Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 186 (2021) 344–351

Procedia Computer Science

www.elsevier.com/locate/procedia

14th International Symposium "Intelligent Systems"

The artificial intelligence influence on real sociality

G. Malinetsky, V. Smolin*

Keldysh Institute of Applied Mathematics, Miusskaya sq., 4, Moscow, 125047, Russia

Abstract

Artificial intelligence (AI) technologies are increasingly affecting society. The individualization of the information presented to the user, the introduction of “smart” devices gives technical capabilities to states and corporations to carry out targeted impact and total control of the digital society members. The AI opportunities balance and the degree of control over their use can be shifted in any direction. Understanding AI capabilities and limitations is necessary to abandon extreme points of view. AI is not good or bad, it’s a technology that can be used for different purposes. The main goals of AI regulation should be not only the desire to create a “beneficial” or “friendly” AI, but also the creation of a system for monitoring the effectiveness of its use. It is necessary to support the use of AI to increase productivity, deepen the labor division, including in scientific activities. AI tools should be used to motivate productive work, for the population's creative potential growth. It is necessary to create social conditions when gadgets that allow developing creative potential will be more in demand by society than means of avoiding reality and germinating destructive abilities. AI is an important component of the digital environment. But while the man remains the main force that ensures the progress of society, the social systems competition will be won by one that gives the best conditions for realizing the creative potential of ordinary people.

Keywords: Artificial Intelligence; neural networks; goals setting; labor division; competition

1. Introduction

The increase in the commercial use of neural network algorithms in devices and systems that solve "intellectual" problems that occurred in 2010-2012 is called the neural network revolution in machine learning. The success was based on the achievements in the effective tuning of a huge number of latent (internal) parameters of neural network algorithms. This made it possible to process complex real-world signals at a qualitatively higher level. Neural

* Corresponding author. Tel.: 7-916- 555-3998; fax: +7-499- 972-0737.
E-mail address: smolin@keldysh.ru

10.1016/j.procs.2021.04.155

network algorithms application decade has led to the creation of an industry with its own technologies, standards, productions and research institutes. Almost all the ideas and methods used in devices and systems with artificial intelligence (AI) appeared as a result of the development of mathematical ideas and the creation of new algorithms, technologies, and devices, and not as borrowing from biology, neurophysiology or psychology. But the experience of using neural network computing allows us to take a fresh look at the vast array of knowledge about a society acquired by the humanities without using ideas about AI.

Knowledge about hierarchical information processes formation in neuron-like structures allows us to understand better the ways of human society development, to look at the social problems from a different angle.

A sociological analysis makes it possible to make forecasts of the interaction of society with AI, to understand better its contribution to the humanitarian and technological revolution [1], Also it will help to formulate requirements for AI devices and systems to integrate safely into society to serve progress of the human civilization.

2. History

Since ancient times, man invented and tried to create devices that can independently carry out actions in a complex environment. After Turing's theorem proving and the advent of computers based on von Neumann schemes, the problem turned into a practical plane. We can execute any algorithm on the computer and now we ask ourselves the question: can we imagine a reasonable, intellectual activity in the form of a set of algorithms?

The term “Artificial Intelligence” (AI) has been proposed by John McCarthy at the Dartmouth seminar held in the summer of 1956. The seminar brought together individual enthusiasts in a professional community with their scientific goals and clear self-determination. John McCarthy suggested that any property of intelligence can be so accurately described that a machine can simulate it. “An attempt will be made to find how to make machines use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves,” wrote John McCarthy in an application for the event [2]. However, these hopes began to be seriously justified only 55 years later.

