who have physical integration of their bodies and machines, living literally in symbiosis with a machine, are the only *projectile*, or a way to bridge the gap between man and smart things created by man. The most significant technological direction is the development of brain-computer interfaces, which will allow us not only to read human thoughts, but also to transfer new information to the brain at a great speed.

## The Turing test and the post-Turing era

In his work on the problem of *another* and the Turing test, A.Y. Alekseev cites the fictional story "Penfriend," which I would like to mention here as well, in order to illustrate the general idea. The essence of the story is as follows. A young man is looking for a life partner and, following today's practice, opens a certain Tinder

application on his smartphone. He sets the request parameters such as age, height, weight, hair color, her Zodiac sign, habits, etc. In response, he gets a choice of candidates. He starts his correspondence. He likes the girl Elisa, of all choices. The correspondence lasts a whole week – you cannot afford to make the wrong choice! They hold conversations on a variety of topics: family matters, children, sex, everyday life, leisure, cuisine, salaries, economics, politics, literature, etc. They play chess with varying degrees of success. The girl amazes the man with her knowledge, ease of communication, and sophistication of style. And her eyes are expressive, judging by the photograph she sent. Finally, the young man cannot stand it and makes a date in the square at the local monument. He brings a bunch of roses. He waits. An hour, two, three. No one is coming. Returning home, he writes an angry letter. In response, he receives this, "I am sorry, darling! I am Electronic Lisa, a computer program" [Alekseev 2008] (6).

It is quite easy to illustrate the whole of the post-Turing era with this example. The only change is that the pen-girlfriend *has attended* a date. The young man took her to a cafe, and there they began to talk. The conversation did not work out. There were no traces of education, humor or sophistication of style in the girl's conversation. Deeply disappointed, the young man left the cafe and opened the computer application where he had kept his correspondence with the girl. He opened her profile, went into a detailed description and only then noticed a small checkmark, "The user has activated the option of automatic responses using a *digital avatar*." Only now he realizes that the girl with whom he had been flirting all this week has never actually replied to him or even read his messages. All this was done by her digital counterpart, who was much smarter and literally knew what the young man wanted to hear, because the system had access to his profile and could adapt its replies to his expectations.

The Turing test can be interpreted as a *duck test* focused on AI&R: "If something speaks like a human person, acts like a person, behaves like a person, then it is a person." But the answers to

it can be fundamentally different. In particular, A.Y. Alekseev and D.I. Dubrovsky propose different ideas about the possibility of passing the Turing test. According to Alekseev, if a computer passes TT, then this AI will at best be a computer zombie [Alekseev 2008]. According to Dubrovsky, passing TT is impossible without transfer of subjective reality (i.e. consciousness) to a non-biological substrate. Agreeing with A.Y. Alekseev in the essential things, this author suggests rejecting negative connotation in the concept of "zombies," because these "zombies" can be quite useful and also pleasant in communication.

**Findings. Then, do chatbots dream of androids?**

The reasoning above leads us to three important conclusions, which may serve as the methodological basis for research in the field of artificial intelligence in the near future.

1. The physical interaction of the AI&R (computer) with the outside world is the main factor limiting the appearance of a general AI. Similar to the way a child's intelligence develops through active games and fine motor activities, computers (i.e., AI&R) must learn how to interact with the physical world, learning from humans and going through all the necessary phases of such learning. Learning based on small data and limited examples (show your child how to open one door, and then he will figure out how to open others) is the most promising way to increase the intellectualization of AI&R. During the last 70 years, research in the field of AI&R has been associated with development of algorithms (psyche). But so far, no significant progress has been made in the field of hardware for motor skills. It is reasonable to assume that, following the evolutionary approach, research should be focused on the development of "motility" to improve the "psyche" of AI&R.

