THE FUTURE OF MAN IN A WORLD OF ARTIFICIAL INTELLIGENCE

Izabela MARSZAŁEK-KOTZUR

Silesian University of Technology; izabela.marszalek-kotzur@polsl.pl, ORCID: 0000-0002-8426-0170

Purpose: The aim of this article is to provide an overview of some of the problems arising from the rapid development and elevation of the status of machines and the increasing disregard for human rights in the context of the use of advanced artificial intelligence. The considerations carried out lead to the conclusion that the pace of development of modern technologies, outstripping the human capacity to comprehend them, poses the danger of machines' actions escaping human control.

Design/methodology/approach: The study is theoretical in nature and based on a critical analysis of the literature on the subject. The literature has been limited to some of the literature on artificial intelligence and related issues.

Findings: The research shows that the pace of development of modern technologies is far ahead of human competence in their control and use. There is also a lack of sufficient reflection on the dangers of surrendering decision-making to machines, for example, or placing responsibility on them.

Originality/value: The technological reality requires the development of new ethics, ethics adapted to the development of artificial intelligence, taking into account the increasingly complex ethical challenges. The article can serve as a starting point for a detailed study of the work of developing proposals for new ethics in this field.

Keywords: humans, artificial intelligence, threats.

Category of the paper: Conceptual paper.

1. Introduction

In the 21st century, it is impossible to imagine life without artificial intelligence. Medicine, communication, logistics, banking, science, automotive, military, management and marketing, journalism, translation, gaming - these are some examples of fields whose functioning is based on various AI solutions. The phenomenon is also increasingly present in the arts, where, on the one hand, it is subject to the influence and expansion of new technologies, and, on the other, it is significantly shaping the 'modern system of culture and technology' (Zawojski, 2019) in a constantly updating way. Artificial intelligence is being used, artificial

intelligence is being talked about, artificial intelligence is being monetised - arguably, this technological and information phenomenon is one of the words that characterise our current reality. the science of artificial intelligence is developing, in technical language AI is evolving. Therefore, the future of humans is inextricably linked to the development of artificial intelligence (Kuzior et al., 2019; Tomaszewska, 2023). Artificial intelligence has entered the world of man, but this statement seems to be increasingly losing its relevance, in favour of the fact that a conversion of sorts is becoming more and more evident and, in fact, man is now living in a world of artificial intelligence. Along with this awareness, there are more and more concerns and questions about man's future, his role in the world and his security. There are therefore voices warning against its uncontrolled use and the need to stop its unverifiable expansion. Already, artificial intelligence developers are proclaiming concerns about the speed at which AI learns and assimilates information, the scale of which is incomparable to human capabilities. New information keeps arriving from cyberspace, which on the one hand opens us up to hitherto unknown cognitive fields, while on the other hand creates disorientation. The reaction to the excessive number of stimuli reaching us on the affective, cognitive and decision-making planes becomes thought chaos. Human life is accompanied by constant exhaustion, both physical and psychological, and finally a whole range of pathological phenomena and behaviours emerge. They are initiated by an internal compulsion to be online, escape from the real world to the virtual world, access to toxic cultural groups, isolation, alienation, network addiction. As robots become smarter and more human-like, the threat posed by them may indeed take on a more real form (Marszałek-Kotzur, 2023). The consequences of humanity's failure to keep up with the potential of artificial intelligence related to the creation of new data and its structures are unpredictable. They relate not only to the inability of humans to understand them, but above all to the inability to control these processes. The impact of technology on human development, including the tremendous advances in artificial intelligence and cognitive technologies, inspires representatives of various fields and scientific disciplines. This problem also calls for urgent humanistic analyses, including philosophical analyses. Indeed, the development of artificial intelligence is not without influence on the formulation of forecasts concerning the future of humanity itself, and even the existence of humanity. The choice of the topic of this study is dictated by the ambivalence associated with the perception of artificial intelligence. On the one hand, we appreciate its revolutionary role in the modern world. On the other hand, one perceives various dangers arising from its operation, which puts humanity to a kind of test of time in response to the effectiveness of intelligent systems. The global growth of interest in AI also highlights the problems and challenges that humanity not only faces in the future, but already today.

