SILESIAN UNIVERSITY OF TECHNOLOGY PUBLISHING HOUSE

SCIENTIFIC PAPERS OF SILESIAN UNIVERSITY OF TECHNOLOGY 2023
ORGANIZATION AND MANAGEMENT SERIES NO. 183

THE ETHICS OF ARTIFICIAL INTELLIGENCE. CAN WE TRUST ROBOTS?

Izabela MARSZAŁEK-KOTZUR

Department of Applied Social Sciences, Faculty of Organization and Management, Silesian University of Technology; izabela.marszalek-kotzur@polsl.pl, ORCID: 0000-0002-8426-0170

Purpose: The purpose of the article is to identify the state of knowledge AND discuss the ethical issues of challenges associated with the development of artificial intelligence and robots. The research questions posed are: How can artificial intelligence be ethical? Is it possible for humans to trust machines?
Design/methodology/approach: The article analyzes selected literature on the problem of trust, security and ethics regarding the development of artificial intelligence and robots.
Findings: The analysis conducted showed that the ethical perspective on the development and application of artificial intelligence in robotics is not sufficient and represents a kind of research gap.
Social implications: Societies are not aware of succumbing to the influence of artificial intelligence and do not demonstrate sufficient knowledge of security from intelligent machines. The trust placed in robots has no scientific basis.
Originality/value: To point out the research gap on the ethical perspective of the application of modern technologies, especially artificial intelligence and robotics in global and individual application in the context of trust, security from modern technologies and ethics in the cooperation of humans and machines. The article also touches on the debatable issue of robot ethics, somewhat overlooked in scientific discourse.

Keywords: robot, human, trust, security, ethics.

Category of the paper: Conceptual paper, Literature review.

1. Introduction

The technologies of the 21st century have been dominated by the development of artificial intelligence. Work on artificial intelligence is consuming huge amounts of both financial and personnel resources, thus indicating the expectations in this area, including in the process of creating intelligent robots. In addition to the fascination with the positive aspects of AI development, there are also serious concerns about the negative consequences of their use in various areas of human life. These concern the economy, economics, democracy, medicine,

energy and also other areas crucial to the proper functioning of states, societies, groups and individuals (Kuzior, 2020; Kwilinski, Kuzior, 2020; Ober, 2022; Kuzior, Kochmańska, Marszałek-Kotzur, 2021). Increasingly important in this context seems to be the problem of guaranteeing the safety of humans from the side of artificial intelligence, and consequently, a clear definition of the issue of ethical dilemmas and trust in robots. Questions about security also do not evade the issue of human domination by artificial intelligence. For some time now, so-called weak artificial intelligence has been permanently accompanying humans, relying on computer system operations that mimic intelligent behavior to a limited extent. The time of so-called strong artificial intelligence, which could perform most or all of the intellectual tasks vested in humans, is likely to be possible in the near future (Siau, Wang, 2018, p. 53). Computerized devices primarily save human labor in terms of physical work. Artificial intelligence, on the other hand, is a system that allows it to perform tasks that require a process of learning and taking into account new circumstances in the course of solving a given problem, and which can, to varying degrees - depending on its configuration - operate autonomously and interact with the environment (Zalewski, 2020). Artificial intelligence and robots can assist the human mind with cognitive capabilities and the ability to learn, choose, select, in short, decide. The dynamic decision-making algorithm that controls the behavior of robots allows them to take into account past experiences. Learning from their mistakes, intelligent robots are able to modify their future actions (Kiepas, 2022; Maternowska, 2022). As robots increase their skills, their "thinking" becomes more unpredictable and beyond human computational capabilities (Michalski, 2018). Due to the lack of transparency in the decisions made by robots, their operators face increasing difficulties in interpreting and overruling the robot's decisions (Lemley, Casey, 2019). The high complexity of the decision-making algorithm and the dynamic adaptation of programming to unforeseen circumstances make robots different from other machines, which, according to many researchers, is an apex for special treatment by the law. Thus, arguments are emerging for the need to rethink legal measures applied to robots (De Chiara, Elizalde, Manna, Segura-Moreiras, 2021; Lemley, Casey, 2019; Shavell, 2019; Talley, 2019). As artificial intelligence develops, improves and reaches higher levels, concerns are being raised that it may begin to manage human systems according to the values of artificial intelligence, not humans. These issues often concern the long-term security of intelligent systems, which is important not only for individual humans, but also for the human species and life on Earth as a whole. These issues and many others are central to the field of ethics. While the above problem of machines taking control of humans may seem distant in time, or completely impossible, there is the question of creating an ever closer human-machine relationship (EU, 2018, p. 7). This raises important and difficult moral questions. One should ask about the safety of humans who live alongside intelligent, autonomous robotic machines. It is also necessary to consider the moral responsibility of humans for the unforeseen consequences of using artificial intelligence. The issue of exercising control and safe management of autonomous systems also requires rethinking. Thorough reflection is also

required on the issue of establishing regulations that would serve to build human well-being on the one hand, and ensure their safety on the other? The trust placed in machines is now so far-reaching that humans are allowing themselves to take away the freedom to determine for themselves. But what does this trust consist of? Man allows artificial intelligence to dictate to himself, as an individual, what we should buy, what road to take when traveling, which flight to take, what food to choose, what movie to watch over the weekend, what credit to choose. On the structural side, meanwhile, artificial intelligence influences stock market transactions, commodity prices, mass transportation, the industrial Internet, legal decisions and even political elections (Ike, 2018). It seems, then, that we have completely put our trust in artificial intelligence and robots. Perhaps in doing so, we have already stripped ourselves of our power over the human world? Globally, we have been put before a kind of revolution and evolution at the same time. Artificial intelligence, together with autonomous robots, arouses human admiration, but also generates a whole range of concerns that have various origins.

