LEGAL ASPECTS OF ARTIFICIAL INTELLIGENCE

Legal Aspects of Artificial Intelligence

Legal Aspects of Artificial Intelligence

edited by Agnieszka Gryszczyńska

Wydawnictwo Instytutu Wymiaru Sprawiedliwości **Warszawa 2024**

The publication was written within the framework of the international scientific project "Polish-Hungarian Research Platform" conducted by the Institute of Justice in Warsaw in 2022

REVIEWERS *dr hab. Joanna Ryszka, prof. UO*
dr Krzysztof Mucha

PROOFREADING *Lingua Lab*

TYPESETTING AND COVER DESIGN *Tomasz Smołka* AT ONCE

ISBN 978-83-67149-37-2

WYDAWNICTWO INSTYTUTU WYMIARU SPRAWIEDLIWOŚCI
ul. Krakowskie Przedmieście 25, 00-071 Warszawa
SEKRETARIAT tel.: (22) 630 94 53, e-mail: wydawnictwo@iws.gov.pl

BOUND AND PRINTED BY *"Elpil", ul. Artyleryjska 11, 08-110 Siedlce*

Table of Contents

Preface

We present a monograph summarising the research conducted in 2022 by an international Polish-Hungarian research team on the legal aspects of artificial intelligence. The team was established as part of the Polish-Hungarian Research Platform project organised by the Institute of Justice in Warsaw. The team was comprised of István Ambrus, Agnieszka Gryszczyńska, Zbigniew Więckowski, Rafał Wielki, and Emőd Veress.

One fundamental issue in artificial intelligence is the question of when a human-created system can be said to be intelligent. In 1950, Alan M. Turing posed the question "Can machines think?" In his response, Turing noted that analysing the meanings of the terms "machine" and "think" would fail to provide a definitive answer to the question at hand due to the ambiguous nature of the terms. John McCarthy, acknowledged as the progenitor of artificial intelligence, described the process in a 1955 proposal for the Dartmouth Summer Research Project on Artificial Intelligence as "that of making a machine behave in ways that would be called intelligent if a human were so behaving".

Despite the passage of almost 70 years since the Dartmouth conference, the definition of artificial intelligence continues to cause difficulties and stir emotions, even though artificial intelligence systems, and in particular machine learning systems (MLS), are now

widely used in practice. Answering the question of what artificial intelligence entails proves highly challenging due to the absence of a widespread consensus on what intelligence is. Aside from this, there is little justification to assume that machine intelligence correlates much with human intelligence, at least at this point.

At present, no legal definition of artificial intelligence has been developed in national legislation or international conventions. Artificial intelligence is defined as a scientific field concerned with the study of the mechanisms of human intelligence and the modelling and construction of systems capable of supporting or replacing intelligent human action, or the ability of a system to correctly interpret data from external sources, learn from it, and use this knowledge to perform specific tasks and achieve goals through flexible adaptation.

Artificial intelligence is also defined as a field of knowledge encompassing, among other things, neural networks, robotics and the creation of models of intelligent behaviour and computer programmes to simulate this behaviour, as well as machine learning, deep learning and reinforcement learning.

However, there is an international consensus to frame the definition of artificial intelligence in terms of the system model, based on the technical development stream of the intelligent agent model. This approach amounts to describing artificial intelligence as an AI system. An AI system, according to the OECD, is a system based on the concept of a machine that can influence the environment by making recommendations, predictions or decisions about a given set of goals. It does this by using input, i.e., machine- or human-generated data to 1) perceive real or virtual environments, 2) manually or automatically assemble these perceptions into models, 3) use model interpretation to formulate outcome options. The systems approach is also evident in the definition in the draft proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (AI Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)). According to the new compromise version of the AI Act agreed during the trilogue in December 2023 an "AI system" is defined as "a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness

after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".[1]

The breakthrough in terms of widespread access to artificial intelligence tools that shaped the Artificial Intelligence Act project was Open AI's release of the ChatGPT advanced language model, which uses machine learning and neural network techniques to generate human-like texts. Generative artificial intelligence (GAI) is a type of artificial intelligence that enables computers to generate new data based on previously given data, such as texts, images, video and sound, which are difficult to distinguish from human-created works. Generative AI uses advances in machine learning and neural networks to learn how to create new content from existing content. Generative AI has many applications. In the administration of justice, large language models, which are advanced artificial intelligence systems that are a type of AI designed to generate and analyse natural language text, appear to be particularly useful and relatively low-risk. Large language models use machine learning and neural network techniques to learn how to create new content from existing content.

Artificial Intelligence is a fast-evolving family of technologies that can bring a wide array of economic and societal benefits across the entire spectrum of industries and social activities. The same elements and techniques that bring socio-economic benefits from the use of artificial intelligence, however, at the same time also involve new risks or negative consequences.

Unfortunately, with the advent of systems using machine learning techniques, existing knowledge is no longer sufficient to create reliable systems. This is mainly due to the fact that traditional systems rely on expert knowledge, whereas in learning systems

[1] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts – Analysis of the final compromise text with a view to agreement, 26.01.2024, https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf [accessed on: 8 March 2024].

the knowledge comes from data of varying quality and sensitivity. It has therefore become necessary to rethink aspects that affect the reliability of the system, such as security, predictability of operation or protection of sensitive data. Furthermore, research has shown that ML information systems can be biased against selected social or ethnic groups. Different defined types of bias (e.g., representativeness bias or aggregation bias) occur at different stages of system development. The effect of the occurrence of at least one type of bias is to produce an unfair model. While it can be very difficult and costly to accurately verify all of the factors that make up the credibility of a system, simply increasing the credibility of ML systems should not only reassure system users, but also lead to a greater willingness to implement new solutions.

In the context of emerging concerns regarding the fairness of ML systems, the interpretability and explicability of the system's operation have become important for building trust. This is why it is so important, especially in the area of AI applications in the administration of justice, to take into account the "Ethics Guidelines for Trustworthy Artificial Intelligence," the result of an expert group established by the European Commission. They propose a framework for the production of trustworthy AI systems, covering both the technical robustness of the system and the legal and ethical aspects of the different phases of the system life cycle. In the application of generative language models to the administration of justice, AI hallucinations may be problematic. "AI hallucinations" is a term used to refer to situations in which AI generates false, inaccurate or illogical information (e.g., citing inconsistent books or court decisions, generating false, irrelevant or simply invented answers). Both Polish and Hungarian legislators ought to consider the problems of explainability and hallucinations when deciding on AI applications in the administration of justice.

Artificial intelligence is additionally a subject of reflection in (the) philosophy (of artificial intelligence) and of interest to the social sciences. Due to its increasingly widespread use in recent years, it has also become the subject of legal debate and normative regulation. The above selected legal dilemmas related to the possibility of using artificial intelligence in the administration of justice have

led to research focusing on the impact of artificial intelligence on the right to a fair trial.

The aim of the research was to test the hypothesis that trustworthy, ethical and human-centred AI can support law enforcement and the administration of justice, thereby contributing to a better fulfilment of the right to a fair trial.

In order to verify such a hypothesis, the first step was to examine how the use of AI in law enforcement and the administration of justice will affect the right to a fair trial. Furthermore, it was examined how the international regulation of AI will affect the legal framework for AI in Poland, and how a legal framework for the use of AI in law enforcement and justice can be created in order to avoid violations of citizens' personal freedoms.

The use of new technologies in the form of artificial intelligence in law enforcement agencies and the judiciary may be controversial, since, on the one hand, new technological solutions can streamline many decision-making processes, however on the other hand, these systems are not infallible, and decision-making with regards to human affairs and freedoms is involved, which hooks into ethical issues. Hence, it is necessary to formulate *de lege ferenda* postulates on the basis of the issues examined within the framework of this project.

