

## ETHICAL-THEORETICAL RECOMMENDATIONS FOR THE AREA OF CASE STUDIES OF NEW TECHNOLOGIES

Daniela FOBELOVÁ<sup>1\*</sup>, Tomáš FORGON<sup>2</sup>

<sup>1</sup> Politechnika Śląska, Wydział Organizacji i Zarządzania, Katedra Stosowanych Nauk Społecznych;  
Daniela.Fobelova@polsl.pl, ORCID: 0000-0003-2972-315x

<sup>2</sup> Cardiocentrum Nitra, s.r.o.; tforgon@gmail.com

\* Correspondence author

**Purpose:** The implementation effort of ethics involves actively entering the current discourse on new technologies, enhancing their credibility and minimizing ethical risks. Ethics can play an important role in enhancing their credibility.

**Design/methodology/approach:** The case method and its application can play an important role in this area. Case studies should not only be the result of specific research and their design but also take into account the current methodological requirements of applied ethics. By presenting a case study, we try to express the optimal variant of a case study and the application of these requirements.

**Findings:** We will try to point towards the establishment and development of the discourse of ethics and new technologies as well as the increasing potential of applied ethics and its constructive role in resolving ethically dilemmatic situations and creating preventive mechanisms for potential ethical failure.

**Originality/value:** The principles of utility minimisation and utility maximisation will trouble us for a long time to come as we introduce them into AI technologies. From this we can see that there will be mainly two dominant ethical theories - utilitarianism with the norm of minimizing the loss of life, health, suffering and deontological ethical theory with the ethical norm concerning the protection of the car passengers, their life. Personally, we would add the ethics of responsibility, which ethicists and lawyers will have to deal with because of not only material responsibility but also, say, loss of life. Here we can be partly inspired by the debate and conclusions on animal ethics.

**Keywords:** case study; new technologies; applied ethics; engineering and medical ethics; artificial intelligence.

## 1. Introduction

The origins of ethics date back to the period before the common era, when it was referred to as practical philosophy. It is important to say that it applied to the actions of so-called free beings. Already in the *Odyssey* (Homéros, 1966) we have the story of how Odysseus, on his return from the Trojan War, hanged a dozen of his female servants for bad behaviour. Because slave women were seen as property, Odysseus' actions were not considered unethical or inadequate. Since those times, ethics have evolved so that the moral attitudes of today are extended to all human beings. This has not stopped the development of ethics. On the contrary, Aldo Leopold (1887-1948) (Aldo, 1949, pp. 221-226; Kuzior, 2006, pp. 266-277), in his work *The Land Ethic*, pushed it further by expanding it to include the land, plants, and animals in addition to human beings. Since the Middle Ages, land as well as plebeians were considered property. Certain rights were exercised towards the land but no duties. Since the 1960s, various scientific initiatives (Peccei, 2005, pp. 39-46) to protect planet Earth and its climate have been emerging (Kuzior, 2014). The end of the 20th and especially the beginning of the 21st century has extended the limits of ethics to human products in the form of new technologies such as the Internet, artificial intelligence (AI), etc. (Kuzior, Kwilinski, Tkachenko, 2019, pp. 1353-1376; Fobel, Kuzior, 2019; Kuzior, 2021). This shift in ethics does not mark the final stop. We must therefore deal with the application of particular ethical theories - in the process of constructing and programming AI - in a way that avoids as much as possible the risks in practice. This is primarily due to the unstoppable progress in AI research and application, which itself brings into the debate questions of value as well as ethical attitudes in the field of rules of conduct, i.e. the use of AI (Kuzior, Sira, Brozek, 2022, pp. 69-90; Kuzior, Sira, Brozek, 2023).

In a short excursus we will try to indicate which of the ethical theories are applied in the construction and programming of artificial intelligence.

