Journal of Digital Economy Research. 2023. 1. P. 61–89
DOI: 10.24833/14511791-2023-1-61-89

Received: 8 October 2022,
Revised 10 November,
Accepted: 3 December 2022

PROBLEMS OF ARTIFICIAL PERSONALITY (ARTIFICIAL INTELLIGENCE) CONTROL

by Oleg N. Gurov, Alexey V. Sherstov

Abstract. Today, a number of researchers representing both technical knowledge and the humanities believe that it is necessary to endow Artificial Intelligence with subjective "human" qualities, which include the ability to self-aware, as well as to make a free choice. In this regard, the problem of the AI autonomy becomes extremely relevant, and further – AI creator's rights and capabilities (or ineligibility) to hold control over AI. Within this framework the Artificial Personality project has been developing over the past 20 years. Given its active scientific and social activities with the involvement of the remarkable interdisciplinary community, the project is far from complete. The presented article summarizes the executed research for Artificial Personality conceptualization and demonstrates that today the fundamental possibility of the creation of Artificial Personality has not yet been convincingly proven. Also, conceptually, there has not been formulated the single generally accepted approach to promising methods and technology for the implementation and the embodiment of the Artificial Personality. So, at the current stage, the study of the Artificial Personality is rather abstract theoretical research. As a result of the study, the authors come to the conclusion that today it is reasonable to use the results of the Natural Personality and Natural Intelligence studies and transfer the methods that have shown their relative effectiveness in various existing manifestations of real social life to the field of creating the concept of Artificial Personality. The proposed approach for the conceptualization of Artificial Personality will help to create a theoretical and methodological foundation for theoretical research and further implementation of Artificial Personality projects.

Keywords: Artificial Intelligence, Artificial Personality, Turing test, AI ethics, Self, Identity, Natural Personality, Natural Intelligence

For citation: Gurov O.N., Sherstov A.V. (2023). Problems of Artificial Personality (Artificial Intelligence) Control. *Journal of Digital Economy Research,* vol. 1, no 1, pp. 61–89. (in English).
DOI: 10.24833/14511791-2023-1-61-89

Information about the authors:

Gurov N. Oleg – PhD in Philosophy, MBA. Senior lecturer at the Educational and Scientific Center for the Humanities and Social Sciences of the Moscow Institute of Physics and Technology, lecturer at the Moscow State Institute of International Relations of the Ministry of Foreign Affairs of Russia, Moscow State University. M.V. Lomonosov, RANEPA.
140180, Moscow region, Zhukovsky, st. Gagarina, 16 (FALT)
gurov-on@ranepa.ru

Sherstov V. Alexey – Master of Economics, PhD student Department of Philosophy of Politics and Law, Faculty of Philosophy, Lomonosov Moscow State University.
119991, Moscow, Leninskie gory, Moscow State University, "Shuvalovsky".
sherstov@miavend.ru

Acknowledgements:

The authors express their gratitude to Doctor of Philosophy, Professor A.Yu. Alekseev, as well as to students E. Nazarova, A. Slepysheva, Ya. Prourzin and A. Treskunova for a fruitful discussion and for help in conceptualizing the issues raised in this study.

Introduction

The rapid development and widespread use of Artificial intelligence projects makes us increasingly think not only about how "intelligent" it is (that is, how quickly and efficiently it is able to study, look for patterns and create connections to solve the proposed problems), but also how much it is independent and really "conscious". With this the modern society is becoming increasingly dependent on digital technologies. Humans are forced to rely on computer systems not only in the fields of production, education, business, but also often have to trust technologies, including those based on Artificial Intelligence, to solve and manage significant and large-scale problems - up to life and death issues [32]. Therefore, some researchers endow Artificial Intelligence with the qualities of a subject's (actor's) identity, including the ability for awareness and freedom of choice in decision-making. This inevitably raises the question of the limits of this independence and the ability of the creator of Artificial Intelligence to control his creation. The question of how far the Humanity can go in the development of Artificial Intelligence, and what we need to avoid losing control, ceases to be only the plot of science fiction films and moves into political, legal and actually practical areas [8, 30, 37].