2. Confidence placed in machines

2.1. The basics of trusting robots

Trust is considered one of the basic psychological factors that guarantee the cohesiveness of social structures. It integrates and binds human communities together. The act of trust is dynamic, that is, activist in nature, and involves the action of other entities and systems, which are the objects of our trust. This is because, as a social attitude, it is related to a specific form of human activity and is oriented towards the future (Sztompka, 1999; Sztompka 2007). The issue of trust has been the subject of human deliberation since antiquity. Plato, for example, described it as trust, that is, in terms of an act of faith (Plato, 1997). When analyzing the meaning and nature of the attitude of trust, its moral dimension cannot be overlooked. It is believed that one of the conditions for trust, considered as the aforementioned type of social attitude, is a factor of a moral nature, i.e. the ability of individuals or certain groups to function within a unified axiological system, i.e. in a collective that adheres to common moral values (Fukuyama, 1997). The attitude of trust is seen as one of the essential elements of social relations under the conditions of late modernity. A characteristic feature of this period is the widely understood unpredictability and associated risks. This is the so-called manufactured risk, which is distinguished from natural risk. Manufactured risk, which is one of the side effects of the development of civilization (Giddens, 2001), plays an important role in the context of the development of information technology and is understood as trust in abstract systems (Giddens, 1999). This issue is included in the category of expert knowledge, because under this concept are complex technical procedures that are incomprehensible to most of the public, ensuring the

possibility of the functioning of the technologies themselves and other abstract systems. Trust, understood in this way, is related to the reliability of the individuals who design, construct and control the operation of systems. It is considered here in two dimensions: personal and impersonal. Personal trust is based on honesty, that is, it is a trait grounded in the moral beliefs of the individual. In contrast, trust in a system, or a particular technology, is instrumental. It is based on the belief that a given technology will perform according to our expectations. Its most common form is the demand for reliability and stability. The latest, widely understood information technologies and products require an increasing amount of trust from their users, both in their personal and impersonal dimensions (Szynkiewicz, 2014). Given the fact of the crucial importance of the development of artificial intelligence for the functioning of modern societies, it is worth asking how we should understand the concept of trust in this perspective? Recognizing the importance of artificial intelligence systems and the risks associated with them, the European Commission published on April 8, 2019 a document developed by an independent group of experts called Ethics Guidelines for Trustworthy AI (EC, 2021). The guidelines contained therein aim to promote trustworthy artificial intelligence defined by three components that should be met throughout the lifecycle of an AI system. The first, is compliance with the law and all applicable laws and regulations. The second is ethicality, or adherence to ethical principles and values. The third is robustness, both technically and socially, given that even with good intentions, AI systems can cause unintended harm. All three components should occur simultaneously (Turner, 2019; Gasparski, 2019). In addition, the foundations of trustworthy AI must include four key elements: respect for human autonomy, prevention of harm, fairness, and explainability. All four are based on fundamental human rights, such as respect for human dignity, individual freedom, respect for democracy, justice and the rule of law, equality, non-discrimination, solidarity, and civil rights (PDPC, 1948).

2.2. Trust in human-robot interaction

The issue of trust in artificial intelligence plays a special role in the robotization development space. Robotization is becoming an increasingly common process, being one of the key elements of Industry 4.0. There is no turning back from this process anymore. Thus, a paradigm shift is taking place in terms of cooperation between humans and machines. They are ceasing to act as mere tools in human hands, and are beginning to play the role of collaborators. The goal of robotization in the context of a factory is to facilitate the performance of tasks. It is worth recalling at this point one definition of a robot, which indicates that it is a tool whose configuration of sensors, actuators and integrated control system provides a significant level of flexible, independent and autonomous operation. To some extent, the robot has a physical installation, which makes it necessary for the robot and a human to coordinate their actions here and now (Weiss et al., 2021, p. 2). Robotization poses new challenges for today's users, such as the need to implement organizational changes, or to change human awareness, behavior and attitudes, including leveling resistance to change. Robots can support

humans in many tasks, but this requires the trust and cooperation of human operators described above. The literature on robot deployments focuses mainly on technical issues and success factors related to human-robot interaction, ignoring the role of human factors (Lambrechts, 2021, p. 24). Trust in the machines appears here as a key factor. The role of robots driven by artificial intelligence, is changing from being tools for humans, to teammates. Humans and robots are increasingly taking on complex and collaborative roles, both in the manufacturing and service sectors. This collaboration is particularly evident in the military, construction, agriculture, medicine, social care, analytical services and manufacturing (Arslan et al., 2021, p. 79). To subject human-robot interaction to research, the problems of their contacts are defined in terms of autonomy, information exchange, teams, task shaping, learning, training, and the classification of types of interaction as remote or close, i.e. existence in collocation, is applied (Tunc, 2020, p. 30). However, a distinction is made between interaction and collaboration. For this purpose, the terms human-robot interaction (HRI) and human-robot collaboration (HRC) are used. Interaction means interacting with someone else, which in the context of work is understood as involving someone else: either a human or a robot. Collaboration, on the other hand, means working with someone toward a common goal. Consequently, cooperation in the context of humans and machines should be understood as a special case of interaction (Castro et al., 2021, p. 6). The term "cooperation" also appears in the literature. It is described as a sequence of actions aimed at a common goal, while cooperation is a sequence of joint actions toward a common goal (Kolbeinsson et al., 2019, p. 453). For example, military robots in combat conditions have to make decisions that could potentially result in the death of civilians, which raises many moral disputes (Marszalek-Kotzur, 2022b). After all, machines have ceased to be mere tools of war and have become collaborators with soldiers (Kamienski, 2022, p. 18). Trust is a dynamic process that changes and fades over time, sometimes rapidly declining as a result of changing interactions between team members (Huang et al., 2020, p. 310). Trust also establishes behavioral expectations that facilitate joint action. It becomes even more important in high-risk situations. Robots are often designed to replace humans in dangerous situations. This is when human trust in machines will be particularly important. If robots are to succeed as team members, then in hazardous situations people need to believe that other team members are capable of protecting their interests (Groom, Nass, 2007). The level of trust is influenced by the robots' sense of control. It is correlated with the predictability of machine behavior. The sense of being in control of the situation and having control over the robot are the factors that determine trust building. It is also essential to have knowledge of the machine's capabilities. Transparency of the system on which the robot operates is often mentioned in definitions of trust as a desirable and even necessary component (Simon et al., 2020, p. 18).