The monograph has been divided into five chapters, relating to the legal problems covered by the research.

The first chapter by Zbigniew Więckowski entitled "The right to a fair trial – the Council of Europe perspective. A critical analysis of the Council of Europe guidelines (CAHAI, CDCJ, CEPEJ)" points out that the essential issue remains to determine what role should be played by AI. Is it going to be an exclusively supportive function, such as generating a proposed version of a ruling, or is it to be perhaps the leading one, and AI will give a ruling in routine cases? This question should not only be answered by developers of IT solutions, but also by judges, lawyers, and representatives of society. The development of a human-centric vision of AI should be pursued, whereby it is always the human being who will make the final decisions on the direction of AI development. Moving on to the work of the Council of Europe, the author pointed out that in

the first stage, specialised bodies of the Council of Europe published guidelines, which belong to the domain of soft law. Currently, an international convention is under development and will be open to non-member states. The implementation of this pioneering convention requires Poland's and Hungary's support and encouragement for other nations to participate.

In the chapter titled "Artificial Intelligence as a Legal or Technological Person, and as a Judge?", Emőd Veress notes that the current scientific consensus, if not always legislative, does not justify granting AI entities that are less than an "artificial human" any personality before the law. The arguments of liability and agency lead to an unwanted complexity in situations that can be resolved by other means. During our research, we have determined that the appropriate time for implementing such a solution has not yet arrived, although regulatory frameworks are already being proposed. In addition, it was underlined that research has implications for the present and future of AI regulation, as we have attempted to explore an element of humanity, which is at times overlooked when discussing proposals for the legal personality for AI entities: the substance of the human condition, the material and cognitive preconditions to participating in economic exchange, not just as a holder of rights, but also of obligations, and of action based on not just practical reason, but rational self-interest as a barer of (even existential) risks associated with actions which an AI devoid of concepts and prerequisites of existence in the physical world may not be required to undertake, and for which reason such an AI should never be grated legal person status.

István Ambrus analysed the legal issues concerning AI in the chapter "Substantive Criminal Law and Artificial Intelligence". These issues are the core of the study. After discussing the concepts related to the object of criminal responsibility, he presented the relevant problems related to the concept of criminal offence and criminal sanction and the part of the Penal Code specific to those problems, and then examined the cases in which AI appears as the object of a criminal offence. In addition to the results from related Hungarian, Anglo-Saxon, and German literature, the author has compiled and evaluated the findings of the most recent domestic legal literature

in the context of the discussion. Given his research, it is apparent that Hungarian criminal law was primarily designed in 1878 to prosecute human criminal activity. Thus, simply applying the law's existing concepts to artificial intelligence violations will no longer be adequate in the future. In light of this, crucial components of offence concepts, such as actions and adherence to statutory definitions, may require reassessment. However, the author highlighted several new challenges to criminality and criminal sanctions, which pose difficulties for both criminal law researchers and legislators.

On the 21st of April 2021, the European Commission presented a proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (AI Act) and amending certain legislative acts of the Union. At the fifth and final trilogue in December 2023, the Council and the European Parliament came to an agreement on all political issues and successfully closed the interinstitutional negotiations. The AI Act is a multifaceted piece of legislation that touches on many areas of law, including civil, criminal and administrative law – both substantive and procedural. Algorithms and modern technological tools have the potential to assist the Polish justice system work more efficiently and timeously. They can in addition contribute to the realisation of the right to a fair trial. However, it is important to highlight the several levels at which the interface between justice and artificial intelligence transpire.

Problems related to the use of artificial intelligence in the administration of justice and the treatment of the principles of the use of artificial intelligence in the administration of justice on the basis of the EU regulations that were analysed by Agnieszka Gryszczyńska in the chapter "The impact of the EU Regulation laying down harmonised rules on artificial intelligence on law enforcement and the administration of justice in Poland" and Rafał Wielki in the chapter "The use of artificial intelligence in law enforcement and criminal justice."

Based on their reports, they pointed out that questions should be raised about the limits of the use of artificial intelligence, if only in the area of predictive policing. Yes, such systems can help identify criminogenic situations before a crime is committed, but they

require real-time data analysis, including of citizens who behave in accordance with legal norms, which makes everyone feel like a potential suspect. Since individual freedom is one of the fundamental rights widely accepted and respected in the European legal systems, it is necessary to agree with the calls to ban the use of predictive policing in the EU law and thus in the regulations of individual member states.

The same applies to biometric identification systems used in public spaces for law enforcement purposes. Legislation Proposed at the European Union level indicates that it should be up to the member states themselves to decide on the use of such solutions. In the case of Poland and Hungary, we do not know the intentions of the legislators as to whether they will decide to implement such systems. If they intend to proceed, it will be necessary to regulate this issue in criminal procedure laws, along the lines of those for, and judicial control of, covert actions within investigations.

There is increasing talk of using statistics in the context of scientific evidence in criminal proceedings. While this is a good methodological step in line with current scientific knowledge, a significant obstacle in Polish law is the lack of a law on experts. Despite numerous attempts over the last thirty years, it has not yet been possible to introduce such a regulation into Polish law. The regulation should include an obligation for the expert to provide information on the databases used and the methods used to calculate the strength of the evidence, at the request of the court, the prosecution or the defence. This type of regulation should also be included in the Code of Criminal Procedure.

Important considerations are being made regarding the use of artificial intelligence in the administration of justice. While there is a great potential for AI in administrative activities, which will have a positive impact on the functioning of various units, the use of AI in criminal decision-making seems to be going too far. Examples of the use of various systems in the field of criminal investigation show that, due to technical shortcomings, the level of risk is still too high in relation to the potential benefits. We should therefore propose that predictive algorithms should be banned at the level of the European Union law, for example when considering parole or

sentencing. However, if the use of such algorithms is allowed in the future, the person being analysed should have access to the calculations or source code of the algorithm. The individual in a criminal case must be able to understand the mechanism of decision-making and the factors that have been taken into account.

Artificial intelligence is a fascinating technology that we already use on a daily basis with varying degrees of awareness, but it generally makes it easier for us to operate and use large datasets. However, law enforcement and the judiciary deal with issues of illegal behaviour by individuals, and judgement and decision-making in these cases requires consideration of numerous emotional factors that artificial intelligence cannot. It seems that until we achieve a sufficient level of trust in the new technologies used by public bodies, and in the highest-quality artificial intelligence systems whose margin of error is limited, there is no room for wider use of these systems in the public sphere, as this may reduce citizens' trust in the government. This, in turn, will cause more losses than the potential benefits of improved crime fighting.

However, Agnieszka Gryszczyńska pointed out that in terms of the Public Prosecutor's Office in Poland, which has a central IT system, PROK-SYS, and digitises case files, the use of a machine learning solution for describing scanned documents could be implemented. Presently, the File Digitisation System demonstrates inadequate recognition of the titles (headings) of scanned documents, leading to a highly labour-intensive process for describing documents. Furthermore, in case of incorrect or incomplete document description, searching for scanned documents in the file repository is impossible. The System's poor recognition of document titles is linked to the need for manual document filling and naming, or correcting metadata suggested by the System, which considerably lengthens the file digitisation process. The integration of machine learning into the process of scanning and describing full-text, scanned and OCR-processed documents delivers the capacity to complete all metadata entries automatically, eliminating the need for manual entry each time. Furthermore, this will enhance productivity, reducing the workload of stakeholders and assisting with more efficient document retrieval from the digital file archive.

