Assoc. Prof. Aleksandra Kuzior, PhD.
Faculty of Organization and Management
Department of Applied Social Sciences
Silesian University of Technology

Katowice, 20.04.2019 r.

Review of the doctoral dissertation on.: „On the Application of Norms within Driverless Cars" by Michael P. Musielewicz.

The submitted to review doctoral dissertation on.: „On the Application of Norms within Driverless Cars" prepared by Michael P. Musielewicz addresses extremely important issues related to the development of modern technologies currently used in the so-called autonomous vehicles. The topic of the dissertation and the considerations it contains are particularly important because they indicate the need to pay attention not only to the technological side, but also, and perhaps above all, to the ethical side of the technological solutions proposed for common use. The discussion about autonomous vehicles, or, as the author of the dissertation underlines, commonly called Driverless Cars began with the appearance of technical capabilities to produce this type of vehicle. Initially, the reflection concerned only the technological side, which did not take into account ethical aspects. However, it was quickly realized that the production and use of such vehicles raises many ethical problems, so similarly as in the case of other modern technical solutions, the ethical view of emerging technological solutions is necessary. Therefore, both in scientific research and in journalism, these topics began to appear more and more often, though far from systematic and full of scientific reflection on the subject. I have the impression that there is still a gulf between the humanities and technical sciences, a chasm analogous to that defined by Snow in the book entitled "The Two Cultures and the Scientific Revolution" between humanistic sciences and natural sciences, so evocatively described by the author. Some humanists and engineers, however, try to build a bridge connecting these two cultures, but it seems to me that, for now, we have only managed to pass a narrow, wobbly footbridge. It seems that the majority of engineers and project teams working on technological innovations, neither at the idea stage nor in further research, take into account ethical aspects. Before implementing innovation, it is more important to research the market to see if the sale of an innovative product or service will be profitable. Ethical, humanistic aspects are omitted, not only at the stage of work on an innovative product, service or process, but also in academic education. On the majority of engineering studies, there are not many philosophy or ethics classes in the grid of hours, usually 15-30 hours in the entire 7-semester cycle, which is not

enough to properly prepare future engineers for the proper valuation of science and technology, also in terms of ethics. Complementary 3-semester MA studies virtually do not include humanities at all. An alarming phenomenon is the replacement of humanistic subjects with social subject, for example, in some fields instead of philosophy, there are introduced subjects like sociology, psychology, negotiations, interpersonal communication are introduced instead of philosophy, however ethics, which is the backbone of human moral activity in the world, is forgotten. Usually, if humanistic subjects appear in the grid of hours, they are assigned 1 ECTS point, which means they are considered to be of little importance. The message addressed to students in this way is clear - it is only a minor addition to studying a given technical field. Of course, we cannot generalize. A positive example is, among others, the Catholic University of Lublin, which introduces the history of philosophy, ethics and logic as separate subjects in engineering studies in information technology or biotechnology. General trends, however, are such that ethics is a subject rarely included in study plans, and yet so necessary in shaping the world of modern human values, among which ethical values should be considered the most important, because they affect all relations of man with other people and the world around him. They set the way of behavior and actions taken by people, show how to live to make life good, wise and beautiful. Thus, they are the foundation of all other values, also in relation to the world of technology and its products.

This general preliminary reflection leads to the conclusion that the doctoral dissertation on „On the Application of Norms within Driverless Cars" by Michael P. Musielewicz is a scientific study that is absolutely necessary, because it is actually the need to include an ethical context in the development of modern technologies that places a lot of emphasis. He points to the need to make ethical considerations in the context of technology development, or more precisely, technical solutions that use / apply artificial intelligence. Besides, in the case of the use of intelligent technologies, we can speak of a specific empowerment of the world of technology, which, having obtained a certain degree of autonomy, is capable of self-control and self-restraint. If this type of empowerment has taken place, then one must take up a scientific reflection on intelligent technologies, as subjects of moral responsibility, based on specific resources of appropriate normativity and look for analogies or differences. The human subject, as Lenk and Spinner indicate, makes decisions and acts on the basis of free will and instrumental, cognitive, normative and reflective reason. But can intelligent subjects of technology work and make decisions in an analogous way? Will they become independent of human intentionality? Originally, the technique was subordinated to human intentionality. It seems that in the era of intelligent technologies the