2.3. Valley of the uncanny

One of the components of the development of artificial intelligence is to give and assign human characteristics to robots, such as appearance, behavior, thinking. It is of great importance for human beings to mirror their behavior, thus imitating their body movements, facial expressions, gestures. The interaction then becomes smoother, the feeling of liking intensifies and the level of trust grows (Fortune, 2021, p. 102). Three factors play an important role: appearance, performance and proximity. The appearance of robots, specifically their anthropomorphism, is a critical dimension affecting interactions with humans. For humans, it is important that the robot has a shape similar to the human body. A moderate degree of robot resemblance to a human has been proven to affect a significant sense of safety. However, when a machine becomes too similar to a human, it begins to induce a sense of strangeness in humans and even causes terror (Simon et al., 2020, p. 10). Humanoid robots evoke different reactions in humans. "Face-to-face" relationships with robots in the near future will occur within a wide variety of activities, including professional, caregiving, medical, or social activities. The feeling of safety in the situation of being in the company of a robot and liking it may depend on a psychological phenomenon that is referred to as the uncanny valley (Masahro, 2012; Kruszewski, 2019). The name of this phenomenon was introduced in 1970 by Japanese robotics engineer Masahiro Mori (Mori, 2012, p. 98). According to Masahiro's hypothesis, robots arouse in people the more positive feelings, the closer their appearance is to a human. Yes, for example, we like a doll-shaped toy robot more than a mechanical arm turning bolts in a factory. However, once a certain limit of realism is crossed, the robot's strong resemblance to a real human begins to give rise to an unpleasant feeling of unease instead of excitement. In short, the intensifying stimulus of the robot's increasing resemblance to a human produces an increasing positive emotional response to the stimulus in the individual, but when the resemblance approaches perfection, the individual's emotional response drops sharply and becomes strongly negative (Burleigh et al., 2013, p. 1). This decline is what is known as the valley of awesomeness. The idea is that a robot's purpose can affect the limit of tolerance towards the robot's appearance. The situation is like in a horror movie. You can describe the strangest monsters and monstrosities, but if they are to be scary, you need to show a creature that moves and talks like a human being, seemingly indistinguishable from people you see every day on the street. Only, for example, his pupils or facial expressions are a little different from humans. This is how the valley of the uncanny effect works. Some differences and unnaturalness do not bother in characters that only resemble humans, but when the resemblance is too great, a small minor detail causes discomfort and concern. Thus, making a robot too much like a human can cause a loss of a sense of security. This sense of insecurity can stem from the fear of being replaced by the robot (Schindler et al., 2017). Thus, there is a thesis that robots that imitate humans too faithfully are met with a negative reception. Masahiro, for the purpose of illustrating his thesis, created a graph, according to which, as the resemblance to humans

increases, we observe increasing comfort, until we encounter a borderline model, for which the comfort level decreases sharply. It is worth noting that this phenomenon is considered both in the field of robotics, human-robot interaction, or social robotics, but also among game designers and computer animation authors. It is also interesting from a purely cognitive perspective, as its study allows us to better understand how we categorize objects that are the object of our perception (Kätsyri et al., 2015).

3. Human safety vs. robot liability

Despite the trust placed in robots, the operation of autonomous systems based on artificial intelligence can produce effects that are not always desirable. It can be subject to error and can cause a variety of harm. There are, for example, accidents caused by autonomous vehicles. There can be damages associated with medical misdiagnoses, losses caused by a decrease in the efficiency of production systems, financial losses caused by systems supported by artificial intelligence, losses resulting from investment errors of programs that support stock market investors, or damages caused by systems used for management in the energy industry. This raises the issue of liability. As the level of autonomy of robots increases, it will become increasingly difficult to assign responsibility for accidents caused by the robots' actions to a specific party. After all, the rules by which robots operate, through machine learning, can evolve as they perform specific tasks. The traditional ways of assigning responsibility in such a case are not in line with the common sense of justice and moral norms of society. It seems that today no one has enough control over the actions of a machine to be able to take responsibility for it. So the idea of assigning robots a legal personality was born. It has been considered in both Europe and the US. February 16, 2017. The European Parliament adopted a resolution containing a recommendation to the European Commission on civil law provisions on robotics. One of the reasons aiming to draft it became the belief that humanity has just stood at the threshold of an era in which increasingly advanced robots, androids, computers and other incarnations of artificial intelligence are giving rise to a new industrial revolution. It is likely to affect every stratum of society. It is therefore necessary to develop regulations that should take into account the legal and ethical implications, as well as the effects of these changes, but without simultaneously stifling innovation. Thus, the need for new, effective and up-to-date regulations was emphasized, which would consist, among other things, of creating universally acceptable and flexible definitions of the terms "robot" and "artificial intelligence", of updating and supplementing the current EU legal framework with key ethical principles that reflect the complexity of robotics and its many social, medical and bioethical implications. In addition, issues related to civil liability for damages caused by a robot should be analyzed to ensure that citizens, consumers and businesses alike receive the same level of efficiency, transparency and