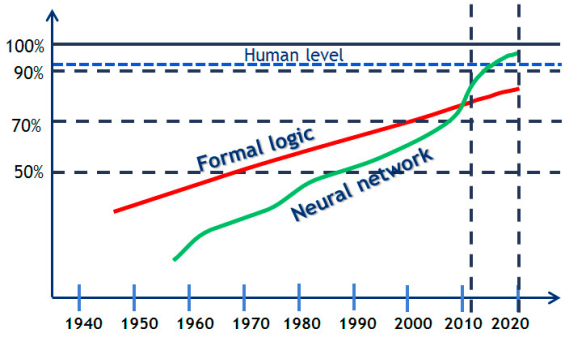

Fig. 1. Percentage increase in error-free solutions of “intellectual” problems when using various approaches

After the Dartmouth seminar, two main directions in approaches to the AI creation were finally formed: formal-logical and neural network. Proponents of the formal logical approach called themselves “Neats”, and neural networks called “Scruffies”. In the 1960s, John McCarthy was the leader of neats, arguing that "mathematically accurate thinking = semantic representations of inference." The leader of the scruffies was Marvin Minsky, who developed the ideas of self-organization, machine learning, shaped semantics.

The imitation of external manifestations of rational activity served as the basis for the success of formal logical approaches in the early stages of solving “intellectual” problems. This approach allowed to develop optimal transformations for the implementation of the required functions. This is a shorter way to solve individual problems than the universal approaches development to solve wide classes of problems that the scruffies have been aiming at. For decades it was not possible to surpass the level of neats, the enthusiasm of scruffies was replaced by pessimism. Two “springs” gave way to two “winters” of AI before neural network approaches in 2010-2012. began to transcend

formal logic. The success of the supporters of neural network approaches was due to the development of specialized schemes for processing images, speech, and other signal modalities (the dynamics of competition between the success of two directions of AI is shown in Fig. 1).

In 1997, in a match with world chess champion Garry Kasparov, the Deep Blue computer won. In general, in some problems (not only in chess), the human level was surpassed even before 2000 by formal logical methods. The use of neural network approaches does not always allow solving "intellectual" problems at least at the human level. But if earlier in most "intellectual" tasks the level of a person seemed unattainable, then starting from 2015-2018 they began the process of mass exceeding human capabilities by various technical devices and systems.

3. The contemporary AI development level

Neural computations require powerful computers, large amounts of data (which the Internet allows to assemble), and a theory of the construction of neural network algorithms. Today, the central way to solve a significant number of AI tasks (although not all) is to use Deep Neural Networks (DNN). The most important property of DNN is their ability to adjust automatically billions of their parameters using the Back Propagation Error (BPE) method. The successes of DNN theory can be conditionally divided into two main areas: the development of new neural network computing schemes and the improvement of methods for automatic parameter setting (training). Figure 2 shows the sequential complication of neural networks, from individual elements to blocks.

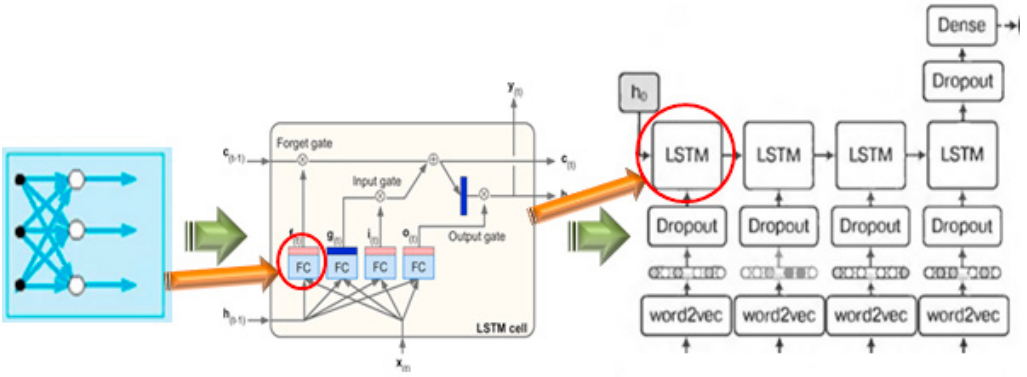

Fig. 2. Stages of complicating the neural network computing structure

In addition to fully connected (FC) multilayer networks, the development has received many fruitful ideas: CNN – convolutional NN; RNN – recurrent NN; GANs – generative adversarial networks; reinforced deep learning networks (Fig. 3). The progress in teaching methods improving is no less significant: it clearly resulted in the fact that over the past ten years the number of layers in DNN has increased from units to hundreds and the number of trained parameters from thousands to billions. But this is an "internal kitchen" that the end-user may not even notice.