situation has changed and the intentionality of man has been subordinated to technology, i.e. that as a result of functioning and communication of specific technological solutions, the technology determines or will determine what is "good" for a human, if a man remains the highest value whose technology is to be used. There are many questions and dilemmas regarding the development of smart technologies. For some of them we find answers in the presented doctoral dissertation. The author of the doctoral dissertation set an ambitious goal to answer the question "whether driverless cars can be bearers of norms and are capable of being normative agents who can follow both legal and ethical norms?". The author, in the next chapters of the work, tries to answer the question, what is very successful. The structure of the work does not raise formal objections. The work consists of six chapters (including introductions and final conclusions) as well as abstract, acknowledgements, table of contents and bibliography. The acknowledgments address the author's family and the professor supporting his research and dissertation. The motto at the beginning of the dissertation – "Thou shalt not make a machine in the likeness of a man's mind" – taken from "The Orange Catholic Bible" also gives it a specific color, because it refers the reader to the world of science fiction, to the Dune universe created by Frank Herbert. As reading the thesis „On the Application of Norms within Driverless Cars" it turns out that this is the real world of technological intelligent solutions and possibilities, although a warning or rather a command not to create machines in the likeness of people gives us a lot to think about, especially that we enter the areas reserved only for human beings, as an intentional subject of moral responsibility .

Chapter 1 of the work is an introduction to the issues. The author draws attention to technological progress in the areas of robotics and artificial intelligence, and also draws attention to the fact that "intelligent" robots have permanently become a part of human life, if only because of the great interest of the media. At the same time, the author claims that attention should be paid to the ethical side of such technological progress. Various institutions in the USA or the European Union have taken up this approach. This issue is also of great interest in the academic world, also among philosophers who are the most inquisitive about ethical issues related to the functioning of cars without a driver in society. The author rightly emphasizes that traditional philosophy treated technics/technology as ethically neutral, and yet only the connection of it with the user requires consideration of ethical issues. In other words, there was a belief that the use of technology might be ethics or unethical, and the technique itself is ethically neutral. It is a pity that on this occasion the author does not mention the so-called "normative turn", which took place at the turn of the 1960s and 1970s,

when the technique ceased to be treated as ethically and axiologically neutral. The traditional philosophical view of the neutrality of ethical technology derives from the position of Aristotle, who distinguished three kinds of human activities: cognition, action and production. The last of them is related to the technique (techne), and in the sense of Aristotle that was not subject to ethical evaluation. As you can see, this view has long been rooted in European culture, and the discourse on the role of ethics as a regulator of activities in science and technology and the factor of legitimization of its results, started only in the 1970s, continues to this day. The development of autonomous cars, however, brings new challenges.

Chapter 1 also contains a brief overview of the content of individual chapters, as it is accepted in this type of scientific studies and a reference to literature, which is the basis for the author's deliberations.

In Chapter 2 of the reviewed doctoral dissertation the author indicates that discussions on the creation and implementation of autonomous vehicles are serious. All over the world, regulators at various levels are trying to convince the society of them. The author mentions that these institutions compete with each other to lead in this technological revolution. The author discusses present and currently implemented legal solutions in various corners of the world. The author indicates also that there is a lively discussion in society about the creation and use of autonomous cars. He also points out that many phrases are used to describe the phenomenon, such as autonomous vehicles, self-driving cars or driverless cars. The author mentions that the creation of autonomous cars is possible thanks to the technological progress (also in the field of software) that took place over the last century. The development of such software suitable for autonomous cars requires the use of a variant of statistical methods and machine learning. The creation of software that will enable the vehicle to anticipate the actions of other objects in its surroundings is the basis for the interactive aspect of autonomous vehicles. The author indicates that the manufacturers of autonomous cars check the operation of a form of deep-learning technology joined with reinforcement learning as methods of learning to drive by autonomous cars, so that they are able to apply the right solution in both typical and unusual situations.