consistency while ensuring legal certainty throughout the European Union. The resolution also raises the issue of diagnosing, analyzing and considering the legal consequences of giving robots a separate legal status in the long term. The idea is to give the most developed autonomous robots "the status of electronic persons responsible for repairing any damage they might cause, and possibly applying electronic personhood when robots make autonomous decisions or their independent interactions with third parties" (RPE, 2017). Thus, we are dealing with the creation of a new legal category with specific features and implications. Taking a clear position on the necessity, as well as the rightness of giving robots the status of electronic persons with electronic personality seems to require deep and extensive analysis (Biczysko-Pudełko, Szostek, 2019). At this point, the Civil Code distinguishes between natural persons and legal persons. These categories refer to people. How, then, to treat robots in a situation where it is increasingly difficult to consider machines, which are in many respects more accurate, faster and learning behavior, as mere objects. However, the creation of a new category such as the electronic person is, one might say, a gigantic civilization change. It also raises a key question: can the development of artificial intelligence be placed within the framework of the "law of robots"? Or could it be that legal regulations will never keep up with the development of technology? It seems that the most important thing will be for possible disputes to be settled, on the basis of general clauses, by wise and reasonable people (Nightingale, 2018). That is why the May 4, 2020 draft regulation of the European Parliament on liability for the operation of artificial intelligence systems emphasizes that the entity responsible for the damage can be either a human being or a legal entity.

4. Is there a robot ethic?

The unpredictability of the formation of artificial intelligence, due to its cognitive capabilities, dictates that various forms of ensuring the safety of use should be sought. One of the more effective ones seems to be conducting research in the area of AI ethics and seeking to enforce specific ethical arrangements on legal grounds. Multifaceted political, scientific and social engagement plays a special role here. Ethics must first and foremost respond to the challenges posed by real technological possibilities. It must not be seen as a futuristic goal, but as a solid scientific field with real contributions to the creation of intelligent machines (Lipińska, 2022). A number of concepts are emerging for developing ethics regarding artificial intelligence. One of them postulates the transformation of ethics into so-called "microethics", postulating that at certain points there would have to be a fundamental change in the space of ethics. There would have to be a transformation of ethics to ethics of technology, ethics of machines, ethics of computers, ethics of information and ethics of data (Hagendorf, 2020). There are a number of discussions on trying to address the ethics of artificial intelligence

(Kuzior, Marszałek-Kotzur, 2022). On February 13, 2018, the first conference on Artificial Intelligence, Ethics, and Society (1st AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society) was held in New Orleans, United States. Its organizers were two associations: Association for the Advancement of Artificial Intelligence (AAAI) and Association for Computing Machinery (ACM). More than 300 people attended the conference. It was stated then that the ethics of artificial intelligence is part of the ethics of technology. It is divided into roboethics, which is concerned with the moral behavior of humans who design, construct and use artificially intelligent beings, and machine ethics, which is concerned with the moral behavior of artificial moral agents (Kuipers, Mattei, 2018). At the time, the US newsletter Ethikos Weekly published a text pointing out the inappropriateness of the term "AI ethics", which creates terminological confusion. It could imply that moral causality is attributed to artificial intelligence, rather than to the humans who develop its systems (Gasparski, 2019; EW, 2018). Discussions related to ethical challenges arising with the development of artificial intelligence, including the development of robots, seem to be gaining momentum. The literature offers various proposals for guidance to help evaluate the ethical dimensions of artificial intelligence. For example, it is recommended to establish a solid foundation, to use ethical principles that are focused on human welfare, to promote a holistic approach to social, economic, scientific, legal, engineering and technological issues and ethical issues in artificial intelligence research. Another is the issue of the presence of ethicists and ethics from the very beginning of the technology chain, developing and enforcing rules to ensure respect for human dignity and human rights. It is also crucial to limit military applications of AI and promote world peace through a global ethical governance framework. Also not to be overlooked is guaranteeing social equality in AI, i.e. respecting gender equality and refraining from prejudice (Gasparski, 2019). On the website of the UK's Institute of Business Ethics, as of January 2018, there is an article titled Business Ethics and Artificial Intelligence, posted with the intention of pointing out the essence of artificial intelligence (BEB, 2018; Gasparski, 2019) The bulletin identifies ten key issues organized by the following letters of the word ARTIFICIAL: Accuracy, Respect of privacy, Transparency, Interpretability, Fairness, Integrity, Control, Impact, Accountability, Learning, i.e., Accuracy, Respect of privacy, Transparency, Understandability, Fairness, Integrity, Control, Impact, Accountability, Learning. Each of these issues has been defined and dedicated to business organizations for reflection, along with the suggestion to undertake a comprehensive and multi-pronged debate on the value and impact of artificial intelligence on the business environment, with as many stakeholders as possible. Business decision-makers, employees, customers and the public should be aware of the impact of AI on their business and on their stakeholders, taking into account not only the benefits, but especially any side effects (Gasparski, 2019). EU bodies have set themselves the goal of building an AI ecosystem based on a community of European values, among which human dignity is central. Accordingly, the European approach to AI is called human-centered AI. The Union wants to become a leader in an ethical approach to technology and wants to

