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 what we really can tell, it will give humanity a lot of challenges to deal with and problems to go through.

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

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"

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 automatobiles 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 automatobile]."

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."

Being alarmed about the perspectives described above, a group of scientists came up with the following initiative. In January 2015, Stephen Hawking, Elon Musk, and dozens of well-known artificial intelligence experts [2] signed an open letter on artificial intelligence calling for research on the societal impacts of AI, in which they had publicly voiced the opinion that superhuman artificial intelligence could provide incalculable benefits, but could also end the human race if deployed incautiously. In short, the essence of this letter can be conveyed by the following phrase: *researchers must not create something which cannot be controlled.* It is better to take an advanced AI system as a "genie in a bottle" that can fulfill almost all of our wishes, but with terrible, unexpected, and unpredictable consequences.

We are decades or even centuries away from the world described in Asimov's "Sally" with fully-autonomous smart cars instead of usual ones. AI revolution is even further, thankfully. Currently, there are no legally operating, fully-autonomous (not to mention *intelligent*) vehicles. There are, however, *partially*-autonomous vehicles – cars and trucks with varying amounts of self-automation, from conventional cars with brake and lane assistance to highly-independent, self-driving prototypes.

But it doesn't mean that there are no ethical problems or moral dilemmas.

Nowadays self-driving cars are considered to be safer than regular cars. Hypothetically, self-driving vehicles can and do reduce the number of deaths in car accidents, because the software could prove to be less error-prone than humans.

But there is one small problem.

When a driver jams on the brakes to avoid running over a pedestrian crossing the road illegally, he or she is making a moral choice: either risking the pedestrian or the people in the car. Driverless cars should also make such ethical judgments on their own. The problem is that every algorithm must be clearly defined which means that the steps in the algorithm must be strictly detailed. But settling on a universal moral code for the vehicles could be a thorny task. It reminds me of a famous trolley problem, that is a thought experiment in ethics modeling an ethical dilemma. The trolley problem has been the subject of many surveys in which approximately 90% of respondents have chosen to kill the one and save the five [3]. If the situation is modified where the one sacrificed for the five was a relative or romantic partner, respondents are much less likely to be willing to sacrifice their life [4]. Some researchers criticized the use of the trolley problem, arguing, among other things, that the scenario it presents is too extreme and unconnected to real-life moral situations. In some way, it is true. The problem is that self-driving cars *should know* how to behave and how to react in such extreme situations. If the driverless system has to choose between two pedestrians, who'd be most likely hitten by car: a little schoolgirl or a sweet granny? But what if one of them is your friend or relative? How can a system prioritize? Or, more precisely, how *we* can set those priorities? And who can and should take upon itself a right to decide whose life is more expensive?

The largest survey of machine ethics (of 2.3 million people from around the world) ever [5] finds that many of the moral principles that guide a driver's decisions vary by country. For example, in a scenario in which some combination of pedestrians and passengers will die in a collision, people from relatively prosperous countries with strong institutions were less likely to spare a pedestrian who stepped into traffic illegally.

No matter their age, gender or country of residence, most people spared humans over pets, and groups of people over individuals. Still, before adding some personal interests. But agreement ends there.

Of course, there are lots of critics of this survey as well, although the authors say that their scenarios represent the minor moral judgements that human drivers make routinely.

So for me, this is the main problem is self-driving cars sphere. The survey described above shows that there are no universal rules, according to which driverless car should act in extreme situations. No one can come up with a perfect set of perfect rules for such situations. In most cases – *normal cases* – self-driving cars are safer, because they act according to rules without any violation.

So for now, people need to understand which risks we are willing to take, since it's impossible to remove them completely.

"If no one ever took risks, Michaelangelo would have painted the Sistine floor" – Neil Simon, an American playwright, screenwriter and author.

СПИСОК ВИКОРИСТАНИХ ДЖЕРЕЛ

1. Asimov I. Sally (short story). [Електронний ресурс] / I. Asimov. // World Heritage Encyclopedia. – Режим доступу: http://www.self.gutenberg.org/articles/Sally_(short_story)

2. Research Priorities For Robust And Beneficial Artificial Intelligence: An Open Letter. [Електронний ресурс] – Режим доступу: https://futureoflife.org/ai-open-letter

3. "'Trolley Problem': Virtual-Reality Test for Moral Dilemma – TIME.com". [Електронний ресурс] // TIME.com. – Режим доступу: https://healthland.time.com/2011/12/05/would-you-kill-one-person-to-save-five-new-research-on-a-classic-debate/

4. Journal of Social, Evolutionary, and Cultural Psychology Archived 2012-04-11 at the Wayback Machine. Volume 4(3), 2010.

5. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A. & Rahwan, I. (2018). The moral machine experiment. Nature, 563(7729), 59-64

REFERENCES

1. Asimov, I. Sally (short story). Retrieved from http://www.self.gutenberg.org/articles/Sally_(short_story)

2. Research Priorities For Robust And Beneficial Artificial Intelligence: An Open Letter. Retrieved from https://futureoflife.org/ai-open-letter

3. "'Trolley Problem': Virtual-Reality Test for Moral Dilemma – TIME.com". Retrieved from https://healthland.time.com/2011/12/05/would-you-kill-one-person-to-save-five-new-research-on-a-classic-debate/

4. Journal of Social, Evolutionary, and Cultural Psychology. Wayback Machine, Archived 2012-04-11, Volume 4(3), 2010.

5. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A. & Rahwan, I. (2018). The moral machine experiment. *Nature*, 563(7729), 59-64

Received Editorial Board 08.10.20

К. О. Коровай, магістр 2-го курсу
Факультет кібернетики та інформатики
Київський національний університет імені Тараса Шевченка,
вул. Володимирська, 60, Київ, 01033, Україна

ДИЛЕМА БЕЗПІЛОТНИХ АВТОМОБІЛІВ.
ЯКІ ЕТИЧНІ ПРОБЛЕМИ МОЖНА ЗНАЙТИ В ПЕРСПЕКТИВАХ БЕЗПІЛОТНИХ АВТОМОБІЛІВ?

К. О. Коровай, магистр 2-го курса
Факультет кибернетики и информатики
Киевский национальный университет имени Тараса Шевченко,
ул. Владимирская, 60, Киев, 01033, Украина

ДИЛЕММЫ БЕСПИЛОТНЫХ АВТОМОБИЛЕЙ.
КАКИЕ ЭТИЧЕСКИЕ ПРОБЛЕМЫ МОЖНО НАЙТИ В ПЕРСПЕКТИВАХ БЕСПИЛОТНЫХ АВТОМОБИЛЕЙ?