COMPUTERS AND HUMANS: WHOSE SOLUTION TO CHOOSE?

Shestakova I.

St. Petersburg Mining University, St. Petersburg, e-mail: Irina_Shestakova@inbox.ru

This paper analyzes the issue of trusting the machine that, having been developed from the simplest device for land processing, has become able to make decisions. Since the machine has beaten the world champion in chess, more and more intense discussions arose around the question: should humans, realizing their defeat, continue to reserve their right to the "last word" or entrust the computer with the ultimate solution? This question also relates to both emergencies and events when humans have enough time to figure out the possible solutions. Despite the fact that ethics is a category not applicable to technology, if we entrust the machines with decision-making, this means that we have accepted the ethical guidelines on which they may ground their decisions. Perhaps, if we perform the "social contract" when programming robots, the machine will serve as ethical regulator able to make the humankind act in accordance with the categorical imperative and the highest standards.

Keywords: machine and man, artificial intelligence, ethics, Kant, robots, chess, Kasparov, Deep Blue, scientific and technological progress, computer

A set of issues of moral and ethical nature remains one of the most important areas of discussion related to scientific and technological progress. We have already addressed some of these issues in our papers [1, 2]. Now we would like to turn to the issue of trusting the machines when making decisions in case of emergencies like military attacks or natural disasters, as well as events when humans have enough time to figure out the possible solutions on their own.

The history of the machine: it is only a mechanical assistant

For a long time man has found artificial assistants, beginning with a stick-digger. At the same time throughout the history of technology development, mankind expresses fears about the negative impact of the new that enters into his life, up to fantasies about the end of the world. [2]. For example, at the beginning of the XVII century appears a legend (based on the more ancient Jewish mythology) about the clay giant Golem, who, according to the plan of the creators, had to fulfill the requirements of man, as from already an ancient time people dreamed of an assistant who can perform hard work. However, Golem, gradually gaining experience, began to exercise his will, becoming dangerous for the person himself [3]. The appearance of cinematography brings to the apotheosis the idea of an uprising of artificial intelligence [1]. But at some point the fantasies about the independent will of the machines begin to come true, at least approach it. A person delegates to mechanisms the possibility of making fundamental decisions.

From the moment of large-scale automation, a person implements its capabilities everywhere, trusting the computer part of its powers. The machine performs work, for which the person needed a long training, the presence of special qualities, such as the speed of the reaction. But it's still just a routine job. For example, in aviation, automatic control is activated when the aircraft has gained altitude. The pilot takes control on takeoff, in course determination, on landing, despite the fact that a modern autopilot can land an airplane. The man by inertia continues to be considered more reliable, therefore during the movement of the aircraft next to the autopilot sits a pilot to prevent unforeseen situations.

In this sense, the machine is still seen as a helpless helper. There is no question of ethics in relation to machines, because it only fulfills the will of man. This mechanism is in the hands of a person who can apply the machine in accordance with their ethical principles.

The computer grows to make important decisions

At some point, the computer in its ability to process information "grows" to making important decisions. And then the question arises about the need to recheck the decision of the automatic system and leave the possibility of a final solution for the person. When these decisions are made in a calm environment, the person has time to analyze the decision of the machine and proceed as the person deems appropriate. But there are cases of force majeure, that is, an emergency situation, when you need to respond instantly and there is no time for analysis. How to deal with this situation? Give control to a person or computer?

Until recently, there was full confidence that the human mind, of course, surpasses the ability of both the animal and artificial world, because the machine does everything only by searching for embedded programs, but the person acts incomprehensibly correctly, because it

contains the spark of God. The grain of doubt was dropped when the grandmaster lost the game as a representative of the intellectual elite to the chess computer. [4].