## 2. Ethical-theoretical recommendations for case studies

We begin with the oldest ethical theory, i.e. virtue ethics. Virtue – “the ability to act on the basis of certain accepted values” (Fobelová, 2002). It refers to a person who acquires it by life's practical experience, habit and practice. In AI, it would be mainly a combination of dianoetic (theoretical i.e. rational) virtues such as the capacity for wise judgment, scientific thinking and ethical (theoretical volitional) virtues that would be at the birth of AI programming. According to Aristotle, the combination of both kinds of virtues would mean that the rational person is the one who is able to find the middle way in his or her actions, and the wise person is the one who is able to pursue true happiness throughout his or her life. This ethical theory is therefore of

particular relevance to the selection of scientists, AI creators, but especially those who will use (pay for research and do business with) artificial intelligence.

The ethics of duty, with hints already appearing in the ancient ethical thought of the Stoics (4th century B.C.) and perceiving duty as a natural moral law or unwritten law for human actions (Tullius Cicero, 1913) above all a theory linked to the ethics of Immanuel Kant (Kant, 1990). For Kant duty meant the rationality of man. Man has a duty to do good for good's sake. This is deontological ethics, which examines man's moral motivation. All this is done against the background of Kant's view that we can substitute all things for other things, but we cannot substitute man for other man. The reason is that things have their price, but people have their dignity. So that man is never an instrument, a means to an end, but always a purpose, led Kant to explore the distinction between moral motive and utility, moral practicality and limited pragmatism. For both motive and purpose are present in moral decision-making. This is an important point for proper application to artificial intelligence if we want to apply an ethical dimension to this product. Compared to other ethical theories the deontological approach to the regulation of artificial intelligence (AI) is more in line with international agreements anchored by human rights and respect for human dignity, freedom (moral choice), equality and solidarity. We can see moral obligations as negative or as positive, but this does not solve the problem, because the universalistic understanding of morality and moral laws that apply universally to everyone in the same situation has long been invalid. Only the rational side of the will can constitute a moral valuation of action. Duty compels one's will and actions to honour moral laws that derive from reason. Such action is what Kant calls legality, in contrast to morality, which presupposes acting out of duty. Practical laws apply to the will, no matter what is caused by their causality. Deontological ethics (ethics of duty) is (should be) one of the supporting ethical theories in the creation of artificial intelligence (e.g. motive, intention to protect life in programming autonomous vehicles).

The ethics of utilitarianism (utility, benefit) has developed dominantly in parallel with deontological ethics (rationalism, rigorism) in a different cultural and mental environment (empiricism and hedonism). It is a type of ethics where the principle of utility from the position of the good for all is preferred as a moral criterion of action. What is ethically significant is not the motive or intention but the act and its outcome. J.S. Mill saw utilitarianism as the “art of living” as a unity of morality, politics and aesthetics. Utilitarianism is built on four moral principles:

1. Principle of Consequence.
2. Principle of utility (usefulness).
3. Hedonism (good as happiness, satisfaction).
4. Social (happiness for all concerned).

These moral principles are also fully applicable to artificial intelligence (autonomous vehicles). Conduct is subject to the rule of the majority at the expense of the minority. Man is responsible for all consequences of actions (even those he did not cause).

Utilitarianism is divided into act utilitarianism - an action is right only when it produces the best possible consequences (e.g. J.J.C. Smart, D. Regan - cooperative utilitarianism, D. Holbrook - qualitative utilitarianism, etc.) and rule utilitarianism (rule utilitarianism) - an action is right only when it follows a certain chosen rule valid in a society, social group. (R.B. Brandt, J.C. Harsanyi, P. Singer - preferential utilitarianism). Preferential utilitarianism considers an act as morally right only when it corresponds as closely as possible to the preferences of all beings affected by the act. A person who chooses an act should be informed about all possible alternatives to a future act. That person should think logically without prejudice or emotion in the decision-making. Consequentialism in its non-utilitarian form is an attempt to solve a problem by minimizing suffering, unhappiness.