Users are impressed by new successes every month in processing real-world signals. For this, it is necessary to express the states of the observed objects and phenomena in the form of vectors $\vec{X}$, and then use the vectors $\vec{Y}$ to generate physical phenomena - light, sound, pressure, motion, etc.

In the age of digital technology, such conversions are made in many devices and are not causing surprise. But sometimes it takes some ingenuity to convey the meaning of a word in a vector or express a mood in vector form. And several neural network approaches have been developed, for example, Wordtovec [3] and Glove, which allow forming a word meanings vector description. Vectorization methods continue to improve, but even now the progress in solving a wide class of "intellectual" problems is significant.

The difficulty is not so much in digital vector description creating, as in forming a successful metric. The modern development of neural network computing schemes allows not only to create an approximation of the transformations $\vec{X} \to \vec{Y}$, but also to form successful metrics of digitalized vectors (by the transformations $\vec{X} \to \vec{Y}$).

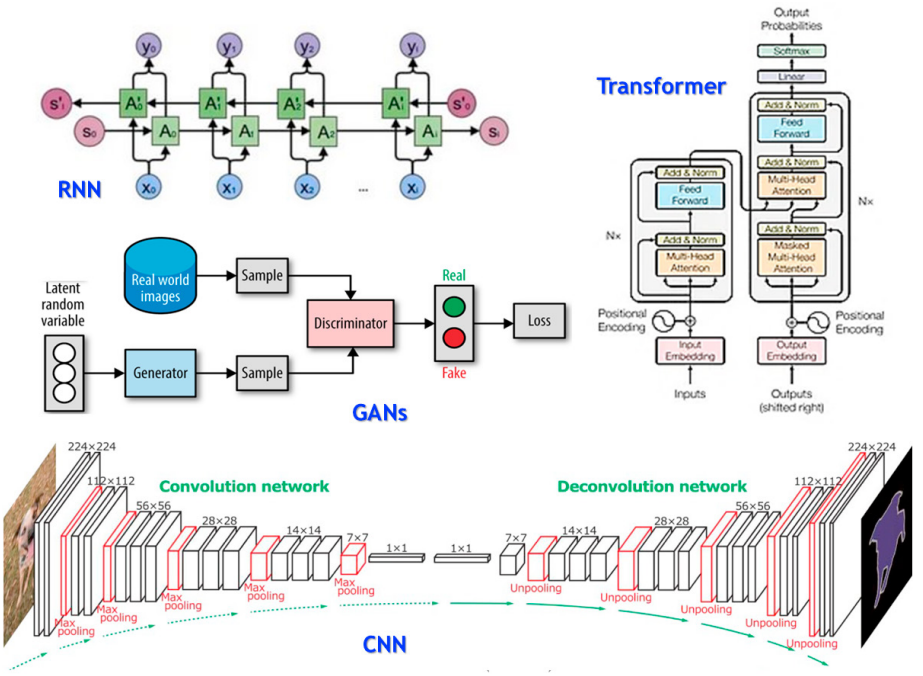

Fig. 3. Formal neurons networks schemes that implement the multilayer (deep) neurocomputing basic ideas

4. Difficulties in AGI defining, creating and supervising

With all the differences in views on creating a strong AI (Artificial General Intelligence, AGI) problem, the main obstacle to creating an AGI is the complexity of describing the real world. Not all problems are effectively solved by parallel processing [4] and are reduced to deep learning. As approaches to solving the problem, for example, the winners of the Turing Prize for 2018, Joshua Benjio, Jan LeCun [5] and Jeffrey Hinton, call the development of such properties as attention, understanding of the causality of events, planning of actions, the formation of high-level concepts. Work in these areas has long been underway and almost all have successes. The problem is that while there is no possibility to assemble a single model that simultaneously satisfies all the desired properties.

