Post-turing Methodology: Breaking the Wall on the Way to Artificial General Intelligence

Albert Efimov[1,2,3(✉)] [iD]

[1] Sberbank Robotics Laboratory, Moscow, Russian Federation
makkawity@gmail.com
[2] Institute of Philosophy, Russian Academy of Sciences, Moscow, Russian Federation
[3] National Research Technology University "MISiS", Moscow, Russian Federation

Abstract. This article offers comprehensive criticism of the Turing test and develops quality criteria for new artificial general intelligence (AGI) assessment tests. It is shown that the prerequisites A. Turing drew upon when reducing personality and human consciousness to "suitable branches of thought" reflected the engineering level of his time. In fact, the Turing "imitation game" employed only symbolic communication and ignored the physical world. This paper suggests that by restricting thinking ability to symbolic systems alone Turing unknowingly constructed "the wall" that excludes any possibility of transition from a complex observable phenomenon to an abstract image or concept. It is, therefore, sensible to factor in new requirements for AI (artificial intelligence) maturity assessment when approaching the Turing test. Such AI must support all forms of communication with a human being, and it should be able to comprehend abstract images and specify concepts as well as participate in social practices.

Keywords: Artificial intelligence · Philosophy of artificial intelligence · Philosophy of mind

1 Introduction. Turing Methodology for Assessment of Artificial Intelligence (1950–2014)

Alan Turing, a British mathematician, laid in his works (1937–1952) a foundation for the research into what we now call "artificial intelligence" (AI) or "artificial general intelligence" (AGI). Relying on the new theory of computability and information, on the one hand, and on the first machines engineered for universal computing, on the other, Turing directly approached the difficult question, "Can machines think?". Certainly, he could not create a model that would completely describe human reasoning or even the work of the brain as a basis for thinking. There was an obvious lack of neurobiological data at that time. Therefore, he simplified the model by reducing it to a machine resembling a communicating person with "suitable branches of thought" as A. Turing put it [3].

This simplification became the basis for A. Turing's thesis about isomorphic features between thinking and computing: "If we consider the result of the work of calculators (that is people employed for computing) as intellectual, then why cannot we

make a similar assumption regarding machines that perform these operations faster than people?" [1].

In this work Turing was also the first one to analyze the role of "embodied intelligence". He believed that a certain creature equipped with microphones, television cameras and loudspeakers could be taught to walk while balancing its limbs and being equipped with a telecontrolled brain. Turing believed that if they had created such a "monster" based on the technologies available at the time, it would have been "certainly enormous" and would have posed a serious threat to the inhabitants. Thus, having recognized the ability to imitate humans as "embodied intelligence", Turing pointed out that "the creature would still have no access to food, sex, sport and many other simple human joys" [1]. As envisioned by Turing, future researchers had to focus on imitating human intelligence in the following five areas: (1) various games, such as chess, tic-tac-toe, poker, bridge; (2) learning languages; (3) translations from one language into another; (4) cryptography; (5) mathematics.

Of these five areas, Turing believed (4) was the most practically useful for AI [1]. Pointing out these exact areas of research has affected the entire subsequent course of AI development up until now; relatively homogeneous tasks, partially solved by Von Neumann's architecture computers, made it possible to obtain new results by simply speeding up computational capabilities. A certain developmental inertia emerged when enormous efforts were devoted to solving a very narrow range of tasks. Human thinking and society, however, deal with a much wider range of "puzzles". As a result, available software AI systems are used in various fields of application but still cannot be safely and applied in the real world for general cases. This builds up unfounded expectation from AI as we want general intelligence from systems which are not designed for the real world.

In his most frequently cited work Turing suggested playing an "imitation game", which, in essence, was an engineering solution to the problem of answering the question "Can a machine think?". Instead of working on definitions of what "machine intelligence" or human intelligence is, Turing proposed a "blind" comparison of a man's key intellectual ability – reasoning and lying – with the actions of a computer. The imitation game became the foundation of the Turing methodology for constructing AGI. In this paper, drawing on the original work by Turing and applying the descriptive methodology proposed by A. Alekseev in [2], we will briefly look into the scheme proposed by Turing.

Having set the directions of the research (languages, translations, games, cryptography and mathematics) in his previous works, in 1950 [3] A. Turing proposed a methodology for determining the achievement of the final result. Only in the mid-1970s this methodology came to be called the Turing test, although essentially it remained the methodology for determining the achievement of the final result (definition-of-done) in the AI research program.

