

Artificial Intelligence – Ethical Issues

Stanisław Ciupka¹, Oliwia Wiczorek²

¹Bielsko-Biala University of Applied Sciences,
Poland

²University of National Commission of Education in Cracov
Poland

Abstract— The accelerating progress in the field of artificial intelligence (AI) is giving rise to increasingly lively philosophical discussions and new ethical challenges that humanity must face. An inevitable element of this progress is the growing autonomy of artificial intelligence in terms of making decisions that are not directly supervised by humans. Many AI decisions give rise and will give rise to moral conflicts and dilemmas. It is worth considering today what measures are necessary to equip future autonomous, self-learning and self-replicating devices equipped with artificial intelligence, and at the same time capable of acting independently and in a large range of variability of external conditions, with a specific type of ethical intelligence. The problem that both the designers and users of objects equipped with artificial intelligence must face is the need to optimally balance the reasons, needs and interests between both sides of human-nonhuman interaction. In a situation of growing autonomy of artificial intelligence increasingly does not fit the dominant anthropocentric ethics. It is becoming necessary to expand and modify the model of ethics, which will allow to predict and encompass the so far insufficient area of mutual relations between humans and artificial intelligence.

Keywords— AI, ethics, bussines, security of life

I. INTRODUCTION

The extraordinary advances in the field of artificial intelligence are having an increasingly significant impact on all human activity, personal and social life, politics and the economy. There is a need for an open dialogue about the importance and possibilities of these new technologies that can have a positive impact. However, the need for such a dialogue also stems from the fact that there is a fear among many researchers, ordinary people about the destructive impact of

artificial intelligence:

" Some argue that intelligent systems will automatically be moral systems. However, in a rational system, achieving goals is completely separated from reasoning and considerations about the model of the world. Benevolent intelligent systems can be loaded with harmful goals. These harmful goals - consisting, for example, in opposing the goals of other players or destroying these players - are unfortunately very easy to define. Therefore, it will be crucial to design an infrastructure capable of detecting and controlling the behavior of harmful systems (Omohundro 2020). "

AG Lisi allows himself to draw our attention to the fact that it is not impossible that artificial intelligence at the head of a sovereign government may bring to life the basilisk Roko (Lisi 2020). Should humanity be subject to the verdicts of cold rationality presented by artificial intelligence, or are we ourselves capable of defining the limits of its introduction, when it comes to our professional and personal lives (Szulczewski 2019). Systems based on AI can use large data sets, perform a huge number of calculations, draw far-reaching conclusions. Today, they can even independently build and optimize new algorithms that allow them to improve their operation, use computing resources more efficiently, and, in addition, unlike people, they do not forget anything (Lipski 2023). It seems that we are emerging with a significant competitor who is to solve problems subject to automation and in a much more efficient way than today's man. In addition, it should be noted that artificial intelligence skills are much easier to develop than human skills, because in order to develop knowledge and skills among people, we need a significant number of years and a specific upbringing and teaching process,

as they say from kindergarten to university. We are aware that this is a very energy-consuming, expensive process, not free from many risks. Meanwhile, the duplication and dissemination of knowledge in the world of artificial intelligence is extremely fast compared to the human process. The emergence of *artificial intelligence* gives humanity a great chance to solve many problems that, due to limitations the possibilities of human reason were unsolvable. However, this fact also creates new problems, because artificial intelligence is, on the one hand, more efficient than human reason and can make independent decisions. Before the development of AI, technological achievements improved many human possibilities, but at the end of this process stood a human and his decisions, and to whom a specific responsibility could be assigned. However, as he notes Jan Andrzej Lipski:

"Today, machines can perform increasingly complex reasoning and make decisions that affect people and the environment. Machine actions are no longer morally neutral" (Lipski 2023).

Therefore, securing influence on the directions of development of artificial intelligence before it goes beyond human control is a special task for philosophers, ethicists, and computer scientists who operate in today's reality, and not relegating this problem to future generations.

II. ETHICAL ISSUES RELATED TO ARTIFICIAL INTELLIGENCE

Already from the initial remarks made above, it can be seen that the field of research on issues related to artificial intelligence can in no case be found in a scientific vacuum. Scientists who reflect on AI issues face problems similar to those to which other scientists are subject, i.e. the responsibility of scientists and the question of whether the development of science should be subject to control (Łupkowski 2005). In a text from 1991, Stefan Amsterdamski pointed out that the increasing professionalization of science leads to a process in which practicing science is no longer just a private matter of the scientist, but an obligation towards the institution, e.g. the one in which the scientist is employed, and in his opinion this professionalization of scientific research makes one aware that the times of a peaceful conscience are passing, when:

"...the scientist could be convinced that his ethics as a scientist are his methodology. If he remains faithful to the proven rules of research procedure, he contributes to the advancement of knowledge, ergo - he does humanity a clear good (Amsterdamski 1991)."