Recently, machine learning scientists have been achieving more and more impressive results. In particular, the invention of new architectures of neural networks (especially BERT, which was released in 2018) and the growth of computing power have led to several breakthroughs in the field of natural language processing (NLP) at once: Artificial Intelligence results have significantly improved on many classical tasks, from various markups to machine translation, generation of meaningful texts and dialogues [36]. Trained on hundreds of gigabytes of texts from books, articles, posts on social networks, neural networks have become able to perfectly preserve the context, highlight the main thoughts and produce meaningful and much the same time interesting and unpredictable results. This is important because high similarity to human text is a priority, and human texts tend to be rather "unpredictable" in terms of perplexity, as opposed to typical machine-generated text, as linguists claim [26].

Such results cannot but make one think about the prospects for the near future [27]. It is quite possible that a further increase in computing power and more and more new architectures will lead to the fact that the neural network will no longer simply operate algorithmically with vector representations of lexemes. Something will appear that has a semblance of emotions, capable of something closer to human thinking [12, 34]. At some point, the line between Human and a similar Artificial Personality will blur so much that Humans themselves will start to recognize it of their own kind. The prerequisites for this are already happening in our time: as an example, we can take the incident with LaMDA, Google's AI speech imitation system [33]. This system is so advanced that one of the company's employees came to believe in its reasonableness and initiated the discussion of the "unethical performance of the company" [15, 35].

Dive into the problem of "consciousness" of Artificial Intelligence brings researchers to the Artificial Personality (or Artificial Intelligence – further we will be referring to this concept as to Artificial Personality) project. Can a computer become a subject? When discussing the topic of Artificial Personality, most scientists and experts proceed from the following premises: an artificial subject is a copy (imitation) of a natural subject and is created from a natural subject using computer technology, while the implementation methods are not limited.

The target of this work is to identify potential management and control problems of the Artificial Personality, based on the current level of knowledge and technology development, as well as on the scientific discussion on this issue.

The problem of defining an Artificial Personality

Since the mid-1990s, the Artificial Personality project has been widely discussed in the interdisciplinary scientific discourse. One of the leading experts, inspirers and significant scientists in this field is the famous Russian thinker, Doctor of Philosophy, A. Yu. Alekseev. This scientist identifies several established definitions of Artificial Personality, based on research position regarding the "personality" of cognitive-computer system [1, 39].

The classification is given in the context of the attitude to the concept of "Philosophical zombie", which is conceptually close or even identical to the concept of Artificial Personality. In general, philosophical zombies are "unconscious systems that are behaviorally, functionally and/or physically identical, indistinguishable and/or similar to conscious beings" [3, 40]. Among the researchers, there are conventional "zombiphiles" who are ready to discuss this theory, having their own view on the very concept of "philosophical zombie". They admit the existence of zombies, or at least the possibility of imagining them. That is, they assume that Humans will be able to create their functionally identical counterparts that do not have consciousness.

"Anti-zombists" consider even the idea of the possibility of zombies' existence as absurd.

And the third, neutral, group is trying, at least, to bring certainty into this confusing problem. "Neutrals" believe that a computer model should include a "pseudo-consciousness" module, a kind of analogue of human consciousness.

Some other researchers, the so-called "azombists", ignore the very issue of "consciousness" of Artificial systems. The main issue for them is the actual presence of a "believable imitation of a Human", and not the presence of characteristics that allow to personify it. According to Alekseev, the advantage of this approach to classification lies in the concretization of the sub'ject area: it is not "intelligence" that is analyzed, but more complex concepts – "consciousness", "personality", "self-consciousness", "Self". Next, we present a few of the main Artificial Personality models developed by researchers.

1. *Artificial Personality as the imitation of Natural Personality.*

This definition is typical for azombists. The concept is that the Artificial Subject is not different from the natural one: neither in function, nor in behavior, or even physically. An important condition that is characteristic of this approach is the indistinguishability of natural and artificial personalities. Both individuals are capable of passing Turing test with equal success [4]. This means that personalities are similar not only in the structure and functioning of external visible systems (physical, physiological-anatomical, verbal-communicative, etc.), but also in the inner spiritual system, which is more complex to reproduce, and which is responsible for abstract concepts inherent in emotional and the sensual, reflective side of the Human (i.e. "love", "responsibility", "the right", etc.) [5].

2. *Artificial Personality as the model of a Natural Personality.*

This concept implies the presence of "pseudo-conscious" complex for the Artificial Personality, which is used to effectively manipulate the "data" and "knowledge" of the intellectual system. By the same analogy, Human may experience and be aware of pain, which is the work of the effective signaling mechanism in case there are problems with the health of body.