The methodological difficulty remains the lack of a universally recognized definition of AGI and intelligence in general. There are several directions in the definition of AGI (and what is intelligence):

The first direction is to imitate the external manifestations of “intellectual” activity. And so far, AI successes are judged primarily by the success of the simulation.

The second direction is that devices should intellectually exceed the level of man. In the absence of a generally accepted definition of intelligence, a comparison in terms of intelligence can't be correct. Today AlphaZero is confidently defeating the world go and chess champions, but no one believes that the AGI level has been reached.

The third area of AGI assessment is the ability to perform rational or reasonable actions. Reasonable actions are not always rational. The main argument of criticism of the third direction is that for rational and reasonable actions some goal is necessary to achieve and optimize the path to it. The norms of ethics and morality were formed, changed and passed through selection, as well as the goals and objectives that were set by individuals and society as a whole. Even in a world riddled with the ideas of globalization, ethical and moral norms have marked variability. This makes the assessment of the reasonableness and rationality of actions dependent on the circumstances of their implementation and can cause critical social problems when AGI appears in the future.

The fourth area is the ability to formulate and select goals for action. For the manifestation of “intelligence”, it is necessary to have freedom of the target parameters values choice, which cannot be achieved if the criterion of choice optimality is rigidly defined and laid down in the AGI system initially. AGI should formulate the criteria for

choosing goals by its own, based on its own experience of interacting with the environment and perceiving the rules of the civilized society in which it operates.

The fifth direction in determining the properties of AGI is the presence of desires to formulate goals, perform actions, excel man or other AGI systems. If the AGI system has all of these abilities but does not seek to show them, then it will remain an automaton that fulfills the instructions. The presence of desires to realize the available opportunities, while assessing on the basis of their experience the degree of compliance with civilizational standards, will allow AGI to deliberately demonstrate its abilities in choosing goals and planning behavior.

In order for a device or system to comply with the AGI concept, it is necessary that all five of the above definitions are met and, possibly, that some additional properties are found that will be revealed in deeper studies. It is important that AGI does not need to have super knowledge, super abilities, super fastness, and other "super" properties. The presence of desire and the ability to create goals of behavior is sufficient to form and implement on their basis a reasonable and rational behavior that is comparable or superior in its effectiveness to human actions.

How much will humanity be able to control the AGI, which in some (or even all ...) parameters will exceed the mental abilities of people? The fantastic idea is popular that it costs one scientist programmer to write a successful code (literally a few lines or several pages ...), as it immediately leaks into the network, captures all the computing meens there and gets complete power over humanity. And then there are three main options:

a. AGI provides absolute happiness to all mankind by completely eradicating the existing injustices;
b. AGI carries out the enslavement and / or destruction of humanity for the realization of its own goals;
c. AGI is not interested in the problems of mankind and flies away from Earth to outer space to develop its own civilization.

In each of these three fantastic options have many combinations of possible smaller storylines. The unreality of such scenarios is based on the magical assumption that in order to solve the complex problem of creating an AGI, over which many serious teams have been working for decades, it is enough for a graduate student to get distracted from computer toys for a week and realize his ideas in the program. Naturally, such a “genius” does not arise any thoughts about the interaction of AGI with human civilization, or these thoughts do not affect the program code.

In practice (as opposed to science fiction), this is not so. Malfunctions in the digital network, of course, do happen, but they are all monitored and quickly eliminated. AI devices (even weak ones) are being developed in large organizations that are concerned not only with the process of creating gadgets but also with the security issues associated with their use. Moreover, not only the problems of the regular use of devices are considered, but also their resistance to external, including unfriendly, influences in terms of ensuring the reliability of their work.

AGI is not the first invention that poses a potential threat to humanity. Nuclear technologies and missile systems, which were not developed by individuals, can be mentioned as the most dangerous. Other types of equipment: ships, cars, planes, rail transport, and most industrial enterprises also have their own means and rules for safe use. In more ancient history, the process of domestication of animals was also accompanied by the clarification of the rules for their (relatively) safe use in the economy and military purposes.