There is a significant specialization among scientists, which can result in a situation where a specialist in a given field does not fully understand the discoveries they are making. Let us once again quote the words of Stefan Amsterdamski, who says that specialists:

"...they do not feel responsible either for the biological or social consequences of these discoveries, nor for their adverse side effects, nor for their misuse (Amsteramski 1991)."

Let us also pay attention to the view expressed in the following words by Monika Torczyńska, who states:

"In contemporary realities, the essence and significance of artificial intelligence are analyzed not only in the considerations of academics and inventors, but are also the object of interest, views and assessments of so-called ordinary people. It is obvious that their understanding and opinions on artificial intelligence will not fully coincide with the image formed in the academic world. However, they are an important voice in the overall social debate on the real place and role that artificial intelligence has to play in individual and collective human existence. Despite the diversity of opinions and positions, one observation seems obvious today: <the world has switched to sixth gear and is rushing at a dizzying pace carried by the power of technology. We are currently witnessing an important dialogue on the significance and impact of artificial intelligence on the development of the economy, science, society and each of us. These discussions are often filled with emotions. The vision of robots taking over the world is mixed with the heavenly perspective of AI for the glory of humanity"

(Malczewski, 2019).

The emerging potential of advanced artificial intelligence, the fear of side effects in its development, are and must be discussed more and more loudly. Many scientists are convinced that the ongoing work on AI requires the formulation of specific guidelines promoting responsible innovation (Ress 2020). One of the ways proposed by some researchers of issues in the field of artificial intelligence is the idea of self-limitation in this research, imposing specific restrictions on the freedom of scientific research (Lem 1996). Some researchers, for example Andrzej Kiepas, postulate that philosophers, and therefore ethicists, scientists from other fields, in the field of artificial intelligence research, must not so much try to provide specific solutions to individual problems, but rather proceed in indicating possible directions of search, or at least, if they can do nothing more in the current situation, to name them (Kiepas 1992). Max Tegmark suggests in the article Let's Prepare!:

"When it comes to artificial intelligence, what interests me most is not what we think about it, but what we do towards it. ... Together with the most important economists, law professors and other experts, we are trying to find answers to all the classic questions:

What will happen to people if machines gradually replace us in the job market?

When, if ever, will machines overtake humans in performing all intellectual tasks?

What happens next? Will there be an explosion of machine intelligence that leaves us far behind?

And if that happens, what role will humans play then? Will they have a role at all (Tegmark 1992)?

The sober calls from many scientific circles about the need to develop a substantive research plan on the issue of artificial intelligence are met with very popular opinions in the public sphere, views on the threats from AI. All that remains for us is to be very consistent in not succumbing to the loud arguments about the negative impact of artificial intelligence on human life. The growing impact of AI on humanity, we are to create real foundations for its positive impact on human society. At this point in our considerations, it is worth recalling the view

formulated by Roger Penrose, indicating the direction of human searches, defining the research field in relation to artificial intelligence:

"After all, if we construct conscious beings, we will thereby become responsible for them. This will clearly be our moral duty. I cannot understand why the supporters of artificial intelligence, who believe that we are all computers, are not concerned about this. After all, they are constantly striving to build a conscious machine. Shouldn't they say to themselves: 'Oh God, what if I succeed? I will not be able to remain indifferent to what this machine thinks or feels! I am responsible for it'(Penrose 2001)?

It is therefore worth noting that it is the scientist's duty to be aware of the possible consequences of his work and to communicate his conclusions and observations to a wider audience. In the rapidly developing field of artificial intelligence ethics, attention must be paid to the fact that it is already making important and practical social decisions. For example, such as the initial selection of applications for studies, for work, referral for inspection by the tax authorities, in the United States it speaks on the issue of conditional release. It is precisely in the matter of conditional release that a situation arose where criminals of African origin were discriminated against using a program based on the COMPAS algorithm, similar cases of discrimination occurred at Amazon regarding the employment of women (Coeckelbergh 2020). Thus, at the moment of our considerations, a very important question arises? Where does the responsibility lie, and at the same time the possibility of holding accountable for decisions made by artificial intelligence. Jan Andrzej Lipski suggests that if we know which programmers created the algorithmic instructions, and the authorities of a given company accepted the above instructions, then responsibility for any errors in this area can be easily enforced. However, the more difficult issue is the responsibility in the case of the above-mentioned applications of algorithms to court decisions and recruitment processes. In the recruitment and court cases mentioned above, it is not easy to determine who is responsible for the mistakes made. However, this situation shows that artificial intelligence systems are already making certain ethical decisions, which according to many AI experts, it means that it is not an object, but a moral subject. If this reasoning is correct, a very important issue becomes what values artificial intelligence will be guided by (Lipski 2020).