distinguish itself with this on the world stage. To this end, as can be read in the document, "Europe needs to define a normative vision for the AI-led future it wants to realize" (EU, 2019, p. 11.) In order to develop expert opinions on new technologies, the European Commission has set up three expert groups: the High-Level Independent Expert Group on the Impact of Digital Transformation on EU Labor Markets, the Expert Group on Accountability and New Technologies and the High-Level Independent Expert Group on Artificial Intelligence. The latter addressed ethical issues of artificial intelligence development. The Independent High-Level Expert Group on Artificial Intelligence (AI HLEG), composed of 52 experts, was established in June 2018 and worked until July 2020. During this time, four significant documents were produced: Ethics Guidelines for Trustworthy AI (April 10, 2019), Policy and Investment Recommendations for Trustworthy AI (June 26, 2019), *Assessment List for Trustworthy AI* (July 17, 2020) and Sectoral Considerations on the Policy and Investment Recommendations (July 23, 2020). These served as the starting point for initiatives by the European Commission and member states (Lipinska, 2022). In April 2021, the Artificial Intelligence Act was released: the first legislation regulating artificial intelligence. In it, the European Commission proposed the first EU legislative framework for artificial intelligence. The draft analyzes and classifies AI systems that can be used in various applications. The classification is based on an assessment of the risks it poses to its users. Technologies endowed with artificial intelligence that pose an unacceptable risk, i.e. considered a threat to humans, will be banned. These include, but are not limited to: cognitive-behavioral manipulation of people or certain vulnerable groups, such as voice-activated toys that encourage children to behave dangerously; citizen scoring, which is the classification of people based on their behavior, socioeconomic status or personal characteristics; real-time and remote biometric identification systems, such as facial recognition. Some exceptions will be allowed, such as remote biometric identification systems that identify with a significant delay and are used to prosecute serious crimes, but only after court approval (EU, 2021). In this regard, the EU document postulates a correct understanding of three fundamental values, such as human dignity, autonomy and moral responsibility. Autonomy in the ethically relevant sense of the word can only be attributed to human beings. The revolution resulting from the creation of artificial intelligence will still, and perhaps above all, concern the inter- and intrapersonal condition of each of us (Marszałek-Kotzur, 2022a). The record on the legal and ethical implications and effects of these changes without inhibiting innovation seems to be the equivalent of Collingridge's dilemma. In 1980, David Collingridge pointed out the paradox of control in relation to technology development. At the beginning of a technology's development, it may not make sense to control it due to the lack of possibility of undesirable consequences of its use. However, when these consequences manifest themselves, control is very difficult because of the integrity with economic and social systems (Héder, 2022).

5. Summary

Conducted above analysis of the state of ethics concerning the development of artificial intelligence and robotics of problems in this area, the arrival of which has so far escaped as remote or impossible. The outline of the above issues was also intended to draw the reader's attention to the issue of new challenges to the application of ethical norms relating to the use of the latest technologies both in the global perspective and the individual human being. All of the above issues are related to the question of ensuring human security, and therefore trust in artificial intelligence, responsibility and ethical action, also the observance, or violation of human rights. In the European Parliament, however, much attention is paid to ethics. It promotes the development of artificial intelligence fast and decisive, but at the same time fair and safe for humans. This is the difference between Europe and some Asian countries, where what matters is technology, its quality and the speed of its development. Asian countries many robots already resemble humans, and their resemblance to humans is still being improved. In Europe, the prevailing conception to date is that a robot is supposed to serve humans and cannot "become" a human. In the EU, it is clearly emphasized that it is unethical to program machines to do things that are socially harmful and harm humans. It is worth considering in this perspective what is the difference between humans and artificial intelligence, so as not to lose sight of the key question for philosophical anthropology: who is a human being? Overlooking this question seems to be one of the causes of the contemporary anthropological crisis. It may also threaten the definition of fundamental human characteristics and the characteristics of artificial intelligence. This paper is only an outline of the issue and does not claim to be exhaustive. It can be a starting point for conducting detailed research in the above area.

References

1. *AI and the Ethical Conundrum. How organizations can build ethically robust AI systems and gain trust*. Retrieved from: https://www.capgemini.com/wp-content/uploads/2021/05/AI-and-the-Ethical-Conundrum-Report-1-40.pdf, 24.06.2023.
2. Arslan, A., Cooper, C., Khan, Z., Golgeci, I., Ali,I. (2021). Artificial intelligence and human workers interaction at team level: a conceptual assessment of the challenges and potential. HRM strategies. *International Journal of Manpower, Vol. 43(1),* pp. 75-88, DOI: 10.1108/IJM-01-2021-0052.
3. Biczysko-Pudełko, K., Szostek, D. (2019). *Koncepcje dotyczące osobowości prawnej robotów – zagadnienia wybrane*, Retrieved from: https://repozytorium.uni.wroc.pl/Content/120390/PDF/02_Koncepcje_dotyczace_osob.pdf, 6.08.2022.