AI researchers and philosophers have been developing various methodologies that could become foundations for a more advanced methodology than that of the Turing test. Unfortunately, in the pursuit of designing more adequate tests, the researchers have been overlooking some important details in the methodology proposed by Turing. This paper attempts to address this shortcoming.

2 Methodology for the Critical Analysis of the Turing Test

After the introduction it seems necessary to indicate the main methodological difficulties in the modern assessment of the Turing test:

a) The test has grown so popular that it pushes many researchers towards a simplified version: "within 5 minutes of a telephone talk you must understand whether you are talking to a machine or a person";
b) any scientific research requires simple and transparent testing, yet a reliable assessment of human consciousness and intelligence is still under debate. Nevertheless, all engineering products tend to be tested, and since "AI" is most often presented in the form of software products, the test boils down to communication with the software. This has formed the perceptive inertia for "intelligent machines".

If we turn to the Turing's methodology proper, it is necessary to pay attention to the following three aspects that are important for our subsequent considerations.

Firstly, all the five areas of research originally proposed by Turing (like chess) are more suitable than others (like gymnastics) to the symbolic approach as we communicate them through symbols to one another and subsequently to machines.

The evolution of digital computers over the last seven decades since Turing original proposal has greatly expanded the scope of their application, but it did not change the approach which still relies on the primitive Turing machines working with symbolic systems. It is the speed of symbolic processing that has changed. As D. Dennett put it, "All the improvements in computers since Turing invented his imaginary paper-tape machines are simply ways of making them faster" [4].

Secondly, the Turing methodology always implies a wall separating the two key participants. All subsequent modifications of the Turing methodology that arose after 1952 implied a comparison by a Judge (J) of the activities by a Human (H) and a Computer (C), but their activities were always separated by an impenetrable wall. J was the only one who interacted with C or H through the "Turing Wall" which was transparent only to symbolic communication. But H and C did not communicate at all and did not solve any problems together.

Thirdly, Turing believed that the problem was "mainly that of programming", and he did not consider the need to accelerate the operating speed of digital computers in order to solve the problem of the "imitation game". In other words, Turing saw the task of creating AI as designing a system of abstractions that could recognize and take into account all the nuances of human communication. Turing was fully aware of the problem of a multi-level symbolic game, noting that an interlocutor's task lies in the most complicated field, noting that it "seems however to depend rather too much on sense organs and locomotion to be feasible" [1]. Unfortunately, this remark was largely overlooked by the subsequent generations of researchers, who considered linguistic behavior and the ability to play games to be enough of an intelligence indicator and took for granted the study of imitating the reasoning of a person or of the ability to play games. Here, we can see the emergence of a paradox: on the one hand, these three aspects of the methodology proposed by A. Turing constituted the cornerstone of all research between 1950–2014 aimed at implementation of "artificial intelligence"; on the other hand, this

methodology was insufficient to solve a whole set of problems that "natural intelligence" solves. Thus, it seems, the Turing test should not be chosen as a reliable criterion for creating "artificial intelligence". All the five Turing's areas of research require solving calculation tasks, whereas human intelligence is not limited to information processing, but also includes formulation of new concepts and finding certain patterns of objects through observation (without necessarily fixating all the rest).

Nevertheless, the Turing methodology has become the basis for a huge family of various AI tests. It is similar to the mechanistic materialism of the 18th century: initially limited, it, nevertheless, made it possible to solve a whole class of specific problems [3].

The object of this article is to make a step forward from the Turing test as a criterion for creating a mature AI. It is necessary to show the fundamental limitations of the Turing methodology and develop an approach to assessing the tests created for situations that are not supposed to pass the Turing test.

The subject of the article is to reject the consciousness modelling paradigm that was based on the use of symbolic systems alone, as well as to reject the contradiction of new approaches in AI assessment with the neopositivist foundations of the Turing test.

Our criterion comes down to a more complete assessment of a personality and agency of an individual.