3. *Artificial Personality as the reproduction of Natural Personality.*

According to this definition, computer system actually reproduces general and individual phenomena of consciousness through the implementation of a complex functional dependence of neurophysiological codes of subjective reality on a substrate that is invariant with respect to the physiological structure of the human brain. This approach is relevant for anti-zombists. From their point of view, zombie is a creature that is functionally completely identical to Human, but completely devoid of consciousness. Dubrovsky defines an important condition for the emergence of Artificial Personality – the "Self" [10]. It includes two important levels that actually shape the personality: genetic and biographical. However, if the genetic level can be easily faked or reproduced, the biographical level is the experience that must be obtained, comprehended and preserved. In this regard, researchers are faced with the question of whether such a robot, whose experience is completely created by programmer, will be considered a subject, or is it just a reflection of the subjective world and the mindset of its creators [14].

4. Artificial Personality – as the creation (formation) of "superpersonality".

According to D. Denette the Artificial Personality creates such qualities (phenomena) that have no analogue with the natural one. The latter definition leads to the idea of "superpersonality", implying that as a result of the fact that computer technologies allow the integration of knowledge and achieve the appearance of "meaning" (in the highest philosophical meaning), this level and scope of meaning exceeds the usual level of "knowledge" of ordinary Humans [7].

It is worth noting that, in our opinion, the reproduction and creation concepts of the Artificial Personality cannot be used for productive research on the prospects for the formation and management of the Artificial Personality, and the significance of these concepts is more broadly philosophical than practically scientific.

Is it possible in principle to give an accurate and unified definition of an Artificial Personality? This seems to be a difficult, if not impossible task at this stage. For example, if we take the definition of its prototype as a starting point, but already at this, initial step, the discussion encounters an almost insurmountable obstacle, since there is still no single, universally recognized understanding of the phenomenon of the Human Personality [11]!

Alekseev emphasizes the "linguistic disunity" of modern projects of the Artificial Personality, and the reason lies in the fundamental impossibility of creating the unified definition of "Human Personality" [2]. But this is not the only issue. The other side of the problem is the variety of ideas about the ways of computer realization of personological phenomena. In this regard, it is practically impossible to accurately answer the question whether, in principle, an exact repetition of the human psyche is real (not only at the formal, but also at the functional and semantic levels).

Basic line of approach to architect Artificial Personality

Scholars give accents to two major directions. In one respect, Artificial Personality may be determined as as a robot endowed with pseudo-consciousness, which makes it functional similarity of human subjective consciousness of reality, and on the other hand, Artificial Personality can be perceived as an expert system with "sense" mechanisms [2, 25].

The first approach involves appeal to the idea of anthropopathism, i.e., in fact, Artificial Personality is endowed with the properties of the human psyche. However, the robot, despite the complete identity in actions and feelings, will never become Human, that is, a Natural Personality [16]. The idea is very organically correlated with the reproductive concept and helps to develop it. Dubrovsky more than once drew attention to the fact that there is an important factor of "Self", which at this stage of development cannot be copied [9, 25]. The result of formal copying of experience is devoid of a large component - feelings and emotions. It is they that allow a Natural Personality, that is, a Subject (Human), to fix memory, draw conclusions from the results and, most importantly, evaluate lessons learned and the experience gained. The assessment in this

case is not numbers and dependencies, but mostly moral satisfaction with the outcome [23]. But the machine is most likely not capable of experiencing feelings, it can only imitate them. Will such an imitation be equivalent to a real human one? To answer these questions, one needs to clearly understand: what does Human give and get, what is that "satisfaction", how to measure it, and most importantly - by what parameters can it be compared with the "satisfaction" of Machine?

The second approach, considering that it involves certain methodological ambiguities, requires return to the topic of defining concepts. It is worth mentioning that many researchers have addressed the problem of modeling "sense". The most productive from the point of view of computer implementation, is the "contextual approach", according to which "meaning" is the context of formalized expert "knowledge" that has parameters of system unity [2]. Alekseev presents the formula of the Artificial Personality project – "meaning + understanding". But what is the "meaning" for the observer, and what is the "understanding" for the Machine itself? If the machine "understands", then what is the "meaning" for it?