Moreover, approximately 1.1 million criminal cases are registered each year in Poland. This figure was relatively stable between 2019 and 2023. The majority of criminal cases concern victims, either individuals or institutions. Initiation of criminal proceedings is typically based on a complaint filed by the victim. Given the informatisation of public administration and the availability of certain tools, it would be useful for victims to be able to file a report online, using a form and chatbot support. This allows for quick collection of the necessary information at the initial stage of the procedure. To ensure the right to a fair trial, a virtual assistant based on machine learning could be used.

Finally, AI should be a tool for people and be a force for good in society with the ultimate aim of increasing human well-being. In the area of justice, artificial intelligence, if it is based on trust, can lead not only to a reduction in costs but also to a significant reduction in the length of a trial and the delivery of a verdict.

We believe that the monograph will spark a lively discussion in the scientific world and will have a positive impact and assist in the implementation of solutions based on artificial intelligence in the administration of justice.

Agnieszka Gryszczyńska

DR ZBIGNIEW WIĘCKOWSKI
Cardinal Stefan Wyszyński University in Warsaw | DOI: 10.32041/9788367149372/1

Chapter 1. The Right to a Fair Trial – the Council of Europe Perspective. A Critical Analysis of the Council of Europe Guidelines (CAHAI, CDCJ, CEPEJ)

1.1. Introduction

The Council of Europe has been shaping the human rights protection system for more than seven decades.[1] Although it is a regional organisation, it often inspires global initiatives. In the case of the ongoing next industrial revolution, digital this time, the Council of Europe has consistently supported the vision of a human-centered artificial intelligence system based on human rights. Although the European Convention for the Protection of Human Rights and Fundamental Freedoms (ECHR) was adopted before the term "artificial intelligence"[2] (hereafter "AI"), not the slightest doubt can arise that

[1] Council of Europe is the oldest European political organisation after the Second World War, founded in 1949. 46 states with a population of over 650 million in total have constituted the organisation in 2022. The Council of Europe was established to promote democracy and to protect human rights and the rule of law in Europe. The organisation has created a number of legal instruments known as treaties, conventions, charters, and agreements. The most significant achievement of the Council of Europe is the European Convention on Human Rights passed in 1950 (enables individuals to appeal to the European Court of Human Rights in Strasbourg).

[2] The term "artificial intelligence" appeared for the first time in the text of John McCarthy from 1955: J. McCarthy, M.L. Minsky, N. Rochester, C.E. Shannon, *A proposal for the Darmouth Summer Research Project on Artificial Intelligence*,

the system of rights and freedoms guaranteed by the ECHR incorporates the implementation of AI-based technologies.

One of the fundamental human rights is access to justice. In the ECHR, this is guaranteed by the provisions of Article 6 (right to a fair trial) and Article 13 (right to an effective remedy).[3]

The application of AI in the justice system has long since ceased to be solely a projection from the realm of science-fiction. AI systems assist the justice process in many countries.[4]

However, defining the scope of application of AI requires further discussion. Doubts also surround the emergence of the so-called "black box" effect,[5] i.e., the impossibility of comprehensive analysis of the decision-making process.[6]

1955, http://www-formal.stanford.edu/jmc/history/darmouth.html [accessed on: 4 September 2022].

[3] Article 6 (1) ECHR: "everyone is entitled to a fair and public hearing within a reasonable time by an independent and impartial tribunal established by law. Judgment shall be pronounced publicly but the press and public may be excluded from all or part of the trial in the interests of morals, public order or national security in a democratic society, where the interests of juveniles or the protection of the private life of the parties so require, or to the extent strictly necessary in the opinion of the court in special circumstances where publicity would prejudice the interests of justice." Article 13 ECHR: "Everyone whose rights and freedoms as set forth in this Convention are violated shall have an effective remedy before a national authority notwithstanding that the violation has been committed by persons acting in an official capacity."

[4] Confer: Estonian experiences: https://e-estonia.com/artificial-intelligence-as-the-new-reality-of-e-justice/ [accessed on: 18 August 2022].

[5] R. Girasa, *Artificial Intelligence as a Disruptive Technology. Economic Transformation and Government Regulation*, Pleasantville 2020, p. 4.

[6] See pt 41 of conclusions of the Council of the European Union: "Outcomes of artificial intelligence systems based on machine learning cannot be retraced, leading to a black box effect that prevents adequate and necessary responsibility and makes it impossible to check how the result was reached and whether it complies with relevant regulations. This lack of transparency could undermine the possibility of effectively challenging decisions based on such outcomes and may thereby infringe the right to a fair trial and an effective remedy, and limits the areas in which these systems can be legally used." Council Conclusions "Access to Justice – Seizing the Opportunities of Digitalisation", https://www.consilium.europa.eu/pl/press/press-releases/2020/10/13/digital-justice-council-adopts-conclusions-on-digitalisation-to-improve-access-to-justice/ [accessed on: 7 September 2022]. Also, the European Commission points out in the AI White

The essential issue, however, remains to determine what role should be played by AI. Is it going to be an exclusively supportive function, such as generating a proposed version of a ruling, or is it to be perhaps the leading one, and AI will give a ruling in routine cases? Regardless of the answer given, in my view, the oversight of the justice system must always be carried out by a human. Otherwise, the civilisational and cultural foundation of justice, which is constituted by the autonomy and independence of the judiciary, is going to be disturbed and undermined.[7]

It seems crucial to include the widest possible array of stakeholders in the discussion regarding the future shape of justice. The new system should be created not only by developers of IT solutions, but also by judges, lawyers, and representatives of society. Changes should take place not only in a transparent manner, but also in an evolutionary way.

Given its achievements to date in the area of the development of the human rights protection system, the Council of Europe has a key role to play in ensuring that AI continues to develop in line with its standards.[8] An AI system in the justice system should not be

---

Paper that "The specific characteristics of many AI technologies, including opacity ('black box effect'), complexity, unpredictability and partially autonomous behaviour, may make it hard to verify compliance with, and may hamper the effective enforcement of, rules of existing EU law meant to protect fundamental rights." White Paper On Artificial Intelligence – A European approach to excellence and trust, European Commission, Brussels, 19.02.2020, COM(2020) 65 final, p. 12.

[7] "What distinguish judicial authorities and judges from other entities that exercise power are autonomy and independence. If these are lacking, there will be neither court nor judge", A. Partyk, *Legitim 2.0., czyli o robocie przyszłości… rozstrzygającym spory zachowkowe*, "Studia Prawnicze. Rozprawy i Materiały" 2019, Vol. 2, No. 25, p. 38.

[8] To date, the regulations that have emerged from the Council of Europe system that address some of the challenges of the AI issues are represented by: Convention 108 – protection of personal data, Budapest Convention – combating cybercrimes, Declaration on the Manipulation Capabilities of Algorithmic Processes, Recommendation on the Human Rights Impacts of Algorithmic Systems, the European Ethical Charter for the use of artificial intelligence in judicial systems, Principles on a human-rights compliant use of digital technologies in electoral processes (the Venice Commission), Recommendation on "Technological convergence, artificial intelligence and human rights, smart cities: the

developed without establishing rules protecting citizens from risks of discrimination, invasions of privacy or security breaches. It seems imperative to define the principles of AI liability and remaining legal aspects of the application of this type of system in the courts.[9]

In this chapter, initiatives of the Council of Europe to date in the area of the implementation of the right to a fair trial in the context of the technological revolution will be presented and discussed. The ponderings will be based primarily on selected guidelines, studies, documents produced as a result of the work of the specialised bodies of the Council of Europe: Ad hoc Committee on Artificial Intelligence (CAHAI), European Committee on Legal Co-operation (CDCJ), European Commission for the Efficiency of Justice (CEPEJ). The aim of the analysis is to find an answer to the question of whether AI employed in the administration of justice can contribute to improvement of the level of accessibility to the courts. What characteristics should artificial intelligence exhibit so that it does not violate the rule of law and, above all, citizens' right to a fair trial?

