Original research article

**The fetish of artificial intelligence.**

**In response to Iason Gabriel’s “Towards a Theory of Justice for Artificial Intelligence”**

**David.I. Dubrovsky**

Institute of Philosophy, Russian Academy of Sciences, Moscow, Russia

**Albert R. Efimov**

Sberbank Innovation and Research, Engineering Cybernetics Faculty NUST "MISIS", Moscow, Russia.

**Vladimir.E. Lepskiy**

Institute of Philosophy, Russian Academy of Science, Moscow, Russia

**Boris B. Slavin,**

Financial University under the Government of the Russian Federation, Moscow, Russia

**Abstract**

The article presents the grounds for defining the fetish of artificial intelligence (AI). The fundamental differences of AI from all previous technological innovations are highlighted, as primarily related to the introduction into the human cognitive sphere and fundamentally new uncontrolled consequences for society. Convincing arguments are presented that the leaders of the globalist project are the main beneficiaries of the AI fetish. This is clearly manifested in the works of philosophers close to big technology corporations and their mega-projects. It is proposed to consider the problem of how to use the capabilities of AI to overcome the growing international conflicts and the global crisis in general. The focus is on the problem of subjectness, the solution of which from the standpoint of an anthropomorphic approach to AI is fraught with serious negative consequences. When endowing AI with subjectness, responsibility is implicitly removed from the person who uses this technology, and the established legislative practice is also destroyed. The presentation of AI as an agent endowed with a set of invariant simplified qualities that natural subjects have is proposed. These qualities include: the ability to purposefulness, a kind of reflexivity, communication and simplified elements of sociality. Such a representation of AI as an agent (pseudo-subject) is consistent with the principle of distributed control in biology and psychology, which was called the principle of a dual subject. In combination with the systems of principles and ontologies specified in the concept of post-nonclassical cybernetics of self-developing environments, this will allow the use of AI as a means of social innovation, while maintaining control over AI technologies. And also to pose and solve the problem of integrating the formations of artificial and natural intelligence while maintaining the basic qualities of carriers of natural intelligence.

**Keywords:** artificial intelligence fetish, globalist project, anthropomorphic approach, subject, pseudo-subject, post-non-classical cybernetics.

**David Dubrovsky** – Doctor of Philosophical Sciences, Professor, Chief Researcher at the Institute of Philosophy of the Russian Academy of Science, Moscow, Russia

ddi29@mail.ru

https://orcid.org/ 0000-0003-4392-2526

**Albert Efimov** – PhD in Philosophy, Vice-president of Sberbank Innovation and Research, The Chair of the Engineering Cybernetics Faculty NUST "MISIS". Moscow, Russia

makkawity@gmail.com

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

**Vladimir Lepskiy** – D.Sc. in Psychology, Main Research Fellow, Department of Interdisciplinary Problems in the Advance of Science and Technology, Institute of Philosophy, Russian Academy of Science.

VELepskiy@mail.ru

https://orcid.org/0000-0002-0590-4020

**Boris Slavin** – Doctor of Economics, Professor of the Department of Business Informatics, Financial University under the Government of the Russian Federation

bbslavin@gmail.com

https://orcid.org/0000-0003-3465-0311

For citation: Dubrovsky D., Efimov A., Lepskiy V., Slavin B. (2022) The fetish of artificial intelligence. Russian Journal of Philosophical Sciences = Filosofskie nauki. Vol. 65, no. 1, pp. .

**Introduction**

As the economy and social communications are digitized, the role of digital technologies is also increasing. Technological innovations have always attracted attention, and so it was during the advent of steam engines, and during the period of electrification, and in the heyday of electronics. The emergence of new technologies led to the transformation of economic and social relations, and therefore technologies were often attributed to subject capabilities for the reconstruction of the world. However, time puts everything in its place, and predictions like "Over time, television will change the life of all mankind. There will be nothing: no cinema, no theater, no books, no newspapers – only television", as a rule, are not justified. Technologies find their place in the general ecosystem of scientific progress and do not break the usual foundations. Therefore, an element of skepticism is always necessary when new revolutionary technologies appear.