In this context, the issue of developing Artificial Intelligence for communication with a Human is rather impressive but ambiguous. To illustrate, we offer the case of Tay Chatbot from Microsoft that is of particular interest [24]. Tay is a chatbot that educates and trains from the messages of the users it talks to and generates new replicas based on the data it receives [33]. Artificial Intelligence was faced with the task of identifying formal communicative and linguistic patterns (in other words, like a child has to learn grammar from scratch, learn certain lexical units, etc.) and apply them. The experiment showed that at first Tay wrote harmless and quite friendly messages to the interlocutors. However, after a few hours, the chatbot learned to write insults, Nazi statements, and even death wishes. Microsoft explained this behavior of Tay by the fact that a "planned attack" organized by Internet trolls was carried out on it. According to the corporation, the attackers took advantage of the "vulnerabilities" of the chatbot, which made it possible to train the algorithm to respond with insults. In turn, the company admitted that it had underestimated the social aspect in the development, and was more focused on the technical implementation and the reliability of the simulation of communication [28].

Is it possible in this case to say that the chatbot consciously responds with cruel expressions? Obviously not. This instance of Artificial Intelligence is not able to filter negative information from the positive one. There are several reasons for this: firstly, the creator did not take into account the peculiarities of Internet culture, which is why he allowed his algorithm to absorb any information indiscriminately. Secondly, Artificial Intelligence is not yet so intelligent as to learn the moral norms itself, which allow you to separate the permissible from the impermissible.

An important pattern is noticeable here - the current prototypes of the Artificial Personality are not independent. They still need a moderator (operator) who manually censors what is good and bad for the algorithm. At the same time, Human, on an instinctive level, is able to notice the reaction of other people and understand how per-

missible what he says is. The Machine has no such "social savvy". In fact, both "good" and "bad" are equivalent for it, it is not able to give them an emotional coloring and the meaning of these concepts is equivalent for the Machine [18]. The programmer simply puts into it the information that there are "Class 1" and "Class 2", which are equivalent for the Machine itself. However, due to the fact that the creator gave "Class 1" a higher value of w (weight), which is primarily a number (!), and a value less than w1 to "Class 2", the Machine ranks objects of the first class higher than objects of the second one, and even can simply completely exclude them from its memory. Simply put, it compares w1 with w2 and returns those results that have higher precedence (which makes the value of w).

In the above example, there is nothing about "satisfaction with the result". It is not the Machine which is "satisfied", but the Human who created this Machine. That is why, most likely, machines, with a complete imitation of a Natural Personality, will never be able to fully do exactly the same as a Human does.

Artificial intelligence is an algorithm that has a certain result for any set of input data. Artificial intelligence is about finding mathematical relationships between data, about the accuracy of measurements, the Machine does not have an error, it always knows the result with an accuracy of nano-units. Natural intelligence, on the other hand, seeks connections not only on the basis of mathematics, but also based on experience and common sense, since "natural" cannot be described mathematically accurately and there is always at least a tiny error. Mathematics is precision built on convention, on the "rounding" of the natural to the whole. This is a natural limitation to simplify understanding. Therefore, a personality created mathematically is simplified, generalized to something concrete, while Natural Personality exists according to the canonical laws of nature.

Discussions

The review of approaches and opinions presented above allows the authors to assert that the only constant in the presented problematics is uncertainty. In this case, first of all, we are talking about the lack of complete confidence that, in principle, it is possible to create an Artificial Personality. However, if we assume that this is still a feasible task, then we are faced with another uncertainty in the definition of what an Artificial Personality is. If we are talking about Artificial Intelligence in general, by analogy with Natural Intelligence, then we are faced with the fact that there is no common understanding of what Intelligence in general is, because representatives of various fields of knowledge and practice, carriers of different cultures put often contradictory characteristics into this concept. Transferring this problematic to the topic of the Artificial Personality, we to some extent produce even more uncertainty. However, this does not mean that we are multiplying entities, since at present Humanity is faced

with the need to simultaneously solve many problems - theoretical, methodological, practical ones, since the acceleration of time makes it necessary to increase the speed of development in all areas of social life, regardless of how large-scale the existing requirements are, and whatever complex systems it concerns.

Given that we have already found out that the main problem that we face when we develop on Artificial Personality is its uncertainty and the imperfection of attempts to implement it, therefore, at the moment it is impossible to identify the ways to operate it, because we simply do not definitely know what we have to manage.

1. Artificial Personality as a controlled subject which has all the features of the Natural Personality.

Firstly, if Human can be controlled both intuitively and based on the large sample of experiments and real cases, then there are no demonstrative cases of controlling Artificial Personality yet. Nevertheless, we can try to "equate" the Artificial Personality with the natural one for the purposes of discussion, and assume that the behavior of the AI, as an imitation of Human, will somehow resemble the behavior of Human. From this follows the first convention in the development of management methodology for an Artificial Personality, described above. The search for methods of managing Artificial Personality must be constantly adjusted: inefficient ones should be removed and new ones added for subsequent testing and application [29].