The author notes also that autonomous vehicles have a specific knowledge base in which they store information. As the author emphasizes, they are stored so that the vehicle can understand the space-time relationship between the vehicle and its surroundings. Thanks to built-in sensors and the interpretation of signals transmitted through them, a car without a driver is able to make better and better decisions about driving and overcoming obstacles. The

author also raises the issue of interactivity of autonomous cars. He emphasizes that this issue largely depends on the vehicle's ability to predict the actions of other objects in its surroundings. Other objects are understood here as: other cars with lorries, bicycles, horse and pedestrian teams.

The author also points to the importance of the psychological factor. Although there is a plan to constantly improve road safety, many people still do not trust vehicles without a driver and are afraid of the driver's inability to intervene in the emergency operation of the vehicle. These fears are justified because existing and having been tested autonomous cars in some cases have not been able to avoid serious collisions, often tragic in consequences.

The chapter contains a number of detailed descriptions and many technical details, as well as the history and now of autonomous cars. Is well prepared in terms of content.

In chapter 3 the author points out that creating ethical norms related to autonomous cars is presenting them as some kind of normative factors that function within a given normative system. Therefore, according to the author, several factors should be considered. The author uses the assumption that autonomous vehicles are really agents, or entities that have the ability to act. The author argues that the issue of defining a normative agency is based primarily on the norm itself. If, therefore, there is a need to make a normative claim against other persons, it is assumed that they act as an agent and, at the same time, have certain obligations related to the claims in question. However, when talking about cars without a driver, it is difficult to determine who such claims should be directed to.

The author indicates that defining standards as rules regulating the behavior of individual agents leads to the conclusion that they have two interrelated aspects. First of all, the types of activities that are obligatory, prohibited or allowed are determined, and then these activities are assigned to agents in the system. The author notes that if we deal with normative principles, we notice that each standard has both objective and subjective content.

The author also points to the issue of responsibility for autonomous cars. Problems caused by moving away from the human factor to artificial intelligence are common especially in the field of corporate law.

The author argues that the role of drivers in autonomous cars should be minimal, and at the higher levels of automation, human activity is unjustified.

The author points out that, more and more often today, there is a discussion about the concept of personality, with particular emphasis on legal personality in relation to

autonomous vehicles. This discussion results to a large extent from the history of the concept itself, which can be seen in popular literature, and it is also propagated in the media.

According to the author, if autonomous vehicles could be given legal personality, such action would change the way in which they are assigned to particular normative systems. In particular, it would allow a vehicle without a driver to be a legal agent under a given legal system. As a result, an autonomous car would itself become a driver and enter into the obligations imposed on drivers by the law.

In the chapter 4. the author draws attention to the need to define ethics as a necessary element to consider the autonomous cars and points out the "ethics of autonomous vehicles".. According to the author, the discussion about the ethics of cars without a driver results in the need to reflect on the ethical consequences for such devices. This discussion would include such elements as: the impact of autonomous cars on the environment, their usefulness to people with disabilities, and the number of lives saved. The author also focuses on the "ethics for autonomous cars", which would include changes in the social order in such a way that these vehicles could be implemented at all, including allocating funds for research or the advantages of building new infrastructure to support such vehicles. The author also mentions that the "ethics in autonomous cars" should be taken into account, i.e. an examination of how the vehicle should work to be independent of the driver. The author points out possible dilemmas that will appear in crisis situations, in particular when an autonomous car will have to make a decision on who or what it strike when there is no option to completely avoid a collision. The author, however, is of the opinion that people will still expect a clear declaration of how autonomous cars will behave in such situations. The author here invokes the example of a trolley, translating it into the reality of automatic cars. He wonders how such a vehicle would behave if there were failures in it and would only be able to stay on the existing lane, which would result in killing five people, or changing the lane to the neighboring one, killing only one person.