2. Methodology of research

The study is theoretical in nature and based on a critical analysis of the literature on the subject. The literature has been limited to some of the literature on cognitive technologies and related issues. The literature study, i.e. the analysis and critique of the literature of selected publications, has made it possible to identify what has been analysed in the topic of interest in this paper and in what way. It also allowed the direction of further research to be outlined. The focus was on the issues of defining what artificial intelligence and algorithms are in essence. Some examples of the benefits of using these technologies were cited and some of their negative effects on humans in selected situations were described. Several plausible reasons why artificial intelligence may become disobedient to humanity are described. Only some of these are presented in this paper. Attempts to justify concerns about the future of humans in a world of machines have been reflected upon. Some exemplary suggestions for work on legal provisions related to the participation of artificial intelligence are also recalled. From the research carried out, it appears that the pace of development of modern technologies is far ahead of human competence in their control and use. There is also a lack of sufficient reflection on the dangers of handing over decision-making to machines, for example, or placing responsibility on them. The considerations presented do not, of course, exhaust the entire issue, but it seems that they may open up new fields and directions of research.

3. Results of research and discussion

How does Artificial Intelligence work?

The 21st century has brought unprecedented developments in technology and, among other things, the ability to create so-called procedures. Hence, today AI has the ability to learn, to create and remember huge data sets, to communicate between databases and even to use the cause-and-effect principle (Cymanow-Sosin et al., 2024, 59). In the context of discussions concerning the operation of artificial intelligence, we speak of so-called Machine Learning and its subgroup called Deep Learning (Kondas, 2022). Deep Learning is one of the foundations of cognitive computing (Kwilinski, Kuzior, 2020). It is a process in which a computer learns to perform tasks that are natural to the human brain, such as speech recognition, image identification, natural language processing or a recommendation system. So, computers are capable of learning without programming new skills directly to them. They use neural networks. This method is based on algorithms that, by analysing the data provided to them, draw conclusions and learn from them in order to use the knowledge thus gained in making the decisions needed to solve specific problems. Over time, as more and more data is processed,

they self-improve without having to be reprogrammed (Kuciński, 2021). There are many definitions of algorithms. It is appropriate to recall some of them. One, dating back to the 1970s, describes them as a combination of logic and control processes (Kowalski, 1979). Another, describes algorithms as procedures for transforming input data into our expected results using mathematical calculations performed by a computer (Neyland, 2018). Algorithms can also be described as a defining technology because, as such, it shapes the way we think and perceive the world (Bolter, 1990). Algorithm can also be thought of as an abbreviated name for a sociotechnical an assemblage containing: an algorithm (in the technical sense), a model, a target group, data, applications and hardware - all linked to a social environment (Gillespie, 2014). Algorithms are capable of processing thousands of complex data in real time, exceeding the capabilities of not only a single individual, but entire expert groups. Thus, they are often used, among other things, to standardise many decision-making processes (Caplan, Boyd, 2918). Algorithms are increasingly making various choices for humans, shaping their taste, flavour and preferences. In doing so, they generate visions of a perfect, error-free world, a world of 100% accuracy, efficiency and predictability (Nunes, 2011). This puts our vigilance to sleep, and just when one thinks that technology is a neutral phenomenon, it slips out of control and exerts its greatest influence on humans (Heidegger, 1977). The operation of algorithms aims to reduce complex reality to simple logical models, allowing to perform fast and complex calculations, correlating extensive data. Their increasing level of complexity makes their operation virtually impossible to grasp by human reason. It is also increasingly difficult to predict the 'behaviour' of a given algorithm in a specific situation (Kosiński, 2021). The better algorithms become at deciding how to solve human problems, the more willing humans are to hand over to them the right to decide and manage their lives. By surrendering decision-making, and thus responsibility, to algorithms, we systematically give them power over humans (Marszałek-Kotzur, 2022). Meanwhile, algorithms, perceived as objective and rational participants in life, by simplifying, fragmenting and trivialising human life, manifest their superiority over humans. Observing the work on artificial intelligence, it is already evident that, although AI does not yet have a self, it exhibits autonomy. This autonomy is a major source of concern. This is because there is no guarantee that it will not get out of control or be used for evil purposes (Hawking et al., 2014; Beck, 2010).