Further, identifying effective methods is an ongoing process. New trends in management are rapidly changing based on the constant changes in the world around us, so the search for working methods in relation to the Artificial Personality should be constantly considered, as well as screening for relevant methods for people.

When developing the first prototype of the methodology, it is worth remembering the frequent problems that arise in the people management (of Natural Personalities). It is easiest to take examples from business. Most often, at the beginning, when developing a strategy, it is necessary to take into account regional economic, political, social and, in particular, cultural characteristics. For example, Russian environment, including the business one, is characterized by the significant role of verbal agreements, which are relatively rarely secured by any documents. However, can such a problem arise for Artificial Personality? There is no single answer to this question, as it may change depending on our attitude towards the Artificial Personality. On the one hand, we can consider the Artificial Personality as a clone-child of its creator, then its features are transferred to the personality. As a result, the Artificial Personality turns out to be a "mirror" of a representative of a Natural Personality, that is, its detailed imitation. However, on the other hand, we can consider the Artificial Personality as an independent unit, which nevertheless received its own unique experience, was able to think it over and emotionally live it. In this case, several outcomes are also possible: either the features of the Artificial Personality will be unified and will not differ from one individual to another, or we will get new and possibly unexpected features that do not arise when interacting with Human [21].

It can already be noted that when applying this model to control Artificial Personality, we ignore possible inconsistencies in understanding the conditions of the task by Artificial Personality. Subsequently, this can lead, first of all, to incorrect results, and secondly, to "interpersonal" conflicts with the Machine. The last aspect is no less important than the first one: assuming that Artificial Personality will act identically to the Observer (Manager), the latter endows it with its own subjective personal qualities and tries to determine its attitude to various situations. However, in reality, they may not correlate, since, according to this definition of Artificial Personality, it is a "believable imitation of Human" and has internal spiritual and emotionally sensitive qualities that may differ significantly from the observer's point of view. At the same time, if in the end the Artificial Personality still does not have the internal culture of the creator, then we a priori declare that its own culture and understanding (or lack thereof) correlate with ours.

This leads us to the next thesis-limitation in the development of the Artificial Personality management methodology:

2. The internal culture of the Artificial Personality and its possible interpretations completely coincide with the interpretation of its manager.

The last convention is that for Artificial Intelligence there is no internal emotional difference between classes: one set of data for it is equivalent to the second one, etc. The set of weights created by the programmer becomes the basis for sorting and prioritizing certain values. That is, what is important or not important is determined by the Human Creator, and not by the Machine itself. As a result, we get that:

3. Human teaches Artificial Personality what is right and what is wrong.

Eventually, it turns out that Artificial Personality is completely dependent on Human. When we try to control her, we are essentially controlling a partial copy of ourselves, since many of the personality traits of the Artificial Personality are determined by its Manager and its Creator, and not by itself. That is, we get an ideal employee whose misfires can only be due to the personality traits of the Creator of Artificial Intelligence.

The problems presented above indicate that the philosophical and methodological understanding of Artificial Intelligence technologies and projects related to it requires special attention and is a promising area of scientific activity. It is especially important to comprehend the presented problems, as the responsibility for the development of technologies is already growing today. Uncertainty, misunderstanding of the fundamentals and lack of consensus on key aspects of the development of Artificial Intelligence technologies can cause threats from its use, and today it is required to formulate guarantees for the safe development and use of these systems. In this regard, it is important to timely and critically analyze the problems of Artificial Personality in order to prevent the negative consequences of Artificial Intelligence performance, at the same time, developing it for the benefit of society [19].

The fundamental possibilities of the technological implementation of the Artificial Personality project that have arisen today give rise to a wide range of sociocultural and humanitarian problems. The attention of many scientists is riveted to this issue, and therefore any format of scientific events at which participants are given the opportunity to express their point of view on the problem under discussion is very important.

In October 2022, major Humanities vs Sciences Congress was held, organized by the Moscow Institute of Physics and Technology. The main goal of the Congress was the synthesis of exact sciences and the humanities, highlighting the most advanced developments and technologies in the field of Artificial Intelligence, creating conditions for the development of the environment and Science Art and identifying new opportunities for open cooperation between them [17].

