2023, Vol. X
No. 2 (24-46)

NAUKA I DRUŠTVO – SCIENCE AND SOCIETY
Časopis za društvene nauke – Journal of Social Sciences

Review Paper
UDC: 007.52:274/278
DOI: https://doi.org/10.5281/zenodo.10432546
Received: 20 August 2023
Accepted: 4 September 2023

Sanja Popović*

University of Belgrade

Faculty of Political Sciences

ETHICAL IMPLICATIONS OF USING ARTIFICIAL INTELLIGENCE TECHNOLOGIES

Abstract

Artificial intelligence is an integral part of our everyday life, and has found its place of application in many different areas. Its multidisciplinarity, as a reflection of its potential, achieves the best effects when combined with human experience and decisions. It is indisputable that with the advancement of intelligent systems, our world is becoming more efficient and its quality better, however, multiple legal and social issues have revealed that the capacity of new technologies can also produce unwanted effects. The fear of "thinking machines" has led AI ethics to focus on problems of different kind, opening up their background and philosophies both in computing and in other sciences. In the following text, the author will try to familiarize the readers with the basic ethical concepts, their theoretical background, as well as their role in the discussion of modern

* E-mail: popovic.sanya@yahoo.com

artificial intelligence. The author will try to explain why ethical principles are a key factor in ensuring safe and responsible implementation, while carefully considering the issues of system autonomy, privacy, surveillance, fairness, bias and transparency in the changed socio - economic circumstances.

***Key words**: autonomous systems, privacy, surveillance, fairness, bias, transparency*

INTRODUCTION

Artificial intelligence (AI) is often described as a new generation of software that can learn, progress, make decisions and perform tasks independently, without human intervention. Technology can predict what the human brain cannot, and its ability to rationalize and take actions that have the best chance of achieving a particular goal represents its greatest potential. However, technology alone is not enough, it achieves the best effects when combined with human experience and decisions.

AI is an integral part of our daily lives through the use of smartphones and virtual assistants, while driving a car with an auto-pilot system or learning a foreign language. It has found its place of application in the most diverse areas, from traffic, energy, medicine, finance, public services to the military sector, and this multidisciplinarity is a reflection of its genuine capacity.

At the same time, multiple legal and social issues have revealed that the potential of artificial intelligence technologies can also produce unwanted impacts. Algorithms can enhance pre-existing prejudices, can discriminate individuals on various grounds, and threaten our security. Technology has raised fundamental questions about what we should do with these systems, furthermore what the systems themselves should do, what risks they involve, and how we can control them.

In this paper, we will single out one of the most important and interesting issues, which has become the subject of many debates - the issue of ethics.

Technology can be used as a tool to improve the quality of life or as a weapon that can completely destroy it, therefore ethics appears as an important discipline that cannot and must not be separated from technology.

The fear of "thinking machines" has led AI ethics to focus on problems of different kind, opening up their background and philosophies both in computing and in other sciences. In the following text we will try to familiarize ourselves with the basic ethical concepts, their theoretical background, as well as their role in the discussion of modern artificial intelligence.

Artificial Intelligence

The term "artificial intelligence" (AI) is broadly understood as any type of artificial computer system that exhibits intelligent, i.e. complex behavior, which leads to the execution of certain tasks and the achievement of specific goals. It seriously began to develop shortly after the Second World War, constituted as a field in direct connection with the establishment of computer science, as a scientific-technical discipline, but it was only at the end of the nineties of the twentieth century that it attracted actual serious attention of the whole world. Namely, computers defeated the world chess champion and proved the theorems that humans had not been able to prove until then, and already in the next century, appeared extremely successful systems based on "deep learning" - systems that successfully recognized faces in photographs, translated natural languages, vehicle navigation, etc.

Some believe that artificial intelligence brings great benefits, and some that it is threatening, while, of course, there are those who believe in both. What is certain is that there is still no agreement on what AI is and what it actually does. Intelligence usually means the ability to acquire, remember and process certain knowledge. Although there is a perception that the central goal of artificial intelligence is to imitate human intelligence, most of its subfields have a different goal (Jančić, Nikolić, 2022, p.9). Despite the widespread lack

of knowledge in this area, AI is a technology, which, as a wide-ranging tool is definitely transforming every walk of life. Artificial intelligence systematizes and automates intellectual tasks and is therefore potentially significant for any sphere of human intellectual activity. (Milosavljević, 2015, p. 1).

