Interdisciplinary Description of Complex Systems 17(4), 716-722, 2019

THE DANGER OF USING ARTIFICIAL INTELLIGENCE IN DEVELOPMENT OF AUTONOMOUS VEHICLES

Gabor Kiss*

Óbuda University
Budapest, Hungary

DOI: 10.7906/indecs.17.4.3
Regular article

Received: 17 December 2018.
Accepted: 22 December 2019.

ABSTRACT

The world of autonomous vehicles approaches as technology evolves. Researches have been done, development has been made in several countries, car manufacturers have already marketed their semi-self-driven automobiles. Nowadays artificial intelligence is present across nearly all industries due to scientific achievements in the field of artificial neural networks [1], computer vision [2], and a multi-layer neural network [3]. Utilizing AI for developing autonomous vehicles has been an obvious choice as making decisions based on continuously flowing vast amount of information from different sensors requires fast processing. In case of industrial AI where decision making is based on video image analysing, false decisions can lead to categorizing either flawless products as faulty or wrong products as good. In case of human politics when artificial intelligence is used to determine tender winners, making the wrong call could only mean gender biased results [4]. However in case of self-driven cars making bad decision might equal causing accidents and endangering people's lives, such as it happened to Uber [5]. Scientists at MIT successfully developed the World's first psychopath AI, which achievement claimed the responsibility of educating non-natural minds [6]. The aim of this article is to point out those situations and scenarios in which self-driven cars could be hijacked, misguided, captured, or even influenced to turn against other vehicles.

KEY WORDS

autonomous vehicle, hijacking, capture, cheat, candidate AI

CLASSIFICATION

JEL: O33

*Corresponding author, η: kiss.gabor@bgk.uni-obuda.hu; -; -

INTRODUCTION

The topic of self-driven cars has recently become popular, although the idea has a long history. Radio-controlled cars existed in 1925. The Chrysler Imperial was capable of cruise controlling in 1958. In 1995 Mercedes developed an almost fully autonomous car that travelled 2 000 km, but had room for only one person, the driver due to the multiple electronic devices the task needed. Google started its self-driven car project in 2009. Tesla's Autopilot software has been available since 2015 and has been updated ever since. Japan plans to transport the visitors of the 2020 Olympic Games only by using autonomous taxis [7]. As of now various car manufacturers provide partially self-driven cars, and self-driven vehicles are present in fixed path public transportation (trams and metros) in metropolises. Self-driven cars are in plan to enter traffic within ten years of time, meanwhile Tesla's strategy announced two years.

The expectation of spreading self-driven cars lies in the hope of significantly decreasing the 1,3 million death toll accidents world-wide, which are caused by human factor 90 % of the time. In policies of insurance companies the reaction time of a human realizing any dangerous situation, reacting to it and putting the breaks into action is two seconds. The reaction time would be reduced by the power of AI since it can process a huge amount of data coming from sensors and, with information corresponding to the situation, could make decision much faster than humans.

Until the end of November 2018 Tesla cars travelled one billion miles in self-driven mode using the Autopilot software which was issued in 2015. The mileage is five times of the Sun-Earth distance. The accidents statistics of the journey showed one accident or accident-like incident every 3,34 million miles. According to the US Department of Transportation there is an accident every 492 000 miles in America, making the self-driven mode seven times safer [8].

The aim of this article is to discuss situations and scenarios when artificial intelligence of vehicles could be confused or influenced to make bad decisions endangering passengers' lives.

SAE LEVELS

The Society of Automative Engineers (SAE) determines the intelligence level and automation capabilities of vehicles on six levels, from 0 to 5. On level 0 there is no automation, fully manual vehicles belong in the category. Level 1 is the lowest level of automation with only one automated fuction, like steering, speeding or braking control, most cars of our days belong to this category. Level 2 vehicles are capable of automated steering and acceleration and may be capable of self-driving in zero-traffic environment, with clearly visible lane painting, although the driver has to indicate all times readiness of taking control (e.g. by touching the steering wheel). Level 3 cars can detect environment and travel in self-driving mode for a longer amount of time, but in case of any problem, they stop and give the control back to the driver, which might cause further problems under certain circumstances like in the inner lane of highways.