1.2. Output of Ad hoc Committee on Artificial Intelligence (CAHAI)

CAHAI[10] was established by a decision of the Committee of Ministers of the Council of Europe, for the period 11 September 2019–31 December 2021. The purpose of the CAHAI was to examine on the basis of extensive multi-stakeholder consultation the legal framework for the development, design, and application of artificial

challenges for democracy (The Congress of Local and Regional Authorities of the Council of Europe), Unboxing artificial intelligence: 10 measures to protect human rights (the Commissioner for Human Rights)."

[9] K. Yeung, *A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework*, November 9, 2018, MSI-AUT (2018) 05, https://ssrn.com/abstract=3286027 [accessed on: 12 August 2022].

[10] https://www.coe.int/en/web/artificial-intelligence/cahai [accessed on: 15 July 2022].

intelligence, based on the Council of Europe standards on human rights, democracy, and the rule of law.[11]

CAHAI work resulted in three key documents:

1) Feasibility study on a legal framework on AI design, development, and application based on CoE standards (accepted in December 2020),
2) Towards regulation of AI systems (accepted in December 2020),[12]
3) Possible elements of a legal framework on artificial intelligence, based on the Council of Europe's standards on human rights, democracy, and the rule of law (accepted in December 2021).[13]

[11] The CAHAI is composed of: representatives of 47 Member States, appointed by their respective governments, who have recognised expertise in digital governance and the legal implications of various forms of AI; representatives of observer states; representatives of other Council of Europe bodies, in particular the Secretariat of the Parliamentary Assembly, the Office of the Commissioner for Human Rights and the intergovernmental commissions dealing with AI issues. Human Rights and intergovernmental commissions dealing with AI issues; representatives of other international and regional organisations operating in the field of AI; representatives of the private sector, including companies and associations with which the Council of Europe has exchanged letters concerning the partnership with digital businesses; representatives of civil society, research and academic institutions who have been admitted by CAHAI as observers. For more: https://www.coe.int/en/web/artificial-intelligence/cahai#{%2266693418%22:[0]}; https://rm.coe.int/list-of-cahai-members-web/16809e7f8d [accessed on: 18 July 2022].

[12] I. Ben-Israel, J. Cerdio, A. Ema, L. Friedman, M. Ienca, A. Mantelero, E. Matania, C. Muller, H. Shiroyama, E. Vayena, *Towards Regulation of AI Systems, Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law*, Prepared by the CAHAI Secretariat, Compilation of contributions DGI (2020)16, December 2020.

[13] *Ad hoc Committee on Artificial Intelligence (CAHAI) Progress report*, 23.09.2020, https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=090000168091ed062 [accessed on: 21 July 2022]; Feasibility study on a legal framework on AI design, development and application based on CoE standards, adopted by the CAHAI on 17 December 2020, https://rm.coe.int/cahai-2020-23-final-eng-feasibility-study-/1680a0c6da [accessed on: 21 February 2021]; Publication: Towards regulation of AI systems, December 2020, https://rm.coe.int/prems-107320-gbr-2018-compli-cahai-couv-texte-a4-bat-web/1680a0c17a [accessed on: 21 July 2022].

The first of the aforementioned documents was preceded by the CAHAI report adopted by the Committee of Ministers of the Council of Europe on 23 September 2020.[14] In the report, not only the opportunities but also the risks posed by the development of AI are listed. In addition, the importance of the joint establishment of guidelines for AI certification by independent bodies, regulating the status of giga-data specialists, developing a catalogue of ethical principles along the lines of the Hippocratic Oath was highlighted[15] as was the need for the creation of a document to validate driving licenses for autonomous vehicles.

A feasibility study on a legal framework for the design, development, and application of AI based on Council of Europe standards was adopted at the third CAHAI plenary in December 2020 (Feasibility study on a legal framework on AI design, development and application based on CoE standards).[16]

In the absence of a univocal definition of artificial intelligence, CAHAI opted to attempt to define it. It seems to be a reasonable approach, though in my opinion too idealistic, that it was assumed that in process of defining AI, it was necessary to find a balance between a definition that would not be too precise and could become obsolete in a short period of time, and a definition that would not leave too much room for interpretation, which in turn could result in ambiguous, uneven applications of AI.[17]

---

[14] https://search.coe.int/cm/Pages/result_details.aspx?ObjectID=090000168090ed062 [accessed on: 18 July 2022].

[15] Also: L. Felländer-Tsai, *AI ethics, accountability, and sustainability: revisiting the Hippocratic oath*, "Acta Orthopaedica" 2020, Vol. 91, issue 1, DOI: 10.1080/17453674.2019.1682850; D. Talby, *Healthcare AI does not need a new Hippocratic Oath*, "Forbes" 2020, May, https://www.forbes.com/sites/forbestechcouncil/2020/05/22/healthcare-ai-does-not-need-a-new-hippocratic-oath/?sh=12c34a541752 [accessed on: 18 July 2022].

[16] CAHAI, Feasibility study on a legal framework on AI design, development and application based on CoE standards, December 2020; hereinafter: Feasibility study…

[17] The only definition of AI to date is the one provided by CEPEJ and incorporated into the Ethical Charter: "A set of scientific methods, theories and techniques whose aim is to reproduce, by a machine, the cognitive abilities of human beings", https://www.coe.int/en/web/cepej/cepej-european-ethical-charter -on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment [accessed on: 21 June 2022].

The Feasiblity study… indicates the opportunities that artificial intelligence brings, above all, in achieving the sustainable development goals outlined by the UN. The study analyses the impact of artificial intelligence on the democratic system. AI, on the one hand, may assist in accessing information, catalysing public discussions, on the other hand, it may lead to the spread of disinformation, propaganda, control of citizens, etc. AI also has an impact on the rule of law. Applied in the reasonable way, AI can help make government more effective through more efficient justice and administration. AI-based systems may assist in eradicating corruption, and in detection and defence against cyber-attacks.

The ground-breaking solution of the commencement of development of an international convention covering AI issues was proposed in the Feasibility study. The enactment of a convention based on existing Council of Europe standards (democracy, rule of law, human rights) would be a globally unique initiative. Indeed, to date, examples of soft law initiatives predominate. It is planned that the system under development will take into consideration such fundamental principles as: a) human dignity, b) the right to be informed about communication with AI and not with another human being; c) minimizing harm caused by AI; d) preservation of human autonomy in the full cycle of AI development; e) counteracting the discriminatory nature of AI.

The revolutionary importance of AI is strongly emphasised in the CAHAI reports. Its potential to take autonomous action and, therefore, its ability to influence almost every area of human life is seen as essential. Leaving the development of AI without human oversight and the definition of a legal framework for its operation could pose a significant threat to the established system of values and principles. In analyses submitted to CAHAI,[18] it is emphasised that ethical standards are not global and are context-dependent. The sole permissible point of reference for a universal artificial intelligence system is human rights.

[18] A. Mantelero, *Analysis of international legally binding instruments. Final report*, [in:] I. Ben-Israel, J. Cerdio, A. Ema, L. Friedman, M. Ienca, A. Mantelero, E. Matania, C. Muller, H. Shiroyama, E. Vayena, *Towards Regulation…*, *op. cit*.