Such an approach could be applied to modern digital technologies, thus postponing lenghthy discussions until further notice. However, digital technologies have one feature that significantly distinguishes them from all other technologies. All the innovations that appeared in the pre-digital era were designed to increase human productivity, or to make our life easier and more comfortable. They transformed the environment, making it more convenient for humans, but at the same time they did not affect the very essence of human beings, their cognitive capabilities. Of course, the improvement of tools led to the fact that humanity accumulated new knowledge, but this knowledge was still in the minds of people. Books were just a tool for transferring knowledge from person to person, but they don't make sense without a person.

The digital age has made it possible to accumulate knowledge in digital form. In the beginning, this, like any innovations, only facilitated the use of knowledge by a person – it became possible to read books and articles from electronic media, search for information in them. But with the development of data processing technologies, it turned out that analytical programs are able to analyze data in such a way (the so-called data mining) that makes it possible to find in them what is not directly in the text or in the data. For the first time in the history of humankind, technology has encroached on the most sacred thing – the cognitive capabilities of a human. If technology can replace at least partially the ability of a human to think, does this not mean that technology can get the very subjectness for the transformation of society, which they have never possessed before? This question today stirs the minds of many scientists in various levels.

**The growing wave of interest in AI and the harbingers of the storm**

Thanks to the increase in computing performance, advances in the development of algorithms for the use of artificial neural networks (primarily machine learning, and especially deep learning) and the emergence of tools for working with big data, artificial intelligence (AI) technologies have moved from the category of promising to the category of the most popular technologies. We must say right away that the existing AI technologies mainly concern the tasks of recognition, prediction and imitation of human activity with machine learning, and do not pretend to create a real analogue of human intelligence.

Current successes are limited to solving problems of certain classes – AI still cannot be trusted to make complex decisions independently, on which a human's life depends. However, even in this form, AI has become widely used in predictive analytics, scoring, face recognition and gaming applications. In 2021, the world 's largest companies invested more than 250 billion US dollars in research and development in AI. At the same time, the global AI technology market will amount to 554.3 billion US dollars by 2024 with an average annual growth rate of 17.5% [Breakthrough innovations 2022].

Simultaneously with the growing success of AI, a series of adoption of strategies for its development at national levels began. In 2016, in the USA, the National Science and Technology Council prepared the reports “Preparing for the future of Artificial Intelligence”, “National Artificial Intelligence Research and Development Strategic Plan” and “Artificial Intelligence, Automation, and the Economy”, which determined the strategy for the development of AI in the country. In July 2017, the China's State Council issued a document, entitled “A New Generation Artificial Intelligence Development Plan”, designed to create an artificial intelligence industry and turn China into a leading power in the field of AI by 2030. At the 2018 annual meeting of the World Economic Forum, British Prime Minister Theresa May announced that she was going to make the UK a world leader in artificial intelligence. In 2019, in Russia, the President V. V. Putin approved the "National Strategy for the Development of Artificial Intelligence for the period up to 2030", and later a project for the development of AI was allocated within the framework of the Digital Economy program. In recent years, a new global trend has been gaining popularity – the development of Artificial General Intelligence (AGI), which in its functions should approach the solution of tasks specific to natural intelligence. This will be discussed in more detail below.

Simultaneously with the amazing progress of AI in business and focused attention of various governments, protests and demands started to limit the use of recognition tools that violated individual rights and carried the risks of making erroneous decisions. Thus, European countries, already at the signing of the Declaration of cooperation on Artificial Intelligence in April 2018, emphasized the need to take into account social, economic, ethical and legal issues. In particular, the European Commission established a Group of Experts who published guidelines on AI ethics in April 2019. In September 2021, the UN Human Rights Council issued a report with recommendations for States and enterprises to develop and implement safeguards to prevent and minimize harmful effects and promote the full use of the benefits that artificial intelligence can provide. The head of the Council (UN High Commissioner for Human Rights) Michelle Bachelet, commenting on the report, generally called for a moratorium on artificial intelligence systems that threaten human rights until governments can establish guarantees.