Three versions of the human future in an AI world

Forecasts of the future fate of artificial intelligence development are sometimes varied. One proposal is a vision of three future scenarios. On the basis of these, an attempt is made to evaluate three ways of assessing the present and future condition of humans and their humanity. The concept is arranged in a triad: Paradise - Hell - Triumph (Garreau, 2005). These scenarios constitute three metaphors, representing the consequences of technological development. The first, called Paradise, is optimistic, associated with the Christian vision of paradise. Among other things, it assumes that man will transcend the limits of his nature, making it

possible to reprogram his body and remove his body's imperfections. According to this scenario, in the coming decades, the ageing process will be radically slowed down, a significant proportion of diseases will be prevented, and treatment will be simple and uncomplicated and successful. Most people will be assisted in their daily lives by devices originally designed for people with disabilities. The global economy will be booming and most people's needs for food, shelter and security will be met. Most importantly, the boundary between the world of humans and the world of intelligent machines will become increasingly blurred. At the same time, artificial intelligence will be modelled on human intelligence and human intelligence will be assisted by machines. With transhumanism, humans will have a greater capacity for logical thinking, memory and perception. In turn, the personal character, skills and knowledge of many machines will be derived from the human mind. In the Edenic vision of the future, therefore, the understanding of humanity will change dramatically (Zawojski, 2016, p. 24). Descriptions of this technological optimism are contained, for example, in Converging Technologies for Improved Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science. This report assumes an increasing integration of man and machine and points to a whole range of resulting benefits (Albus, 2003).

The second scenario, called Hell, predicts that a high degree of technological development will pose a threat to humanity. New, intelligent means will be used by various extremists to pursue their evil intentions. The development of genetics, robotics, information technology and nanotechnology, for example, will make it possible to cause diseases or epidemics that kill many specific victims, such as people of a particular race. This would give the possibility to manipulate human evolution, while undermining the idea of equality and democracy. In such a world, robots endowed with super intelligence would reduce the lives of their creators to the level of vegetation, and thus the scenario of the so-called 'grey slime', the 'grey slime', would become real (Drexler, 1986, pp. 127, 208). It signifies the hypothetical point of an apocalyptic human scenario in which self-replicating nanomechanisms break out of control and transform the entire biosphere into copies of themselves, killing everything that lives on earth. The very name 'grey ooze' is meant to suggest an amorphous, spreading mass reproduced by machines. The essence of the infernal scenario is the loss of humanity and, consequently, the annihilation of humanity. The only realistic alternative to ensure that such a scenario does not occur is to reject overly dangerous technical solutions and abandon further research in certain areas of knowledge (Garreau, 2005, pp. 139-193). The third scenario is Triumph. It involves embracing the gains of new technologies while being aware of the risks involved. In May 2014, an article co-authored by Stephen Hawking appeared in The Independent. The text included a warning about the dangers of the rapid development of artificial intelligence (Hawking et al., 2014). According to the authors, the creation of an independently thinking machine, or superintelligence, could be the greatest event in human history. It could surpass humans in intelligence, take control of them and even create weapons that we cannot understand. The scientists also cautioned readers that considering their conundrum as a science-fiction-type statement could prove to be the biggest mistake in history.

Probable causes of AI disobedience

The literature on the extensive research into the problems associated with the development of artificial intelligence distinguishes at least several plausible reasons why artificial intelligence may become disobedient to humanity. This paper presents only some of them. One important reason is the difficulty of target selection. Currently, most machines, or software called artificial intelligence, do not reason in the same way as humans. To force artificial intelligence to perform tasks, so-called utility functions are defined (Jablecki, Wozniak, 2022). Artificial intelligence systems are not constructed according to rules that are implemented after the first run. Utility functions generally follow the value of utility. In short, AI is oriented to perform certain predefined tasks so as to optimise profits (LeViness, 2018). The situation is well illustrated by the theory of Stuart Russel, a specialist in artificial intelligence. This theory states that although utility functions are selected by humans so that the AI performs tasks to the satisfaction of humans, the utility function may not perfectly match the values of the human race, which are very difficult (or impossible) to determine (Brockman, 2014).