III. THE WORLD OF ARTIFICIAL INTELLIGENCE VALUES

In the context of equipping AI with moral values, the question arises about the basis of this equipping. It should come as no surprise that equipping AI with a specific ethical theory is an extremely complicated matter. Nick Bostrom states:

"...even if we could be reasonably certain—and we cannot be—that we have identified the correct ethical theory, we would still be in danger of making mistakes in working out the important details of that theory. There may be enormous hidden complexity underlying seemingly simple moral theories."

(Bostrom 2021)

The solution to this difficult dilemma, posed by Nick Bostrom, led to the situation in which, in the work on artificial intelligence, a concept was born regarding equipping this artificial intelligence with moral values based on the term – objective function (objective). funktion). This term means a function whose value will be maximized using the means available to the algorithm. The correct definition of this function is an extremely important matter and raises the following questions, even important ones: Should artificial intelligence be guided by the good of people (which ones?), or rather the good of the planet or the Universe? Perhaps it should take into account other abstract values? (Lipski 2023). The questions posed above therefore seem important, because the task of the well-known researcher of artificial intelligence issues, already often cited, Nick Bostrom, AI will soon surpass human intelligence in certain areas. In his opinion, one of the key issues in the discussed issue is to instill the proper world of values into artificial intelligence (Bostrom 2021). However, we also have another, as they say, difficult nut to crack. Nick Bostrom argues (and rightly so) that we do not have any ethical system that would be binding among philosophers and ethicists, and this creates serious inconveniences if equipping artificial intelligence with these values, in the short or long term, would be harmful to humanity. Solving this dilemma is not an easy matter. Unable to develop a single ethical theory for artificial intelligence, we cannot therefore open the way to writing a program for AI using uniform guidelines (Szulczewski 2019). One of the researchers of artificial intelligence proposes an interesting solution to this matter. Namely, his proposal aims to gradually teach AI such a world of values that their application by artificial intelligence would create the desired effects for the largest possible number of people, i.e. he postulates being guided by the general principle of utilitarianism (Russel 2020). Yet another concept of equipping the world of artificial intelligence with moral values is proposed by the aforementioned Mark Coeckelbergh, who claims that the European world should draw from ethical thinking based on concepts from Eastern religions. This boils down to using the concepts of Eastern religions, for example, that we treat artificial intelligence as a certain part of nature or the world and then AI is perceived by us as support for people and additionally with artificial intelligence, we should try to have various interactions, such as, for example, people often interact with animate or inanimate nature. In Far Eastern religions, the place of man in the world is less privileged than it is in monotheistic religions, especially in Christianity, so basing on the religious concepts of the Far East, according to this researcher, allows for a different shaping of the relationship between man and artificial intelligence. AI could, in his opinion, make decisions that could positively affect the well-being of the planet, the Universe and this sometimes contrary to people's expectations. Man would no longer be the measure of everything. Mark Coeckelbergh does not give a clear answer, who would have the decisive voice? Nature, or maybe artificial intelligence. There was also another proposal to introduce ethical issues into the functioning of artificial intelligence, that since it is not possible

to create unambiguous directives based on existing ethical theories, then perhaps we should reach for intermediate solutions, i.e. try to isolate such values that should guide the AI in its decision-making process. Of course, a fundamental difficulty immediately arises here, namely who will be responsible for the selection and hierarchy of establishing these values? Nick Bostrom, wanting to illustrate the difficulties in this area, formulates the following suggestions, if, for example, we assume that the primary value with which we want to equip artificial intelligence is the value of happiness. He states that then:

"...programming languages do not contain concepts such as <happiness> as primitives. If such a concept is to be used, it must be previously defined. It is not enough to define it in terms of high-level human concepts [...] The definition must go down to the lowest level in terms of the AI programming language and ultimately in primitives such as mathematical operators and addresses referring to the contents of individual memory registers (Bostrom 2021)."

The above, presented above, in principle, in a modest dimension, the above concepts of equipping the world of artificial intelligence with moral values, seem to be varieties of utilitarian ethics. In the circle of European culture, it is difficult not to reach for comments on building ethical values in artificial intelligence based on Christianity. Important in this respect is the Roman Appeal published in 2022, i.e. the declaration of the President of the Pontifical Academy of Life, representatives of the Italian Ministry of Innovation, the Director General of the FAO, representatives of Microsoft and IBM, representatives of Judaism and Islam. As part of this declaration, the above-mentioned representatives draw attention to the changes that are taking place due to the increasing use of artificial intelligence by the human world. Additionally, these changes are not only quantitative in nature, but above all they are qualitative changes that re-evaluate the ideas about reality, and especially about human nature. The declaration defines six principles on the basis of which the moral, ethical order should be built in the process of creating artificial intelligence:

- 1) Artificial intelligence systems should be clearly explained.
- 2) The needs of all people must be taken into account so that everyone can benefit from the achievements of artificial intelligence.
- 3) Designers and those who implement it must act responsibly.
- 4) You can't build AI on bias, but you can definitely protect justice and human dignity.
- 5) Artificial intelligence systems must function reliably.
- 6) Artificial intelligence must operate securely and respect user privacy.