Within the framework of the Congress, the Panel Discussion "Humanitarian and technical in the project of Artificial Personality" was held. At this event, the possibilities and prospects of computer implementation (imitation, modeling, reproduction) of a wide range of personal phenomena were discussed, as well as the project of Artificial Personality as an Artificial Intelligence system that convincingly passes the Turing complex test for computer implementation of a wide range of Humanities phenomena – "Self", semantic, cognitive, existential, creative, communicative, motivational, strong-willed, moral, etc. Part of the presentations at the panel discussion was devoted to the history of creation, the current state and prospects for the technological implementation of the Artificial Personality project. The panel discussion was attended by programmers, mathematicians, engineers, as well as philosophers, psychologists, lawyers, linguists, political scientists, sociologists, culturologists, art historians and representatives of other sciences [13].

Thanks to the multidisciplinarity of the panel discussion, it became possible to systematically study both the general worldview and methodological problems of Artificial Personality, according to specific information technology, legal, ethical, aesthetic and other aspects of sciences [15, 20, 38].

Conclusions

At this point, in scientific discourse there is still no ground for agreement considering definition of Artificial Personality. Moreover, there is no consensus whether it is practically possible to create it in principle. At the current stage of scientific discussion, it is necessary to find a solution to many theoretical and methodological issues that lie in the interdisciplinary field.

To date, the situation in philosophical studies, which subject is Artificial Intelligence does not allow us to unambiguously determine the possibility and working tools for managing Artificial Personality. First, we do not fully know what Natural Personality is, and most likely we will never be able to unambiguously define what it is. From here it is impossible to give a single designation of Artificial Personality, which is the second problem. In addition, Artificial Personality can be practically implemented in a

variety of ways, which is why there are also many different interpretations. As a result, a situation is created when in the scientific community there are many interpretations of both conceptual understanding and different visions of development paths. Thus, we get an indefinite control object, which makes it impossible to say exactly which methods will be effective and productive for it.

However, based on the existing experience with Natural Personality, it is possible to suggest the use of methods that have shown their relative effectiveness in the most diverse situations of social life. At the same time, it should be taken into account that there is no guarantee that these methods will eventually become applicable, since, as mentioned above, Artificial Personality is only a theoretical project at this stage, and we do not have, first of all, statistical data that could allow us to predict the behavior of Artificial Personality in practice, and convincingly predict at least any approximate expected result [22].

To sum up, we would like to propose for discussion the hypothesis that the control problem may not arise at all: since Artificial Personality is an algorithm, it is mathematically limited: if we represent solutions in the form of a graph, then despite the huge number of possible solutions, this number is still finite, since each set of input data will have its own final answer. Human is capable to bypass the enumeration method, using internal sensations to ignore the mechanical selection of existing data to obtain a solution. To demonstrate the above assumption by an example, consider the following situation. You can teach Artificial Personality the basics of music (notes, rhythm, etc.) and make it write a melody. According to the laws of combinatorics, there is an infinite number of variations, and for sure among the selected compositions there will be exact copies of Tchaikovsky, Bach, Wagner and others. However, the machine "composed" this not because it has the inner ability to feel the harmony and compatibility in music, but because it can mathematically create any combination and select it from trillions of created ones. When a Human composes music, there is something inside him that allows him to intuitively find successful solutions and combine them into a composition. Human can create a single masterpiece, and at once it will be qualitative and brilliant, and Machine can create an infinite amount, but most of what is created will be mediocrity and cacophony.

In conclusion, the authors would like to present their general opinion regarding the prospects for the development of the presented problems. Undoubtedly, the conceptualization of Artificial Personality seems to be a promising study for further understanding of Artificial Intelligence in general [31]. This Terra Incognita, into which Mankind is still quite blindly entering today, requires the conceptualization of new ideas - technobiological symbiosis, Artificial life, computational creativity, Artificial Societies and many others [6, 7]. The authors of the article come to the conclusion that it is necessary to form a new scientific language at the intersection of philosophical anthropology, philosophy of consciousness, philosophy of technology, philosophy of Artificial Intelligence. Perhaps a unifying factor and a common theme may be the study of ethical issues that arise in this new space. Therefore, the prospect of interdisci-

plinary approach and the conduct of interdisciplinary research and expertise with the participation of both traditional specialists in the field of technology and the humanities seems obvious. Participation in this discussion of experts from various fields will allow reformatting meanings, values and rules for the new reality, so that the further development of technology benefits humanity and ensures a prosperous future.