In CAHAI's view, the principal areas of human activity affected by AI from the perspective of the realisation of human rights are personal data,[19] democratic systems,[20] and the administration of

[19] CAHAI highlights the need to implement appropriate legal safeguards to counteract the identifiability of anonymized data. Developers of AI systems should critically analyse the quality and origin of the personal data employed. The use of synthetic data should be considered as a way to minimise the amount of data processed by the AI system. The Council of Europe encourages AI developers to consult their enterprises with independent committees of experts representing a broad spectrum of research areas (cooperation may also involve research institutions e.g. universities) to help identify potential risks and create a human rights-based AI system. Individuals as well as groups of people who would be affected by the implemented applications should be involved in the risk analysis process. All AI-based products and services must allow for human control over them. The AI system should be developed in such a way that it is possible to replicate each stage of the product life cycle (transparency of the process).

[20] The right of citizens to participate in public affairs is encompassed in the broad concept of "public affairs" which includes dialogue with stakeholders and public debate. Both are linked to the right to freedom of expression, assembly, and association. AI systems may assist in developing public/civic engagement. However, the prerequisite for further development is to ensure system transparency, universal accessibility, and interoperability between multiple services and platforms. AI systems applied for public purposes should be continuously audited and the results made publicly available. A vital issue is the impact of the AI system on the electoral process. The problem should be analysed in two ways: a) e-voting, predicting the outcome, and b) targeting, profiling, propaganda, and generating false information. The first area is not controversial, while the second raises serious questions (more: A. Beatrice, *Driving political campaigns with artificial intelligence technologies*, Analytics Insight, 8 September 2020, https://www.analyticsinsight.net/transforming-political-campaigns-artificial-intelligence-technologies/ [accessed on: 18 July 2022]). A decisive response is needed to the problem of spreading propaganda and disinformation. Disinformation phenomena are on the rise, which is particularly evident throughout the period of the fight against the COVID-19 pandemic and the war of the Russian Federation against Ukraine. International organisations, including the EU and the Council of Europe, are taking specific steps to counter the phenomenon of spreading misinformation: https://www.consilium.europa.eu/en/policies/coronavirus/fighting-disinformation/; https://ec.europa.eu/info/live-work-travel-eu/coronavirus-response/fighting-disinformation_en; https://www.coe.int/en/web/freedom-expression/news/-/asset_publisher/thFVuWFiT2Lk/content/tackling-disinformation-in-the-global-media-environment-new-council-of-europe-report?_101_INSTANCE_thFVuWFiT2Lk_viewMode=view/ [accessed on: 18 August 2022].

justice. Given the subject of the study, it is worth paying more attention to the latter. Unlike human decision-making, in particular legal adjudication decisions, the logic of artificial intelligence is not based on legal reasoning, but on mathematical analysis.[21] The aforementioned factors heighten the risk of possible errors in the decision-making process in cases requiring legal decisions. In addition, the consequences of such a mistake when it affects the sphere of freedom of a particular person are much more severe than in the case of a mistake concerning other spheres of life. It is worth adding that in the case of a human error, the consequences usually affect individuals. An improperly designed AI algorithm may lead to discrimination against many people in the same or similar situations.

The specific nature of the cases decided by the courts makes it impossible to entrust them all to be decided by an AI system. The application of AI assistance by the court (for example, to analyse documents), should be known to the parties. Full transparency should apply to the indication of the data sources employed by AI for training purposes.

The necessity of providing continuity of work on a legal framework for an artificial intelligence system has meant that the tasks of CAHAI have been taken over by the standing committee on artificial intelligence (Committee on Artificial Intelligence – CAI[22]). The first meeting of the CAI was held on 4–6 April 2022, and was strictly organisational in nature. The next one is scheduled for the end of September 2022. In addition to its many ancillary tasks, the primary objective of the CAI is to develop an international convention on AI.

[21] For instance, mediation conducted between parties is based on psychological elements such as guilt or motivations to act. An AI system will never have emotional intelligence.

[22] The composition of the CAI's personnel is identical to that of the CAHAI.

1.3. Output of European Committee on Legal Co-operation (CDCJ) concerning the impact of AI on justice

The activities undertaken by the two committees – CAHAI & CAI – have dealt with general issues related to artificial intelligence (health, social issues, democracy, ethics). CDCJ is a specialised body, established in 1963, with wide-ranging competences in public and private law. The CDCJ's tasks include: drafting conventions, protocols, and guidelines, adopting opinions on legal matters, proposing standards on the protection of personal data and the right to private life. On the subject of issues involving artificial intelligence, CDCJ published two pivotal documents:

1) Guidelines on online dispute resolution (ODR – ZW) mechanisms in civil and administrative court proceedings (June 2021) plus Explanatory Memorandum to the Guidelines of the Committee of Ministers of the Council of Europe on online dispute resolution mechanisms in civil and administrative court proceedings (hereinafter referred to as: ODR guidelines);
2) Guidelines on electronic evidence in civil and administrative proceedings (January 2019) plus Explanatory Memorandum (hereinafter: guidelines on electronic evidence).

The ODR Guidelines are based on the principles developed in the case law of the European Court of Human Rights on the basis of Articles 6 and 13 ECHR.[23] The choice of issues addressed in the Guidelines is consistent with the principle that the provisions of the

[23] Guide on Article 6 of the European Convention on Human Rights, Right to a fair trial (civil limb), updated on 31 August 2019, Council of Europe/European Court of Human Rights, 2019. See also: D. Vitkauskas, G. Dikov, *Protecting the right to a fair trial under the European Convention on Human Rights: A handbook for legal practitioners*, Council of Europe 2017, https://rm.coe.int/protecting-the-right-to-a-fair-trial-under-the-european-convention-on-/168075a4dd [accessed on: 15 August 2022].

ECHR should be interpreted taking into account the economic and social conditions currently prevailing.[24]

The dynamic digitisation of courts of general jurisdiction is of essential importance in terms of ensuring access to justice. The commonplaceness of remote solutions provides an opportunity for persons with disabilities, the elderly, and those living in localities situated far from the court premises to actively participate in the justice system.

Extensive justification for the ODR Guidelines is provided in the official commentary to them, i.e., the *Explanatory Memorandum.* It indicated that the introduction of AI tools in civil and administrative proceedings allows for automated decisions to be taken,[25] acceleration of proceedings, and reaching more predictable and fair dispute settlements.[26] Many countries are already applying AI tools to anonymise court rulings or translate documents. New AI tools may assist judges in other activities, such as advanced data analytics.[27] In some cases, subject to the reconstruction of the national civil procedure, it may even be possible to consider replacing the judge with an IT system for processing and analysing data. In some cases, providing that the reconstruction of the national civil procedure will take place, it may even be possible to consider replacing the judge with an IT system for processing and analysing data.[28] The increasing employment of AI tools in the courts should be taken into account in basic procedural rules.[29]

---

[24] Marckx v. Belgium, 13 June 1979, § 41, Series A No. 31; Tyrer v. the United Kingdom, 25 April 1978, § 31, Series A No. 26.

[25] D. Carneiro *et al.*, *ODR: an Artificial Intelligence Perspective*, "Artificial Intelligence Review" 2014, Vol. 41, pp. 211–240.

[26] M. Scherer, *Artificial Intelligence and Legal Decision-Making: The Wide Open?*, "Journal of International Arbitration" 2019, Vol. 36, No. 5, pp. 539–574.

[27] S. Samoili, M. López Cobo, E. Gómez, G. De Prato, F. Martínez-Plumed, B. Delipetrev, *AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*, EUR 30117 EN, Publications Office of the European Union, JRC118163, Luxembourg 2020, pp. 7–8.

[28] J. Gołaczyński, *E-sąd przyszłości*, "Monitor Prawniczy" 2019, Vol. 2, p. 97, DOI: 10.32027/MOP.19.2.7.