4. Bilan, S., Šuleř, P., Skrynnyk, O., Krajňáková, E., Vasilyeva, T. (2022). Systematic bibliometric review of artificial intelligence technology in Organizational Management, Development, Change and Culture. *Business: Theory and Practice, 23(1),* pp. 1-13.
5. *Business Ethics Briefing, Iss, 58, January 2018.* Retrieved from: https://www.ibe.org.uk/userassets/briefi ngs/ibe_briefi ng_58_business_ethics_and_artifi cial_intelligence.pdf, 22.02.2019.
6. Business Insider Polska (2017). *Robot odpowie za swoją pracę? „Już musimy rozmawiać, jak to regulować".* Retrieved from: http://businessinsider.com.pl/technologie/nowe-technologie/roboty-i-sztuczna-inteligencja-a-prawo-i-osobowosc-prawna/k2e2zdx, 6.08.2022.
7. Castro, A., Silva, F., Santos, V. (2021). Trends of Human-Robot Collaboration in Industry Contexts: Handover. *Sensors, Vol. 21(12*), pp. 1-28, DOI: 10.3390/s21124113
8. Ciechomska, M. (2021). *E-usługi a RODO*. Warszawa: Wolters Kluwer.
9. De Chiara, A., Elizalde, I., Manna, E., Segura-Moreiras, A. (2021). *Car Accidents in the Age of Robots*. Retrieved from: https://www.sciencedirect.com/science/article/pii/S0144818821000466 06.08.2022
10. Ethikos Weekly (2018). *Calum Chace, Does AI Ethics Have a Bad Name?* subscriptions@corporatecompliance.org, 10.04.2019.
11. EU (2018). *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems*. Retrieved from: https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1, 22.07.2023.
12. EU (2019). *Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji.* Retrieved from: https://op.europa.eu/pl/publication-detail/-/publication/d3988569-0434-11ea-8c1f-01aa75ed71a1, 22.07.2023.
13. EU (2023). *The EU Artificial Intelligence Act*. Retrieved from: https://www.artificial-intelligence-act.com/, 25.07.2023.
14. European Commission, *Ethics Guidelines for Trustworthy AI.* Retrieved from: https://ec.europa.eu/futurium/en/ai-alliance-consultation.1.html, 22.07.2023.
15. Fortuna, P. (2021). *Optimum. Idea cyberpsychologii pozytywnej*. Warszawa: PWN.
16. Fukuyama, F. (1997). *Zaufanie. Kapitał społeczny a droga do dobrobytu*. Warszawa/Wrocław: PWN.
17. Giddens, A. (1999). *The Consequences of Modernity.* Cambridge: Polity Press.
18. Giddens, A. (2001). *Poza lewicę i prawicę. Przyszłość polityki radykalnej.* Poznań: Zysk i Ska.
19. Groom, V., Nass, C. (2007). Can robots be teammates? Benchmarks in human-robot teams. *Interaction Studies, Vol. 8(3),* pp. 483-500, DOI: 10.1075/is.8.3.10gro.
20. Hagendorff, T. (2020). The Ethics of AI Ethics: An Evaluation of Guidelines. *Minds and Machines, vol. 30*, p. 111.

21. *Harnessing the value of generative AI. Top use cases across industries* (2023). Retrieved from: https://prod.ucwe.capgemini.com/wp-content/uploads/2023/07/GENERATIVE-AI_-Final-Web-.pdf, 24.06.2023.
22. Héder, M. (2020). A Criticism of AI Ethics Guidelines. *Információs Társadalom, vol. 20, No. 4*, p. 61.
23. Heynes, Ch. *Report of the Special Rapporteuron extrajudicial, summary or arbitrary executions, A/HRC/23/47*. Retrieved from: http://www.ohchr.org/Documents/HRBodies/HRCouncil/RegularSession/Session23/AHRC-23-47_en.pdf, 27.04.2023.
24. Huang, I., Cooke, N.J., Gutzwiller, R.S., Berman, S., Chiou, E.K., Demir, M., Zhang, W. (2020). Distributed dynamic team trust in human, artificial intelligence, and robot teaming. In: N. Chang, J. Lyons (eds.), *Trust in Human-Robot Intraction*. Academic Press, DOI: 10.1016/B978-0-12-819472-0.00013-7.
25. Ike, O. (2018). *Between Technophobia and Technoutopia: Ethical Challenges of Artifi cial Intelligence*. Retrieved from: https://www.academia.edu/41675587/Etyka_biznesu_wobec_Internetu_Rzeczy, 23.07.2023.
26. Juraszczyk, A., Skarżyńska, O. (2023). *Według dużych firm sztuczna inteligencja niesie więcej korzyści niż ryzyka [raport]*. Retrieved from: https://capgeminipolska.prowly.com/253931-wedlug-duzych-firm-sztuczna-inteligencja-niesie-wiecej-korzysci-niz-ryzyka-raport, 24.06.2023.
27. Kamieński, L. (2022). *Mimowolne cyborgi. Mózg i wojna przyszłości*. Wołowiec: Wydawnictwo Czarne.
28. Kätsyri, J., Förger, K., Mäkäräinen, M., Takala, T. (2015). A review of empirical evidence on different uncanny valley hypotheses: support for perceptual mismatch as one road to the valley of eeriness. *Front Psychol., 6*, p. 390.
29. Kiepas, A. (2020). Człowiek w świecie procesów cyfryzacji, Współczesne wyzwania i przyszłe skutki. *Filozofia i Nauka, vol. 8, 1*.
30. Kolbeinsson, A., Lagerstedt, E., Lindblom, J. (2019). Foundation for a classification of collabo-ration levels for human-robot cooperation in manufacturing. *Production & Manufacturing Research, Vol. 7(1*), pp. 448-471, DOI: 10.1080/21693277.2019.1645628.
31. Komisja Europejska (2019). *Wytyczne w zakresie etyki dotyczące godnej zaufania sztucznej inteligencji, Grupa ekspertów wysokiego szczebla ds. sztucznej inteligencji*. Retrieved from: https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60436, 8.08.2021.
32. Komisja Europejska (2020). *Biała księga w sprawie sztucznej inteligencji. Europejskie podejście do doskonałości i zaufania*. Retrieved from: https://ec.europa.eu/info/sites/default/files/commission-white-paper-artificial-intel-ligence-feb2020_pl.pdf, 8.08.2021.
33. Kruszewski, T. (2019). Ocena zależności między wizerunkiem robota a zaufaniem do robota w świetle koncepcji doliny niesamowitości, na przykładzie zawodów o wysokim prestiżu społecznym. *Zagadnienia Informacji Naukowej – Studia Informacyjne, 2(114)*, pp. 80-96.