References

1. Alekseev A. Difficulties of Artificial Intelligence Project // Artificial societies. – 2008. – V. 3. – issue 1. URL: https://artsoc.jes.su/s207751800000077-3-1/

2. Alekseev A. Yu. Cognitive technology projects of Artificial Personality. Human: Image and essence. Humanitarian Aspects, no. 1, 2014, pp. 156–174.

3. Alekseev A. Yu. The concept of zombies and problems of consciousness // Problems of consciousness in philosophy and science / Ed. prof. D. I. Dubrovsky. - M.: Kanon +: ROOI "Rehabilitation", 2009. - S. 195–214.

4. Alekseev A., Efimov A., Finn V. The Future of Artificial Intelligence: Turing or Post-Turing Methodology? // Artificial societies. – 2019. – V. 14. – Issue 4. URL: https://artsoc.jes.su/s207751800007698-6-1/ DOI: 10.18254/S207751800007698-6

5. Barrett, Lisa & Adolphs, Ralph & Marsella, Stacy & Martinez, Aleix & Pollak, Seth. (2019). Emotional Expressions Reconsidered: Challenges to Inferring Emotion From Human Facial Movements. Psychological Science in the Public Interest. 20. 1-68. 10.1177/1529100619832930.

6. Belokhina Y., Gurov O. (2020). Ethics of the Digital Twin of Society. Artificial societies. vol. 15, no. 4 DOI: 10.18254/S207751800012583-0

7. Dennett, Daniel. (1994). Artificial Life as Philosophy. Artificial Life. 1. 291-292. 10.1162/artl.1994.1.291.

8. Dregger, Alexander. (2020). More than Intelligence? The role of artificial personality in the user experience of artificial intelligent agents.

9. Dubrovsky D. The value of neuroscience research of consciousness for the development of general artificial intelligence (methodological issues) // Problems of Philosophy. 2022. V. No. 2. S. 83–93.

10. Dubrovsky D.: "Cybernetic immortality. Fantasy or scientific problem?" // 2045 URL: http://2045.com/articles/30810.html (accessed 02.11.2022).

11. Esterwood, Connor & Essenmacher, Kyle & Yang, Han & Robert, Lionel & Zeng, Fanpan. (2021). A Meta-Analysis of Human Personality and Robot Acceptance in Human-Robot Interaction. 10.1145/3411764.3445542.

12. Geher, Glenn & Betancourt, Kian & Jewell, Olivia. (2017). The Link between Emotional Intelligence and Creativity. Imagination, Cognition and Personality. 37. 027623661771002. 10.1177/0276236617710029.

13. Glukhikh, V. & Eliseev, S. & Kirsanova, N.. (2022). Artificial Intelligence as a Problem of Modern Sociology. Discourse. 8. 82-93. 10.32603/2412-8562-2022-8-1-82-93.

14. Guggemos, Josef & Seufert, Sabine & Sonderegger, Stefan. (2020). Pepper: A humanoid robot with personality?

15. Gurov O. (2020). Ethical Interaction with Intellectual Systems. Artificial societies. vol. 15, no. 3 DOI: 10.18254/S207751800010905-4

16. Hosseinzadeh Dizaj, Mehran. (2022). The effect of artificial intelligence in robotics.

17. Humanities vs Sciences & the Knowledge Accelerating in Modern World: Parallels and Interaction // MIPT URL: https://humanities-vs-sciences.events/ (Accessed 02.11.2022).

18. Inzlicht, Michael & Bartholow, Bruce & Hirsh, Jacob. (2015). Emotional foundations of cognitive control. Trends in Cognitive Sciences. 19. 10.1016/j.tics.2015.01.004.

19. Jirak, Doreen & Aoki, Motonobu & Yanagi, Takura & Takamatsu, Atsushi & Bouet, Stephane & Yamamura, Tomohiro & Sandini, Giulio & Rea, Francesco. (2022). Is It Me or the Robot? A Critical Evaluation of Human Affective State Recognition in a Cognitive Task. Frontiers in Neurorobotics. 16. 882483. 10.3389/fnbot.2022.882483.

20. Kadri, Faisal. (2014). Understanding and learning to reconcile differences between disciplines through constructing an artificial personality. Kybernetes. 43. 1338-1345. 10.1108/K-07-2014-0152.

21. Kambur, Emine. (2021). Emotional Intelligence or Artificial Intelligence? Emotional Artificial Intelligence. Florya Chronicles of Political Economy. 7. 10.17932/IAU.FCPE.2015.010/fcpe_v07i2004.