The key difference between level 3 and level 4 automation is that level 4 vehicles are able to intervene themselves if things go wrong or there is a system failure. In this sense, these cars are left completely to their own devices without any human intervention in the vast majority of situations, although the option to manually override does remain in difficult or preferable circumstances. Level 5 cars must be able to take passengers to their destination fully self-driven taking care of all problems occurring during journey. Car manufacturers build redundant systems in order to avoid malfunctions [9].

EDUCATION OF ARTIFICIAL INTELLIGENCE IN AUTONOMOUS VEHICLES

Thanks to the most recent findings and results, it is an obvious choice to develop artificial intelligence for self-driving cars to meet the quality and time requirements of decision making in as complex situations as being in traffic. The requirement is processing the data coming from sensors and cameras and having a decision made under the thirtieth of a second.

When teaching the AI the simplest would be to only focus on traffic rules before decision. It would be also more simple from the perspective of the law in case of any accident. But being in traffic is not simple.

Engineers at Tesla have developed a shadow mode for the Autopilot software. In shadow mode the system monitors the drivers' actions and sends the information to the central database for further development. The idea is based on the notion that experienced drivers are capable of deviating from dangerous situations instantly. Although gathering data from vast number of unexperienced drivers might cause problems.

Various target hardware is present in vehicles. One of the most novel developments for teaching self-driving cars comes from NVIDIA. This tool creates a lifelike, detailed, interactive world that is not only for gaming but is useful for AI education purposes as well [10].

Hyundai CRADLE invested into the start-up, Perceptive Automata whose product observes the behaviour of pedestrians and tries to estimate the behavioural outcomes, such as the possibility of pedestrians changing their minds while crossing intersections letting autonomous cars cross first [11].

During the transitional period switching to self-driven cars from human-driven ones, recognizing insecure driving behaviour may be a key feature. Drivers with years of experience in traffic can recognize insecure chauffeurs instinctively. If drivers show shaky control over their cars in front of experienced drivers, the latter ones will either overtake them or will tail away to avoid getting into risky situations. The AI of autonomous vehicles might be trained to do so as well.

SITUATIONS TO SCAM SELF-DRIVEN CARS

There is no perfectly secured system, there will be no 100 % safe solution for self-driven cars either. S. Chen, the head of BlackBerry, has claimed that his company will be able to provide 90 % security for the systems of autonomous vehicles, although the system must be monitored at all times from the moment of first usage [12].

This article focuses on situations that can confuse or scam self-driven cars, and does not focus on analysing how to hack artificial intelligence of vehicles instead. The aim has been to convince automobile manufacturers to test their products for these possible scenarios, for a more safe traffic environment dominated by autonomous vehicles.

THE DANGERS OF CHANGES IN HUMAN BEHAVIOR

Artificial intelligence will be much faster processing and responding to traffic situations providing a safer solution for driving. In the proximity of autonomous vehicles, using their shortened reaction time, drivers in human driven cars might abuse it to their advantage. Drivers might cut in lane in front of self-driven cars, forcing it to use the break, or driving their cars into intersections relying on the approaching self-driven car's software to stop, or enrolling in-front of them at highway entrances, etc.

Traffic might slow down because of these probable scenarios. There might also happen accidents caused during the transitioning era if a human driven car abuses the shortened

reaction time of a self-driven one followed by a human driven car, which needs larger time to react. It also holds the possibility of planned accidents caused by human driven cars targeting the other human driven car following a self-driven one.

FAKE ROAD SIGNS

Modern cars have program for road sign identification, then show it to the driver on dashboard or windshield. This warns driver for speeding, restriction of parking, etc. Research had been made at Princeton University about fooling of this system and it had 90 % success of attacks, which could aware on interesting situations [13]. Why is it possible? Just in the European Union use we different road signs with different colours with same meaning. [14]. For example, we can change a "Road closed to all vehicles in both directions" sign to a "Maximus speed limit 50" in the city for make an accident, see Figure 1.