2. Things around us are gradually learning to hear and understand us, and to respond to us. To Anton Chekhov and the audience of his play *The Cherry Orchard*, addressing a piece of furniture as "My dear and honored case" seemed an act of great transcendence, designed to emphasize that the world of things remains silent and

indifferent to the appeal of the personage. However, in our times, getting access to the wooden case (cabinet) is not only possible, but also quite relevant [Peckham 2019]. In fact, in the very near future, any device that can be embedded in things around us (and potentially, in people) will be built in (where this is commercially viable or improves security). Let us conduct a mental experiment, "a man with a stone ax." If, at the beginning of the Stone Age, a thinker of Alan Turing's level (and there certainly were such people among those who inhabited the Earth) wondered whether universal use of a stone ax was necessary, then after a few hundred years (progress then was much slower) he would have to abandon this idea, because, despite the fact that one type of ax could not satisfy all possible needs, the incredible variety and simplicity of manufacturing stone axes would make the "stone ax test" completely irrelevant. We can now transpose this analogy to the Turing test. While in the era when the computer remained a computer, i.e. as a separate subject, a thing that can be addressed personally (recall the famous salute "Computer!" from the sci-fi series *Star Trek*), such testing was possible, then later, where such technology has become widely spread (pervasive computing, universal computing, ubiquitous computing), such testing simply loses all meaning, as now people can choose from a vast variety of forms of communication, including creation of special simplified combined languages. In other words, the heroes of *Star Trek* could now provoke laughter if they started their request with this form of address, "Computer!" Each of our gadgets would think that this appeal applies to it (because there is a computer inside each of them) and would start to fulfill the request, each in its own way.

3. Besides the fact that computers are literally ubiquitous, they also acquire very personal properties in communicating with us. Returning to the analogy described above [Alekseev 2008], the electronic interlocutor, the 'girl' who turned the young man's head, was able to do this because she had access to his social profile, his entire search history, all the photos, including photos taken by his parents when he was only three years old. Also, the electronic girl (or rather, the AI system behind her personality) can access all his

transactions, including financial ones, made through authorized applications on the smartphone, visits to all restaurants, museums and parks, etc. Even now, our electronic interlocutors calling us from product support services of retail companies will often have information about us that is as personal as the information we share with close family members. Despite the fact that our electronic interlocutors can sometimes respond inappropriately and tend to use the telegraph-command style of speech, they know more about us and can be more useful than living people. This again raises the above issue: why can communication with an electronic interlocutor, despite all its limitations, be much more interesting, useful and pleasant than communication with a living person. An artistic illustration of this idea was provided in the film *Her* (2013). For the first time in the history of homo sapiens, we face devices that we have created, yet they know much more about our everyday life than we know ourselves. Moreover, this knowledge does not come from the experience of human-machine communication, but from the experience that we acquired in completely different circumstances. In other words, you do not even need to get acquainted with the device, because it already knows enough about you. Universal and instant knowledge about the interlocutor makes even minimal attempts to hide the fact that our interlocutor is a computer completely redundant. It is quite possible that in the near future we will be more willing to communicate with a computer. Thus, research efforts should not be focused on simulating human behavior in the implementation of AI&R, but rather on creating the most effective and ethical model of AI&R behavior in communication with a person.

The paragraphs 1–3 above fully confirm that we have entered the era of the development of AI&R which can be called "post-Turing." This era is characterized by a transition from attempts to simulate human intelligence, to creation of non-human intelligence, no less interesting and useful for humans. The efforts of researchers and engineers should be focused on the formation of AI&R models that evolutionarily master human skills and apply them in a format that complements rather than imitates humans.

Conversational AI (in chatbots) will gradually develop in the direction of their physical embodiment, as only the development of motor skills (manipulation, locomotion) can give impetus to the development of the psyche (algorithms and programs). Figuratively speaking, digital chatbots dream that they inhabit the body of an android (gynoid), and our task is to give them such a body, ensuring that we do not regret it later.

*NOTES*

(1) In the movie *Blade Runner* (1982), based on this novel, the action takes place in 2019, i.e. in our time.

(2) One of the most widespread examples of the application of artificial intelligence is chatbots (voice or text).

(3) One of the most popular ideas about a robot is anthropomorphic (humanoid) androids (ginoids).

(4) The concept of the "duck test" refers to the words of the 19th-century poet J.W. Riley: "When I see a bird that walks like a duck and swims like a duck and quacks like a duck, I call that bird a duck."

(5) The competition between armor and projectile is a metaphor for the constant competition of confronting technologies. The origin of the metaphor is associated with the confrontation of military equipment: an increase in the power of artillery ammunition led to designing heavier tank armor, which in turn again demanded an increase in artillery power. The confrontation has its limitations (engine power of tanks, dimensions of artillery systems, use of alternative technologies, etc.). This metaphor illustrates continuous confrontation of approaches or technologies.

(6) Adapted to the realities of today.