Another reason may be instrumental convergence assuming that an intelligent agent with unlimited but seemingly harmless goals can act in surprisingly harmful ways. A distinction is made here between so-called instrumental goals and final goals. For example, a machine whose only unlimited (final) goal is to solve a difficult mathematical problem might try to turn the entire Earth into one big computer in order to increase its computing power in order to succeed in its calculations (instrumental goal). According to the researchers, the proposed basic drives, or instrumental goals, of artificial intelligence resulting from this theory include: integrity of purpose content, self-protection, freedom from interference, self-improvement, and unmet acquisition of additional resources (Russel et al., 2003).

Another possible reason is the so-called orthogonality thesis, which states that an artificial intelligence can have any combination of intelligence level and purpose, that is, its utility functions and general intelligence can vary independently of each other. This contradicts the belief that, because of their intelligence, all AIs will pursue a common goal. This thesis was originally defined by Nick Bostrom in his article Superintelligent Will. It suggests that it would be a mistake to assume that superintelligence will share any of the values stereotypically associated with human wisdom and intellectual development. These include, for example, scientific curiosity, kindness, renunciation of material greed, humility, selflessness, etc. Bostrom also warns against anthropomorphism. Humans want to carry out their projects in a way that is consistent with human values. An artificial intelligence may have no regard for their existence or the well-being of the people around it, and instead may only care about getting the job done (Bostrom, 2012).

The above examples assume the accidental production of an AI hostile to humans. However, it seems equally likely that malicious AI could be created deliberately and intentionally. There is a fairly substantial list of entities that could be interested in creating such

AI, and these include: state governments that could use AI to control people and assert power, the criminal world, individuals who want to demonstrate that malicious AI is not so malicious, the military that could use AI as a weapon. (Pistono, Yampolskiy, 2016). Malicious AI pulling down the risk of extinction for humanity could therefore be designed by humans themselves. With this in mind, countries, or alliances, may suspect each other of working on such technology. It seems, therefore, that the main problem to be faced in the context of artificial intelligence is the problem of controlling its development once it is able to autonomously and recursively develop and improve itself (Bostrom, 2014).

Are the fears justified?

Fears of machines taking over the world are nothing new. The first references of this kind could be encountered as early as the 19th century, when the writer Samuel Butler warned that the moment when machines would take over the whole world was only a matter of time (Breuer, 1975). In later years, these discussions gained momentum. As early as the 1990s, the American philosopher, media scholar and cultural critic Neil Postman and the Canadian communication theorist Herbert Marshall McLuhan wrote about the dangers of technology's triumph over culture and perhaps even humanity. They shared a belief in technological determinism, assuming that technical progress determines the direction of development. Postman was of the opinion that the total symbiosis with various aspects of human life he called this subordination of all forms of cultural life to the reign of technology and technique He defined the current state of society, for which the goal of human labour is productivity and standard procedures, technical calculations, meters and tests are in many respects better than human judgement. He drew attention to the need to teach how to use technology, which, if unsupervised, can pose dangers to humans (Postman, 1995). In 2014, philosopher and proponent of transhumanism, Nick Bckstrom, published a controversial book entitled Superintelligence (Bostrom, 2014), to which the likes of Elon Musk and Bill Gates, for example, have not been indifferent (Metz, 2018). In turn, in 2016, the scientific journal Nature posted a warning against machines that are superior to humans in every way and can evolve beyond human control, having divergent interests with humans (Nature, 2016).

The direction of artificial intelligence depends on how it was designed, the data used to do so and the purpose of its activities. The information produced by artificial intelligence may be intentionally or unintentionally biased. The mere use of only numerical data and their statistical representations to search for solutions in a complex social reality can lead to serious consequences of humans becoming lost in a dehumanised world. The developers of artificial intelligence themselves recognise the need to develop better and more effective mechanisms for its implementation than before. In 2023, hundreds of businessmen, investors and AI experts, including Elon Musk among others, issued a letter published by the Future of Life Institute centre calling for at least a six-month pause in the development of artificial intelligence. Its authors believe that the current race led by AI labs is uncontrollable and, as a result, proper

planning and management processes are not keeping up, posing many risks to humanity (Business Insider, 2023).

Artificial intelligence today supporting human activities can lose or extend its purpose beyond human-defined boundaries. This could be due to both a lack of adequate oversight and an inability to anticipate possible pathways for AI transformation. Artificial intelligence today learns from information, generated by humans. However, it is constantly creating new facts based on these, from which it will process data, already without human participation. The consequences of this process are impossible for us to predict, but it is certain that the structures produced by AI will become increasingly complex, and that it will be impossible for humans to understand them without the help of technology (Jaskuła, 2023).