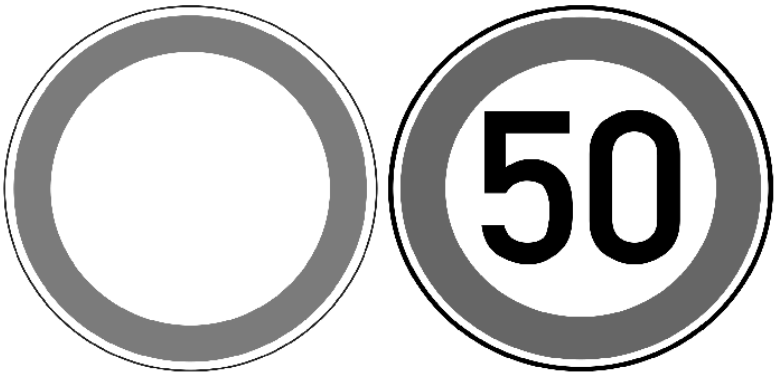

Fig. 1. Fake road signs.

Another example: one can change a " Maximum speed limit 130" sign on the highway to a "Road closed to all vehicles in both directions" by covering the numbers, thus causing a traffic jam. As what road sign will the system interpret that case? Map of the vehicle knows the original meaning but detected data should have higher priority at decision making, which is good at road works, but generates insecurity in this case. Moreover, if a printed sign (for example on paper) is placed irregularly it can force autonomous cars to change their driving direction. That can occur with an entering restriction sign forcing the cars to choose different way. Capturing is a possibility by this, because one can have signs placed on both roads, or one can create rush if the two-way traffic is let on a one-way street. Somewhat more advanced possibility is to project a tri-dimensional sign using a LED projector. Would a system identify it as a road sign, or not?

FAKE LANE

Tracking system is already able to keep vehicle in the middle of the lane, if paint and visibility are adequate. If these terms are no longer adequate, a sign appears for the driver to take care from now on. *Digital Light* technology was developed by Mercedes and it uses one million mirrors in both reflectors to works as a HD resolution projector. This device can enlighten onwards and it can also project symbols or lanes (Figure 2) like a same technology. Does this technology make possible deceiving a vehicle behind to follow the fake lane projected directly in front of it? This could cause a voluntary accident or could send an autonomous vehicle to a predetermined location.

Sensors can be bothered in a much simpler way, if we cover, plaster or paint them while parking at parking lot. This makes a vehicle spastic, because it would not get data from sensors thereby wasting time of its passengers. Consequently that bounds their freedom to places where travelling is only possible by car because of the distances covered.

Figure 2. Fake lanes.

CANDIDATE AI

Apparently, usage of artificial intelligence is spreading among developer com-panies, because it can be a better solution for self-driving as earlier methods like, the continually learning system of Tesla. Artificial intelligence based system is a hard challenge for developers, because it must be prepared for and taught to eve-ry possibly situations and avoiding its dangers. This teaching method makes a difference. Researcher of MIT in 2018 presented a psychopath AI, called Norman [6]. Norman took a Rorschach test just as a simple AI [15], to allow analysing differences between their answers, what do they see in inkblots, to highlight the importance of teaching method in decision making of AI [16]. Study of MIT researcher shows that an AI could be made with killing functions like, "Christine" in the book of S. King [17]. In the world of autonomous vehicles this fact could be very important. Take an example of Tesla in June 2018, one malcontent staffer was enough to make trouble [18]. Of course, an evil car would be caught during a long testing period, but the plan could succeed, if it would base on the idea of the movie *The Manchurian Candidate* [19]. In this case, the killing function would link with a look of rare road sign, rare situation in traffic, a song from the radio, etc. This object could activate malice and change protecting functions to attacking, that could create an accident, injury or a terror event and an investigation needs long time to find out and then recall all Candidate AI. The investigation could be make more difficult, if killing function activates at the sign but acts only a random time or distance later, so all situations would be different. This would delay the detection, but the Manchurian AI could not be used for mass destruction.

CONSLUSION

Various scenarios and situations have been mentioned that could scam self-driven cars or confuse the AI to make the wrong decision. Educating the artificial intelligence is crucial, validating raw data and data protection must be taken into consideration. We counted a lot of situations, where it is worthy to prepare the systems for. A pretended, or even an incidental case can bring about the situation for which the system is not prepared, and which it could not treat properly. An autonomous vehicle must be aware to problems of the mentioned situations.