As we encounter today, in the conditions of a complex global socio-economic system, a multitude of different problems ensued as a result, their division into those that can be easily solved using computers, on the one hand, and those that can be solved by humans more easily than machines, on the other. These problems include various games, vision, shape recognition, understanding and classification of documents, semantic analysis of texts, motor skills, driving vehicles, intuitive and heuristic reasoning and decision-making in conditions of uncertainty.

This second group of problems is in the field of artificial intelligence precisely for this reason, that the way humans solve them is still superior to automated computer systems. Therefore, the field of artificial intelligence is dynamic and open. Some problems are excluded from it, such as the previously mentioned chess playing (where the possibilities of technology have exceeded human), while some new ones are included, such as data mining or medical diagnostics (2015, p. 2). Most of the tasks of artificial intelligence can be described in terms of algorithms, search, deductive and inductive reasoning and other precise mathematical concepts, and only a small number of researchers deal with methods that aim to achieve general human-style reasoning (Jančić, Nikolić, 2022, p.9).

AI consists of many different subfields, but what they all have in common is the general structure of the problem-solving process. Artificial intelligence software systems make decisions that typically require a human level of expertise, helping people anticipate problems or solve them by working in a deliberate, intelligent, and adaptive way. Intentionality is reflected in AI algorithms that are designed to make decisions using real-time data – using sensors and digital data, they combine information from various sources, analyze

the results obtained and act on the insights obtained from such data. With vast improvements in storage systems, processing speeds, and analytical techniques, they are capable of a high level of sophistication in analysis and decision-making, thus distinguishing them from passive machines capable only of mechanical or predetermined reactions.

Artificial intelligence is essentially implemented along with machine learning and data analytics. Machine learning uses specific data and finds underlying trends from it, while the data itself have to be robust enough for the algorithms to discern useful patterns. This data can be in different forms – digital information, satellite images, visual information, text or unstructured data.

AI systems also have the ability to learn and adapt during decision-making, which can best be seen in the example of autonomous and semi-autonomous vehicles. Semi-autonomous vehicles have various tools that inform drivers about possible traffic obstacles, traffic jams, road works, etc.

The advanced algorithms of these vehicles, various sensors and cameras incorporate experience into current operations and use dashboards and displays to present real-time information to drivers, who need to gain insight into the current traffic situation. On the other hand, in fully autonomous vehicles, advanced systems can fully control the vehicle and make all navigational decisions instead of the driver. As you can see, artificial intelligence is almost ubiquitous today and is not a futuristic vision but something that exists here and now. As a broad and complex field, it is integrated and distributed in various sectors, such as finance, national security, health care, transportation, etc. There are numerous examples in which artificial intelligence is already affecting the world and significantly increasing human capabilities, and it is generally believed that the aforementioned ubiquity will continue to expand and deepen in the future.

Of the many questions and discussions that artificial intelligence raises, in this paper we will point out and look at one of the most important, and

perhaps the most interesting - questions of ethics, triggered by the fear of "thinking machines".

Ethics of Artificial Intelligence

The ethics of artificial intelligence and robotics is, one might say, focused on problems of different kind, which is a typical response to new technologies. Some of the most common ethical dilemmas that can be heard even in everyday conversation relate to the selection and filtering of information on the Internet and social networks that reach users, autonomous driving systems, the famous "killer robots", as well as the replacement of people and activities. which are performed by machines. There are many questions that need to be answered in this area, and the two most important, which form the very essence of AI, are: How humans can treat AI ethically and How AI can treat humans ethically?

Some previously developed technologies, such as nuclear energy, cars or plastics, have provoked serious ethical and political discussions and significant political efforts to control the trajectory of these technologies, which usually happens when some damage has already occurred.