Concerns about the development of artificial intelligence also relate to human freedom under threat. After all, with the help of AI, human thinking and behaviour can be shaped. It can be programmed according to the intentions of politicians, who are not always guided by the real good of human beings. We are already subjected to mass surveillance through the Internet, smartphones, cameras, etc.

Threats also come from technologies that encourage an emotional connection with them. By definition, these are deceptive and damaging to interpersonal relationships. The development of nanorobots is both hopeful and worrying. Placed in the human body, they can fight disease, but can be used to control and even cause death. Of great concern is the use of AI in military operations in terms of its ethical aspects (Skorupka, 2022). AI has been used for eavesdropping, spying, assassination and military operations. Thus, the premise of Asimov's so-called three laws of robotics, which proclaim that robots will always be friendly to humans and will not harm humans seems to be a fiction (Asimov, 1976; Anderson, 2008). AI can be very dangerous because it is connected to a network and we are practically unable to turn off the Internet. Therefore, decisions made by AI can be completely contrary to what is good for humans (Warwick, 2020). Warnings about uncontrolled technological advances have been around for many years. However, they are not widely published and publicised. Perhaps a key reason for this is that describing the risks is not conducive to making a profit, and cutting-edge technologies have a clearly commercial application and are developed almost exclusively in corporations (Joy, 2003).

Attempts to maintain control over AI

In times of such rapid development of artificial intelligence, human subjectivity should not be lost sight of. The conscious use of technology is at the same time linked to a partial rejection of it. Technologies should first and foremost serve to establish human connections (Ball, 2022). The question of how the ongoing development of technology can be properly regulated as soon as possible is increasingly being raised. Some sound the alarm and argue that this should be done as soon as possible, while others believe that artificial intelligence is at the threshold of its development and there is no need to rush. Still others advocate only the introduction of

generally accepted standards for testing and transparency of algorithms. In 2020, The Global Partnership on Artificial Intelligence (GPAI) was founded with the overarching goal of making efforts to ensure that the development of artificial intelligence takes place in accordance with human rights and democratic values (GPAI, 2020). Australia, Canada, the European Union, France, Germany, India, Italy, Japan, the Republic of Korea, Mexico, New Zealand, Singapore, Slovenia, the US and the UK were founding members. The GPAI deals with four main topics. The first is data management. The second is responsible artificial intelligence. Another is the future of work and innovation and commercialisation. These themes have also been addressed in many international forums and conferences like the World Economic Forum, dedicated to integrated science. The need to adhere to ethical as well as legal principles in AI research is highlighted, so that these technologies can contribute to positive causes and protect society (Kuzior, Marszałek-Kotzur, 2022). Another key area related to the threat posed by artificial intelligence is the military use of the technology (Marszałek-Kotzur, 2022b). Many experts are already talking about the phenomenon of an arms race related to the introduction of artificial intelligence, where countries will compete in which of them will have access to the best artificial intelligence, as this will potentially provide a huge advantage and world domination (Geist, 2016). In this regard, international norms are being developed to prioritise security and to prioritise the development of technologies that minimise the risk of threats. The last decade has also seen several major international discussions on the approach to autonomous weapons, with the aim of developing common universal principles in this matter. Whether AI can indeed threaten humanity in the future has been the subject of much debate for many years. Articles addressing this topic have also been published. Statements by well-known visionaries such as Steven Hawking (Hawking, 2015), Elon Musk (Musk, 2017), Bill Gates (Rawlinson, 2017) are the most likely to reach the public, Yann LeCunn (Forbes, 2017), Eric Horvitz or Noam Chomski. Chomsky also warns against losing control of artificial intelligence, but according to him, artificial intelligence is merely a voracious statistical machine for speech recognition or the most likely to answer a scientific question. He therefore proposes to call AI simply "plagiarized software". The human mind, on the other hand, is, according to him, a surprisingly efficient and elegant system that works with limited information. It does not try to corrupt unedited correlations from data, but tries to create explanations (Chomsky, 2023).