The report of the UN Human Rights Council states that decision-making processes in many artificial intelligence systems are not transparent. The complexity of the information environment, algorithms and models underlying the development and operation of artificial intelligence systems, as well as the deliberate secrecy of public and private actors are factors that undermine meaningful ways for the public to understand the impact of artificial intelligence systems on human rights. Interestingly, it is machine learning that causes the greatest concern among human rights defenders, since the result of such computations is unpredictable: machine learning systems add an important element of opacity; they may be able to identify patterns and develop recipes that are difficult or impossible to explain. This is often referred to as the “black box" problem. Opacity makes it difficult to scrutinize an artificial intelligence system and can become an obstacle to effective accountability in cases where artificial intelligence systems cause damage [Human Rights Council 2021]. In addition, it is of great concern that any machine learning depends very much on the data on which the training itself takes place. If these data are not checked for compliance with elementary ethical standards, or worse, they are specially selected so that they contradict them, then machine learning systems will inevitably issue deliberately unethical recommendations.

**The main** **interested parties and beneficiaries of the AI fetish**

Recently there appeared in print an article "Toward a Theory of Justice for Artificial Intelligence" by Iason Gabriel, who is a Staff Research Scientist and an expert in philosophy at a company owned by Google - DeepMind [Gabriel 2022]. The article is devoted to the development of humanistic principles of using artificial intelligence technologies. Let's analyze the philosophical and methodological foundations of the author's position, their influence on the ideas about the potential consequences of the development of AI, as well as, most importantly, about the main interested parties of the proposed approach.

First of all, attention is drawn to the ideological attitude of the author of the article, based on the works of John Rawls devoted to his "Theory of Justice" [Rawls 1999]. The essence of this approach is depicted as follows: people are interested in increasing their own profits and decreasing the common share of benefits. This is a pronounced position of the ideology of social liberalism. To represent the role and place of AI, the author suggests understanding the basic structure of society as a set of sociotechnical systems, the functioning of which develops under the increasing influence of AI. Such a representation of AI can be characterized as a vivid manifestation of technocratic reductionism in the organization of social processes. Here, as a consequence, the growing decline in the role of the state and society under the influence of AI is implicitly proposed. The leaders of the globalist project are interested in such results. But who exactly? This needs to be specified.

It is easy to see that the leaders of the globalist approach and the bright exponents and defenders of the ideology of Western liberalism are the owners of big technology corporations who, like an octopus, cover the global information space. Owners of companies such as Facebook, Twitter, Amazon, Google, etc. and their hired theorists, attributing the quality of subjectness to AI systems, are trying to create the impression that they are endowed with some "divine" functions. After all, they are not just developing new software products, but creating tools that are supposedly able to establish justice and protect human rights. Such tools are utopia, but proclaiming faith in them, imitation of such faith is a good ground for protecting globalist ideas, and at the same time, the innermost interests of the world's largest technology manufacturers. Manipulating the mass consciousness, using all forms of social communication, all means of AI, they declare that they will "sow reasonable, good, eternal" for all of us. It is no coincidence that Facebook founder Mark Zuckerberg decided to create a Metaverse based on his technology products, in which a new, more «correct and prosperous» life will be built. However, it is extremely difficult to assume that Mark Zuckerberg is really concerned about such high humanistic aspirations. Judging by all his business activities and his experience of manipulating mass consciousness in order to achieve maximum profit, the tales of the "Metaverse" serve only as a disguise for the implementation of the same goals.

The globalist approach, based on the ideology of Western liberalism with its pseudo-humanistic cliches, convincingly showed its inconsistency in the conditions of the COVID-19 pandemic [Lepskiy 2020] and thereby demonstrated its negative role in assessing and developing ways to overcome the growing threats to humanity. Including threats related to digital transformations and the development of AI.

**Prospects of AI in the context of the global crisis of world civilization**

It is this context of considering the process of AI development and its social significance, often pushed to the background (especially in publications of the type discussed above), that is now of paramount, fateful importance for humanity. The steady increase in the global crisis of our consumer civilization leads to the inciting of increasingly large-scale economic, political, and all kinds of social conflicts, an uncompromising struggle for resources, for spheres of influence, and ultimately for the redistribution of the entire global structure of socio-economic and political self-organization. And in this regard, AI becomes an instrument of struggle in the confrontation of various forces.