[29] See also E. Nissan, *Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement*, "AI

The basic principles, defined in the Guidelines should be respected by Member States, legislators, courts, manufacturers, and service providers of digital solutions in order to ensure that the technologies deployed do not violate human dignity, human rights, and fundamental freedoms.[30]

The ODR Guidelines include technologies such as: (1) online filing systems/platforms that are accessible to parties and/or their representatives for the purpose of making procedural submissions, (2) online systems for storing, processing, and evaluating electronic evidence, (3) artificial intelligence, big data analytics techniques, and automation, to the extent that they have an impact on court proceedings, (4) platforms for online court hearings and online trials, such as audio and video conferencing, including oral testimony by witnesses and experts.

The term 'ODR' refers to a technology or a mechanism employed to resolve legal disputes that are conducted at a distance using computers, including mobile devices and the Internet. The definition is general and vague. Further clarification may be found in the Explanatory Memorandum to the Guidelines. According to it, ODR is not a dispute resolution method per se, but rather a technology or a mechanism that is used within existing court proceedings. It is therefore not a new type of procedure and is not an alternative

& SOCIETY" 2017, Vol. 32, pp. 539–574; M. Scherer, *Artificial Intelligence and Legal Decision-Making: The Wide Open?*, "Journal of International Arbitration" 2019, Vol. 36, No. 5, pp. 539–574.

[30] Guidelines represent not a "hard" but a "soft" law instrument. Their purpose is not to establish binding legal standards. They serve as a practical "toolkit" for Member States to ensure that the practice of national digital courts complies with the requirements of Articles 6 and 13 ECHR. The Guidelines are intended to provide practical advice and guidance to Member States. The Guidelines are an example of modern regulation of digital tools in the administration of justice. If Member States follow the Guidelines, they can reasonably expect that the information technology they have implemented will not be challenged under the ECHR. The Guidelines are the result of collaboration and exchange of experiences between Member States. Both successes and failures of individual IT implementations in the most experienced Member States were taken into account during the preparation of the document.

to any court proceedings. ODR merely provides new ways to access or implement existing types of court proceedings.[31]

ODR represents different technology from artificial intelligence, and not all ODR technologies include AI components. ODR is a broader concept encompassing all types of online dispute resolution mechanisms, including automation tools that do not necessarily rely on AI tools.[32] This distinction between ODR and AI is correctly retained in the Council of Europe guidelines. However, while the requirements to meet the judicial guarantees under the ECHR apply to all ODR techniques, some issues are more relevant to AI. This applies in particular to the possibility of fully automating the decision-making process (without human intervention), as well as the possibility of reviewing or overturning these decisions. The guidelines provide for a number of separate rules dedicated to AI tools.

The ideal ODR mechanism should be characterised by speed, simplicity, egalitarianism, inexpensiveness, and efficiency. The physical presence of the parties at the hearing should not be required. To ensure universal accessibility, for both developed and developing countries, ODR should include elements of online and offline dispute resolution.[33] Reduced costs, shortened time to resolve a dispute, and no need for the parties to meet directly are undeniable advantages of ODR. The application of the ODR mechanism should be governed by the principles of: independence of the determining authority; transparency of the proceedings; adversarial nature of the dispute; efficiency of the proceedings; legality of the decision, the liberty of the parties and the right to representation of the parties.[34] At

[31] P. Loutocký, *Online dispute resolution and the latest development of UNCITRAL model law*, [in:] *Cofola International 2015: Current challenges to resolution of international (cross-border) disputes. Conference proceedings*, K. Drličková (ed.), Brno 2015, pp. 243–256.

[32] D. Carneiro, P. Novais, F. Andrade *et al.*, *Online dispute resolution: an artificial intelligence perspective*, "Artificial Intelligence Review" 2014, Vol. 41, https://doi.org/10.1007/s10462-011-9305-z, pp. 211–240.

[33] K. Karasiński, *Online Dispute Resolution*, "Przegląd Prawa Handlowego" 2016, No. 8, pp. 41–46.

[34] J. Mucha, *Alternatywne metody rozwiązywania sporów konsumenckich w prawie unijnym – nowe rozwiązania prawne (dyrektywa 2013/11/UE w sprawie ADR oraz rozporządzenie nr 524/2013 w sprawie ODR)*, IKAR 2014, No. 4, p. 7.

this point, it should be mentioned that it is of utmost importance to strengthen the confidence of citizens in the ODR mechanisms used by the courts. Additionally, the idea of creating a new dispute resolution system could work particularly well in those Member States where the justice system is not working efficiently. ODR methods could then indeed be an important alternative to protracted court proceedings and judgements of questionable quality.[35]

ODR is not intended to fully replace the existing judicial model, but rather to complement it and create additional opportunities for access to justice. ODR should be viewed as a kind of support for judicial decision-making and as a facilitator of the judge's work, not as a constraint.[36] ODR must also be tailored to the needs of judges and other users and should never infringe on procedural guarantees and rights, such as, *inter alia*, the right to a fair hearing by a court of competent jurisdiction.

ODR contributes to more effective and efficient access to justice. However, the main obstacle to the wider use of ODR is access to technology. A certain part of the population does not have the necessary skills (lack of familiarity with the application) or capacity (lack of network access or lack of a device) to use ODR and resolve disputes remotely. This problem is called "digital exclusion" and Member States should take this into account when developing ODR. For example, authorities could set up support points in court buildings or legal aid offices.

Parties should be notified of the intention to process their case through AI-based ODR. In particular, parties to the proceedings have the right to be informed of the justification of the AI-based processing operations applied to them. This information shall also

---

[35] *Ibidem*.

[36] A similar conclusion follows from pp. 58 and 60 of the combined cases C-317/08 – C-320/08, Rosalba Alassini and Filomena Califano v. Wind SpA, Lucia Anna Giorgia Iacono v. Telecom Italia SpA and Multiservice Srl v. Telecom Italia SpA, ECJ judgment of 8 March 2010. J. Kleinberg *et al.*, *Human Decisions and Machine Predictions*, National Bureau of Economic Research, February 2017, https://www.cs.cornell.edu/home/kleinber/w23180.pdf; S. Wachter *et al.*, *Transparent, explainable, and accountable AI for robotics*, "Science Robotics" 2017, Vol. 2, No. 6, http://robotics.sciencemag.org/content/2/6/eaan6080.full.

include the consequences of the tool used.[37] In essence, this is a requirement for transparency formulated by numerous international organisations.

The parties must be provided with justification for decisions made with AI tools. Decisions that make it impossible to see how the result was achieved are as much a threat to transparency and the principle of due process as decisions that do not contain a statement of reasons at all.[38] Parties are entitled to an explanation of the processing operations applied to them. This should include the consequences of such reasoning. If, due to the nature of the AI tool used, no information can be provided, courts should refrain from issuing decisions made with AI whose reasoning results cannot be reproduced.

According to the Council of Europe, adjudications based on AI tools should be subject to review. This issue is controversial insofar as it suggests the possibility of AI tools replacing the judge. Is this justified in light of the ECHR and the position of other international organisations? It would seem so. Indeed, in adopting the Ethical Charter on the Use of Artificial Intelligence in Judicial Systems, the Council of Europe stressed the importance of the principle of "under the control of the user," i.e., "preclude a prescriptive approach and ensure that users are informed actors and in control of their choices."[39]

The main issue is the method by which automated adjudications should be verified. The Guidelines do not provide a solution to this problem. This issue is going to become essential when ODR instruments start to take the shape of fully automated decision-making tools. The key provision is Article 13 ECHR. In this regard, it seems that parties should not only be able to challenge decisions made in a fully automated manner, but also to request that such a review

[37] J. Burrell, *How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms*, "Big Data & Society" 2016, Vol. 3, No. 1, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2660674 [accessed on: 18 August 2022].