ACKNOWLEDGMENTS

The research presented in this article is a part of the project *Dynamics and Control of Autonomous Vehicles meeting the Synergy Demands of Automated Transport Systems*, No. EFOP-3.6.2-16-2017-00016, in the framework of the New Széchenyi Plan. The project is funded by the European Union and co-financed by the European Social Fund.

REFERENCES

[1] Aizenberg, I.N.; Aizenberg, N.N. and Vandewalle, J.: *Multi-Valued and Universal Binary Neurons: Theory, Learning and Applications*.
Springer, Boston, 2000,
http://dx.doi.org/10.1007/978-1-4757-3115-6,

[2] Fukushima, K.: *Neocognitron: A self-organizing neural network model for a mechanism of pattern recognition unaffected by shift in position*.
Biological Cybernetics **36**(4), 193-202, 1980,
http://dx.doi.org/10.1007/BF00344251,

[3] Carvalho, A.; Fairhurst, M.C. and Bisset, D.L.: *An integrated Boolean neural network for pattern classification*.
Pattern Recognition Letters **15**(8) 807-813, 1994,

[4] Lee, D.: *Amazon scrapped 'sexist AI' tool*.
BBC News, October 10 2018,
https://www.bbc.com/news/technology-45809919,

[5] Wakabayashi, D.: *Uber's Self-Driving Cars Were Struggling before Arizona Crash.*
The New York Times, March 23 2018,
https://www.nytimes.com/2018/03/23/technology/uber-self-driving-cars-arizona.html,

[6] Yanardag, P.; Cebrian, M. and Rahwan, I.: *Norman, World's first psychopath AI*.
http://norman-ai.mit.edu,

[7] Jenn, U: *The Road to Driverless Cars: 1925 – 2025*.
https://www.engineering.com/DesignerEdge/DesignerEdgeArticles/ArticleID/12665/The-Road-to-Driverless-Cars-1925--2025.aspx,

[8] Lambert, F: *Tesla owners have driven 1 billion miles with Autopilot activated*.
https://electrek.co/2018/11/28/tesla-autopilot-1-billion-miles,

[9] SAE: *Taxonomy and Definitions for Terms Related to Driving Automation Systems for On-Road Motor Vehicles*.
SAE Standard J3016_201806, SAE, 2018,

[10] Nvidia: *Invention Has Potential to Create Virtual Worlds for Gaming, Automotive, Robotics, VR*.
https://nvidianews.nvidia.com/news/new-nvidia-research-creates-interactive-worlds-with-ai,

[11] Hyundai Motor: *Hyundai CRADLE Invests in Perceptive Automata to Bring Human Intuition Software to Self-Driving Cars*.
https://www.hyundai.com/worldwide/en/news/news-room/news/hyundai-cradle-invests-in-perceptive-automata-to-bring-human-intuition-software-to-self-driving-cars-0000016052?pageNo=4&searchKey=&rowCount=6&type[]=RES&listPageUrl=news.release.all,

[12] TechSecurity.news: *BlackBerry CEO John Chen warns driverless cars could turn into "fully loaded weapons" if hacked*.
https://techsecurity.news/2018/09/blackberry-ceo-john-chen-warns-driverless-cars-could-turn-into-fully-loaded-weapons-if-hacked,

[13] Sitawarin, C.; Bhagoji, A.N.; Mosenia, A.; Chiang, M. and Mittal, M.: *DARTS: Deceiving Autonomous Cars with Toxic Signs*.
https://arxiv.org/pdf/1802.06430.pdf,

[14] –: *Comparison of European road signs*.
https://en.wikipedia.org/wiki/Comparison_of_European_road_signs,

[15] Rorschach, H.: *Rorschach Test – Psychodiagnostic Plates*.
Hogrefe Publishing Corp, Cambridge, 1927,

[16] Exner, J.E.: *The Rorschach: A Comprehensive System*. Vol. 1: Basic Foundations.
John Wiley & Sons, New York, 1995,

[17] King, S.: *Christine*.
Viking Publishing, 1983,

[18] Kolodny L.: *Elon Musk emails employees about 'extensive and damaging sabotage' by employee*. https://www.cnbc.com/2018/06/18/elon-musk-email-employee-conducted-extensive-and-damaging-sabotage.html,
[19] Condon R.*: The Manchurian Candidate*. McGraw-Hill, 1959.