3   The Continuum of Turing-Like Tests and Its Limitations

Almost seventy years have passed since Turing expressed his revolutionary philosophical ideas about the possibility of creating "thinking machines" in his fundamental work published in the journal Mind [3]. Several generations of mathematicians, philosophers and researchers of AI have devoted multiple articles to his mental experiments. As a result, a whole set of Turing-like tests have been designed. However, if one carefully considers this set of mental experiments and engineering solutions aimed at determining the definition-of-done approach to AI (summarized in Alekseev's work [3]), one can identify two axes that are orthogonal to each other, and we call them the dimensions of the "Turing-like testing continuum". All tests are grouped around them.

3.1   From Verbal to Non-verbal

Verbal interaction with AI involves the exchange of meaningful information messages, abstractions and images in a specific linguistic context. The meaning of the messages is set precisely by their verbal semantics. These messages can refer to everyday life ("What day is it today?") or bear imaginative content ("What if the universe were closed?").

Non-verbal (one might say, non-linguistic) interaction with AI involves the exchange of information messages without using a language. This may include facial expressions, gestures, movements, motor skills and even emotions that are expressed in specific actions (laughter, crying, sadness, suffering).

3.2   From Virtual to Physical

Virtual interaction with AI happens exclusively via computer interfaces available to us, including traditional (and becoming outdated) hardware such as monitor displays, keyboards, augmented/virtual reality devices and even exciting brain-computer interfaces.

Physical interaction with AI (although the word "robot" can be used in this context meaning an "actuated computer with AI") occurs in the physical world and involves its active transformation by AI itself. It requires a specific ability to affect other physical objects. A robot operating in the kitchen can wash the dishes, an unmanned autonomous motorcar drives us from point A to point B. All these actions necessarily occur in the physical world.
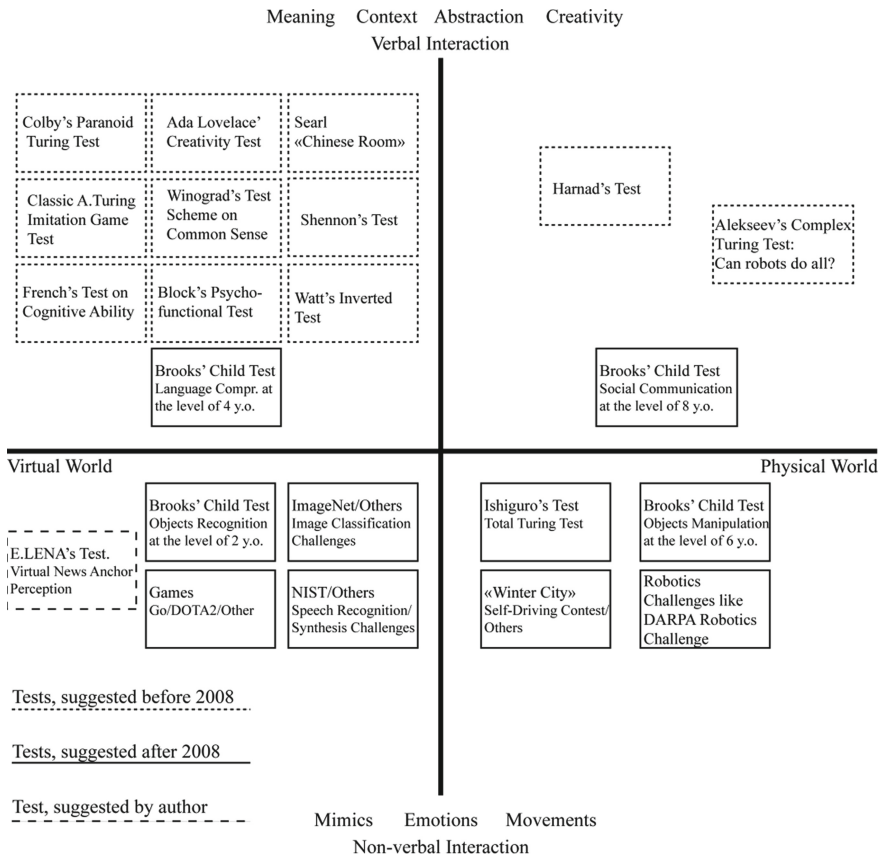

Fig. 1. Shows the continuum of Turing-like tests correlated on the virtual-physical and verbal-nonverbal axes

3.3 Four Areas for AGI Development

The two dimensions described above have given us four areas. Let us consider the four areas of this continuum as shown in Fig. 1 in more detail.