The following three aspects of utilitarian ethics are essential:

1. Consequentialist - acting to bring about the best possible consequences.
2. Eudaimonistic - maximum happiness for maximum number of people.
3. Hedonistic - maximum pleasurable or maximum satisfaction of desire, happiness, pleasure, delight.

These principles include the principle of impartiality, which states that a moral subject should (must) attribute the same value to the needs of all moral subjects with the same consequences in his/her decision-making.

Based on the above, we can conclude that this ethical theory represents a type of ethics that is relatively easy to apply in practice, albeit with some difficulties. The principle of impartiality may run into a possessiveness problem. If someone owns an AI product (e.g. a robot or an autonomous vehicle) who is it supposed to serve? Would it be ethically acceptable for it to serve only the owner? Or should it constantly evaluate in the spirit of the theory in question and act on the basis of maximizing utility, benefit for all? Because in practice, we would expect the norm to apply that when we are the owners the AI will serve us. We would have to program the AI not with a pure version of utilitarian ethics but to combine it with a deontological norm (the duty to always favour the AI owner).

If we combine a form of non-utilitarian consequentialism i.e. negative utilitarianism (K.R. Popper) emphasizing the minimization of suffering, misfortune, and the use of autonomous vehicles (AV in practice), it may mean that in the spirit of ethical theory the AV will sacrifice the crew in case of their crash if the utility of their sacrifice is even slightly greater than the utility of not sacrificing them. If it's not supposed to protect me and the people present in the car, then why should we acquire it? The research and production of autonomous vehicles has a highly humane goal of reducing the number of road casualties ideally by up to 93%. However, assuming it highly protects the car's occupants at the expense of pedestrians, the moral credit of the autonomous car with humans vanishes. The designers of AV attempt to solve this moral dilemma by combining in some proportion a type of utilitarian and deontological ethical theory. It is necessary to identify the limit, the boundaries of acceptability of using both theories for the sake of the objective i.e. preservation of human life, health as well

as property. It would be a technical and ethical hybrid of the autonomous vehicle. This is not quite feasible in practice, because it is difficult to predict what will happen when the limit of one or the other ethical theory is exceeded even minimally. Hence the difficulties in determining the consequences of an action (these concern mainly the utilitarian ethical theory because the deontological one is directed to the motive of the action), which mainly concern the quantification of the maximization (minimization) of the good (evil, harm) - life, health, death, fractures, amputations, etc. Ultimately, such “bargaining” sounds absolutely immoral, inappropriate. One healthy person can save 5 lives by transplanting 5 organs, and so killing or letting die this healthy person is theoretically consistent with utilitarianism. We will not encounter a pure classical or non-classical form of utilitarian ethical theory in the field of artificial intelligence for the above reasons, even though it appears that it would be applicable to its control algorithms.

In the second half of the 20th century, deontological ethics was followed by the ethics of responsibility (H. Jonas), especially in the field of environmental protection. It is an ethical responsibility based on the voluntariness of the commitment we make, based on four components: who is responsible, to whom, for what and according to what criteria. With this ethical theory, some problematic questions about artificial intelligence (autonomous vehicles - AV) come to the fore Who will be responsible - the producer, the owner, the AI?

Finally, but not excluding all ethical theories, we will focus on the ethics of principles by V.R. Potter (Potter, 1971), one of the oldest in applied ethics, which emerged in bioethics and is currently experiencing its twilight. The ethics of principles is based on what are known as - *prima facie* - principles, namely:

1. Beneficence acts as a moral norm in a positive aspect. As far as artificial intelligence (AI) is concerned, it will be required to behave and act beneficially towards humans at all levels.
2. Non-maleficence as a moral norm says that if you cannot help at least do no harm, i.e. do not cause evil, misfortune, suffering, which also applies to AI in relation to human life, their health - physical and mental, protection (quality) of the environment, animals, plants and climate in general.
3. Autonomy - this is the free informed choice to lead a good life according to one's wishes. Artificial Intelligence (AI) is ordered in the form of moral norms not to lie, not to restrict movement, freedom, etc.
4. Fairness - a moral norm addressing the issue that everyone must get what is due to them while maintaining fairness in various spheres of life provided by the AI.
5. Transparency - accountability as a moral norm specifies this principle of AI ethics in requiring auditability and intelligibility by humans. People's life with artificial intelligence is present therefore we need to increase people's trust for it.