In this regard, as already noted above, special attention is drawn to the task of creating a AGI, the solution of which has become in recent years the subject of competition between the largest bigtechs, and more broadly between the leading states in the field of AI, among which are our strategic opponents. This obliges us to concentrate our efforts in this direction as much as possible and to get ahead of our competitors. Recently, there was published first book in Russian by Russian scientists, dedicated to AGI [Strong Artificial Intelligence 2021]. It analyzes in detail the main theoretical and methodological issues of the development of AGI and the scientific approaches necessary for this, identifies those specific functions that must be created and, most interestingly, the supposed large-scale changes that it is able to produce in the emerging new world order.

The authors identify two key abilities of AGI, which are characteristic of natural intelligence. Unlike a "narrow" AI that solves one specific task, it must be integral, i.e., capable of solving many types of different tasks. And it must acquire the quality of autonomy, i.e., the ability to act effectively independently in a wide range of environments. All this qualitatively increases the operational capabilities of AI systems, their use for military equipment and military operations, for solving problems of production, management, planning, organization of economic processes, optimization of a wide variety of spheres of public life and scientific research, which is extremely important for our country in the current conditions. And this is equally important for understanding and implementing those future historical changes in the development of the Earth's civilization, which are associated with the collapse of the principle and practice of the monopolar world.

At the same time, the development of AGI raises new complex theoretical questions concerning its interactions with natural intelligence, the creation and prospects of hybrid intelligence, possible competition with natural intelligence, possible risks and threats to humans and society. Achieving a high degree of autonomy of the General Intelligence creates the likelihood of the appearance of such types and methods of its "amateur activity", which may pose a danger to humans and society, will require the development of new methods of ensuring security. Here we will face a new aspect of the same problem of subjectness of AI systems, which retains its high relevance.

**Is AI a subject or an agent controlled by subjects with natural intelligence?**

Let us turn once again to the interpretations of theoretical questions about the subjectness of AI systems in relation to real human subjects. The approach of the author of the above-mentioned article, Iason Gabriel [Gabriel 2022], is to transfer to AI the same principles that are established for humans. The author writes that "AI increasingly shapes elements of the basic structure” of society, and, "the development and deployment of AI systems represent a new site for the operation of principles of distributive justice". According to the author, " AI interacts with the behavior of human decision-makers to shape the character of these practices, including how they distribute benefits and burdens across the population."

It turns out that for the author technologies seem to "come alive": " What is important for our purpose," he writes, "is that in modern societies, background justice is increasingly mediated algorithmically." Here the concept of background (apparently, mass market) justice is introduced, and it is implicitly assumed that this justice is carried out by an algorithm. And further: "By making assessments or predictions based upon an individual’s past choices, and by providing decisions or recommendations that then shape that person’s opportunity set, these systems exert a strong influence on the unfolding relationship between individual choices and collective outcomes. " As we can see, it is again implicitly assumed that the systems are doing something "intelligently" and providing something. The author implicitly gave subjectness to AI technologies, and demands to extend the laws of social relations to the use of AI, which should "support citizens’ basic liberties, promote fair equality of opportunity, and provide the greatest benefit to those who are worst-off."

This kind of anthropomorphic approach to artificial intelligence is fraught with serious consequences. Firstly, by transferring subjectness to AI, responsibility is implicitly removed from the person who uses this technology, which levels legislative practice. Any technology is imperfect, and cannot make a decision by itself. The error in the operation of the lie detector will never be zero, as well as when using AI. The task of people is to take into account the limitations of technology, and not just try to put them in some given framework. It is not by chance that in judicial practice, with all the possibilities of criminology, people make the final decision. Artificial intelligence systems are no different from other technologies, so technologies should not remove responsibility from a person, and therefore should not have human characteristics – i.e. be fair, humane, etc.

The success of AI turned out to be so significant that many decided that this technology could replace a human, only it had to be put in a certain framework. In fact, artificial intelligence has become a fetish of the XXI century, which some began to extol as our future, while others began to fight against it. In fact, AI, although it can recognize what the human eye cannot, or identify a correlation that a person cannot find, at the same time remains far from the real cognitive capabilities of a human. Maybe in future it will be possible to create an AI integrated into the social environment, the so-called strong AI, but now we are far from this, and giving it subjectness is not only wrong, but dangerous.