Verbal Interaction in the Virtual World. For historical reasons, most of the tests (mental experiments) developed before 2008 fall into this area. In fact, the classic Turing test, Lady Lovelace's creativity test, Colby's paranoid test, Shannon's social test, Watt's test (Turing's inverted test), Searle's Chinese room experiment, and Block's psycho-functional test are focused on testing verbal abilities in human/AI interaction. In this case, a person interacts with the virtual world environment (a display, a keyboard, a mouse).

Verbal Interaction in the Physical World. This area was not popular among researchers, as it was rejected by Turing from the outset. Only S. Harnad [5] and A. Alekseev [2] proposed complex tests demonstrating verbal interaction of humans and AI in the physical world. Although there is a related field of research where the emotional trace of the transmitted message and the study of its subtlest aspects are of great importance.

Non-verbal Interaction in the Virtual World. This area of the Turing-like tests continuum was overlooked by researchers for a long time, although it was Turing himself who, for the first time, drew attention to its importance for AI when he said that intelligent machines can play chess at the human level. After all a game (chess or any other) between AI and humans is a non-verbal manifestation of intellectual abilities in the virtual space. However, a game of chess remains to be a codified form of interaction. The next in the same area of this continuum are the tests related to recognition of images [6] and recognition or synthesis of human speech [7]. These tests, which played a huge role in the advancement of AI technologies, are nothing more than human-machine interaction in the virtual environment. In this case, AI does not change the physical world in any way, and at the same time there is no semantic verbal interaction; even in case with speech recognition a machine can only identify the correct words but does not understand their meanings.

Non-verbal Interaction in the Physical World. This area is the hardest to master for AI, since it depends the most on the development level of robotics, sensorics and AI technologies. If the virtual world possesses standard characteristics of the external environment, then the reality is inexhaustible, the role of chance is high, while abstracting is hampered. From the outset, this area has been ignored by researchers, including Turing himself, although its importance in human communication is emphasized by all researchers of communication. Ishiguro [8] suggests checking the technological maturity of robotics and AI by contrasting an android robot and a person in simple acts of communication: the robot only says the pre-programmed human phrases, even though bearing the maximum resemblance to a person. Another example of a test where AI and robots performed the tasks that people would generally do was the large-scale DARPA Robotics Challenge held in 2015. At this competition robots interacted with the physical world eliminating the consequences of a nuclear disaster at the training ground,

although there was no verbal communication with the people. The latest example of this is numerous driving contests where robots compete with humans in speed, accuracy and safety [9].

In 2018, R. Brooks [10] suggested a number of new tests for AGI. He proposed to see child capabilities as an indicator of technological achievement in AGI and robotics, drifting away from the Turing "conversational" paradigm of AGI and people communicating through walls. He called it "a competency-based" approach: (1) robots should be taught to recognize any objects in the physical world at least at the level of a two-year-old child; (2) robots should be taught to recognize natural language at least at the level of a four-year-old child; (3) robots should possess manual dexterity and fine motor skills of at least a six-year-old child; (4) robots should have social communication skills of at least an eight-year-old child.

With these requirements in view, the Brooks' test is divided into four parts (1–4) and is placed sequentially in all the areas of the Turing-like test continuum in Fig. 1.

E.LENA Test. In 2019, a specialized platform was developed at Sberbank Robotics Laboratory in order to convert text into a video image of a television presenter. The platform is called E.LENA (*Electronic* Lena) [11]. The idea of assigning visual forms to AI first became popular in science fiction. Yet, researchers did not embrace AI visualization as an object of study, since the appropriate technology has not existed up until now. We are the first to propose a perception test for identification of a digital television announcer by comparing it to a human announcer. This approach helps researchers to embrace a twofold improvement of AI technology – while testing is being done on the verbal interaction in the virtual world, it is simultaneously conducted in the non-verbal-virtual world.

We need to emphasize the two observations from above. Firstly, the majority of tests invented by the researchers, starting with A. Turing, implied performance in one specific area, which, according to the researchers, was best suited to the task of creating AGI. Setting tests' goals for engineering research by designing 'definition of done' for AGI (the best performance of certain robots or AGI in one of the four particular areas) defined their approach to designing programs, computers architectures and robots. Researchers and engineers build machines that perform at their best only in one specific area (like verbal interaction in the virtual world): the technology and computer architecture used for a chat-bot that excels in deceiving humans are utterly useless for a self-driving application. Various AGI/AI systems are designed and evolve only within their enclosed areas separated by the Turing walls from other areas of application.