This ethical theory only works when all the principles are positively fulfilled, which is impossible in practice (as the practice of bioethics shows).

Situational ethics - deals with real, concrete phenomena, processes that cannot be predicted. Man exists in each situation uniquely unrepeatably, and therefore general, universally valid ethical norms cannot apply. A normative view is only deducible from a particular situation. Using this type of ethical theory would mean producing every single model of, for example, an autonomous vehicle in custom form, which would be costly and would not satisfy most people.

#### **4. Case study: Caterpillar or calculator - autonomous vehicle**

Statistics show that the majority of road accidents is caused by the human factor. Worldwide, 268,087 people have died in these accidents as of March 15, 2022 ([www.worldometers.info/sk/14.3.2022](http://www.worldometers.info/sk/14.3.2022)). It is therefore a moral challenge for the engineers - the designers of artificial intelligence (autonomous vehicles) - to work on this project so that the responsibility is not shifted to artificial intelligence alone. A few autonomous vehicles have already been produced and are on the roads around the world.

In one of the world's metropolitan areas lives a more financially well-off family XY. The wife CD of a wealthy but mainly busy businessman ED, with their two minor children, longed for an autonomous vehicle that would make her life easier, but mainly ensure her safety, given that she is not an experienced driver. The husband ED agreed to the suggestion and subsequently bought the autonomous vehicle for his wife CD. She used it without any problems until she has a collision with another vehicle and both vehicles burst into flames. Fortunately, this collision resulted in minor injuries to the passengers in the autonomous vehicle, but unfortunately one person from the non-autonomous vehicle died. The wife CD began to be troubled by a moral dilemma regarding the safety of non-autonomous vehicles and the extent of her safe car's liability.

Hypothesis: *We assume that the wife CD will be more interested in the safety of the autonomous vehicle (assumed ideal safety is 93%) than in the consequences of collisions with other vehicles.*

##### **Solution alternatives**

1. The wife CD refuses to continue using the autonomous vehicle due to the consequences that remain with non-autonomous vehicles.
2. The wife CD, although frightened by the deaths of other passengers, decides to continue to use the autonomous vehicle because of the desirable consequences to her from collision with non-autonomous vehicle.
3. The wife CD learns from the experience and seeks to communicate the experience of using an autonomous vehicle in practice, so that conclusions are not drawn from laboratories or a single case only, but become paradigmatic.

The first alternative - to abandon the use of an autonomous vehicle - means fleeing, i.e. from the position of virtue ethics, it is a certain, if partly understandable, cowardice. From the position of the ethics of responsibility - who? - the wife CD is responsible (for what?) for the safe use of the autonomous vehicle (with respect to whom?) with respect to herself, her family and society (according to what criteria?) according to the supreme value, which is life, which every normal person cherishes, this is also assessed as a negative action. Given that she is one of those people who can afford this type of vehicle it is (should be) her duty - an ethic of duty based on reasonableness to help create safer transport.

The second alternative of getting scared and so preferring the use an autonomous vehicle so that no one threatens her and the children is commendable but supremely selfish from a position of virtue ethics. From the point of view of the ethics of utilitarianism - utility maximisation, happiness maximisation - this is also a negative attitude. Ensuring the greatest possible safety for oneself and loved ones may be a duty but it should not be at the expense (against the categorical imperative) of other road users.

The third alternative is balanced. The mother attempts to provide security for herself and the children, which from the aspect of virtue ethics we evaluate positively as bravery. From the perspective of responsibility ethics, this is a positive attitude towards herself and the children on the part of the mother CD (who?) with respect to her family and society (with respect to whom?) for safe transport (for what?) according to the expected benefits of this autonomous vehicle (according to what criteria?). It is also a reasonable duty of the mother CD in the spirit of the rules of the ethics of duty.

