UDC 316

https://doi.org/10.23939/sjs2024.01.072

Mgr. Sláva Gracová
Univerzita sv. Cyrila a Metoda v Trnave
Fakulta masmediálnej komunikácie
Nám. J. Herdu 2, 917 01, Trnava
SLOVENSKÁ REPUBLIKA
slava.gracova@ucm.sk
ORCID: 0000-0002-3485-4333

Mgr. Martin Graca
Univerzita sv. Cyrila a Metoda v Trnave
Fakulta masmediálnej komunikácie
Nám. J. Herdu 2, 917 01, Trnava
SLOVENSKÁ REPUBLIKA
martin.graca@ucm.sk
ORCID: 0000-0002-7451-7497

THREATS TO THE USE OF ARTIFICIAL INTELLIGENCE AND ITS LEGISLATIVE

We are encountering the term artificial intelligence more and more often. In everyday life, we are almost completely unaware that we come into contact with it and use it. It is found in various areas of human life. Artificial intelligence does not have comprehensive legislation to date. There are a number of states active in this area, with their own leaders. These include the United States of America, the United Kingdom and the European Union. There are a number of signed declarations, regulations and procedures in this area, but a complete regulatory framework is still lacking or is only at the beginning and needs to be implemented in society. In this paper, we pay attention to the basic concepts and breakdown of artificial intelligence, which we define in the theoretical part. We then characterize the risks that artificial intelligence may pose. In addition to its relatively large contribution in various fields of development, education, or streamlining administrative affairs, artificial intelligence poses risks that humans can exploit for their own enrichment, information acquisition, or political influence. AI can work efficiently and relatively quickly with large volumes of data. It can analyse it and learn from it. It can therefore be exploited, for example, in the context of censorship, the creation of false content and disinformation, phishing, Ramsomware, cyber-attacks on various companies or institutions, deepfake videos and so on. In the conclusion of the present study, we will analyse the activities of the European Union in the field of laying the legislative framework for artificial intelligence from 2020 to the present. These regulatory activities can contribute to positive developments in eliminating the misuse of artificial intelligence for various activities dangerous to society. The speed of implementation and the quality of regulation will be an important factor for the future direction of AI.

Key words: Artificial Intelligence. Risks of AI. Legislation. Law. European Union. Ransomware. Deepfake. Voice-cloning. Cyber-attack. Spear Phishing.

Introduction. The term artificial intelligence is not unfamiliar; it is now being used more and more frequently.To begin with, it is necessary to define it and characterize the main types.

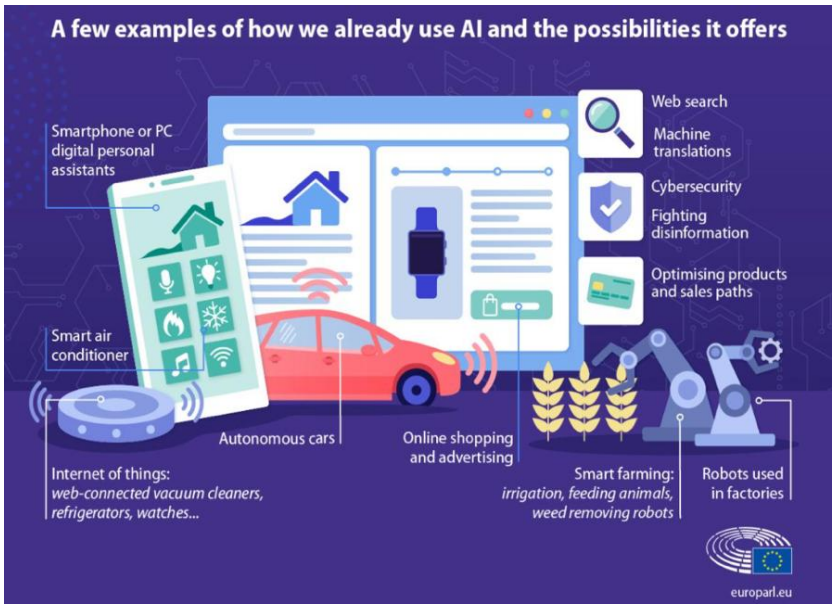
It is the science and engineering of making intelligent machines, e.g. special intelligent computer programs.It is related to the similar task of using computers to understand human intelligence, but artificial intelligence need not be limited to methods that are biologically observable [10].

