M. Stavniychuk, Bachelor Student
O. Pasichnyk, Ph.D., As. Prof., language advisor
Khmelnytsky National University

ARTIFICIAL INTELLIGENCE: EXPECTATIONS AND RISKS

Artificial Intelligence (AI) is one of the most popular but contradictive subjects in Computer Science. From computer vision to game playing, it has made a lot of progress in the past few years. Big and influential companies all over the world have already placed huge bets on this technology, and over the next decade AI is expected to gain all possible spheres of human life. AI ranges from machines truly capable of thinking to search algorithms used to play board games. It has applications in nearly every way we use computers in society.

It takes its roots from the study of non-learning artificial neural networks in the researches of Walter Pitts and Warren McCullouch. One more pioneer was Frank Rosenblatt who developed and extended the idea of perceptron, a learning network with a single layer, similar to the old concept of linear regression [3]. Neural networks were applied to the problem of intelligent control (for robotics) or learning, using such techniques as Hebbian learning, GMDH or competitive learning. Later, Alan Turing wrote a paper on the notion of machines being able to simulate human beings and the ability to do intelligent things, such as play chess. The main achievements over the past sixty years have been advances in search algorithms, machine learning algorithms, and integrating statistical analysis into understanding the world at large.

In computer science, the field of AI research defines itself as the study of "intelligent agents": any device that perceives its environment and takes actions that maximize its chance of success at some goal [4].

Scientists are now debating the implications of AI, a fast-moving technology that enables machines to perform tasks that could previously be done only by humans. In the near term, the goal of keeping AI's impact on society beneficial motivates research in many areas, from economics and law to technical topics such as verification, validity, security and control. It becomes very convenient that an AI system does what you want it to do if it controls your car, your airplane, your pacemaker, your automated trading system or your power grid. The potential benefits are huge as everything that civilization has to offer is a product of human intelligence. It's difficult to predict what we might achieve when this intelligence is magnified by the tools that AI may provide.

In the long term, an important question is what will happen if one day an AI system becomes better than humans at all cognitive tasks. As Irving Good realized in 1965, designing smarter AI systems is itself a cognitive task. Such a system could potentially undergo recursive self-improvement, triggering an intelligence explosion leaving human intellect far behind. It raises up one more concern whether we will be able to align the goals of the AI with ours before it becomes super intelligent and control it.

There are some who question whether strong AI will ever be achieved, and others who insist that the creation of super intelligent AI is guaranteed to be beneficial. In my opinion people should be aware of both of these possibilities and direct all their efforts to

prevent such potentially negative consequences in the future, thus enjoying the benefits of AI.

Thus, most experts [2] attract our attention to these most possible scenarios:

1) The AI is programmed to do something devastating. Autonomous weapons are artificial intelligence systems that are programmed to kill. In the hands of the wrong person, these weapons could easily cause mass casualties. Moreover, an AI arms race could inadvertently lead to an AI war that also results in mass casualties. To avoid being thwarted by the enemy, these weapons would be designed to be extremely difficult to simply "turn off," so humans could plausibly lose control of such a situation. This risk is one that's present even with narrow AI, but grows as levels of AI intelligence and autonomy increase.

2) The AI is programmed to do something beneficial, but it develops a destructive method for achieving its goal. This can happen whenever we fail to fully align the AI's goals with ours, which is strikingly difficult. If you ask an obedient intelligent car to take you to the airport as fast as possible, it might get you there chased by helicopters, doing not what you wanted but literally what you asked for. If a super intelligent system is tasked with an ambitious project, it might wreak havoc with our ecosystem as a side effect, and view human attempts to stop it as a threat to be met.

As these examples illustrate, the concern about advanced AI isn't malevolence but competence. A super-intelligent AI will be extremely good at accomplishing its goals, and if those goals aren't aligned with ours, humanity will get in trouble.

To sum up, it is believed that AI won't stop its development due to all advantages it can offer for people. It means that in a few years we will get independent thinking machines and will have to distinguish the ways of cooperating with them. These powerful mechanisms require control and people's oversight. Thus, it's up to us to determine our future.

REFERENCES

1. Глибовець М.М., Олецький О.В. Штучний інтелект. – Київ : «Києво-Могилянська академія», 2002. – 364 с.

2. Benefits and risks of artificial intelligence [електронний ресурс] https://futureoflife.org/background/benefits-risks-of-artificial-intelligence/

3. Frank Rosenblatt Principles of Neurodynamics: Perceptrons and the Theory of Brain Mechanisms Spartan Books . – 1962. – 248 p.

4. Hutter, Marcus Universal Artificial Intelligence. – Berlin: Springer. – 2005. – 198 p.