There is no doubt that, the human mind has long been surpassed mechanically: in counting speed, in the speed of reactions. The dramatic ending of Kasparov's game with Deep Blue [4] reveals that he is outnumbered in analytics, hitting in that great game of a man – to chess, which is a measure of intellect throughout history. It would seem, is it possible to find more intellectually powerful people in such parameters as memory and analytics than top-level chess players? At a time when machines started to play chess, there was an opinion that machines can only play at the level of the first category (pure combinatorics, when machines simply sort out the combinations they put in), and the real masters think creatively, but the machine is incapable of it.

Chess as a game of super-intellectuals, like the plot of human history, showing the superiority of man, has come to its dramatic end. It is destroyed. Mankind from this drama leaves with a sad face. With this victory, the computer struck a blow to the grounds for admiring people with themselves and their main property – the intellect. The computer has surpassed man not in a primitive account (with this man has long resigned himself), but on the very top of the chess game, where it was believed that only unique people occupy the first positions, because this game requires creative thinking, intuition and that's why people win against the computer. It was convinction that a man would never ever lose that game [5].

In this case it is important for us not only, that the superiority of the computer and the shaming of the human mind in its intellectual game, mean that it reconsiders the notion of the mind as a concept of a fundamental and inherent only to a person. It leads us to a situation where a person apparently needs to rethink their own role in this universe and develop a new vision of perspective on their superiority. What is important is that a chess loss undermines the basis of trust to human in comparison with a computer.

Moral and ethical casus. Metaphysical thinking

Thus, today there is no doubt that in the game the machine can make the optimal move, which will be either at the level of choice of the best chess player or even better. This understanding breeds metaphysical apprehension. From now on, the person is faced with the need to give an answer to the question of give the computer a vote of confidence, including in solving problems even, where the cost is the human life – in the event of a threat of military attack or natural disasters. The modern machine is a "social fact", and we need to recognize and accept this [6: 195].

The defeat of humanity in chess history forms an understanding that computer decisions are more correct and expedient. Man gradually gets used to the idea of imperfection of his analytical apparatus in comparison with a supercomputer. A person may disagree with a computer, but is already ready to put his superiority to doubt. The sophisticated person understands that the computer has long surpassed a person in the speed of analysis. The computer reacts much faster, so in a situation of force majeure, we must rely on artificial intelligence (Here, by artificial intelligence (AI), we mean the not the being described by Turing, in communication with which we cannot understand that we communicate with the machine, but the AI, the elements of which are already used practically with us, such as GPS navigator, Siri. They are helping us not only to find information, but also to read the text, recognize and translate speech into printing, etc. In this sense, each person can face such a dilemma).

But what if this decision, issued by a computer, from the point of view of a person leads to possible death? Do I need to interrupt the action? Or, perhaps, the computer has calculated all the risks, and any other solution contains an even greater percentage of probability of death? Should a person make their decision or trust the decision of the computer? How would a person act in this situation when making decisions?

The categorical imperative of Kant, or the question of justice

With regard to a given question, one of the most popular and, therefore, common tasks is the problem of choosing the actions of an autopilot in a car in a situation that puts machine before choosing (dilemma) the destruction of one (driver) or many. This is an important issue for automakers because they understand how difficult it will be to sell a car with an autopilot programmed for the "right" (The world annually the transport accidents kill 1 million 250 thousand people [8]. Taking all this complicated dilemmas into account, the introduction of unmanned vehicles reduce the mortality on the roads) choice [7].

This casus, of course, occurs right now, when it is a question of programming

the machine. But this question is also not new. This, in fact, is an eternal question of ethical choice. Every person at every moment of time is faced with a choice: what to do?

To this question, for example, Kant answers, deriving the famous categorical imperative: "Act according to such a maxim that it can become a universal law" [9, 10]. There is still much more ancient "Golden Rule of Morality", with slight variations in Buddha, Confucius, Christ, Mohammed, sounding like "Do to others what you would want them to do to you".