Artificial intelligence (AI) is a concept that has been known in the public domain for several years.Generative AI has been developed over decades and its emergence can be dated to the second half of the 20th century.Century.Its first form transmogrified into the form of chatbots.

In the 1960s, Professor Joseph Weizenbaum created the first chatbot called Elliza.The chatbot simulated a natural language conversation between a human and a computer [15]. The real advancement of generative AI came almost 50 years later after machine learning became popular in the scientific world. Specifically, Generative Adversarial Networks (GANs), which was developed by Ian J. Goodfellow and thus provided generative AI with new opportunities to create authentic content [7].

The European Parliament document states that artificial intelligence is the ability of a device to manifest human-like abilities such as reasoning, learning, planning, and creativity [2–4].

Artificial Intelligence, or AI, has three basic types, but only one is commonly used – Artificial Narrow Intelligence.The others have no practical use at the moment.



Artificial Intelligence, Everyday Uses and Potential
Source: Spravodajstvo | Európsky parlament (2023)

Types of artificial intelligence:

a) Artificial Narrow Intelligence (ANI) – Narrow AI tasks can be driven by highly complex algorithms and neural networks, yet are unified and goal-oriented. Facial recognition, web browsing, and self-driving cars are examples of narrow AI. It is categorized as weak not because it lacks scale and power, but because it is a long way from having the human components we attribute to true intelligence.

b) Artificial General Intelligence (AGI) – should be able to successfully perform any intellectual task that a human can. Like narrow AI systems, AGI systems can learn from experience and can spot and predict patterns – but have the ability to take it a step further. AGI can extrapolate this knowledge across a wide range of tasks and situations that are not addressed by previously collected data or existing algorithms.

c) Artificial Superintelligence (ASI) – ASI systems are theoretically fully self-aware. In addition to simply mimicking or understanding human behaviour, they understand it at a fundamental level. With these human characteristics, and further enriched by processing and analytical power that far exceeds our own, AI may seem to represent a dystopian, sci-fi future in which humans become increasingly obsolete [12].

The term artificial intelligence is used in two ways. In the first case as the name of the whole technology, in the second as a type subset of the technology. We also encounter another type division in which the well-known machine learning figures:

1. **Artificial Intelligence (AI)** – as the technology itself is able to analyse the task, define and retrieve the dataset it needs to solve the task and then process the dataset to obtain the result for the task. Thus, to train on the dataset, i.e. to learn. AI usually gets to the results by iteratively processing the data, which, in addition, it is able to add to and even improve the processing itself in the process of solving. In the process of learning, AI can develop procedures, rules, patterns used to arrive at the correct result, which are new, either because they did not exist before or because they have not been discovered yet. The problem is that humans are often unable to figure out what procedure the AI used to arrive at the result, whether it is correct or incorrect. Therefore, he cannot correct the AI procedure, only give feedback on the correctness of the result.
2. **Machine Learning (ML)** – learns patterns from historical data (dataset), which it then uses to process future data. Procedures and rules are inserted into ML by humans. However, a huge amount of qualified data is needed to train it. The learning (ML training) technology is implemented by cyclically repeating the same process. The dataset for the ML application is defined and prepared by a human. If even the algorithms in the ML application are correct, but the input dataset is wrong (e.g., we label some cats in the pictures as dogs and label some dogs as cats), or there were no inputs in the dataset from which the ML could learn the correct answer (it will not learn to recognize a cow from the pictures of cats alone), we will not train the ML to give the correct results. Thus, the success of an ML project depends significantly on the quality of the dataset, since the ML technology is unable to recognize and correct the erroneous parts of the dataset. With ML, we know the process of how the ML application got the results as contrasted with AI where the process is unknown to humans. Therefore, if conditions change in ML algorithms, these must be supplied to ML as new inputs. AI can react to changes automatically. Both technologies (AI and ML) need some degree of feedback, where a human evaluates the correctness of partial results in the learning process. In ML, the feedback is more interactive and leads the application more directly to the expected outcome.
3. **Deep Learning (DL)** – uses certain kinds of ML techniques where simpler process/computing units are networked into a single DL system. The result of one unit's processing data are input to another unit. The units are arranged in layers by default. The input information when processed by a DL system must pass through multiple layers of processing units before the DL arrives at any result. The architecture is inspired by the human brain, so we call these DL systems neural networks. In practice, neural networks consist of a software representation of neurons and layer-by-layer virtual interconnections between them. Each network has at least an input layer and an output layer, i.e., an input layer through which the input data arrives at the network and an output layer through which the processed results get out of the neural network. In between these base layers are commonly nested inner layers that process the data in the same way into some intermediate results for the next layer of neurons. This ML technical design allows the DL network to learn more complex structures without requiring unrealistically large amounts of data [5].