*REFERENCES*

Alekseev A.Y. (2008) Another Problem in Computer Communications. *Russian Journal of Philosophical Sciences = Filosofskie nauki*. 2008. No. 6, pp. 114–136 (in Russian).

Alekseev A.Y. (2013) *The Comprehensive Turing Test: Philosophical, Methodological and Socio-Cultural Aspects.* Moscow: IInteLL (in Russian).

Beskaravayny S.S. (2018) *Genesis of Technology and Singularity.* Moscow: RIPOL classic (in Russian).

McAfee A. & Brynjolfsson E. (2017) *Machine, Platform, Crowd. Harnessing Our Digital Future* (Russian translation: Moscow: Mann, Ivanov and Ferber, 2019).

Brooks R. (2019) *A Better Lesson. Robots, AI, and Other Stuff.* Retrieved from https://rodneybrooks.com/a-better-lesson/

Deutsch D. (2012, October 3) Creative Blocks: The Very Laws of Physics Imply That Artificial Intelligence Must Be Possible. What's Holding Us Up? *Aeon*. Retrieved from https://aeon.co/essays/how-close-are-we-to-creating-artificial-intelligence

Dreyfus H.L. (1972) *What Computers Can't Do: The Limits of Artificial Intelligence* (Russian translation: Moscow: Progress, 1978).

Dreyfus H.L. (1991) *Being-in-the-World: A Commentary on Heidegger's Being and Time, Division I.* Cambridge, MA: The MIT Press.

Dubrovsky D. (2018) The Hard Problem of Consciousness. Theoretical Solution of Its Main Questions. *Neuroscience.* Vol. 6, no. 2, pp. 85–103.

Finn V.K. (2018) *Artificial Intelligence: Methodology, Application, Philosophy.* Moscow: Krasand (in Russian).

Fodor J. (1990) *Theory of Content and Other Issues.* Cambridge, MA: The MIT Press.

Hernandez D. (2017, November 10) How to Survive a Robot Apocalypse: Just Close the Door. *The Wall Street Journal.* Retrieved from https://www.wsj.com/articles/how-to-survive-a-robot-apocalypse-just-close-the-door-1510327719

Krol M. (1999) Have We Witnessed a Real-Life Turing Test? *Computer.* Vol. 32, no. 3, pp. 27–30.

Krotkov E., Hackett D., Jackel L., Perschbacher M., Pippine J., Strauss J., & Orlowski C. (2016) The DARPA Robotics Challenge Finals: Results and Perspectives. *Journal of Field Robotics.* Vol. 34, no. 2, pp. 229–240.

Moor J.H. (Ed.) (2003) *The Turing Test: The Elusive Standard of Artificial Intelligence.* Dordrecht: Kluwer Academic.

Moravec H. (1988) *Mind Children: The Future of Robot and Human Intelligence.* Cambridge, MA: Harvard University Press, 1988.

Peckham J. (2019, January 9) Your Mirror May Soon Be Able to Decide the Clothes You Wear. *Techradar.* Retrieved from https://www.techradar.com/news/your-mirror-may-soon-be-able-to-decide-the-clothes-you-wear

Purdy M. & Daugherty P. (2016) Why Artificial Intelligence is the Future of Growth. In: *Accenture*. Retrieved from: https://www.accenture.com/t20170524t055435__w__/ca-en/_acnmedia/pdf-52/accenture-why-ai-is-the-future-of-growth.pdf

Putnam H. (1975) *Philosophical Papers. Volume 2: Mind, Language and Reality.* Cambridge, UK: Cambridge University Press.

Rasmussen J. (1979) *On the Structure of Knowledge – A Morphology of Mental Models in a Man-Machine System Context.* Roskilde: Riso National Laboratory.

Shaw I.G.R. (2017) Robot Wars: US Empire and Geopolitics in the Robotic Age. *Security Dialogue.* Vol. 48, no. 5, pp. 451–470.

Shoham Y. et al. (2018) *The AI Index 2018 Annual Report.* Stanford, CA: AI Index Steering Committee. Retrieved from http://cdn.aiindex.org/2018/AI%20Index%202018%20Annual%20Report.pdf

Solow R.M. (1987, July 12) We'd Better Watch Out. *New York Times Book Review.* P. 36.

Turing A. (1950) Computing Machinery and Intelligence. *Mind.* Vol. 59, pp. 433–460.

Warwick K. & Shah H. (2016) *Turing's Imitation Game: Conversations with Unknown*. Cambridge: Cambridge University Press.