4. Summary

At present, opinions on the degree of existential threat to humanity that may result from the global development of artificial intelligence are sharply divided. There are extremely optimistic and extremely pessimistic opinions. There seems to be a general consensus of opinion that there are risks associated with artificial intelligence, as with most technological advances,

but that different individuals perceive these risks differently. The specific situations evoked in the above reflections, as well as visions for the future of humans in a world filled with artificial intelligence, open up new horizons for research and the search for answers to the question: what is next for humans? Artificial intelligence is making its mark on adults, young people and children. For the generation of people born just after the war, artificial intelligence seems more like an abstract construct. Later generations view artificial intelligence as a natural consequence of developments based on algorithms and zero-sum code, which underpin the functioning of computer programmes used in work and everyday life by their users. Perhaps the fears of artificial intelligence are exaggerated and independent artificial intelligence will never pose a threat to humans. Perhaps it will even allow humanity to enter a whole new level of civilisation? It is therefore worthwhile to continue research into artificial intelligence systems capable of self-adaptation and autonomous evolution, but the specific scope of the legal and ethical framework must be established. Contemporary artificial intelligence development processes should not only be subject to an analysis of current problems, they should more broadly attempt to identify risks that may escalate in the future. The purpose of this article is to encourage the reader to reflect on the problem raised and perhaps raise new questions on the issue.

References

1. Albus, J.S. (2003). Engineering of Mind for Enhancing Human Productivity. In: M.C. Roco, W.S. Bainbridgec (Eds.), *Converging Technologies for Improved Human Performance: Nanotechnology, Biotechnology, Information Technology and Cognitive Science* (pp. 281-293). New York: Kluwer Academic Publishers.
2. Anderson, S.L. (2008). Asimov's "Three Laws of Robotics" and Machine Metaethics. *AI & Society, vol. 22, No. 4*, pp. 477-493.
3. Asimov, I. (1976). *Bicentennial Man.* New York: Doubleday & Co.
4. Ball, M. (2022). *Metawersum. Jak Internet przyszłości zrewolucjonizuje świat i biznes.* Warszawa: MT Biznes.
5. Beck, S. (2010). Roboter, Cyborgs und das Recht. Von der Fiktion zur Realität. In: T.M. Spranger (Ed.), *Aktuelle Herausforderungen der Life Science* (pp. 95-120). Berlin/Münster: LIT Verlag.
6. Bolter, J.D. (1990). *Człowiek Turinga. Kultura Zachodu w wieku komputera*. Warszawa: PIW.
7. Bostrom, N. (2012). Superintelligent Will: motivation and instrumental rationality in advanced artificial agents. *Minds and Machines, Vol. 22, Iss. 2*, pp. 1-16.

8. Bostrom, N. (2014). *Superintelligence: Paths, Dangers, Strategies*. Oxford University Press.

9. Breuer, H.-P. (1975). Samuel Butler's "the Book of the Machines" and the Argument from Design. *Modern Philology, Vol. 72, No. 4*, pp. 365-383.

10. Brockman, J. (2014). *The Myth Of AI. A Conversation With Jaron Lanier*. Retrieved from: https://www.edge.org/conversation/the-myth-of-ai#26015, 24.05.2024.

11. Business Insider (2023). Retrieved from: https://businessinsider.com.pl/technologie/nowe-technologie/elon-musk-i-setki-biznesmenow-wezwali-do-przerwy-w-rozwoju-sztucznej-inteligencji/me2v3sn, 21.05.2024.

12. Caplan, R., Boyd, D. (2018). Isomorphism through algorithms: Institutional dependencies in the case of Facebook. *Big Data & Society, vol. 1-12*.

13. Chomsky, N. (2023). *Noam Chomsky dla „New York Times"*. Retrieved from: https://agencja-informacyjna.com/noam-chomsky-chatgpt/, 21.05.2024.

14. Cymanow-Sosin, K., Tenerowicz, K., Cymanow, P. (2024). *Media przyszłości – wybrane aspekty komunikacyjne, ekonomiczne i edukacyjne*. Kraków: Homini.

15. Drexler, E. (1986). *Engines of Creation: The Coming Era of Nanotechnology*. Doubleday: Anchor Books.

16. *Elon Musk on Potential Risks of Artificial Intelligence*. Retrieved from: https://www.youtube.com/watch?v=bzAXifQIUZ8&ab_channel=WPRI, 26.06.2024.