Large industrial enterprises and gigantic ocean vessels have economic advantages, although they pose an increased danger in their operation. But namely high economic efficiency allows us to allocate more funds to ensure their safe operation. Considerations of economic efficiency force us to create powerful giants, the violation in the exploitation of which is fraught with major troubles. However, when sizing a device or manufacturing, safety issues are always considered.

The safety control of using AGI is complicated by the fact that if all previous inventions of mankind surpassed man only in strength, power, speed, and other physical characteristics, while AGI is able to give more rational solutions. But the usefulness of the direct and, especially, the long-term consequences of such decisions for a person is difficult to evaluate. It’s like in a game of chess with a grandmaster: he makes moves whose purpose is often not clear to a simple chess fan. To play on equal terms, a chess fan needs a consultant at the grandmaster level or chess program, at the level of a world champion. This simple example shows that the creation of a unique world-wide AGI implemented on all of the Earth’s networked computing resources is bound to be uncontrolled. Since it will not be possible to create a tool comparable to such an AGI that can understand the consequences of the chosen AGI action.

The main way to solve this problem is to use the agent approach in creating the AGI. That is, one should not strive to create superintelligence, many orders of magnitude greater than the mental capabilities of an individual. Even more important is not the degree to which human capabilities are exceeded (the degree will continue to grow

with progress), but the presence of a significant number of independent AGI devices that can be used for their mutual control. And, in the violation event of the friendliness of individual AGI devices towards a person, most of the other devices that have remained friendly will be able to compensate for possible malicious actions.

The agency approach to creating an AGI is a natural development of the modern situation, when there are many AI development centers. It is only necessary to follow the direction of development towards the creation of less powerful AGI when one center creates many interacting smart devices. Instead of creating more powerful unique AGIs that use all the computing capabilities of the combined resources of several centers.

In order to provide opportunities for mutual control of the AGI agent actions, it is necessary to legislatively fix the requirement to record the information exchange of these devices with such technologies as blockchain. It is also useful to limit the permissible maximum computing power of AGI agent devices, which may increase over time as the system of checks and balances is developed.

Perhaps the agent-based approach to AGI has some drawbacks caused by the need for more information exchange in comparison to more powerful, unique AGI devices. But it is precisely the large flow of information exchange that will provide the ability to monitor the activities of AGI. And, most importantly, human civilization is built precisely on the principle of combining a large number of interacting intelligent agents.

5. Civilization as a superintelligence

Most often, as the most important difference between humans and animals, articulate speech is mentioned. At first glance, it might seem that the difference is quantitative: many animals use sounds and gestures to transmit information, they just have a poorer dictionary. But this difference gives qualitative results of its use - the human population is many orders of magnitude superior to other species in the labor division degree, the only species that creates multistage hierarchies for managing large territories (states) and uses speech not only to directly coordinate actions, but also to create legends and myths. That is very useful in forming large communities [6].

Humans are also significantly superior to all other animals in the degree of this ability to create and use tools realization. The use of tools and the development of speech contributed to each other, but it seems that the improvement of articulate speech was faster than the expansion of the tools use (and means of coercion to work ...) than vice versa. Based on this, articulate speech became the main physiological factor distinguishing mankind from the animal world. The role of speech in the formation of human society was understood long ago, but was usually considered in a number of other properties: upright posture, developed grasping abilities of hands and others. The idea of the need for hierarchical processing of information in neural networks allows us to distinguish speech from these properties. In all animals, including humans, the number of processing levels and the associated maximum degree of abstractness of the analyzed concepts is limited by brain structure.

