

## НАУКОВИ ЕСЕ

A. M. Zuiev, a Second-year Master's Student  
Faculty of Cybernetics and Computer Science  
Taras Shevchenko National University of Kyiv,  
60, Volodymyrska Street, Kyiv, 01033, Ukraine  
ant0n3uev@gmail.com

### ARTIFICIAL INTELLIGENCE AND THE PROBLEM OF REDUNDANT PEOPLE

Last five-ten years was very productive in development of machine learning creating some kind of artificial intelligence in the foreseen future starts looking realistic. And creating and development of artificial intelligence will raise a lot of ethical questions as well.

Let's start with what we mean when we say "artificial intelligence". I think that the best definition is next. Artificial intelligence is some kind artificial (partially – computer) system developed and in some kind programmed to perform some tasks that can be performed by humans. Also, ability to learn and modify approaches to task is also often considered to be important part of AI. If we want to call something an AI it also should be able to "think" in some way, which means – being able to approach tasks in a way that wasn't directly put in the system by its creators.

Can we say that AI exists in some way now? No. But we definitely can say that development of AI is very fast. Google and Amazon are already selling devices which can perform simple tasks, as well as developing rival driverless cars. AI systems are being designed for supermarkets which allow customers to choose their shopping and exit without going to a checkout. Soon, robots will be stacking the shelves and running the entire show. No wonder unions are worried.

As for driverless cars, what are the ethics involved in deciding how they should respond to obstacles in their path? How do they differentiate a dead pheasant or a deer from a person who may have fallen down? And who is responsible for incidents with them – person in the car for not interacting with car system in some way, system of the car or maybe the developer of this system?

But nevertheless a lot of systems like "smart house", self-driving cars, even some applications in your phone are already much more than just a sequence of deterministic instructions. Also, a lot of big facilities are already mostly automated and require only a few human supervisors to ensure everything is fine. And how much time we have before AI systems advance so far that human supervisors will be replaced by some kind of decision system?

There is book *"I, Robot"* by A. Asimov and same-named film. They give very fine perspective on how our life will look like if AI and robotics development continue the way they're going now. In this book humans have robot servants and basically main question is "can robots develop consciousness?". We'll return to this as we get some questions this book and film have some answers.

This fast development of AI and grooving implementation of programmed system in places of human work gives us a lot of questions to think about. Let's talk about some of them.

**Is this ethically correct to replace humans by AI/robots?** In first glance we can think that answer is obviously no. Lots of people will lose their works just because AI is cheaper, not because they are bad in what they are doing. But, let's think about it more. We can ask similar question – "Is this ethically correct to replace cabs with cars?" Yes, both horses and cab driver might be doing their work fine. But imagine our world with a lot of horses

and horse cabs around the streets. It's not only about ethics but also about progress. Each improvement and invention will probably hurt some people but afterwards we'll find ourselves in slightly better world.

**Can we believe AI can do their work correctly?** No, we can't. But as well we can't be 100% sure that humans doing their work good. There will probably be some quality problems. But who can say we don't have them now. And, AI's ability to learn will guarantee that same mistake will not happen again. And there probably will be some systems to ensure AI correctness.

**Is replacing everything with AI safe for humans?** In general, I think this can be dangerous. Basically allowing AI to control important parts of our lives can result in some very bad consequences. We need to be sure that such system can't fail and can't harm people. A. Asimov suggested 3 laws of robotics which can be good examples of how task priority of AI can be established

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey the orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Laws.

This rules are made primarily for robot servants with AI elements.

Also if we face AI so intelligent it can realize humanity is a threat for it, consequences might be dire. A lot of films were devoted to this theme, like *"Matrix"*, *"Terminator"*, etc. This show that humanity is already slightly afraid of possibility of AI existence.

Another good question, what would humanity do if all their life needs are supplied by robots and AI. Experiments on mice shows that in utopian overpopulated environment mice "society" slowly degenerated which lead to extinction. And who knows what can happen to humanity in this circumstances?

**Can very advanced AI have human right? Can AI have consciousness?**

This probably the hardest and complex question amongst all. Can be system designed to copy human behavior and way of thinking consider human? Can this system have equal rights, ability to vote, place in society, etc.? Where is the line between intelligent AI and mindless robot? Is it "humanely" to limit capabilities of AI?

I'd probably say that the definitely, advanced enough UI can be considered equal to human. Though, edge between human-like AI and robot is unclear for me now. And it's definitely hard to believe that humanity would gladly share it rights to something so non-human in its nature.

In *"I, Robot"* we're told that robot consciousness can be result of sequence of errors which resulted into forming character. In some other cases, like *"Detroit: Become Human"* we see that AI made to be human-like becomes human. In some films and books AI could still remain "cold-hearted" machine but perform to act like they more human than they really are. It's hard to tell in

advance how humanly would we will try to make an AI and how humanly it will really be.