In addition to such "ethical concerns", new technologies certainly challenge existing norms and conceptual systems, and when we understand technology in its proper context, there is a need to shape our social response, regulations and law. All of these features exist in the case of new AI and robotics technologies, which are further supported by the fundamental fear that they could end the era of human control on earth (Müller, 2020, p. 2).

The profound impact that these technologies have on everyday life is not without risks, and the possibility of using them for malicious purposes, accompanied by unwanted negative consequences, causes a certain amount of anxiety and fear. Accordingly, we have a moral obligation to use such a powerful technology well, promote its positive aspects and avoid or mitigate the negative ones.

It is undeniable that there are many positive aspects. For example, AI can save lives through improved medical diagnostics, new medical discoveries, better prediction of extreme weather conditions, safer driving with driver assistance and (eventually) self-driving technologies. There are also many opportunities to improve life. Microsoft's AI for Humanitarian Action program applies AI to recovering from natural disasters, addressing the needs of children, protecting refugees, and promoting human rights. Google's AI for Social Good program supports work on rainforest protection, human rights jurisprudence, pollution monitoring, measurement of fossil fuel emissions, crisis counseling, news fact checking, suicide prevention, recycling, and other issues. The University of Chicago's Center for Data Science for Social Good applies machine learning to problems in criminal justice, economic development, education, public health, energy, and environment. Optimization of business processes using machine learning will make businesses more productive, increasing wealth and providing more employment. Automation can replace the tedious and dangerous tasks that many workers face, and free them to concentrate on more interesting aspects (Russell, Norvig, 2021, p. 986). By using different tools based on machine learning, technology allows us to communicate in an easier way with other people of different cultures who speak different languages. These are just some of the positive sides of AI, which, in addition to the above, are many, but there is also the other, negative one, which we must not ignore.

Many new technologies have had unintended negative side effects: nuclear fission brought Chernobyl and Negative side effects the threat of global destruction; the internal combustion engine brought air pollution, global warming, and the paving of paradise. Other technologies can have negative effects even when used as intended, such as sarin gas, AR-15 rifles, and telephone solicitation (2021, p. 987). It is also expected that artificial intelligence will replace humans in many jobs - starting from translation services, copywriting, legal research, to complex ones, such as design, programming, etc. For example, many of us are certainly familiar with and use

ChatGPT extensively – conversational artificial intelligence, which is one of the many innovations of hyperautomation and AI.

It will form a part of architected solutions that automate, augment humans or machines, and autonomously execute business and IT processes. It will also likely be used to replace, recalibrate and redefine some of the activities and tasks included in various jobs (McCartney, 2023). But amidst all the excitement about this tool, many questions remain about what generative AI really is and what it can do, both for people and in enterprise use cases.

When we talk about the potential problem of future unemployment, counterarguments can be made that rely on the claim that our needs will evolve into complex forms that will require real people and that AI will eventually employ people with a different set of abilities. What is challenging is finding suitable alternative jobs for humans replaced by artificial intelligence (Taumar, 2021, p.10). In addition to unemployment, a possible unequal distribution of wealth is a serious problem. Automation will undoubtedly generate large revenues, however, most of the revenues will go to the owners of the automated systems which will lead to an increase in income inequality and thus disrupt the current economic conditions. Developing countries, which are oriented towards growth, through cheap production for export, may be greatly affected if developed countries turn to fully automated production capacities, which raises the extremely important question of whether the wealth created will be distributed based on need, capacity and information or general well-being?

What is certain is that primarily ethical and national government decisions will determine the level of inequality that artificial intelligence can cause.

All scientists and engineers face ethical considerations of what projects they should or should not take on, and how they can make sure the execution of the project is safe and beneficial. In 2010, the UK's Engineering and Physical Sciences Research Council held a meeting to develop a set of Principles of Robotics. In subsequent years other government agencies,

nonprofit organizations, and companies created similar sets of principles. The gist is that every organization that creates AI technology, and everyone in the organization, has a responsibility to make sure the technology contributes to good, not harm. The most commonly-cited principles are solicitation (Russell, Norvig, 2021, p. 987):

Ensure safety	Establish accountability
Ensure fairness	Uphold human rights and values
Respect privacy	Reflect diversity/inclusion
Promote collaboration	Avoid concentration of power
Provide transparency	Acknowledge legal/policy implications
Limit harmful uses of AI	Contemplate implications for employment

In addition to the previously mentioned main negative effects of AI, in the following text we will look at the most significant ethical issues of human use of AI and robotic systems that can be more or less autonomous, and occur with certain uses of technologies, which would not appear with others.