The emergence of the developed speech made it possible to reproduce concepts in speech communication and again pass them through the hierarchical structure of processing. Thus, articulate speech allowed a person to increase greatly (compared to other animals) the level of abstract concepts considered. This enabled people to form complex hierarchical social and industrial relations, which became the basis for the creation of large economic, military, religious, and other organizations that form the structure of the state. An individual is extremely weak in the face of organizations that have significant potential for impact not only on society but also on nature. The organization's strength is determined not only by combining the physical efforts of individuals, but by the possibility of a hierarchical analysis of information when making decisions about the actions performed. In this way, organizations and states form superintelligence based on a hierarchical combination of the individuals intelligence

6. Labor division and competition

For the transformations, fulfilling by neural networks, the decomposition of the tasks performed into several simpler ones allows to achieve better learning outcomes in solving complex problems. This is effective even in the case of using the same computing resources as in solving a complex problem without decomposition. But efficiency increases especially when, after decomposition of a complex problem into several more simple ones, additional resources are involved in solving each subproblem. This property constitutes the theoretical basis for the labor division, the degree of which has increased throughout the history of the development of human society.

Another important theoretical property based on ideas about the training of neural networks is the local optimality of the solutions found. To find deeper local minima, it is necessary to start the search for solutions from other initial values of the parameters of the neural network and / or use a different data set for training. The practical implementation of this property in the history of the development of human society is competition in almost any kind of activity. Which helps to overcome local minima in an search for optimal solutions.

Being the engines of human progress, the labor division and competition are contrary to the ideas of freedom and equality, popular in recent centuries. Person has less freedom with his own niche in the labor division in a civilized society. He depends on all other workers producing the products necessary for human life. Society should not only build relationships between people involved in various specialties but also ensure that all vacancies are filled. Otherwise, no one will perform some types of work and society will not have everything for its normal life.

To fill all vacancies, coercion to work is necessary, which negatively affects the level of freedom. Three main methods of constant coercion can be distinguished: the threat of the force used, material incentives, and the creation of myths. All three methods, as a rule, are used together; we can only talk about different contributions degree of each method to coercion to work. From the three methods of enforcing to work within the labor division framework, myths are the most cost-effective and the most widely used[6]. It should be expected that in the future AGI will be more involved in the fulfillment of tasks on the ideological front for managing social processes in the labor market.

In the conditions of the labor division, given the need for coercion to certain types of labor, it is difficult to ensure full equality. Purely theoretically, it can be imagined that for the entire long list of various work types and labor conditions, rules are written that provide approximately equal remuneration for equal work. But at the same time, one should either abandon competition with its search for more efficient solutions to production problems or accept the inequality expressed in higher remuneration for those who use more efficient production methods.

The labor division and competition are not the main sources of inequality in modern society. Significantly more inequality occurs from such phenomena as monopolism, monopsony, and rent. But complete freedom and equality cannot be achieved by elimination of these phenomena.

Nevertheless, it seems necessary to maintain and deepen the processes of labor division and competition, as the means of ensuring the progress of civilization checked by history. The agent approach to AGI is most suitable for these ideas

The introduction of AGI in the process of labor division will leave a person more time to carry out a high-level mental activity that is not related to the performance of routine work, which will change the landscape of the labor and education market. It was man's mental activity that served as the basis for the emergence of new methods of production, labor and social relations, and generally ensured the progress of society. AGI will help to accelerate the development of all social processes and rapid social transformations, but progress will be more effective if it will be provided in cooperation human-AGI, and not by replacing of humans.

Most promising is the development of an agent-based approach to AGI, where multiple independent AGI devices interact with each other to solve complex problems. Numerous and less powerful AGI systems are not only easier for humans to control, but they can also exercise mutual control. In addition, the AGI agent's self-improvement process is also easier to direct towards the development of friendliness and usefulness for human civilization.

7. AGI possibilities to influence on policy and economy

The following can be singled out as the main consequences of the introduction of AGI:

a. displacement of a person from routine work;
b. potential control over the managers actions;
c. the complexity of the decisions interpretability, made by AI;
d. the possibility of losing control over the actions of AGI.

Replacing a person with automatic machines that tirelessly carry out routine work, while making fewer mistakes and lowering production costs, does not cause any objections. But if you massively replace employees with automatic machines, this will create a powerful wave of unemployment and spill over into a social crisis. The high cost-effectiveness of AI-based solutions allows unconditional basic income (UBI) payment. UBI can temporary solve some social crisis problems, but will result in crowding out the population from the economy. More promising way is a massive increase in the educational level and the involvement of the population in creative work when the faceless degrading mass turns from an unnecessary burden into a powerful productive force with the growth of

natural intellectual potential. This way will give countries that will choose the development of the creative potential of the population instead of only paying the UBI noticeable advantages in international competition.