#### Who's responsible for AI errors?

If we consider AI something that has at least remotely human-like consciousness, is it able to response for its actions or errors? If error in AI leads to consequences would creator of this AI be responsible? What if this AI was developed by another AI? This questions are hard to answer as well. We will need to develop deeper understanding of who's really considered the performer of some action in cases we're dealing with AI. AI definitely should follow some rules (laws of robotics are good example of building such set of rules). And we should establish some formal limit of development and each AI, more develop than this limit must be able to response for its deeds.

Creation of real AI will probably take at least 10-15 years and it's hard to tell now, what it would be. But

А. М. Зуєв, магістр 2-го курсу  
Факультет кібернетики та інформатики  
Київський національний університет імені Тараса Шевченка,  
вул. Володимирська, 60, Київ, 01033, Україна

#### ШТУЧНИЙ ІНТЕЛЕКТ І ПРОБЛЕМА "ЗАЙВИХ ЛЮДЕЙ"

А. Н. Зуев, магистр 2-го курса  
Факультет кибернетики и информатики  
Киевский национальный университет имени Тараса Шевченко,  
ул. Владимирская, 60, Киев, 01033, Украина

#### ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ПРОБЛЕМА "ЛИШНИХ ЛЮДЕЙ"

#### СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Jobs for the bots: will AI make humans redundant? [Електронний ресурс] – Режим доступу: <https://www.irishtimes.com/special-reports/artificial-intelligence/jobs-for-the-bots-will-ai-make-humans-redundant-1.4095412>
2. How 1960s Mouse Utopias Led to Grim Predictions for Future of Humanity. [Електронний ресурс] – Режим доступу: <https://www.smithsonianmag.com/smart-news/how-mouse-utopias-1960s-led-grim-predictions-humans-180954423/>

#### REFERENCES

1. Jobs for the bots: will AI make humans redundant. Retrieved from <https://www.irishtimes.com/special-reports/artificial-intelligence/jobs-for-the-bots-will-ai-make-humans-redundant-1.4095412>
2. How 1960s Mouse Utopias Led to Grim Predictions for Future of Humanity. Retrieved from <https://www.smithsonianmag.com/smart-news/how-mouse-utopias-1960s-led-grim-predictions-humans-180954423/>

Received Editorial Board 08.10.20

K. O. Korovai, a Second-year Master's Student  
Faculty of Cybernetics and Computer Science  
Taras Shevchenko National University of Kyiv,  
60, Volodymyrska Street, Kyiv, 01033, Ukraine

#### SELF-DRIVING CAR DILEMMAS.

#### WHAT ETHICAL PROBLEMS CAN YOU FIND IN SELF-DRIVING CAR PROSPECTS?

Self-driving cars are a completely new way forward in mobility. They are just the latest in a long list of examples of Sci-Fi becoming a *Sci-Fact*.

Self-driving vehicles were the ordinary stuff from science fiction since the first roads were paved. But now they are real, and they are going to radically change what it's like to get from point A to point B.

Science fiction has been successfully predicting the capabilities that can be seen now in modern autonomous cars since the early 1930s. Unfortunately, it is still silent about the legal and ethical implications.

Isaac Asimov, who is famous for his "Three Laws of Robotics", was the first one who predicted the public's anxiety about the bounds of artificial intelligence in terms of driverless cars in "Sally" (1950) [1], a short novel about an autonomous car. The story ends with Jake, the main character, losing trust in his cars, thinking about what the world will become if cars realize that they are effectively enslaved by humans, and therefore revolt.

"There are millions of automobiles on Earth, tens of millions. If the thought gets rooted in them that they're slaves; that they should do something about it... [...] I don't get as much pleasure out of my cars as I used to. Lately, I notice that I'm even beginning to avoid Sally [his favorite automobile]."

The problem described above is one of the general problems in artificial intelligence since all the self-driving systems and driverless cars also rely on AI. This problem

has been a matter of concern to the number of scientists, philosophers, researchers, and the general public for decades. What if artificial intelligence itself (and the driverless cars in particular) turned against people, its creators? This doesn't mean by turning "evil" in the way a human might do it, or the way AI disasters are usually represented in Hollywood movies or Sci-Fi. Still, it is rather a dangerous scenario that people are afraid of. One source of this concern is that controlling a superintelligent machine, that can appear if AI surpasses humanity in general intelligence, may be a harder problem than naïvely supposed. Being a part of human species that currently dominates other species, we used to overestimate ourselves. But what if we just have not had a worthy adversary until recently? The likelihood of this type of scenario is widely debated.

For example, in "Sally" the main antagonist Gellhorn was killed by his autonomous bus that was treated brutally by this person.

"Lord, what a way to die! They found tire marks on his arms and body. [...] The doctor reported he had been running and was in a state of totally spent exhaustion. I wondered for how many miles the bus had played with him before the final lunge. [...] Gellhorn had been a criminal. His treatment of the bus had been brutal. There was no question in my mind he deserved death. But still I felt a bit queasy over the manner of it."