Autonomous systems

Autonomous artificial intelligence systems represent complex software systems that have the ability to recognize the environment and, based on the analysis of certain data, independently make decisions and perform tasks - they can function without human control**.**

These systems use machine learning algorithms, deep learning, natural language and other combinations of advanced technologies to process and analyze enormous amounts of data, identify patterns, and then based on

them, make certain decisions and take actions. Some of the most common and well-known examples are autonomous vehicles, autonomous weapons, drones, robots, data analysis software and other similar systems.

When we talk about autonomous vehicles, the fact is that these vehicles can reduce the number of traffic accidents, since they are programmed to be safer than human drivers, they can improve the efficiency of transport, moving faster and more precisely than human drivers, etc. However, the main question is, what is the safety of these vehicles and what would be their reaction in unexpected situations. Michael Wooldridge presents it vividly in the following way: What should a driverless car do that has to choose between driving straight and killing five people or swerving and killing one? We should first ask ourselves, is it reasonable for us to expect an AI system to do this? Second, none of us had to take an ethics test to get a driver's license, because driving a car does not require deep ethical reasoning. Therefore, it seems somewhat absurd to demand that autonomous vehicles can solve this kind of problem, and one cannot even expect that they will develop ethical reasoning of this type in the coming decades. So what would a real autonomous vehicle do in practice, faced with this situation? The most likely outcome would be that the car would just brake, and maybe that's probably all we'd be able to do in the same situation (Wooldridge, 2021, p. 141 – 143). On the other hand, the aforementioned dilemma is inseparably connected with the question of responsibility for possible human and material damage that may occur. Although autonomous vehicles are still in the development phase, there is an undeniable great potential for their application in the future, however, before the full adoption of these systems, it is necessary to first solve the challenges related to safety, regulatory and economic issues.

Most of the public debate over autonomous weapons has focused on ultraadvanced technologies not yet in operation that have been featured in futuristic movies, such as Ex Machina, the Terminator series, and Avengers: Age of Ultron, in which robots run amok (Kurth Kronin, 2020, p. 238).

Sensationalized as these depictions are, serious debate is in order. Autonomous weapons systems up the ante on potential dangers, because with greater involvement of AI the course of conflict becomes increasingly unpredictable (Kurth Kronin, 2020). Autonomous weapon systems (LAWS) select and engage targets without human intervention and become lethal when those targets include humans.

Much of the discussion on autonomous weapons has arisen from the increasing use of drones in warfare. Drones are unmanned aircraft, and in military settings, they carry weapons such as missiles. Because they do not have to carry human pilots, drones can be smaller, lighter, and cheaper than conventional aircraft, and because flying them poses no risk to those remotely piloting them, they can be used in situations that would be considered too risky for piloted vehicles. All these features naturally make drones an attractive option for military organizations (Wooldridge, 2021, p. 160).

Technologies have reached a point at which the deployment of such systems is — practically if not legally — feasible within years, not decades. The stakes are high: LAWS have been described as the third revolution in warfare, after gunpowder and nuclear arms. Existing AI and robotics components can provide physical platforms, perception, motor control, navigation, mapping, tactical decision-making and long-term planning. They just need to be combined (Russel, 2015, p. 415).

When this type of weapon is used in war, there is a great danger of errors in target estimation, and related to this , an uncontrolled increase in mortality and violence. This further raises the issue of liability in cases of damage and errors in target identification. What is particularly worrying is that LAWS could be misused by terrorists and other criminal organizations on the one hand, and by states with the aim of aggressive military action and expansion of influence, on the other.

International humanitarian law — which governs attacks on humans in times of war — has no specific provisions for such autonomy, but may still be

applicable. The 1949 Geneva Convention on humane conduct in war requires any attack to satisfy three criteria: military necessity; discrimination between combatants and non-combatants; and proportionality between the value of the military objective and the potential for collateral damage. (Also relevant is the Martens Clause, added in 1977, which bans weapons that violate the "principles of humanity and the dictates of public conscience.") These are subjective judgments that are difficult or impossible for current AI systems to satisfy. (Russel, 2015, p. 416).