Consequences b) and c) are related: both are explainability problems. Managers don't want to be held accountable for their own mistakes, and engineers - for AI decisions. Although the explainability of AI is now seen as one of the most important constraints to its implementation, consequences b) and c) show that there are two problems: too high and too low explanatory skills. Today's AI systems can already generate verbal explanations for these decisions. As with humans, this explanation does not necessarily correctly describe the reasons for choosing such a decision. But unlike humans, AGI devices are more suited for mutual control of decision making.

Technically, the development of an agent-based approach to AGI will contribute to the preservation of cooperation between living people and AGI. In this case, individual AGI agents will merge as part of the modern "superintellects" – state, public and commercial structures and will be able to develop organically the system of checks and balances that exists in today's world. Competition can be extended to the interaction of AGI agents, including monitoring their activities, which will reduce the risk of loss of control.

Achieving the superhuman capabilities of AGI will make it possible to use it not only for solving technical problems, but also for determining the development goals of science and society. Failure to use AGI in these matters will result in a loss in international competition. The benefits of using AGI include participation in economic competition, more informed choice of economic and political decisions. The authorities should not limit the social program to UBI payments. This will lead to the people ousting from the development of civilization with a further reduction in the number of people and their degradation. Only the population creative potential development can help people to participate together with AGI in the scientific, technical and social development of society.

8. Conclusion

We can identify the main tasks in the development of AI and AGI technologies.

1. Priority support for the development of AI and AGI, as technologies that provide a powerful impetus to accelerate scientific and technological progress. Creation of conditions to support researchers and sectors implementing AI pilot technologies in society. These ideas have been developed in a number of national AI development programs, including Presidential Decree No. 490 of 10/10/2019 [7].
2. Formation of a balanced approach to the use of AI and AGI and restrictions on the use of these technologies in the management of armaments, dangerous objects and in areas where control errors can pose a threat to human life and freedom, which is discussed in detail, for example, in [8].
3. Development of rules for the use of AI and AGI in the competition. Creating a legal framework for AGI agents.
4. Intensification of the development of conditions for the realization of the population creative potential, since the creative profession's role will increase as the development and implementation of AGI.

Acknowledgements

The work was supported by Russian Foundation for Basic Research under the project #20-511-00003.

References

[1] Malinetsky, G.G. (). Digital reality at the bifurcation point and the strategic objectives of the Union State in the context of the humanitarian and technological revolution Projecting the future. In: *Digital reality problems: proceedings of the 2nd International Conference*. M.: KIAM RAS, 2019. (in Russian)

[2] McCarthy, J., Minsky, M.L., Rochester, N. and Shannon, C.E. (1955). *A proposal for the dartmouth summer research project on artificial intelligence*. [online] Available at: http://www.formal.stanford.edu/jmc/history/dartmouth/dartmouth.html

[3] Mikolov, T., Sutskever, I., Chen, K., Corrado, G. and Dean J. (2013). *Distributed representations of words and phrases and their compositionality*. [online] arXiv:1310.4546[cs]. Available at: https://arxiv.org/abs/1310.4546.

[4] Minsky, M. and Papert, S. (1972). *Perceptrons: An Introduction to Computational Geometry*. Cambridge MA: The MIT Press.

[5] LeCun, Y. (2017). *Power & Limits of Deep Learning*. [online] Available at: https://www.youtube.com/watch?v=0tEhw5t6rhc.

[6] Harari, Y.N. (2017). *Homo Deus: A Brief History of Tomorrow*. London: Vintage.

[7] (2019). Presidential Decree No. 490 of 10/10/2019 «On the development of artificial intelligence in the Russian Federation».

[8] Tegmark, M. (2017). *Life 3.0: Being human in the age of artificial intelligence (first ed.)*. New York: Knopf.