22. Kraus, Kateryna & Kraus, Nataliia & Hryhorkiv, Mariia & Kuzmuk, Ihor & Shtepa, Olena. (2022). Artificial Intelligence in Established of Industry 4.0. WSEAS TRANSACTIONS ON BUSINESS AND ECONOMICS. 19. 1884-1900. 10.37394/23207.2022.19.170.

23. Lambert, Alexis & Norouzi, Nahal & Bruder, Gerd & Welch, Gregory. (2020). A Systematic Review of Ten Years of Research on Human Interaction with Social Robots. International Journal of Human-Computer Interaction. 36. 1-14. 10.1080/10447318.2020.1801172.

24. Learning from Tay's introduction // Microsoft URL: https://blogs.microsoft.com/blog/2016/03/25/learning-tays-introduction/ (accessed 11/2/2022).

25. Lee, Kwan & Peng, Wei & Jin, Seung-A & Yan, Chang. (2006). Can Robots Manifest Personality?: An Empirical Test of Personality Recognition, Social Responses, and Social Presence in Human–Robot Interaction. Journal of Communication. 56. 754 - 772. 10.1111/j.1460-2466.2006.00318.x.

26. Leichtmann, Benedikt & Nitsch, Verena & Mara, Martina. (2022). Crisis Ahead? Why Human-Robot Interaction User Studies May Have Replicability Problems and Directions for Improvement. Frontiers in Robotics and AI. 9. 838116. 10.3389/frobt.2022.838116.

27. Lennox, John. (2020). 2084: Artificial Intelligence and the Future of Humanity. Perspectives on Science and Christian Faith. 72. 254-255. 10.56315/PSCF12-20Lennox.

28. Lombard, Matthew. (2021). Social Responses to Media Technologies in the 21st Century: The Media are Social Actors Paradigm. 2. 29-55. 10.30658/hmc.2.2.

29. Mileounis, Alexandros & Cuijpers, Raymond & Barakova, Emilia. (2015). Creating Robots with Personality: The Effect of Personality on Social Intelligence. 10.1007/978-3-319-18914-7_13.

30. Minbaleev, A.V.. (2022). The Concept Of "Artificial Intelligence" In Law. Bulletin of Udmurt University. Series Economics and Law. 32. 1094-1099. 10.35634/2412-9593-2022-32-6-1094-1099.

31. Mou, Yi & Shi, Changqian & Shen, Tianyu. (2019). A Systematic Review of the Personality of Robot: Mapping Its Conceptualization, Operationalization, Contextualization and Effects. International Journal of Human-Computer Interaction. 36. 1-15. 10.1080/10447318.2019.1663008.

32. Natarajan, Manisha & Gombolay, Matthew. (2020). Effects of Anthropomorphism and Accountability on Trust in Human Robot Interaction. 33-42. 10.1145/3319502.3374839.

33. O'Leary, Daniel. (2022). Massive data language models and conversational artificial intelligence: Emerging issues. Intelligent Systems in Accounting, Finance and Management. 29. 182-198. 10.1002/isaf.1522

34. Pessoa, Luiz. (2017). Do Intelligent Robots Need Emotion?. Trends in Cognitive Sciences. 21. 10.1016/j.tics.2017.06.010.

35. Robert, L. P., Alahmad, R., Esterwood, C., Kim, S., You, S., & Zhang, Q. (2020). A Review of Personality in Human-Robot Interactions. Foundations and Trends in Information Systems, 4(2), 107-212.

36. Sabharwal, Navin & Agrawal, Amit. (2021). Hands-on Question Answering Systems with BERT, Applications in Neural Networks and Natural Language Processing. 10.1007/978-1-4842-6664-9.

37. Shi, Yuwen. (2022). On Negativism of Legal Personality of Artificial Intelligence. Journal of Education, Humanities and Social Sciences. 1. 90-96. 10.54097/ehss.v1i.645.

38. Solaiman, S M. (2017). Legal personality of robots, corporations, idols and chimpanzees: a quest for legitimacy. Artificial Intelligence and Law. 25. 10.1007/s10506-016-9192-3.

39. Spatola, Nicolas & Monceau, Sophie & Ferrand, Ludovic. (2020). Cognitive Impact of Social Robots: How Anthropomorphism Boosts Performances. IEEE Robotics & Automation Magazine. 27. 73-83. 10.1109/MRA.2019.2928823.

40. Tucker, Christopher. (2018). Artificial personality demonstration. 10.13140/RG.2.2.33133.72167.