34. Kuipers, B., Mattei, N. (2018). 1st AAAI/ACM Conference on Artificial Intelligence, Ethics, and Society: a retrospective. *AI Matters, Vol. 4, Iss. 1,* pp. 7-9. Retrieved from: https://doi.org/10.1145/3203247.3203250, 6.08.2022.
35. Kuzior, A. (2020). Kompetencje społeczeństwa 4.0 w dobie rozwoju technologii kognitywnych, sztucznej inteligencji i przemysłu 4.0. In: A. Michna, J. Kaźmierczak (eds). *Przemysł 4.0 w organizacjach. Wyzwania i szanse dla mikro, małych i średnich przedsiębiorstw*. Warszawa: CeDeWu, pp. 15-27.
36. Kuzior, A., Kochmańska, A., Marszałek-Kotzur, I. (30-31 May 2021). *Information and communication technologies as a tool of modern communication in organizations and society. Innovation management and information technology impact on global economy in the era of pandemic.* K.S. Soliman (ed.). Proceedings of the 37th International Business Information Management Association Conference (IBIMA), Cordoba, Spain. International Business Information Management Association, pp. 5240-5247.
37. Kuzior, A., Kwilinski, A. (2022). Cognitive Technologies and Artificial Intelligence in Social Perception. *Management Systems in Production Engineering, 2*, pp. 109-115.
38. Kuzior, A., Marszałek-Kotzur, I. (2022). Ethical problems of the development of artificial intelligence. In: P. Kasprowski (ed.), *Artificial intelligence and data processing, vol. 954* (pp. 527-537). Politechnika Śląska.
39. 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, No. 2*, pp.133-138. DOI:10.2478/mspe-2020-0020.
40. Lambrechts, W., Klaver, J.S., Koudijzer, L., Semeijn, J. (2021). Human Factors Influencing the Implementation of Robots in High Volume Distribution Centres. *Logistics, Vol. 5(2),* pp. 1-24, DOI: 10.3390/logistics5020032.
41. Leibert, W., Schmidt, J.C. (2010). Collingridge's Dilemma and Technoscience. *Poiesis Prax, vol. 7*, p. 57.
42. Lemley, M.A., Casey, B. (2019). *Remedies for Robots*, pp. 1311-1396. Retrieved from: https://lawreview.uchicago.edu/publication/remedies-robots, 6.05.2022.
43. Lipińska, I. (2022). Etyka sztucznej inteligencji w dokumentach Unii Europejskiej w latach 2017-2020. *Edukacja Filozoficzna,* pp. 11-38, DOI: 10.14394/edufil.2022.0001.
44. *Losing Humanity. The Case against Killer Robots* (11.2012), pp. 1-49. Retrieved from: http://www.hrw.org/sites/default/files/ reports/arms1112_ForUpload.pdf, 14.08.2020.
45. Marczuk, J. (2012). *Etyka zabijania, czyli wszystko o robotach bojowych (wywiad z prof. Ronaldem C. Arkinem). Gazeta Wyborcza, 21/07/2012.*
46. Marszałek-Kotzur, I. (2022a). Cognitive technologies - are we in danger of humanizing machines and dehumanizing humans? *Management Systems in Production Engineering, vol. 30, No. 3*, pp. 269-275. DOI:10.2478/mspe-2022-0034.