**Ethical issues of AI development**

The ethics of using AI are now widely discussed not only by public figures, politicians, but also by scientists. Much attention was drawn to the situation with the article prepared for publication on the ethics of artificial intelligence by Google employee Timnit Gebru, one of the world's leading experts on the problems of bias in algorithms and data mining. Scientific and commercial positions fundamentally diverged, as a result, the researcher left Google.

Discussions are also underway among Russian scientists. Along with constructive proposals, some reputable scientists offer very dubious ideas, call for the politicization of science and technology, endowing technology with the properties of "patriotism". So, on November 23, 2021, a meeting of the Presidium of the Russian Academy of Sciences was held, at which academics discussed, among other things, the ethical side of artificial intelligence and the possibilities for AI to be trusted. One of the speakers stated: "From my point of view, artificial intelligence should not only be trusted, as was discussed today, but also patriotic, that is, it should primarily work for the interests of the country, and not against it" [Slavin, 2021].

Today, many people are protesting against facial recognition technologies (in some US cities such recognition is prohibited, the European Parliament also proposed banning facial recognition technologies). However, it is not the recognition itself that is harmful, but its use for illegal purposes. People do not hide their illnesses in front of a doctor because they trusts this specialist and count on help. It is necessary to regulate the laws of the use of technologies, including artificial intelligence, and seriously punish if they were used to harm a person or illegally. Russia today has turned out to be in many ways poorly protected against fraudsters who use means of communication to deceive gullible citizens. All that the authorities can do today is to warn the population about new ways of fraud. At the same time, the government is quite effective, including with the use of AI, in fighting political opponents. The ethical problems of AI should be solved not by limiting technologies, but primarily by limiting the actions of those people who misuse them.

Solving ethical issues of AI application is a very complex matter that cannot be accomplished only by adopting declarative codes of AI Ethics. It is necessary to take into account that even our ethical principles, whatever we take – from the biblical commandments to the code of the builder of communism – are only points in the infinite space of moral and ethical decisions in which we move daily. Discussions about the ethics of AI are just beginning to unfold and communities of philosophers and engineers should work closely together to develop answers to the new challenges of the time.

It should be emphasized that it is extremely important now to consider ethical problems in relation to the development of AI in a broader conceptual sense – from the angle of the real features of the functioning of ethical norms in social life, the real state of morality of mass consciousness, individual, group, institutional subjects. After all, there has always been a gap between the knowledge of ethical norms and their implementation at all stages of human history. Let's recall the ancient Roman saying: "Video meliora, proboque, deteriora sequor[[1]](#footnote-1)." Too often, personal interest turned out to be above moral precepts, and deception suppressed the truth and made cozy nests for itself in the highest ethical instructions. There are no sufficient grounds to talk about moral progress in the development of mankind (detailed materials on this subject are presented in many philosophical studies [Dubrovsky 2007]. All these circumstances must be taken into account when we discuss the topic of "AI Ethics", both in relation to the creators of AI systems and to their users.

When trying to model the principles of ethics and implement them in the work of AI, the matter is complicated by the fact that the set of ethical norms cannot be ordered in the form of a clear hierarchical structure that allows an alternative choice. Here, the choice can almost always be made only when considering and evaluating specific conditions. Therefore, the specified modeling is possible only in specially defined special cases.

At the same time, the problem of subjectness in the field of AI development remains worthy of close attention. After all, one way or another, the ability of an AI system to solve complex problems is associated with descriptions of some functions of natural intelligence. If it is not advisable to represent AI as a subject similar to a person, then how to understand and define AI, which is given the ability to make decisions in certain situations and which defeats the world chess champion or in Go game. The most adequate approach, in our opinion, may be the representation of AI as an agent endowed with a set of invariant simplified qualities possessed by natural subjects. Such qualities include: a kind of purposefulness, a kind of reflexivity, communicativeness and simplified elements of sociality. This representation of AI as an agent (pseudo-subject) is consistent with the principle of distributed control in biology and psychology, which was called the principle of dual subject [Lepskiy 1998]. This representation of AI, combined with systems of principles and ontologies set in the concept of post-non-classical cybernetics of self-developing environments, allows using AI as a means of social innovation, while maintaining control over AI technologies, as well as stating and solving the problem of integrating artificial and natural intelligence formations while preserving the basic qualities of natural intelligence carriers [Lepskiy 2018; Lepskiy 2021].