[38] Samek *et al.*, *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models*, "ITU Journal: ICT Discoveries" 2017, Special Issue, No. 1, pp. 1–10.

[39] *European Ethical Charter…*, *op. cit.*

is carried out by a human judge. The European Court of Human Rights does not specify at what level of case law this remedy is to be exercised.

Two approaches may be distinguished: the control should be carried out in the same instance or higher. Thus, we see that the use of AI tools may open up new possibilities for redress in cases of infringements in national court systems. Given the exceptional nature of ODR, a Member State may decide, irrespective of the existing mechanisms for challenging adjudication, to establish an additional means of redress. Alternatively, a Member State may leave the possibility to review automated adjudications according to existing appeal mechanisms.

Another issue that is not fully addressed in the guidelines is the design of ODR with AI tools. Developers of AI technologies should strive to better comprehend the justice system. Collaborating with judges and court staff ensures that the ODR architecture meets the needs of society.[40]

Despite its almost comprehensive nature, it seems reasonable to formulate some *de lege ferenda* conclusions. The Guidelines employ the term ODR both in their title and in their content, which is confusing for readers, in particular judges. There is a risk that judges will not even attempt to read the Guidelines, assuming that they refer to ADR proceedings. Perhaps a better wording would be "cyberjustice" or "digital courts."

Some of the basic principles in the Guidelines were formulated incorrectly. For example, the focus of the third principle should be on the particularities arising from the specific application of ODR and its potential impact on procedural issues. The Guidelines should address cyber-security to a greater extent (consistency with EU standards in this regard would make sense). The Guidelines should take greater account of the needs of the judiciary arising from the use of AI tools, rather than electronic communication itself,

[40] See Guidelines on how to drive change towards Cyber justice [Stock-taking of tools deployed and summary of good practices] of 7 December 2016, European Commission for the Efficiency of Justice, CEPEJ(2016)13.

which is already the standard.[41] An important condition for creating modern courts is to be able to process as much data as possible and to allow the court to create digital data (including electronic protocols, digitisation of all documents).[42] It is important to decide clearly at the national level whether AI tools should only prepare a draft judgment with justification and the final decision can only be taken by a human judge or whether, in some cases, e.g., based on simple factual circumstances, a full replacement of the judge will be considered acceptable. The latter solution requires a detailed analysis as to whether, in such a case, we would still be dealing with a court within the sense of the ECHR and other provisions.

The Committee of Ministers of the Council of Europe on 30 January 2019 adopted guidelines on electronic evidence in both civil and administrative proceedings (electronic evidence guidelines). Their aim is to provide practical guidance to Member States on the employment of electronic evidence in civil and administrative proceedings. The guidelines represent an important stage in the process of adapting the judiciary to the IT revolution in the administration of justice. The guidelines organise the legal terminology on electronic evidence.[43] The guidelines address, among other things: the rules for the use, collection, storage, and archiving of electronic evidence. It also raises the issue of increasing public awareness of the importance of electronic evidence and the need for training in this area in Member States.

The CDCJ guidelines serve to increase the confidence of judges and other legal practitioners in the use of cloud-based information technology (cloud computing). With regard to the effective preservation of electronic evidence, blockchain technology is considered to be the optimal choice.

When undertaking the evaluation of electronic evidence, the three main principles outlined in the CDCJ guidelines should be followed: a) it is the role of the court to determine the relevance of the electronic evidence in question (in particular, this decision

[41] J. Gołaczyński, *E-sąd…*, *op. cit.*, p. 98.
[42] *Ibidem*.
[43] The terms of electronic evidence, metadata and trust services were defined.

should not be delegated to an IT expert), b) the principle of neutrality of electronic evidence implies that it should not be discriminated against, as well as privileged over other means of evidence, c) the parties should be treated equally, this means, *inter alia*, that the authenticity of electronic evidence should be open to challenge.

The methods used by courts to examine witnesses at a remote hearing should protect the transmission of video or audio from loss of data, distortion or unauthorised disclosure. As far as technically possible, remote evidence should be conducted in the same manner as it is conducted in court.

Where the testimony requires confidentiality, it is necessary to use measures or technical solutions limiting access only to authorised persons.[44]

Courts should be aware of the potential evidentiary value of metadata. Electronic evidence should be presented in its original form. Metadata present in the original (digital) version of electronic evidence can provide the context necessary to properly evaluate the evidence.

Courts should follow the CDCJ's guidance on procedures for managing the collection, preservation, and archiving of electronic evidence. Electronic evidence requires special precautions because of the ease with which it can be altered, damaged or destroyed through improper handling. The collection and storage of electronic evidence requires Council of Europe member states to adopt specific tools and procedures to ensure its integrity, confidentiality and security.

In the case of electronic evidence, there is an increased risk of generating unnecessary amounts of data due to the ease of obtaining it. This may hinder or even prevent effective preparatory inquiry. It is therefore important to apply the principle of proportionality.

Courts should take a proactive approach to protect the integrity of electronic evidence from cyber threats, including damage or unauthorised access. Unauthorised persons should not have access

[44] For security reasons, the communication systems employed, both public and private, should provide encryption of the video signal to protect it from interception, by unauthorised persons.

to electronic evidence. Stored electronic evidence may be linked to standardised metadata.

The CDCJ guidelines regulate data migration, which involves changing the storage medium in order to maintain accessibility to electronic evidence. Neglecting to properly supervise the migration process can result in unreadable data.

It is recommended that when handling cross-border electronic evidence, the courts should work closely together on this issue, taking into account the existing output of EU regulations in this area.

Optimisation of the transfer of electronic evidence by electronic means may be achieved by implementing common technical standards and file formats and by digitising national judicial and administrative systems.

Awareness of electronic evidence should be promoted among judges and other legal professionals.

In the forthcoming revision of the CDCJ guidelines, a *de lege ferenda* proposal is to define the terms "blockchain" and "cloud computing," given their close relationship with electronic evidence and their increasing importance in legal transactions.

1.4. The output of the European Commission for the Efficiency of Justice (CEPEJ) regarding the impact of AI on justice

The CEPEJ was established on 18 September 2002 by Resolution Res(2002)12 of the Committee of Ministers of the Council of Europe. The purpose of the CEPEJ is to improve the efficiency and functioning of the administration of justice in Council of Europe member states. The CEPEJ's task is primarily to collect and analyse data, develop benchmarks, draft reports, guidelines, action plans, and develop contacts with external stakeholders. One of CEPEJ's important tasks is to conduct research on improving the efficiency of the judiciary through the use of information technology (IT) solutions. On the one hand, the new digital possibilities appear as an opportunity to improve efficiency, on the other hand, they pose a challenge to respect the principles developed so far (including

the adversarial principle, protection of fundamental rights and freedoms, the role of the judge).

The CEPEJ (in December 2019), decided to establish a new working group: Working Group on Cyber Justice and Artificial Intelligence. The CEPEJ entrusted the group with the task of developing solutions for the application of artificial intelligence mechanisms and other digital solutions in the justice system, in order to improve its efficiency and quality. The group's work should be carried out in coordination with other structures in this field, in particular the European Committee for Legal Cooperation (CDCJ) and CAHAI. The CEPEJ-GT-CYBERJUST has been tasked with developing training programs in the field of cyberjustice and artificial intelligence.

CEPEJ's output to date is impressive:

1) European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment – (December 2018),
2) Feasibility study on the establishment of a certification mechanism for artificial intelligence tools and services (in the sphere of justice and judiciary) – (December 2020),
3) Guidelines on videoconferencing in judicial proceedings (June 2021),
4) Guidelines on electronic court filing (e-filing) and digitalisation of courts (December 2021),
5) Revised roadmap for ensuring an appropriate follow-up of the CEPEJ Ethical Charter on the use of artificial intelligence in judicial systems and their environment (December 2021).