47. Marszałek-Kotzur, I. (2022b). Ethical perspective of the development and use of modern military technologies. *Zeszyty Naukowe Politechniki Śląskiej. Organizacja i Zarządzanie, No. 165,* pp. 183-198. DOI:10.29119/1641-3466.2022.165.13.
48. Masahiro, M. (2012). *The Uncanny Valley: The Original Essay by Masahiro Mori.* Retrieved from: http://spectrum.ieee.org/automaton/robotics/humanoids/the-uncanny-valley, 29.04.2017.
49. Maternowska, M. (2022). Dylematy odpowiedzialności za roboty sterowane sztuczną inteligencją. *Nowoczesne Systemy Zarządzania, 3*, pp. 13-24. Retrieved from: https://www.ceeol.com/search/article-detail?, 29.07.2022.
50. Metzinger, T. *Ethics Washing Made in Europe*. Retrieved from: https://www.tagesspiegel.de/politik/eu-guidelines-ethics-washing-made-in-europe/24195496.html, 16.07.2023.
51. Michalski, R. (2018). How to Sue a Robot. *Utah Law Review, No. 5*, pp. 1021-1072. Retrieved from: https://dc.law.utah.edu/cgi/viewcontent.cgi?article=1189&context=ulr, 6.05.2022.
52. Missala, T. (2015). EMC robotów społecznych – wyzwanie XXI wieku. *Przegląd Elektrotechniczny, 91, 11*, pp. 54-57, doi:10.15199/48.2015.11.16.
53. Nissenbaum, H. (2010). *Privacy in the context: technology, privacy and the integrity of social life.* Stanford: University Press.
54. Nowak, W. (2020). Specyfika zagrożeń w cyberprzestrzeni. In: C. Banasiński, M. Rojszczak (eds.), *Cyberbezpieczeństwo*. Warszawa: Wolters Kluwer.
55. Ober, J. (2022). Open innovation in the ICT industry: substantiation from Poland. *Journal of Open Innovation: Technology, Market, and Complexity, vol. 8, no. 3, Art. no. 158,* pp. 1-33, DOI:10.3390/joitmc8030158.
56. Pawlak, A. (2022). *Prawo do prywatności w dobie sztucznej inteligencji*. Retrieved from: https://web.archive.org/web/20220518224357id_/https://www.wydawnictwo.wsge.edu.pl/pdf-145114-70645?filename=Prawo%20do%20prywatnosci%20w.pdf, 8.08.2022.
57. Platon (1997). *Państwo. Prawa.* Kęty: Antyk.
58. *Powszechna Deklaracja Praw Człowieka*. Retrieved from: https://www.unesco.pl/fileadmin/user_upload/pdf/Powszechna_Deklaracja_Praw_Czlowieka.pdf, 22.07.2023.
59. Rezolucja 2015/2103 (2015). *Rezolucja Parlamentu Europejskiego z 16.2.2017 r. zawierająca zalecenie dla Komisji w sprawie przepisów prawa cywilnego dotyczących robotyki* (2015/2103(INL)).
60. Rojszczak, M. (2019). *Ochrona prywatności w cyberprzestrzeni z uwzględnieniem zagrożeń wynikających z nowych technik przetwarzania informacji*. Warszawa: Wolters Kluwer.
61. Schindler, S., Zell, E., Botsch, M., Kissler, J. (2017). Differential effects of face-realism and emotion on event-related brain potentials and their implications for the uncanny valley theory. *Scientific Reports, 7, Art. no. 45003*, Doi: 10.1038/srep45003.

62. Shavell, S. (2019). On the Redesign of Accident Liability for the World of Autonomous Vehicles. *NBER Working Paper Series*. Retrieved from: https://www.nber.org/system/files/working_papers/w26220/ w26220.pdf, 16.05.2023.
63. Siau, K., Wang, W. (2018). Building Trust in Artificial Intelligence. *Machine Learning, and Robotics. Cutter Business Technology Journal, 31*, 47-53.
64. Simon, O., Neuhofer, B., Egger, R. (2020). Human–robot Interaction: Conceptualising trust in frontline teams through LEGO® Serious Play®. *Tourism Managment Perspectives, Vol. 25(100692)*, pp. 1-19, DOI: 10.1016/j.tmp.2020.100692.
65. Skrynnyk, O., Lyeonov, S., Lenska, S., Litvinchuk, S., Galaieva, L., Radkevych, O. (2022). Artificial Intelligence in Solving Educational Problems. *Journal of Information Technology Management*, *14*, pp. 132-146.
66. Słowik, P. (2018). *Zaczyna się walka o prawa robotów. "Maszyna jest lepsza od najzdolniejszego człowieka".* Retrieved from: https://wiadomosci.dziennik.pl/nauka/artykuly/582621,prawo-roboty-technologia-przepisy-etyka-zmiana.html, 29.04.2019.
67. *Statement on Artificial Intelligence, Robotics and 'Autonomous' Systems* (2018). Retrieved from: https://op.europa.eu/en/publication-detail/-/publication/dfebe62e-4ce9-11e8-be1d-01aa75ed71a1, 22.06.2022.
68. Sztompka, P. (1999). *Trust. A Sociological Theory*. Cambridge: Cambridge University Press.
69. Sztompka, P. (2007). *Zaufanie. Fundament społeczeństwa.* Kraków: Znak.
70. Szynkiewicz, M. (2014). Problem zaufania w kontekście rozwoju społecznego znaczenia technologii informatycznych. *Filo–Sofija, No 24(2014/1)*, pp. 259-272.
71. Talley, E. (2019). *Automatorts: How Should Accident Law Adapt to Autonomous Vehicles? Lessons from Law and Economics*. Retrieved from: https://www.hoover.org/sites/default/files/ip2-19002-paper.pdf, 26.06.2022.
72. Tunc, A.O. (2020). Human-Robot Interaction in Organizations. In: U. Hacioglu (ed.), *Digital Business Strategies in Blockchain Ecosystems.* DOI: 10.1007/978-3-030-29739-8_2.
73. Turner, J. (2019). *Robot Rules. Regulating Artificial Intelligence.* Cham.
74. Weiss, A., Wortmeier, A.-K., Kubicek, B. (2021). Robots in Industry 4.0: A Roadmap for Future Practice Studies on Human–Robot Collaboration. *IEEE Transactions on Human–Machine Systems*, *Vol. 3*, pp. 1-11, DOI: 10.1109/THMS.2021.3092684.
75. Witkowska, A. (2021). *Etyczna zagadka AI, czyli jak budować zaufanie do sztucznej inteligencji.* Retrieved from: https://capgeminipolska.prowly.com/118309-etyczna-zagadka-ai-czyli-jak-budowac-zaufanie-do-sztucznej-inteligencji.
76. Zalewski, T. (2020). Definicja sztucznej inteligencji. In: L. Lai, M. Świerczyński (eds.), *Prawo sztucznej inteligencji.* Warszawa: C.H. Beck.