Secondly, the Turing wall separating the subject of the test (a human judge) from the test object (a computer, a robot) only continued to solidify. Researchers could not even think of a computer/robot meeting face-to-face and interacting with each other (a typical estimate of the timing of an AI creation considers the time-out of this event, but not the specifics of programming or computer architecture [12, 13]). A computer or a robot compete with a human in each of these areas. If AI is doing better than a tested human, then we have arrived to our goal.

To sum up, each of the tests from the past seventy years has only strengthened the Turing wall, which separated the area of verbal-virtual communication between a machine and a person from the huge and incredibly unpredictable world beyond this

wall. This leads to a situation where human knowledge and experience mastered by AI in one area (non-verbal in the virtual world) cannot be transferred to another area (non-verbal in the physical world) because they are ultimately separated by 'the Turing wall'. By original design, our AI systems do not have the capability to learn and act in more than one of the areas from Fig. 1. All these concerns are the deficiencies of the Turing methodology.

4   Empirical Identification of Inadequacy of the Turing Test

Over the past ten years two important trends have shattered the Turing wall so much that it gave a deep crack and is about to collapse.

The first trend became obvious in the summer of 2014, when the Royal Society in London carried out the "Turing test" competition. The winner was a chatbot named Eugene Goostman that imitated the identity of a thirteen-year-old boy from Odessa. This chatbot fooled over 30% of the judges.

This Turing-inspired test invoked much criticism. The main point of it was that despite overcoming the symbolic barrier in deceiving people no significant breakthrough occurred either in research or in applied technologies: chatbots still remained quite limited in their capabilities, so declaring that they understand a person is possible only in a figurative sense. According to the cognitive scientist G. Marcus, this test did not show that one can consider AI as created, but merely revealed "the ease with which we can fool others" [14], thus reducing the Turing test to a psychological measure of human narcissism, rather than of AI development. Chatbots can go off topic embarrassing the interlocutor and thereby giving themselves away. The philosopher A. Sloman speaks about the irrelevance of the Turing test method as a behavioristic approach to assessing the intelligence of any system, as well as to assessing the solvability of any true problem [15].

In other words, chatbots outplay humans when dealing exclusively with abstractions, but the concretization of the gain and its correlation with reality is only possible with human intervention. Chess programs or chatbots have been beating humans in purely symbolic competitions for several years now. But they do not become full-fledged agents, and they cannot adapt the skills they acquired to other tasks like driving.

The second trend relies on the popular approach based on "brute force" and "greedy" (for data) neural networks but it will not help to answer the original question "Can a machine think?". Let's conduct a mental experiment which we might call an "ultimate imitation game". Suppose that we have limitless computing power and our neural network architecture is capable of processing texts without human supervisors (this condition does not alter the results but makes the experiment longer). Then, imagine that we have managed to recruit (for a short time) volunteers to imitate all men and women of the Earth and have divided them into two groups. The first group will consist of an equal number of men and women, and the second group will consist of men or women acting as judges (the gender does not matter here). If we assume that the number of adult inhabitants of the Earth is 6 billion, then there will be exactly 4 billion people in the first group (equally men and women) and 2 billion people in the group of judges. After that

both groups begin playing the classic imitation game and record all their dialogues and results with the judges. Now, let's suppose that we have all the computing power for an unsupervised deep learning neural network which enables us to train a neural network to answer any conceivable question based on the previous imitation games. It seems likely that if such a computer starts a game in tandem with a woman and claims to be a woman (as described above, following A. Turing), the judge will most likely be unable to distinguish the computer from a woman, and the judge will be equally likely able to identify the AI or the person in this game. Will this mean that the Turing's criteria are observed, and the true General AI is achieved? It does not seem so, since Turing said that a computer should imitate the reasoning of a man who is pretending to be a woman. In this mental experiment the computer is literally reproducing some of the most successful phrases of men who managed to fool the judges and won the game. However, this computer is uncapable of acquiring any "reasoning" faculty. It only demonstrates the ability to quickly find a relevant phrase based on the training set. As a result, this mental experiment supplies us with a dialogue interface capable of skillful imitation, but the computer interface is completely devoid of intelligence.

It seems that this conclusion of the mental experiment is the main reason why the approach based on the Turing method (the Turing test) ceases to be relevant and should give way to another approach based on a post-Turing methodology.