The above declaration has received much attention in the Christian community and beyond and may become the basis for further discussion on creating the foundations for universal ethics relating to artificial intelligence. The principles of the Roman Appeal indicated above have a very general content, and the lack of references to the Holy Scriptures and the teachings of the Church is striking, which may open the way for their acceptance by people who do not refer to religious beliefs.

Nick Bostrom, who was cited, strongly emphasizes that an important task for the community of philosophers, ethicists, and computer scientists is the urgent task of influencing, and this at present, the directions of development of artificial intelligence, and not passing this task on to future generations.

IV. CONCLUSION

The above comments on the ethics of artificial intelligence have tried to touch upon the limits of the use of artificial intelligence in decision-making processes, but also the various moral dilemmas arising in this area in the mutual relations between humans and artificial intelligence. Humanity should be characterized by the ability to perceive the moral consequences of making specific decisions, and this is associated with the emergence of various types of ethical dilemmas and conflicts. At first glance, one may be convinced that the issues discussed go too far into the future and are excessive theorizing. However, problems related to moral beliefs and the creation of ethical principles for artificial intelligence should not be underestimated. We are aware of the fact that ethical argumentation is not able to completely stop scientific work in any field. It seems, however, that we, as humanity, still have time to discuss issues related to the issue of the world of moral values of artificial intelligence.

V. REFERENCES

- Amsterdamski S., Science , [in]: Concepts and problems of knowledge about culture . Encyclopedia of Polish Culture , ed. A. Kłosowska, Knowledge about Culture, Wrocław, 1991, p. 160.
- Bostrom N., Superintelligence: scenarios, strategies, threats , Helios Publishing House, Gliwice, 2021, p. 304
- Coeckelbergh M., AI Ethics , Cambridge 2020, p. 127.
- Kiepas A., Moral challenges of science and technology , Oficyna Wydawnicza Warszgraf , Warsaw-Katowice, 1992, p. 5.
- Lem S., Summa technologiae , Vol. 1, Interart Publishing House , Warsaw 1996, p. 34.
- Lem S., The future open to good and evil , [in:] Znak, 47 (1995), no. 9 (484), p. 83.
- Lipski JA, Can artificial intelligence be equipped with virtues?, [in:] Więź, summer 2023, 2[692], pp. 24-36.
- Lisi AG, I enjoy the vision of machine power , [in:] J. Brockman , (ed.), And you, what do you think about thinking machines?, PWN SA Scientific Publishing House, Warsaw 2020, pp. 39-42.
- Łupkowski Ł, The role of ethics and anthropology in considerations about artificial intelligence , [in:], Ethos R. 18, No. 1-2=69/70 (2005), pp. 240 – 251.
- Omohundro S., A turning point in artificial intelligence , [in:] J. Brockman , (ed.), And you, what do you think about thinking machines? , PWN SA Scientific Publishing House, Warsaw 2020, pp. 28-31.
- Penrose R., Whoever constructs a thinking being will have all the rights and duties of God , [in:], Knowledge and Life, 2001, no. 9, p. 23.
- Ress M., In the long term, organic intelligence has no future , [in:] J. Brockman , (ed.), And you, what do you think about thinking machines? , PWN SA Scientific Publishing House, Warsaw 2020, op. cit., pp. 25-27.

Russell SJ, *Human Compatible: Artificial Intelligence and the Problem of Control*, New York, 2020, p. 178.

Szulczewski G., *Artificial Intelligence and Moral Intelligence. Introductory Issues of Cybernetics*, [in:] *Annales . Ethics in Economic Life*, 2019, Vol. 22, No. 3, p. 20.

Tegmark M., *Let's get ready!*, [in:], J. Brockman, (ed.), *And you, what do you think about thinking machines?*, PWN SA Scientific Publishing House, Warsaw 2020, pp. 61- 63.

Torczyńska M., *Artificial intelligence and its socio-cultural implications in everyday life*, [in:] *Culture and History (Artificial intelligence – action, thinking, consciousness)*, no. 36(2019 – 2), editor-in-chief A. Radomski, UMCS, Lublin, p. 107.

Dastin J., *Amazon scrapes the secret of AI recruiting tool that he showed bias against women*, [in:], www.reuters.com/article/us-amazon-com-jobs-automation-insight-idUSKC-N1MKo8G[accessed: 12/01/2024].