Since 2014, the United Nations in Geneva has conducted regular discussions under the auspices of the Convention on Certain Conventional Weapons (CCW) on the question of whether to ban lethal autonomous weapons. About 30 nations, ranging in size from China to the Holy See, have declared their support for an international treaty, while other key countries—including Israel, Russia, South Korea, and the United States—are opposed to a ban. The debate over autonomous weapons includes legal, ethical and practical aspects (Russell, Norvig, 2021, p. 988)

The debate has many facets. Some argue that the superior effectiveness and selectivity of autonomous weapons can minimize civilian casualties by targeting only combatants. Others insist that LAWS will lower the threshold for going to war by making it possible to attack an enemy while incurring no immediate risk; or that they will enable terrorists and non-state-aligned combatants to inflict catastrophic damage on civilian populations. (Russel, 2015, p. 416). LAWS are actually capable of violating fundamental principles of human dignity, leaving it up to the machines to decide who to target and potentially eliminate, which can result in anyone that exhibits the slightest bit of "threatening behavior".

Almost all CCW member states agree on the need for meaningful human control over these types of weapons, as well as over targeting and engagement decisions made by robotic weapons, however, the meaning of the word "meaningful" in this context has not yet been established. What is needed is

for scientists, i.e. the professional public in the field of artificial intelligence and robotics, to take a clear position on this issue, as physicists, chemists and biologists did before when it came to the use of WMD in war, because doing nothing means a vote for continued development and application. Even if there is public and political will to control or even ban the development and use of LAWS, difficulties may arise in the formulation and implementation of such legislation, however, what is encouraging is that there are indications that governments are willing to try.

Privacy and Surveillance

The digital sphere has widened greatly: All data collection and storage is now digital, our lives are increasingly digital, most digital data is connected to a single Internet, and there is more and more sensor technology in use that generates data about non-digital aspects of our lives. AI increases both the possibilities of intelligent data collection and the possibilities for data analysis. This applies to blanket surveillance of whole populations as well as to classic targeted surveillance. (Müller, 2020, p. 5). As a result, there are concerns that large amounts of data used by AI systems can be used to monitor and violate people's privacy. For example, data about our internet searches, purchases, movement and communication can be collected and analyzed, and then misused for various purposes - marketing is one example that almost all of us have probably come across in our daily activities. The use of artificial intelligence for surveillance, on the other hand, becomes a very big problem because it can recognize faces and track people's movements, which can further be misused in order to monitor activists, journalists or other people who oppose the government, for example. It can also very easily be used to discriminate based on race, gender, ethnicity or some other characteristic.

As of 2018, there were as many as 350 million surveillance cameras in China and 70 million in the United States. China and other countries have begun exporting surveillance technology to low-tech countries, some with

reputations for mistreating their citizens and disproportionately targeting marginalized communities. AI engineers should be clear on what uses of surveillance are compatible with human rights, and decline to work on applications that are incompatible (Russell, Norvig, 2021, p. 990).

It should not be ignored that the technological revolution and the rapid expansion of the Internet and its users has brought with it improved and widespread means of committing various criminal acts. Many essential services of various institutions and entities that we use to meet our daily needs, such as electronic commerce and banking, are increasingly dependent on the Internet, and are therefore a frequent target of cyber attacks, and our personal data is becoming increasingly vulnerable.

Machine learning can be a powerful tool for both sides in the cybersecurity battle. Attackers can use automation to probe for insecurities and they can apply reinforcement learning for phishing attempts and automated blackmail. Defenders can use unsupervised learning to detect anomalous incoming traffic patterns and various machine learning techniques to detect fraud (2021, p. 990).

As attacks become more sophisticated, the responsibility of engineers and security professionals to build stable and secure systems increases, while the corporations and institutions that collect our data have a moral and legal responsibility to be good stewards of the data they hold.