5   Post-turing Methodology Principles for the Study of AI

It seems quite logical to establish a new methodology for assessing the achievements in AI by taking into account both the experience of the last seventy years and the newer technological capabilities. In fact, the first attempts were made right after the 2014 Turing test competition in London [16–22]. However, they are all lacking a practical implementation across the entire Turing continuum, outlined in Fig. 1.

Firstly, in our concept of an intelligent computer we should reject anthropomorphism. The wall constructed by Turing is bound to separate the J and the tested H or C and essentially stimulates a person to evaluate AI in contrast to oneself, creating excessive technological anthropomorphism. However, man has learned how to fly by using the technologies that were totally different from the bird wings. Creating AI capable of reasoning and communicating like a person is probably not the most potent answer to the Turing's question, "Can machines think?". It is counterproductive to discuss the ethical limitations of precisely humanoid robots [23]. If we evaluate the design of modern robots, then the simplest question – "How many fingers should a manipulator hand have?" – can generate multiple answers, and the two-finger solution becomes a widespread type of "hand" [24].

Secondly, we can talk about a variety of forms and methods of cognition available to computers. AI should use abstraction and concretization on a broad scale. Here, the ideal is an independent formulation of new concepts and modeling of its own worldview – of course, with restrictions considering human safety. Now numerous attempts are being made not only to improve recognition of images but also, on the basis of I. Lakatos' theory of games and concepts, to compile a conceptual apparatus for a more flexible interaction of computers and mathematicians [25].

Thirdly, there should be a diversity of the same forms of communication that are available to humans. Machines have widely mastered computerized communication in symbolic structures, while robots' motor skills remain imperfect. *Virtual-non-verbal*, *physical-non-verbal* and *physical-verbal interactions* are still hampered. Probably, the ideal that machines should strive for is an emotionally colored communication involving "the five senses", so that a robot could convey information in any set of sensations available to humans. Here, a good example would be an automated translation from the sign language of the deaf to the test and vice versa. For now, we can only see it on the displays, but it should soon become accessible to robot operators.

Fourthly, a robot should participate in human social practices as a junior partner, but nonetheless possessing an agency. R. Brooks in his tests compared AI with the levels of child development – yet still what could be a better assessment criterion for communication skills than life in society? After all, child development is inseparable from socialization.

As to the Turing-like tests continuum in Fig. 1, we should advise other researchers and engineers to design and develop AI (be it robots or AI-enabled computers) capable of attaining to the human expertise and acting similarly to humans in more than one area. This approach breaks the walls between the areas and makes AI more useful and robust for real life applications as well as useful for human-to-machine interactions. Moreover, the post-Turing methodology requires no blind comparison of a human and machine performance (like in the Turing test) but demands a higher overall performance from a human and a machine learning and acting together.

6 Conclusion

The Turing test has virtually lost its relevance and meaning as even computer software falling short of being called AGI in the full sense of the word can pass such tests in systems of symbolic communication. Moreover, applications can practice abstraction only in minimal forms, which puts a limitation on their cognitive abilities.

Overcoming anthropomorphism and the Turing approach to assessing AGI will allow us to focus on creating the systems that can demonstrate various skills in the four main areas: shaping the system for labor operations; proper formulation of new concepts (abstracting) and their use (concretization); communication with a person involving all the five senses; and, finally, possessing a personal social agency.

The suggested post-Turing methodology might be a good foundation for the future research and engineering efforts because it does not oppose a human to a machine but makes a human and a machine act together in various areas of their interaction irrespective of either the physical or the virtual worlds. Such approach will provide more safety and security for the humankind as the advent of artificial general intelligence is inevitable.