17. Forbes (2017). *Could Artificial Intelligence Ever Become A Threat To Humanity?* Retrieved from: https://www.forbes.com/sites/quora/2017/02/09/could-artificial-intelligence-ever-become-a-threat-to-humanity/?sh=4b732d502142, 26.06.2024.

18. Garreau, J. (2005). *Radykalna ewolucja. Czy człowiek udoskonalony przez naukę i technikę będzie jeszcze człowiekiem?* Poznań: Prószyński i S-ka.

19. Geist, E.M. (2016). It's already too late to stop the AI arms race — We must manage it instead. *Bulletin of the Atomic Scientists, 72(5),* pp. 318-321.

20. Gillespie, T. (2914). The Relevance of Algorithms. In: P. Gillespie, J. Boczkowski, K. Foot (Eds.), *Media Technologies: Essays on Communication, Materiality, and Society* (pp. 167-193). Cambridge, MA: MIT Press.

21. GPAI. Retrieved from: https://gpai.ai/, 26.06.2024.

22. Hawking, H., Russell, S., Tegmark, M., Wilczek, F. (2014). *Stephen Hawking: 'Transcendence looks at the implications of artificial intelligence - but are we taking AI seriously enough?* Retrieved from: http://www.independent.co.uk/news/science/stephen-hawking-transcendence-looks-at-the-implications-of-artificial-intelligence-but-are-we-taking-9313474.html, 22.02.2021.

23. Hawking, S. (2015). *AI could spell end of the human race* Retrieved from: https://youtu.be/fFLVyWBDTfo?t=245, 26.06.2024.

24. Heidegger, M. (1977). *Budować, mieszkać, myśleć*. Warszawa: Aletheia.

25. Jabłecki, P., Woźniak, K. (2022). *Nieposłuszeństwo sztucznej inteligencji – problem ryzyka egzystencjonalnego.* Retrieved from: http://architeles.eu/ethics/index.php/2022/06/02/nieposluszenstwo-sztucznej-inteligencji-problem-ryzyka-egzystencjona 24.06.2024.

26. Jaskuła, S. (2023). Sztuczna inteligencja w edukacji. *Perspektywy Kultury, 3(42),* pp. 13-26. doi: 10.35765/pk.2023.4203.04

27. Joy, B. Dlaczego przyszłość nas nie potrzebuje? In: G. Yeffeth (Ed.), *Wybierz czerwoną pigułkę: Nauka, filozofia i religia w Matrix* (pp. 229-245)*.* Gliwice: Helion.

28. Kondas, A. (2016). *Wstęp do Machine Learning.* Retrieved from: http://itcraftsman.pl/wstep-do-machine-learning/, 5.03.2024.

29. Kosiński, M. (2021). Jak daleko sięga władza algorytmów. In: J. Szomburg, J.M. Szomburg, A. Leśniewicz, M. Wandałowski (Ed.), *Człowiek vs. algorytmy i sztuczna inteligencja – kto kogo zaprogramuje, 4(7)* (pp. 12-21). Pomorski Thinkletter. Retrieved from: www.kongresobywatelski.pl, 22.05.2024.

30. Kowalski, R. (1979). Algorithm = Logic + Control. *Communications of the ACM*, *vol. 22(7),* pp. 424-436.

31. Kuciński, M. (2021). Sztuczna inteligencja a godność człowieka. *Fides, Ratio et Patria. Studia Toruńskie, No. 14*, pp. 210-224.

32. Kuzior, A., Kwilinski, A., Tkachenko, V. (2019). Sustainable development of organizations based on the combinatorial model of artificial intelligence. *Entrepreneurship and Sustainability Issues, 7(2).*

33. Kuzior, A., Marszałek-Kotzur, I. (2022). Ethical problems of the development of artificial intelligence. In: P. Kasprowski (ed.), *Artificial intelligence and data processing. The monograph presenting the achievements of the Silesian University of Technology research staff, T. 954* (pp. 527-537).

34. Kwilinski, A., Kuzior, A. (2020). Cognitive Technologies in the Management and Formation of Directions of The Priority Development of Industrial Enterprises. *Management Systems in Production Engineering, Vol. 28, Iss. 2*, pp. 133-138, 2020.10.2478/mspe-2020-0020

35. LeViness, I. (2018). *What is a Utility Function?* Retrieved from: http://www.ifthen.ai/2018/05/01/what-is-a-utility-function/211/, 24.05.2024.