The author points out that at least two disciplines take part in considerations on ethical issues related to autonomous cars. The first of these is philosophy, and in its framework also ethics, while the second one is IT. Therefore, the considerations concern in particular the importance of ethics in computer science. According to the author, the definition of what we mean under the term "good" is of great importance here.

The author indicates, that it is commonly accepted that human activities should be oriented towards usability. The same ethical reasoning should therefore be used in autonomous cars.

The author calls it a kind of ethical school for cars without a driver. Such an ethics school could be the stage of programming an autonomous car, when the owner or producer would have the opportunity to minimize the risk of failure and optimize the work of an autonomous car, as well as program the desired result to be achieved in a specific situation.

The example of the "trolley" cited by the author also touches on the concept of sacrifice. This is particularly interesting when it is considered that dedication will be the right solution to the problem. In this case, the typical desired result according to the author is to minimize losses and maximize the good.

The author refers to two concepts of "teaching" autonomous cars, how to behave ethically. The first concept consists in the default programming of a mandatory ethics setting (MES), according to which each vehicle in an analogous situation will work exactly as predictably. An alternative is to program a personal ethics setting (PES) in which autonomous cars are programmed in accordance with the ethical standards of users. The author indicates that the main advantage of PES is that it enables the vehicle user to autonomously choose the ethics of their car. Such ethics may sometimes include more selfish behaviors that prefer non-autonomous car drivers. In turn, the advantage of the MES system will be that the public will have knowledge of how each vehicle will behave in a given situation, how to reduce unwanted social effects, minimize the risk, and allow to sacrifice the user of the vehicle, if this translates into saving more people. Will this approach, however, convince the public to accept autonomous cars that will put the good of others above the welfare of their owner? What about the category of confidence in technology? There are many additional questions and dilemmas, which the author undertakes in the next chapter.

In chapter 5. The author, referring to the subject of autonomous vehicles, describes, inter alia, a consequentialist approach. The author raises the issue of egoistic consequentialism. An egoistic perspective indicates that we should undertake activities that maximize our individual happiness. The author distinguishes here two categories: ethical (a set of rules according to which the agent should act) and psychological (including agents' thoughts and the way the agent actually worked). Referring this analysis to autonomous vehicles, one can conclude that a car should behave in such a way as to create as little risk as possible for itself and its user. The author, however, rejects egoistic consequentialism because - in his opinion - an autonomous car does not have its own world and its own psyche to be able to make such decisions. Next, the author discusses a utilitarian approach, based on actions and rules, and raises the issue of altruism. However, this last approach is not without

drawbacks. In this case, as the author points out, the vehicle values the utility of other traffic participants more than its own, which means that it can take action to the detriment of itself or its user. This generates a number of problems, especially when the potential damage that an autonomous vehicle user may experience is greater than the damage of other traffic participants. The author also analyzes hedonistic and pluralistic consequentialism as well as the possibility of using them in autonomous cars. In the hedonistic approach, the greatest difficulty for the car will be to determine the degree of pleasure and pain. The author indicates that if the car works in the interests of its owner or current user, it should know what actions can bring them pleasure or pain. However, this is not possible because they are highly subjective and impossible to analyze for the car. However, if you consider a pluralistic approach, there are also dilemmas, how the vehicle can know the ethical standards known to man, and how they will be perceived by other agents in the system. According to the author - even if the car recognizes certain circumstances, it will have a lot of difficulty in determining their weight and priority.

The author also discusses the principles of deontological ethics and its significance for autonomous cars. This kind of ethics puts a lot of emphasis on the obligation and adherence to the rules or programming of fixed rules in the vehicle. However, the author does not fully approve positions indicating the important role of deontological ethics in this context. He formulates two objections to its application, referring to the category of duty and free will of autonomous vehicles. The first objection is pragmatic and concerns the nature of artificial intelligence, while the second refers to the meaning of will.