The use of AI is being met with enthusiasm and, as mentioned above, we are encountering it more and more often in everyday life. It brings with it a number of risks and it is therefore important that this sector is also regulated.

Methodology. The main topic of the paper is to define the basic concepts in the field of artificial intelligence. Specifically, we discuss the types of artificial intelligence namely: artificial narrow intelligence (ANI), artificial general intelligence (AGI) and artificial superintelligence (ASI). A separate type division includes artificial intelligence (this term is used doubly), machine learning and deep learning.

The aim of this paper is to characterize artificial intelligence in general and to give its division in theoretical terms. The paper provides an update on the emerging legislative framework for the regulation of artificial intelligence in the European Union. It also analyses the risks of misuse of artificial intelligence in various fields.

In our paper we have used several scientific methods. In the theoretical part, mainly induction and deduction in the treatment of available sources on the classification of artificial intelligence. Subsequently, we used analysis to process in particular the available information of the European Parliament on the legislative regulation of artificial intelligence. We complemented the analysis with information on activities in this field outside the EU. We have also analysed the dangers that artificial intelligence can pose or the risks that can be exploited to gain political or economic influence, blackmail, etc.

The object of investigation is the emerging law on AI (Artificial Intelligence) in the European Union environment.

Discusion. Artificial intelligence is going through different stages of development and implementation in various societal and other sectors. But as with all advances in the modern world that have been created with good intentions, AI can be misused to the detriment of citizens, states and the like.

Manipulating the masses with misinformation and hoaxes is now the norm, not just on social media. It is part of the hybrid warfare used by Russia, for example. AI can also be used very effectively to spread hoaxes, propaganda and disinformation. It can create this type of information in a variety of ways using tools for creating deepfake videos, voice-cloning and so on. Voice-cloning can create a very realistic clone of a person's voice after analysing a short audio track within a monologue. As a result, the AI can then generate audio tracks of that person with any text, in any foreign language, with the color and tone of voice that the particular person uses. Combined with Deepfake video, which can create fake videos of different people on a similar principle after analysing the videos, they are the perfect tool for propaganda, disinformation and hoaxes.

AI is also being exploited to enrich various groups of fraudsters. Spear phishing is a type of attack called cybercrime that targets specific individuals. Unlike classic phishing, which targets the masses. Spear phishing sends messages through social networks, emails and the like. It wants to inspire trust in the recipient of these messages. To do this, he needs enough personalised data to be able to communicate with the person in a relevant way. He can get these from social networks, media or online messages. He then works with the victim's trust and uses spear phishing to obtain either a sum of money or other sensitive information that he can evaluate. Artificial intelligence can upgrade and streamline Spear Phishing.