For example, the European Union's General Data Protection Regulation (GDPR) mandates that companies design their systems with protection of data in mind and requires that they obtain user consent for any collection or processing of data.

Despite all efforts, cybercrime remains a major problem, as new ways are constantly being found to exploit technology for malicious purposes. Therefore, it is important not only to continue to develop new strategies and technologies to improve security measures and reduce risk, but also to educate people about the safe use of computer networks and the Internet.

Fairness and bias

A very common question is whether the algorithms used in artificial intelligence make decisions that are fair and whether there is bias in those decisions.

Though artificial intelligence is capable of speed and capacity of processing that's far beyond that of humans, it cannot always be trusted to be fair and neutral (Bossman, 2016). AI algorithms should not discriminate against individuals and groups, but make decisions that are fair in every sense. However, in certain situations, algorithms can be biased, on the one hand, due to the way they are programmed or, on the other hand, due to an insufficiently diverse sample of data used to train the algorithms.

The most common example of bias in AI is discrimination based on race, gender, or ethnicity.

For example, a company's candidate selection algorithm may be programmed to favor candidates who have similarities with previous employees, which may further lead to discrimination on various grounds. Next example is Google and its parent company Alphabet are one of the leaders when it comes to artificial intelligence, as seen in Google's Photos service, where AI is used to identify people, objects and scenes. But it can go wrong, such as when a camera missed the mark on racial sensitivity, or when a software used to predict future criminals showed bias against black people (Bossman, 2016).

When AI is fed with information like race, gender, or language, AI learns to give differential treatment based on these features: on average, men between 20–40 years old are expected to take lesser leaves than women because of pregnancy, therefore, AI might learn to suggest employing more men in the future — this makes AI sexist (Taumar, 2021, p.11).

We should not forget, after all, that artificial intelligence systems are still created by people who themselves can be biased and discriminatory. Therefore, if AI is used properly and used by those who strive for social progress, it can certainly become a catalyst for positive change.

In order to ensure that the algorithms used in different areas do not discriminate and do not threaten the rights of individuals and groups, it is necessary to include the diversity of the sample, ensure the development of neutral algorithms, their transparency in terms of decision-making methods, as well as testing before its application. We must not neglect the creation of an appropriate regulatory framework and guidelines, in accordance with ethical principles and values, which will regulate the application of artificial intelligence in various fields.

Transparency

Transparency is indeed a multifaceted concept used by various disciplines. Recently, it has gone through a resurgence with regards to contemporary discourses around artificial intelligence (AI). For example, the ethical guidelines published by the EU Commission's High-Level Expert Group on AI (AI HLEG) in April 2019 states transparency as one of seven key requirements for the realisation of 'trustworthy AI', which also has made its clear mark in the Commission's white paper on AI, published in February 2020. (Larsson, Heintz, 2020, p. 2).

Building on fairness and bias, the transparency of AI systems is primarily reflected in the ability to understand how algorithms make decisions, what factors influence them and how they are applied in practice.

While traditionally algorithms had to be programmed 'by hand' with rules to follow and weights to attach to specifc data points, machine learning algorithms have changed the way patterns are extracted from data sets and how predictions are made (Felzmann, Fosch Villaronga, Lutz, Tamò Larrieux, 2020, p. 3334).

One notorious recent example concerns the way in which Google Translate handles languages in which the third person singular is gender-neutral, such as Turkish, Swahili and Finnish. Until recently, when translating from such languages into English, for example, the tool tended to return answers that

conform to stereotypical gender roles and characteristics (for example, the Turkish "O bir doktor" was translated as "he is a doctor", whereas "O bir hemşire" returned "she is a nurse"; "O bir mühendis" was translated as "he is an engineer", whereas "O bir aşçı" gave "She is a cook"). This has, appropriately enough, been dubbed "machine bias" but it's important to note that it is not so much that the AI system (or even its designer) is biased; rather the repetition of sexist stereotypes arises as an unintended artefact of the way in which the system is trained. Google Translate works by learning from the patterns found in millions of bilingual texts scraped from the world wide web, and since those patterns ("in the wild" as it were) tend to exemplify more general societal patterns, including prejudices about gender roles, characteristics and occupations, so the translation engine comes to replicate them (Walmsley, 2020, p. 587).