**Conclusion**

AI is experiencing a phase of rapid growth. The large-scale figures of the effects of its introduction should not mislead us - the current period is only the beginning of the total penetration of AI into our lives. That is why we should be very attentive to possible cognitive distortions in the study of emerging phenomena. For example, anthropomorphism applied to AI can make us easily believe in the false subjectness of the machine. This work by the authors who are specialists in philosophy and methodology is a call for a broader dialogue and a transition from creating codes of conduct for AI to creating the next generation of AI acting together with humans and for them.

***REFERENCES***

Dubrovsky 2007 – Dubrovsky D.I. O nravstvennom progresse i nravstvennom regresse (K problematike razvitiya moralnogo soznaniya) // Filosofskiye nauki, № 11, 2007. (On moral progress and moral regression (On the problems of the development of moral consciousness) [in Russian]

Dubrovsky 2021 – Dubrovsky D.I. Zadacha sozdaniya Obshchego iskusstvennogo intellekta i problema soznaniya // Filosofskiye nauki. 2021. Vol. 64. No. 1. p. 13-44. (The task of creating a general artificial intelligence and the problem of consciousness) [in Russian]

Lepsky 2020 – Lepskiy Vladimir E. Refleksiya pandemii COVID-19: subyektno-oriyentirovannyy podkhod // Ekonomicheskiye strategii. 2020. № 8 (174). p. 66-71. (Reflection on the COVID-19 pandemic: a subject-oriented approach) [in Russian]

Lepsky 1998 – Lepskiy Vladimir E. Kontseptsiya subyektno-oriyentirovannoy kompyuterizatsii upravlencheskoy deyatelnosti. M.: Institut psikhologii RAN, 1998. (The concept of subject-oriented computerization of managerial activity) [in Russian]

Lepsky 2021 – Lepskiy Vladimir E. Iskusstvennyy intellekt v subyektnykh paradigmakh upravleniya // Filosofskiye nauki. 2021. Т. 64. № 1. pp. 88-101. (Artificial intelligence in subjective management paradigms) [in Russian]

Breakthrough Innovations 2022 – Proryvnyye innovatsii: chelovek 2.0: doklad k XXIII Yasinskoy (Aprelskoy) mezhdunarodnoy nauchnoy konferentsii po problemam razvitiya ekonomiki i obshchestva, Moskva, 4–8 aprelya 2022 g. / S.A. Gavrish, L. M. Gokhberg, D.E. Gribkova i dr.; pod red. L.M. Gokhberga, A.R. Yefimova, YU.V. Milshinoy; Nats. issled. un-t «Vysshaya shkola ekonomiki», PAO Sberbank. – M.: NIU VSHE, 2022. – 56 p. (Breakthrough Innovations 2022) [in Russian]

Strong Artificial Intelligence 2021 – Silnyy iskusstvennyy intellekt: Na podstupakh k sverkhrazumu / Aleksandr Vedyakhin i dr. – M.: Intellektualnaya Literatura, 2021. – 232 p. (Strong artificial Intelligence: On the approaches to the supermind) [in Russian]

Slavin 2021 – Slavin Boris B. Mojet li iskusstveniy intellect bit spravedlivim? BIT. 2021. №10 (113). pp. 32-35. (Can artificial intelligence be fair?) [in Russian]

Gabriel Iason (2022) Toward a Theory of Justice for Artificial Intelligence, Daedalus, Spring 2022, Vol. 151, No. 2, AI & Society (Spring 2022), pp. 218-231.

Human Rights Council (2021). A/HRC/48/31: The right to privacy in the digital age - Report of the United Nations High Commissioner for Human Rights // https://www.ohchr.org/. 2021. URL: https://www.ohchr.org/en/documents/thematic-reports/ahrc4831-right-privacy-digital-age-report-united-nations-high (date accessed: 18.05.2022).

Lepskiy Vladimir. E. (2018) Evolution of Cybernetics: Philosophical and Methodological Analysis, Kybernetes, vol. 47, No. 2, pp. 249–261.

Rawls John (1999) A Theory of Justice, rev. ed. Cambridge, Mass.: Harvard University Press, 1999.

1. I see better things, and approve, but I follow worse [↑](#footnote-ref-1)