References

1. Turing, A.: Intelligent machinery (1948). In: Alan Turing: His Work and Impact, pp. 501–516. Elsevier (2013)
2. Alekseev, A.: Kompleksnyj test T'juringa: filosofsko-metodologicheskie i socio-kul'turnye aspekty [Turing Comprehensive Test: Philosophical, Methodological and Socio-Cultural Aspects]. IInteLL, Moscow (2013). (in Russian)
3. Turing, A.: Computing machinery and intelligence. Mind **59**, 433–460 (1950)
4. Dennett, D.: Intuition Pumps and Other Tools for Thinking, 1st edn. W.W. Norton, New York (2013)
5. Harnad, S.: Minds, machines and Turing: the indistinguishability of indistinguishables. J. Logic Lang. Inf. **9**(4), 425–445 (2000)
6. Stanford Vision Lab ImageNet. http://www.image-net.org. Accessed 28 Sept 2019
7. NIST 2019 Speaker Recognition Evaluation. https://www.nist.gov/itl/iad/mig/nist-2019-speaker-recognition-evaluation. Accessed 02 May 2020
8. Ishiguro, H.: Toward a new cross-interdisciplinary framework. In: Thrun, S., Brooks, R., Durrant-Whyte, H. (eds.) Robotics Research. STAR, vol. 28, pp. 118–127. Springer, Berlin (2007). https://doi.org/10.1007/978-3-540-48113-3_11
9. Efimov, A.R.: Tehnologicheskie konkursy adresovany ne startapam, a konsorciumam [Technology contests are addressed not to startups, but to consortia]. https://sk.ru/news/b/articles/archive/2017/08/03/albert-efimov-tehnologicheskie-konkursy-adresovany-ne-startapam-a-konsorciumam.aspx. Accessed 28 Sept 2019. (in Russian)
10. Rodney Brooks' Homepage: Steps Toward Super Intelligence IV, Things to Work on Now. https://rodneybrooks.com/forai-steps-toward-super-intelligence-iv-things-to-work-on-now. Accessed 02 May 2020
11. Efimov, A.R.: Tekhnologicheskie predposylki nerazlichimosti cheloveka i ego komp'yuternoj imitacii [Technological background of the indistinguishability of man and his computer simulation]. Artif. Soc. **14**(4) (2019). https://doi.org/10.18254/s207751800007645-8. https://artsoc.jes.su/S207751800007645-8-1
12. Müller, Vincent C., Bostrom, N.: Future progress in artificial intelligence: a survey of expert opinion. In: Müller, V.C. (ed.) Fundamental Issues of Artificial Intelligence. SL, vol. 376, pp. 553–570. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-26485-1_33
13. Baum, S., Goertzel, B., Goertzel, G.: How long until human-level AI? Results from an expert assessment. Technol. Forecast. Soc. Change **78**(1), 185–195 (2011)
14. Marcus, G.: What comes after Turing test? New Yorker. https://www.newyorker.com/tech/annals-of-technology/what-comes-after-the-turing-test. Accessed 02 May 2020
15. Sloman, A.: Judging chatbots at Turing test 2014. University of Birmingham, School of Computer Science. http://www.cs.bham.ac.uk/research/projects/cogaff/misc/turing-test-2014.pdf
16. Marcus, G., Rossi, F., Veloso, M.: Beyond the Turing test. AI Mag. **37**(1), 3–4 (2016)
17. Clark, P., Etzioni, O.: My computer is an honor student — but how intelligent is it? Standardized tests as a measure of AI. AI Mag. **37**(1), 5–12 (2016)
18. Adams, S.S., Banavar, G., Campbell, M.: I-athlon: toward a multidimensional Turing test. AI Mag. **37**(1), 78–84 (2016)
19. Zitnick, C.L., Agrawal, A., Antol, S., Mitchell, M., Batra, D., Parikh, D.: Measuring machine intelligence through visual question answering. AI Mag. **37**(1), 63–72 (2016)
20. Morgenstern, L., Davis, E., Ortiz, C.: Planning, executing, and evaluating the winograd schema challenge. AI Mag. **37**(1), 50–54 (2016)
21. Kitano, H.: Artificial intelligence to win the Nobel prize and beyond: creating the engine for scientific discovery. AI Mag. **37**(1), 39–49 (2016)

22. Jarrold, W., Yeh, P.Z.: The social-emotional Turing challenge. AI Mag. **37**(1), 31–38 (2016)
23. Chakraborty, S.: Can humanoid robots be moral? Ethics Sci. Environ. Politics **18**, 49–60 (2018)
24. Birglen, L., Schlicht, T.: A statistical review of industrial robotic grippers. Robot. Comput.-Integr. Manuf. **49**, 88–97 (2018)
25. Pease, A., Lawrence, J., Budzynska, K., Corneli, J., Reed, C.: Lakatos-style collaborative mathematics through dialectical, structured and abstract argumentation. Artif. Intell. **246**, 181–219 (2017)