It can be said that modern machine learning systems are like "black boxes", which means that we cannot see how they work. Henry Kissinger in the article "How The Enlightenment Ends" pointed out that there is a fundamental problem for democratic decision-making if we rely on a system that is supposedly superior to humans, but cannot explain its decisions (Kissinger, 2018).

This "opacity" or lack of visibility becomes a problem if these systems are used to make decisions that affect individuals. Transparency therefore enables obtaining certain information regarding their internal functioning. The question arises, what information is it about and whether it is ethically relevant, which further depends on the ethical question we are trying to answer. In this sense, transparency is relevant to at least the following three issues:

- Do we know whether and to what extent a certain algorithmic decision is justified and on what basis it was made?
- Do we know what is being concluded about us and in what way?

• To what extent are we responsible for the actions of the system and how much do we need to know about its inner workings in order to be able to take responsibility for possible consequences?

Essentially, achieving transparency in AI systems requires an approach that includes all factors in the process of developing and deploying algorithms, not just developers, engineers, and data experts but also users of the system, to ensure the fair and ethical application of AI.

Singularity

The idea of singularity is that if the trajectory of artificial intelligence reaches up to systems that have a human level of intelligence, then these systems would themselves have the ability to develop AI systems that surpass the human level of intelligence, i.e., they are "superintelligent" (Müller, 2020, p. 18).

There are several theories as to how singularity could arise.

One theory suggests that this could happen when artificial intelligence reaches "general artificial intelligence," which is considered the ultimate goal, as opposed to technical or "narrow artificial intelligence." This means that AI would be capable of performing a wide range of different tasks in almost the same way as the human mind. Another theory states that the singularity could occur when artificial intelligence could independently update itself, improve its performance and evolve without human intervention.

Singularity also represents one of the biggest fears, namely that "the robots we created will rule the world", which could have enormous socio-economic consequences for the human civilization, including a potential threat to human existence.

Some of the concerns include the possibility that intelligent machines, if developed this way, could become unpredictable and unmanageable, and therefore a threat to human safety, and could lead to many jobs being taken

over by artificial intelligence. This raises a serious question about AI, will it one day prevail over us? Will we reach a point where humans are no longer the most intelligent beings on the planet?

CONCLUSION

Intelligent machine systems are undoubtedly transforming our lives for the better. As these systems become more advanced and capable, our world becomes more efficient and therefore richer. Artificial intelligence and robotics have raised fundamental questions, what should we do with these systems, what the systems themselves should do and what consequences and risks they have in the long term.

In the overall development of artificial intelligence, ethical norms and principles are a key factor to ensure safe and responsible application, which means first of all the careful consideration of issues of autonomy, privacy, bias, transparency, accountability and security.

Transparency in the functioning of AI systems allows us to understand how these technologies work, what data is used and how.

As artificial intelligence is an extremely complex technology with many subfields that require specific knowledge and skills, it is important to provide appropriate training and education for all those involved in its development and application. This entails not only programming and upgrading various software, but also a deep understanding of ethical and legal issues, as well as the ability to think critically about potential consequences.

This is further followed by the need to adopt laws and regulations that will enable AI to be used in a responsible and safe way, which includes defining clear rules on data use and protection, as well as setting clear quality standards for development and its application. On the other hand, the creators and producers of AI technologies, moving within the set normative frameworks,

must be responsible for its improvement and use. Therefore, in order to prevent or at least reduce the unwanted consequences that may arise from the use of AI, it is important to ensure the supervision of its development and operation. Development should be inclusive with the cooperation of all relevant factors - users and experts from different fields, as well as cooperation between the private and public sectors.

It is this inclusivity that helps in understanding and accepting different perspectives and possible consequences of AI application, which further leads to the development of better solutions and the elimination of negative impacts, or at least their reduction.

Some of the ethical issues addressed in this paper relate to the alleviation of suffering, and some to the risk of negative consequences. As we consider these risks, we should keep in mind that, taken as a whole, these technological advances mean a better life for everyone. Artificial intelligence undoubtedly has enormous potential, and its responsible use is entirely up to us.