The author, pointing out the disadvantages of consequentialist and deontological theories in applying to autonomous cars, at the same time is looking for an ethical approach that would be the best and indicates the ethics of virtues. Referring to the Aristotelian roots of the ethics of virtues, he also points to the possibility of adapting the "old ethics to new technologies", referring to the studies on the ethics of virtues for robots. The author of the dissertation agrees with the position of Berberich and Diepold that one should look for ways to build a "virtuous machine". The author recommends a teleologically oriented ethics of virtues, which becomes the subject of in-depth analyzes of the author of the dissertation. Considering the issues of virtue ethics with regard to autonomous cars, the author mentions that many times an autonomous car may find itself in a seemingly impossible situation. It is not able to work perfectly, so it must work in order to choose the best possible action, taking into account all the circumstances. Is that possible? The author creates a list of virtues that an autonomous car should have in order to be able to make the best decision. Ultimately,

however, he focuses on two virtues: justice and kindness. At the same time, the author indicates that this is an introduction to future work and changes, not imposed guidelines. When the vehicle becomes more technologically advanced, it will be enabled to take on more and more tasks as well as better decisions in emergency situations. The acquisition of virtues by autonomous vehicles takes place through the use of machine learning techniques in which evil actions are "punished" and good "rewarded".

Chapter 5 is a very reliable analysis of the possibilities and limitations of the implementation of selected ethical theories to the ethical layer of autonomous cars. The cited arguments pointing to the inadequacy of the discussed consequentialist and deontological theories are convincing. Is the implementation of the teleologically oriented ethics of virtues to the needs of shaping the "virtuous machine" sufficiently argued and convincing? In my subjective opinion, not entirely. However, the method of argumentation and great insight and accuracy deserve a high grade. The author makes very detailed analyzes of potential cases of violation of moral norms by autonomous cars. He analyzes various hypothetical and real situations in which ethical dilemmas arise or may appear. In a very competent way, he discusses the various threads of ethical tradition, referring to new challenges related to the development of modern technologies. In a very consistent manner, he also conducts his scientific dissertation, pursuing the goals set at the beginning and answering the research questions posed. The construction of the dissertation is closed with a content buckle. Issues included in the introduction (Chapter 1) are referred to (repeated) also in Chapter 6, the summary, containing the conclusions of the scientific argument. Nevertheless, the summary chapter leaves some dissatisfaction, although compensated by partial conclusions in individual chapters, but still unsatisfied.

At the end, the author introduces an important remark, referring to the need for further research on this subject. In my opinion, such research in the era of dynamically developing technologies based on artificial intelligence is very necessary and should be carried out systematically. The brief characteristics of the content of individual chapters outlined above do not, of course, reflect the complexity and insight of the author's inquiries, but they show the general outline of the work's content. Apart from thorough ethical analyzes, the author also refers to the philosophy of law and the theory of law, deontic logic or computer science. However, there is a lack of a broader context and reference to the philosophy of technology and technical ethics, which would be desirable due to the issues raised. In addition: the bibliography is arranged chaotically, and should be arranged alphabetically; in analyzes, the author does not always refer to source texts, as in the case of Aristotle, the proportions

between chapters are somewhat disturbed (too laconic Chapter 6). These comments do not discredit the substantive value of the work.

Aiming to conclude and evaluate the final review of the dissertation titled "On the Application of Norms within Driverless Cars" prepared by Michael P. Musielewicz, it should be stated that: first, the author realized the goal set at work and responded to the research questions posed; secondly, thoroughly and diligently analyzed the scientific literature used for the preparation of the work on the issues discussed; thirdly, the dissertation is an original study containing substantive and logical arguments; fourthly, the author's scientific presentation shows his good substantive preparation both in the field of technical and IT knowledge as well as philosophical knowledge, mainly ethical. Therefore, I declare that the reviewed doctoral dissertation by Michael P. Musielewicz I consider to meet the statutory requirements and conditions set for doctoral theses, specified in the Act of 14.03.2003 on academic degrees and academic title, and on degrees and title in the field of art (Journal of Laws No. 65, item 595, as amended) and I assess positively. Taking into account the importance of the subject matter and the originality of the research, I apply for the acceptance of the doctoral dissertation and admitting it to the public defense.

Aleksandra Kaziór