The answer to the question of trust in a computer is implicit in human instinct – the instinct of self-preservation. Self-preservation not only for itself, but humanity as a species. This instinct is given in the animal community. For example, a bee stings immediately when its instinct tells that there is a threat to the community to which it belongs – for its swarm. At the same time, she dies. She sacrifices herself to the life of her community. So in our case, the machine would choose the answer according to the moral and ethical code of the human community.

Philosophers of different ages expressed the opinion that slaves do not have ethics [see, for example, 11, 12, 13: 407-524]. But ethics are not applicable to robots. No tool in itself is neither evil nor good. Everything is poison, and everything is medicine. Ethics is not applied to robots, but to people – their creators. This is a question for social institutions. And if according to universal agreements, for example, the theory of a social contract, we would create a machine based on the "natural state", i.e. Programming cars, we use the categorical imperative of Kant, then everyone gets their advantages. The machine will work within the framework of a decision that is the result of a set of moral rules. People using the car, even if they are not disposed to good behavior, will not break the law, as they will not run over anyone on the crossings and even alcohol will not hurt them.

Such a machine is conceived as something impeccable and beautiful from the point of view of its actions, since it has long been noted that "a reasonable impartial observer can never be pleased even with the appearance of the constant prosperity of a man who is not adorned with a single feature of pure and good will; Thus, goodwill is, apparently, an indispensable condition" [9], and this "goodwill" will be inherent in the machine. In the moral sense, the machine will become perfection, because it will act not according to one's own convictions or from a sense of one's own selfishness, but according to universal imperatives (if properly programmed).

And if at this stage there is still skepticism about the possibility of full implementation of this project, soon the person will be cradled into the robotic environment and there is no longer any doubt about the correct choice that the machine will make, a person will not have. After calculating all the pros and cons, the public for safety reasons is likely to decide not to allow a human to drive the car, since this will endanger other participants in the movement.

Chess and military actions: Do we have to trust a machine?

If for the above problem we apply the concept of the categorical imperative of Kant, then there are tasks that go to the level of humanity, to a planetary level, where the price of choosing a machine is to destroy a huge territory, a group of the population or even the whole planet.

For example, imagine a local battle, in which two opposing armed groups participate. Military officers who build up a plateau of actions must take into account a huge number of facts: their own forces and the enemy, the presence of aviation, etc. However, the battle takes place according to known "rules", everything is as in the game, as in chess, where there are figures. For example, we have "horse" and we know how it can walk, and on the battlefield there is a tank, and we know how it can move. The question is whether it is possible in this situation to not turn to the machine, given that it is able to compare and analyze all data in the most optimal way. You enter the necessary conditions: geography, the availability of forces, etc., and ask which decision is correct.

You can consider the chessboard as a foreign policy life, turning at the same time to the historical reality: Cuba has placed our missiles, the Caribbean crisis is arising. This task of getting out of the complex political situation was solved without computers. Now the decision-making would have been done differently. On each side would use a supercomputer, in which all the data would be laid. It's no secret that the gun has been running for a long time from a computer. If in this situation the computer issues a decision according to which it is necessary to immediately press the "button", and the person responsible for the decision making knows an example about chess and has the understanding that the computer's response is always the best, then for all the reluctance of the disaster, a person still has to press "Button", because he realizes that another solution is even worse.

This example, we give because it is close to chess, to games. Such a dilemma of choice can be endlessly applied to any situation with business, etc.

Such collisions.

Thus, we increasingly trust the machine in making decisions. A large-scale experiment began. In the minds of the mankind for a long time, perhaps from the moment of chess defeat, the background is being formed, the created understanding that the person's decision is most likely not the most optimal one, and, perhaps, a more accurate solution, that which the computer gives out. Gradually begins to occur a shift of consciousness. "Today ICT has a tremendous revolutionizing influence on a person's consciousness, which without false pathos can be viewed as a revolution of the consciousness itself" [16]. There is every reason to assume that with further development, the artificial intelligence can gain an advantage in choosing the optimal action.