REFERENCES

Bossmann J. (2016). Top 9 Ethical Issues in Artificial Intelligence. *World Economic Forum*. https://www.weforum.org/agenda/2016/10/top-10-ethical-issues-in-artificial-intelligence/

Felzmann H., Fosch Villaronga E., Lutz C. & Tamò Larrieu A. (2020). Towards Transparency by Design for Artifcial Intelligence. Science and Engineering Ethics. Vol. 26(6). 3333 – 3361.

Jančić P. & Nikolić M. (2022). Veštačka inteligencija (Artificial Intelligence). Univerzitet u Beogradu: Matematički fakultet.

Kissinger H. A. (2018). How the Enlightenment Ends. *The Atlantic*. https://www.theatlantic.com/magazine/archive/2018/06/henry-kissinger-ai-could-mean-the-end-of-human-history/559124/

Kurth Cronin A. (2019). Power to the people: How Open Technological Innovation is Arming Tomorrow's Terrorists. Oxford University Press.

Larsson S. & Heintz F. (2020). Transparency in Artificial Intelligence. *Internet Policy Review*. Vol. 9(2). 1 – 16.

McCartney A. (2023). Your 7 Biggest Questions About ChatGPT, Answered. *Gartner*. https://www.gartner.com/en/articles/your-7-biggest-chatgpt-questions-answered

Milosavljević M. (2015). Veštačka inteligencija (Aritificial Intelligence). Univerzitet Singidunum.

Müller V. C. (2021). Ethics of Artificial Intelligence and Robotics. In E. N. Zalta (Ed.). Stanford University. *Stanford Encyclopedia of Philosophy*. Summer 2021. Edition, 1 – 31.

Russel S. & Norvig P. (2021). Artificial Intelligence: A Modern Approach. Fourth Edition. Pearson.

Russel S. (2019). Human Compatible: Artificial Intelligence and the Problem of Control. Penguin Books

Russel S. (2015). Robotics: Ethics of Artificial Intelligence. *Nature*. Vol. 521. 415 – 418.

Taumar S. (2021). Artificial Itelligence: Revolution, Definitions, Ethics and Foundation. In Kumar P., Vinod Kumar J. & Kumar D. *Artificial Intelligence and Global Society: Impact and Practices*. (pp. 1 – 15). CRC Press.

Walmsley J. (2020). Artificial Intelligence and The Value of Transparency. AI & Society. Vol. 36. 585 – 595.

Wooldridge M. (2021). A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going. Flatiron Books.

Sanja Popović
Univerzitet u Beogradu
Fakultet političkih nauka

ETIČKE IMPLIKACIJE KORIŠĆENJA TEHNOLOGIJA VEŠTAČKE INTELIGENCIJE

Sažetak

Veštačka inteligencija je sastavni deo našeg svakodnevnog života, a mesto primene je pronašla u više različitih oblasti. Njena multidisciplinarnost kao odraz njenog potencijala postiže najbolje efekte kada se kombinuje sa ljudskim iskustvom i odlukama. Nesporno je da napredovanjem inteligentnih sistema, naš svet postaje efikasniji i kvalitetniji, međutim, višestruka pravna i društvena pitanja, otkrila su da kapacitet novih tehnologija može da proizvede i neželjene efekte. Strah od "mašina koje misle", naveo je etiku AI da se fokusira na probleme različitih vrsta, otvarajući njihovu pozadinu i filofoziju kako u računarstvu, tako i u drugim naukama. U daljem tekstu, autor će nastojati da čitaoce upozna sa osnovnim etičkim konceptima, njihovom teorijskom pozadinom, kao i njihovom ulogom u diskusiji o savremenoj veštačkoj inteligenciji. Autor će pokušati da objasni zbog čega su etička načela i principi ključan faktor za obezbeđivanje sigurne i odgovorne primene uz pažljivo razmatranje pitanja autonomije sistema, privatnosti, nadzora, pravednosti, pristrasnosti i transparentnosti u uslovima izmenjenih društveno – ekonomskih okolnosti.

Ključne reči: *autonomni sistemi, privatnost, nadzor, pravičnost, pristrasnost, transparentnost*