### **Solving the ethical dilemma**

From the position of normative ethical theories, we consider the optimal solution to the moral dilemma to be the alternative listed as the third. In the ethics of the 21st century we observe a certain retreat from absolute universalism and at the same time the emergence of particularism, pluralism or discourse ethics, norms of contextualism and coherentism.

Therefore, the reasonable position of the wife-mother CD to use an autonomous vehicle, but to share the experience of it with the designers not only technical but also moral, is to be valued highly especially from the position of virtue ethics - as wisdom. From the aspect of the ethics of utilitarianism, we will especially highlight the maximization of utility, happiness and the minimization of suffering. So the moral algorithm driving the autonomous vehicle should accept, according to this ethical theory, the minimization of suffering, unhappiness, etc. If we choose the standards of the ethics of duty - even on the basis of rationality we would remain in the plane of the protection of the passengers of the autonomous vehicle, in other words, the owner (or those to whom he/she would give the vehicle to use) and his/her life would be taken into account. The prima facie standards (harmlessness, beneficence, autonomy, justice, responsibility) would only partially work with this AI. In terms of the ethics of responsibility, the mother-wife CD (who?) acted responsibly for safe transport (for what?) with respect to

herself, her family and society (with respect to whom?) according to her conscience and the values (especially the value of life) recognized by society.

This position is also viewed positively from the perspective of an ethic of fairness, which would ensure equal opportunity for all those involved in transport, and beyond, without guilt or remorse for doing the wrong thing.

The hypothesis was not confirmed because wife CD approached the solution wisely and sensibly.

## 4. Conclusion

If all interested parties would like autonomous vehicles to fulfil a moral status, we have no choice but to seek and find a balance (a norm of coherence) between the ethical requirements of their potential users and the regulation of the “behaviour” of autonomous vehicles in non-standard situations - in an accident, a collision, etc.

From this we can see that there will be mainly two dominant ethical theories - utilitarianism with the norm of minimizing the loss of life, health, suffering and deontological ethical theory with the ethical norm concerning the protection of the car passengers, their life. Personally, we would add the ethics of responsibility, which ethicists and lawyers will have to deal with because of not only material responsibility but also, say, loss of life. Here we can be partly inspired by the debate and conclusions on animal ethics.

The principles of utility minimisation and utility maximisation will trouble us for a long time to come as we introduce them into AI technologies.

## References

1. Callicot, J.B. (1989). *In Defense of the Land Ethic, Essays in Environmental Philosophy*. Albany: State University of New York Press.
2. Cicero, M.T. (1913). *De Officiis*. With An English Translation. Walter Miller. Cambridge: Harvard University Press.
3. Fobel, P. (2011). *Prípadosť – aplikácie – etika*. Banská Bystrica, FHV UMB.
4. Fobel, P. et al. (2013). *Organizačná etika a profesionálne etické poradenstvo*. Banská Bystrica.
5. Fobel, P., Kuzior, A. (2019). *The future (Industry 4.0) is closer than we think. Will it also be ethical?* AIP Conference Proceedings, 2186, 080003.
6. Fobelová, D. (2000) *Kultúra v živote človeka*. Banská Bystrica: FHV UMB.