This is a fundamental change in views on the tasks of intellectual activity, as well as on the perception of decisions issued by artificial intelligence as something of the best.

Conclusion

It should be noted that when creating machines, very few people thought of them as a regulator of ethical relations in the human community.

The curiosity of the situation is especially attached to the fact that in the minds of most thinkers, scientific and technological progress, if relevant to ethics, is only with a minus sign. This negative attitude is especially aggravated by the development of the Internet. Reflections on robots generally end with apocalyptic predictions for humanity.

Contrary to all negative forecasts, a perspective is seen in which, most likely, everything will turn out in such a way that robots will make a person live according to the "golden rule" and the categorical imperative of Kant.

Realizing this, it is possible today to set the goal to predict the ethics of the future of a robotic society, trying to formulate it and create for it specific new conditions that humanity has never encountered.

References

1. Shestakova I.G. The price of progress: is there a need for a bible for robots? // Almanac of modern science and education. 2016. No. 10 (112). P. 100-103.

2. Shestakova I.G. The price of progress: everything that is done, it is all for the best // Almanac of modern science and education. 2016. No. 10 (112). P. 98-100.

3. Zinger I. B. Golem / Trans. I. Bershtein. M .: Text, 2011. 48 p.

4. Deep Blue wins match // https://www.research.ibm.com/deepblue/home/may11/story_1.html; A Computer Called Watsons. URL: http://www-03.ibm.com/ibm/history/ibm100/us/en/icons/watson/ (circulation date is March 28, 2017).

5. Kasparov G. Deep Thinking: Where Machine Intelligence Ends and Human Creativity Begins. Public Affairs, 2017. — 304 p.

6. Latour B. On the Interobjectivity, Transl. With the English. A. Smirnova; Under the scientific. Ed. V.S. Vakhshtayna // Sociology of Things: Sat. Art. Ed. V.S. Vakhshtayn. M: Territory of the Future, 2006. P. 169-199.

7. Worstall T. When Should Your Driverless Car from Google Be Allowed To Kill You? // URL: www.forbes.com/sites/timworstall/2014/06/18/when-should-your-driverless-car-from-google-be-allowed-to-kill-you/#40c8a1ffa5b6 (circulation date 29.03.2017).

8. Raspopova A. On the roads, young and poor die. // URL: https://www.gazeta.ru/auto/2015/10/19_a_7830845.shtml (circulation date 24.02.17).

9. Kant I. Fundamentals of the metaphysics of morality // URL: http://www.gumer.info/bogoslov_Buks/Philos/kant_omn/01.php (circulation date 28.03.2017).

10. Kant I. Critique of Practical Reason // URL: http://iakovlev.org/zip/kant2.pdf (circulation date 28.03.2017).

11. Varro. Agriculture / Trans. M. E. Sergeenko. M.-L .: Publishing House of the USSR Academy of Sciences, 1963. 220 p.

12. Wallon A. The history of slavery in the ancient world / Trans. With frants. S. P. Kondratieff. M., 1941 // URL: http://www.sno.pro1.ru/lib/vallon_istoriya_rabstva/14.htm (circulation date is March 29, 2017).

13. Nietzsche F. Compositions: In 2 vols T. 2. M .: Thought, 1990.-829p. – P.407-524.

14. The results of research on "techno-optimism" are presented at the Gaidar forum. // URL: http://gaidarforum.ru/news/na-gaydarovskom-forume-predstavleny-rezultaty-issledovaniy-o-tekhnooptimizme/ (circulation date 31.03.2017).

15. Petrov I. Circular defense // URL: https://rg.ru/2014/12/01/centr-upravleniya-site.html (circulation date on March 31, 2017).

16. Sergeichik E.M. New Media: The Revolution of Consciousness // URL: http://www.phisci.ru/files/issues/2016/03/RJPS_2016-03_Sergeychik.pdf (circulation date 31.03.2017).