Another group of dangerous threats are AI-assisted cyberattacks on, for example, companies or state institutions. By penetrating their security systems, attackers are able to obtain sensitive information or a certain amount of money through extortion. Ransomware is one such method of attack. This malware can encrypt a computer's disk and prevent people from accessing their important data. Through extortion, the attacker gets money. After paying it, the disk is made available again. Artificial intelligence increases the effectiveness of a ransomware attack by personalizing the target, analyzing files and data, or automating attacks.

There are many areas where AI can harm society more than help it. Of course, AI can also combat similar threats. It can't respond promptly to all situations that arise from Deepfake video abuse, spear phishing, ransomware, and other cyberattacks, but at the moment the positive uses of AI outweigh the negative ones.

AI has no regulation by law anywhere in the world yet. The European Parliament is preparing a regulation that should be the first. In April 2021, the Commission proposed the first EU regulatory framework for AI. Part of the proposal was that AI systems that can be used in different applications should be analysed and classified according to the risk they pose to users.

Back in October 2020, the European Parliament adopted three reports outlining ways to regulate AI. According to the MEPs, the rules must be set in a way that puts humans at the centre. One of the reports suggested how to ensure safety, transparency and accountability, prevent bias and discrimination, strengthen social and environmental responsibility and ensure respect for fundamental rights. Other MEPs addressed issues of civil liability in the context of AI. Another topic was issues related to intellectual property rights.

As artificial intelligence is used in different sectors, the European Parliament has been formulating progressively different guidelines, for example in January 2021 it proposed guidelines for military and civilian use of AI. The gist of it was that AI systems used for military purposes must always be supervised by a human.

In October 2021, MEPs called for strict safeguards when using AI tools in law enforcement. This included a strict ban on automated recognition of people in public spaces and transparency of algorithms to combat discrimination.

For example, the Committee on Culture is examining the use of AI in education, culture and the audiovisual sector.

There are many areas, and artificial intelligence is making its way into each of them. As we mentioned above, the European Parliament's proposals included analysing and classifying the risk that arises from the use of artificial intelligence. Different levels of risk will mean more and less regulation.

On 14 June 2023, the European Parliament adopted its negotiating position on the AI law. The priority of the law is to ensure that AI systems used in the EU are safe, transparent, traceable, non-discriminatory and environmentally friendly [2, 3, 4].

The new rules set out obligations for providers and users depending on the level of risk posed by AI:

a) **Unacceptable risk**: AI systems with unacceptable risk are systems that are considered a threat to humans and will be banned. These include e.g:

- Cognitive manipulation of the behaviour of individuals or specific vulnerable groups: for example, voice-activated toys that encourage unsafe behaviour in children,
- social scoring: classifying people based on behaviour, socio-economic status, personal characteristics,
- real-time and remote biometric identification systems, such as facial recognition.

b) **High risk:** AI systems that have a negative impact on security or fundamental rights will be considered high risk and will be divided into two categories.

1. AI systems used in products covered by EU product safety legislation. These include toys, aerospace, cars, medical devices and lifts.
2. AI systems falling into eight specific areas that will need to be registered in an EU database:

- biometric identification and categorisation of natural persons,
- management and operation of critical infrastructure,
- education and training,
- employment, workforce management and access to self-employment,
- access to and use of essential private and public services and benefits,
- law enforcement,
- management of migration, asylum and border control,
- assistance in legal interpretation and enforcement.

All high-risk AI systems will be assessed before they are placed on the market and throughout their lifecycle.

Generative AI – this type of AI, such as ChatGPT, will need to meet transparency requirements:

- disclose that the content was generated by artificial intelligence,
- design the model to prevent the creation of illegal content,
- disclose summaries of the copyrighted data used to train the AI.

c) **Limited risk:** AI systems with limited risk should meet minimum transparency requirements to enable users to make informed decisions. After interacting with the applications, the user can decide whether they want to continue using them. Users should be informed that they are interacting with AI. This includes AI systems that generate or manipulate image, audio or video content (e.g. deepfakes) [2–4].

The European Union is not alone in addressing this issue. The United States and the United Kingdom are also showing their efforts to regulate artificial intelligence. In their case, this involves organising a number of summits and formulating initiatives.