7. Fobelová, D. (2011). *Prípadová štúdia ako vzdelávacia metóda v aplikovaných etikách. In Prípadovosť - aplikácia - etika*. Banská Bystrica: Univerzita Mateja Bela, Fakulta humanitných vied. ISBN 978-80-557-0213-1, pp. 83-100.
8. Kant, I. (1990). *Kritika praktického rozumu* (Filozofické odkazy). Bratislava: Spektrum. ISBN 80-218-0051-8.
9. Kant, I. (2004). *Základy metafyziky mravov* (Filozofia do vrecka). Bratislava: Kalligram. ISBN 80-7149-635-9.
10. Kuzior, A. (2006). Obowiązki człowieka wobec środowiska naturalnego w świetle Kantowskiej metafizyki moralności. In: A.J. Noras (ed.), *Pytania i perspektywy transcendentalizmu. W dwusetną rocznicę śmierci Immanuela Kanta* (pp. 266-277). Katowice: Wyd. Uniwersytetu Śląskiego.
11. Kuzior, A. (2014). *Aksjologia zrównoważonego rozwoju*. Banska Bystrica: Belianum.
12. Kuzior, A. (2021). *Applied ethics*. Lublin: Wydawnictwo Naukowe Tygiel.
13. Kuzior, A., Kwilinski, A., Tkachenko, V. (2019). Sustainable development of organizations based on the combinatorial model of artificial intelligence. *Entrepreneurship and Sustainability, Iss. 7(2)*, 1353-1376.
14. Kuzior, A., Sira, M., Brozek, P. (2022). Using Blockchain and Artificial Intelligence in Energy Management as a Tool to Achieve Energy Efficiency. *Virtual Economics, 5(3)*, 69-90.
15. Kuzior, A., Sira, M., Brozek, P. (2023). Use of Artificial Intelligence in Terms of Open Innovation Process and Management. *Sustainability, 15(9)*, 7205.
16. Leopold, A. (1949). *A Sand County Almanac and Sketches Here and There*. New York: Oxford University Press.
17. Leopold, A. (1949). The land ethic. In: L. Aldo, Ch.W. Schwartz (eds.), *A Sand County almanac, and sketches here and there* (pp. 221-226). New York: Oxford Univ. Press.
18. Meadows, D.H., Meadows, D.,L., Randers, J. (1992). *Beyond The Limits. Envisioning a Sustainable Future Confronting Global Collapse*. Post Mills, Vermont: Chelsea Green Publishing Company.
19. Meadows, D.H., Randers, J., Meadows, D.L. (2004). *Limits to Growth. The 30-Year Update*. White River Junction, Vermont: Chelsea Green Publishing Company, ISBN 1931498512, ISBN 978-1931498517.
20. Meadows, D.L. (1970). *The Dynamics of Commodity Production*. Cambridge, Massachusetts: Wright-Allen Press.
21. Meadows, D.L., Behrens, III W.W., Meadows, D.H., Naill, R.F., Randers, J., Zahn, E.K.O. (eds.) (1974). *Dynamics of Growth in a Finite World*. Cambridge, Massachusetts: Wright-Allen Press, ISBN 978-0960029440.
22. Meadows, D.L., Meadows, D.H. (eds.) (1973). *Toward Global Equilibrium: Collected Papers*. Productivity Press, Inc., ISBN 978-0262131438.

23. Meadows, D.L., Meadows, D.H., Randers, J., Behrens, III W.W. (1972). *The Limits to Growth. A Report for the Club of Rome's Project on the Predicament of Mankind*. Washington, DC: Potomac Associates Books; New York: New American Library; New York: Universe Books, ISBN 0876631650.
24. Meadows, D.L., Meadows, D.H., Zahn, E., Milling, P. (1972). *Die Grenzen des Wachstums. Bericht des Club of Rome zur Lage der Menschheit*. Reinbek bei Hamburg: Deitsche Verlag-Anstalt GmbH; Stuttgart: Rowohlt Taschenbuch Verlag GmbH, ISBN 3499168251.
25. Patro, T. (2017). Umelá inteligencia: Čo to je, ako funguje a prečo je dobré sa o ňu zaujímať? *Časopis FIT ČVUT*, 2017.10.15.
26. Peccei, A. (2005). The Club of Rome: Agenda for the End of the Century. In: P. Malaska, M. Vapaavuori (Eds.), *Club of Rome. Dossiers 1965-1984* (pp. 39-46). Vienna: Finnish Association for the Club of Rome (FICOR), European Support Centre of the Club of Rome. ISBN 952-99114-1-6.