36. Marszałek-Kotzur, I. (2022a). Cognitive Technologies – Are We in Danger of Humanizing Machines and Dehumanizing Humans? *Management Systems in Production Engineering, Vol. 30, Iss. 3*, pp. 269-275, https://doi.org/10.2478/mspe-2022-0034

37. Marszałek-Kotzur, I. (2022b). Ethical perspective of the development and use of modern military technologies. *Zeszyty Naukowe Politechniki Śląskiej. Organizacja i Zarządzanie*, 183-198. https://doi.org/10.29119/1641-3466.2022.165.13

38. Marszałek-Kotzur, I. (2023). The ethics of artificial intelligence. Can we trust robots? *Zeszyty Naukowe Politechniki Śląskiej. Organizacja i Zarządzanie*, 353-368. https://doi.org/10.29119/1641-3466.2023.183.21

39. Metz, C. (2018). *Mark Zuckerberg, Elon Musk and the Feud Over Killer Robots*. Retrived from: https://www.nytimes.com/2018/06/09/technology/elon-musk-mark-zuckerberg-artificial-intelligence.htm, 24.05.2024.

40. Nature. *Anticipating artificial intelligence. Nature, 532, 413, 28 April 2016*, doi:10.1038/532413a, Retrieved from: https://www.nature.com/articles/532413a, 22.06.2024.

41. Neyland, D. (2018). *The Everyday Life of an Algorithm*. Cham: Palgrave Pivot.

42. Neyland, D., Möllers, N. (2017). Algorithmic IF... THEN Rules and the Conditions and Consequences of Power. *Information, Communication & Society*, *20(1),* pp. 45-62.

43. Nunes M. (2011). Error, Noise, and Potential: The Outside of Purpose". In: M. Nunes (Ed.), *Error: Glitch, Noise, and Jam in New Media Cultures* (pp. 3-23). New Haven, CT / London: Continuum.

44. Pistono, F., Yampolskiy, R.V. (2016). Unethical Research: How to Create a Malevolent Artificial Intelligence. *Ethics for Artificial Intelligence Workshop (AI-Ethics-2016),* pp. 1-7, doi: 10.48550/ARXIV.1605.02817

45. Postmann, N. (1995). *Technopol: Triumf techniki nad kulturą*. Warszawa: PIW.

46. Rawlinson, K. (2017). *Microsoft's Bill Gates insists AI is a threat*. Retrieved from: https://www.bbc.com/news/31047780, 26.06.2024.

47. Russell, S.J., Norvig, P. (2003). Section 26.3: The Ethics and Risks of Developing Artificial Intelligence. *Artificial Intelligence: A Modern Approach.* Upper Saddle River, N.J.: Prentice Hall.

48. Skorupka, A. (2022). Ogólna sztuczna inteligencja z filozoficznego punktu widzenia. *Transformacje, 1(112)*, pp. 130-142.

49. *The New York Times.* Retrieved from: https://www.nytimes.com/2018/06/09/technology/elon-musk-mark-zuckerberg-artificial-intelligence.html, 26.06.2024.

50. Tomaszewska, R. (2023). Piekło – Triumf. Sztuczna inteligencja w scenariuszach przyszłości. *Ethos, Vol. 36, No. 3(143),* pp. 253-274, doi: 10.12887/36-2023-3-143-14

51. Warwick, K. (2020). Inteligentne maszyny są zagrożeniem dla ludzkości. In: R.V. Yampolskiy (Ed.), *Sztuczna Inteligencja. Bezpieczeństwo i zabezpieczenia* (pp. 523-532). Warszawa.

52. Zawojski, P. (2016). *Technokultura i jej manifestacje artystyczne. Medialny świat hybryd i hybrydyzacji*. Katowice: Wydawnictwo Uniwersytetu Śląskiego.

53. Zawojski, P. (2019). Maszynom (inteligentnym) wbrew? O sztuce w czasach sztucznej inteligencji. *Kultura Współczesna, 1(104)*, pp. 53-66, doi: 10.26112/KW.2019.104.05