The United Kingdom hosted the Al Safety Summit in November 2023, in which leaders from China, the United States and the European Union signed a declaration committing to joint action and risk management. They are planning another similar meeting in France. At the same time, G7 leaders in turn committed to a separate non-binding code on safe AI in a document called the Hiroshima Process. Washington has gone ahead with an executive order on AI, claiming it is the strongest set of measures in the world. It orders various agencies of the federal government to adopt standards and commits companies to new safety standards. Its funding, however, will depend on congressional decisions, and implementation in fragmented offices is questionable [16].

Efforts can be seen in several places, the European Union, as mentioned above, is so far the only one that is trying to define rules and regulations in the field of artificial intelligence that would be legislatively binding. From the available sources, it can be seen that several country leaders are trying to emphasise a unified approach and global coordination, but everyone is trying to make their approach decisive.

Conclusion. There are a number of pitfalls in the development of AI. The possibilities for its use are wide, as are its abuses. Developers and researchers have a precise goal of where they would like to go with artificial intelligence. Often, their good intentions will be exploited by various interest groups and used for influence, economic gain and so on. In the present article, we have highlighted several ways, techniques and tools how AI can be misused. Therefore, it is in the interest of all states and society as a whole to reach a compromise and establish effective regulation of AI not only at the state level but also at the global level.

In 2023, the European Union is closest to laying the legislative basis for AI. As we noted in the article, their efforts are evident from 2020, and in mid-2023 they adopted a negotiating position on an AI law. It formulated rules in which it divided the risks arising from the use of AI into three groups: unacceptable, high and limited risk.

An interesting finding is that China, which has signed a declaration on joint action and risk management, is itself using the technology for social scoring of the population, one of the prohibited uses of AI under the forthcoming European Binding Act on Artificial Intelligence [16].

It is questionable when the legislation binding on a global level will come into force and whether the many summits, declarations and measures will make the situation clearer. The final form of regulation will show whether companies and states abusing AI will escape scrutiny and face the consequences of their possible illegal actions in the use of AI.

REFERENCES

1. Bontridder, N., & Poullet, Y. (2021). The role of artificial intelligence in disinformation. *Data & Policy*, *3*. https://doi.org/10.1017/dap.2021.20

2. *Formovanie európskej legislatívy v oblasti umelej inteligencie.* (2023, October 21). Spravodajstvo | Európsky parlament. https://www.europarl.europa.eu/news/sk/headlines/society/20201015STO89417/formovanie-europskej-legislativy-v-oblasti-umelej-inteligencie

3. *Zákon o AI: prvá regulácia umelej inteligencie.* (2023, June 12). Spravodajstvo | Európsky parlament. https://www.europarl.europa.eu/news/sk/headlines/society/20230601STO93804/zakon-o-ai-prva-regulacia-umelej-inteligencie

4. *Umelá inteligencia: definícia a využitie.* (2023, June 21). Spravodajstvo | Európsky parlament. https://www.europarl.europa.eu/news/sk/headlines/society/20200827STO85804/umela-inteligencia-definicia-a-vyuzitie

5. Frasch, P. (2020, May 7) *Umelá inteligencia*. ALEF. https://www.alef.com/sk/umela-inteligencia.c-531.html

6. Gooding, M. (2021, March 17). *Ransomware payouts top $300,000 with "double extortion" attacks on the rise*. Tech Monitor. https://techmonitor.ai/technology/cybersecurity/double-extortion-ransomware

7. Gonog, L., & Zhou, Y. (2019). A Review: Generative Adversarial Networks. *2019 14th IEEE Conference on Industrial Electronics and Applications (ICIEA)*. https://doi.org/10.1109/iciea.2019.8833686

8. Graca, M., Proner, J. (2023). Umelá inteligencia a jej využitie v audiovizuálnej tvorbe. In Prostináková Hossová, M. Graca, M, Labudová, L. (Eds.), *Marketing & Media Identity: AI – Budúcnosť súčasnosti.* (pp. 37-44). Faculty of Mass Media Communication, University of Ss. Cyril and Methodius in Trnava.

9. Krishnanm A. (2023, December 18). *Generative AI is making phishing attacks more dangerous*. TechTarget | Security. https://www.techtarget.com/searchSecurity/tip/Generative-AI-is-making-phishing-attacks-more-dangerous

10. McCarthy, J. (2007, November 12) *What is Artificial Inteligence?* https://www-formal.stanford.edu/jmc/whatisai.pdf

11. Ryan-Mosley, T. (2023, October 4). *How generative AI is boosting the spread of disinformation and propaganda*. MIT Technology Review. https://www.technologyreview.com/2023/10/04/1080801/generative-ai-boosting-disinformation-and-propaganda-freedom-house/

12. SAP (2023) Čo je umelá inteligencia? https://www.sap.com/sk/products/artificial-intelligence/what-is-artificial-intelligence.html

13. Stone-Gross, B. (2023, October 24). *Bracing for AI-enabled ransomware and cyber extortion attacks*. Help Net Security. https://www.helpnetsecurity.com/2023/10/24/ai-enabled-attacks/

14. Villasenor, J. (2020, November 23). *How to deal with AI-enabled disinformation*. Brookings. https://www.brookings.edu/articles/how-to-deal-with-ai-enabled-disinformation/

15. Weizenbaum, J. (1966). ELIZA---a computer program for the study of natural language communication between man and machine. *Communications of the ACM*, *9*(1), 36–45. https://doi.org/10.1145/365153.365168

16. Zmušková, B. (2023, November 3) *Briti, Spojené štáty a EÚ súperia, kto bude globálnym lídrom v regulácií AI.* Euractiv.sk. https://euractiv.sk/section/digitalizacia/news/briti-usa-a-eu-superia-kto-bude-globalnym-lidrom-v-regulacii-ai/

Слава Грацова
Мартін Граца

ЗАГРОЗИ ДЛЯ ВИКОРИСТАННЯ ШТУЧНОГО ІНТЕЛЕКТУ І ЙОГО ЗАКОНОДАВЧА БАЗА

Ми все частіше вживаємо термін штучний інтелект. У повсякденному житті ми майже не усвідомлюємо, що контактуємо з ним і використовуємо його. Він трапляється в різних сферах життя людини. На сьогодні штучний інтелект не має вичерпного законодавства. У цій сфері діє низка держав зі своїми лідерами. До них належать Сполучені Штати Америки, Велика Британія та Європейський Союз. У сфері використання штучного інтелекту є низка підписаних декларацій, регламентів і процедур, але повної нормативної бази ще немає або вона перебуває лише на початковому етапі і потребує впровадження в суспільстві. У цій роботі ми зосереджуємо увагу на основних поняттях та аналізі штучного інтелекту, які ми визначаємо в теоретичній частині. Характеризуємо ризики, які може становити штучний інтелект. Окрім відносно великого внеску в різні сфери розвитку, освіти чи оптимізації адміністративних справ, штучний інтелект створює ризики, які люди можуть використати для власного збагачення, для отримання інформації чи політичного впливу. Штучний інтелект може ефективно та відносно швидко працювати з великими обсягами даних. Він може аналізувати дані та вчитися на цьому. Отже, його можна використовувати, наприклад, у контексті цензури, створення неправдивого вмісту та дезінформації, фішингу, кібератак на різні компанії чи установи, діпфейкових відео тощо. На завершення цього дослідження ми проаналізували діяльність Європейського Союзу у сфері створення законодавчої бази для штучного інтелекту з 2020 року по теперішній час. Ця регулятивна діяльність може сприяти позитивним зрушенням у використанні зловживання штучним інтелектом для різних видів діяльності, небезпечних для суспільства. Швидкість впровадження та якість регулювання будуть важливими факторами для майбутнього напрямку штучного інтелекту.

Ключові слова: штучний інтелект, ризики ШІ, законодавство, закон, Європейський Союз, діпфейк, клонування голосу, кібератака, фішинг.