European Ethical Charter on the use of artificial intelligence (AI) in judicial systems and their environment was the first European text setting out ethical principles relating to the use of artificial intelligence (AI) in judicial systems. The Charter provides a framework of principles that can guide policy makers, legislators and justice professionals when they grapple with the rapid development of AI in national judicial processes.

The five principles of the Ethical Charter are:

1) respect for fundamental rights: ensure that the design and implementation of artificial intelligence tools and services are compatible with fundamental rights;

2) non-discrimination: specifically prevent the development or intensification of any discrimination between individuals or groups of individuals;
3) quality and security: with regard to the processing of judicial decisions and data, use certified sources and intangible data with models elaborated in a multi-disciplinary manner, in a secure technological environment;
4) transparency, impartiality, and fairness: make data processing methods accessible and understandable, authorise external audits;
5) "under user control": preclude a prescriptive approach and ensure that users are informed actors and in control of the choices made.

At its plenary meeting on 16–17 June 2021, the CEPEJ adopted new guidelines on the conduct of remote hearings in judicial proceedings. The guidelines contain the basic principles that States should follow to ensure that remote hearings comply with the right to a fair trial guaranteed by Article 6 of the Convention and Convention 108+.

The guidelines define videoconferencing as a system that allows bidirectional and simultaneous transmission of image and sound providing audio and verbal interaction during a remote hearing.

According to the Council of Europe, Member States should provide instruction to judges, court staff and legal practitioners regarding the IT solutions applied in the courts and related international human rights standards. Training should be continuous and compulsory for legal practitioners. It is also necessary to supplement the curriculum of legal studies with elements related to the use of technological solutions in courts.

The national legal framework allowing courts to conduct hearings remotely could be clarified through soft law instruments such as recommendations or guidelines, based on CEPEJ guidance.

The parties should be fully at liberty to consult the court on specific technical issues relating to the conduct of the remote hearing, to receive detailed information, to share their concerns about the security of the remote connection.

The parties should retain the possibility to consult the court on specific issues related to the conduct of the remote hearing, to obtain detailed information, and to share their concerns about the security of the remote connection.

Moreover, a court should enable the participants of the remote hearing to check the audio and video quality before the hearing commences. The court should have the possibility to continuously monitor the video and audio quality during the remote hearing. The court should ensure that the transmission is visible and audible to the participants and to the public where the proceedings are public.

The participants of the remote hearings should be identified properly by the court. Means of identifications should not be invasive or burdensome.

The court should preserve the public character of the remote hearing by allowing the participation of the public. The participation of witnesses and experts in remote hearings should correspond to the practice adopted in traditional hearings. The public nature of the remote hearing can be ensured, e.g., by providing for the public to attend the remote hearing in real time or by posting relevant recordings on the court's website.

The court should provide guidance to participants in the proceedings on the procedure for the presentation of evidence or other material during the remote hearing.

Furthermore, the organisation of a remote hearing in criminal proceedings should be based on values such as the protection of public order, public health, the prevention of crime, and the protection of the right to life, liberty, and security of witnesses and victims of crime. The video link provided should allow the accused to see and hear the participants in the remote trial, including the judge, witnesses, and experts. Participants in the proceedings should be able to see and hear the accused. Before the start of the trial, the court should inform the accused how he or she should report technical malfunctions. The accused should retain free access to their legal representative before and during the remote hearing, including the right to confidential conference before the start of the hearing.

Adequate financial resources need to be allocated to ensure that videoconferencing is organised appropriately and effectively so that

remote hearings imitate traditional hearings as closely as possible, including by ensuring that all participants in the hearing are able to communicate in a fully intelligible manner. The conduct of remote hearings should be based on the principles of fairness, efficiency, speed of proceedings, cooperation, security, and legality of the processing of personal data.

The court should provide participants with accessible instructions or tutorials on videoconferencing and remote hearings. It is advisable to prepare briefing materials not only as printed texts but also as short films.

Adequate measures should be taken in advance to mitigate the risk of security breaches of court infrastructure, in particular possible cyberattacks on videoconferencing hardware and software. Courts should develop procedures to cover emergency situations such as sudden technical failures, power outages or data security breaches.

Finally, videoconferencing hardware and software should meet minimum technical standards to facilitate interoperability of the solutions employed and to reduce delays in video and audio transmission. Judges, parties, court staff, and other participants should have immediate access to IT support during remote hearings to avoid delays and technical difficulties when using the videoconferencing system.

The *de lege ferenda* proposal is to delineate specific steps that have to be taken regarding persons threatened with digital exclusion.

1.5. Conclusions

The digital revolution is no longer a projection of the future. The already commenced process of rapid changes no longer can be halted. New digital technologies are transforming almost every area of our lives as well as the sphere of justice. Holding back from taking any action in the face of the revolution taking place does not seem to be a reasonable solution. Moreover, allowing the unfettered development of AI mechanisms may constitute a risk, the consequences of which are difficult to assess today.

The initiatives taken by national governments and international organisations, including the Council of Europe, towards organising the development of AI are necessary. It should be considered particularly essential that the digital revolution does not lead to the erosion of human rights. The development of a human-centric vision of AI should be pursued, whereby it is always the human being who will make the final decisions on the direction of AI development. One of the fundamental human rights is access to justice. AI may, on the one hand, vastly enhance access. On the other hand, it may lead to unfair judgements made by algorithms. Therefore, any legislative initiatives, both at the national and international levels, should focus on how AI can be helpful in facilitating access to justice. The work of the Council of Europe seems to be moving in the right direction. In the first stage, specialised bodies of the Council of Europe published guidelines that belong to soft law domain. Currently, an international convention is being developed that will be open not exclusively to member states. The enactment of an international convention will be of a pioneering nature. Poland and Hungary should support the process of passing such Convention and encourage other countries to join. Solely in the case of domination of human-centric vision of AI, based on the well-established standards of the Council of Europe, citizens' right to a fair trial will not only be safe from any potential threats but, due to AI capabilities, even strengthened.

REFERENCES

Beatrice A., *Driving political campaigns with artificial intelligence technologies*, Analytics Insight, 8 September 2020, https://www.analyticsinsight.net/transforming-political-campaigns-artificial-intelligence-technologies/.

Ben-Israel I., Cerdio J., Ema A., Friedman L., Ienca M., Mantelero A., Matania E., Muller C., Shiroyama H., Vayena E., *Towards Regulation of AI Systems, Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and*

the rule of law, CAHAI Secretariat, Compilation of contributions DGI (2020)16, December 2020.

Burrell J., *How the Machine 'Thinks': Understanding Opacity in Machine Learning Algorithms*, "Big Data & Society" 2016, Vol. 3, No. 1, https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2660674.

Carneiro D. *et al*., *ODR: An Artificial Intelligence Perspective*, "Artificial Intelligence Review" 2014, Vol. 41.

Carneiro D., Novais P., Andrade F. *et al*., *Online dispute resolution: An artificial intelligence perspective*, "Artificial Intelligence Review" 2014, Vol. 41, https://doi.org/10.1007/s10462-011-9305-z.

Felländer-Tsai Li, *AI ethics, accountability, and sustainability: Revisiting the Hippocratic oath*, "Acta Orthopaedica" 2020, Vol. 91, issue 1, DOI: 10.1080/17453674.2019.1682850.

Girasa R., *Artificial Intelligence as a Disruptive Technology. Economic Transformation and Government Regulation*, Pleasantville 2020.

Gołaczyński J., *E-sąd przyszłości*, "Monitor Prawniczy" 2019, Vol. 2, DOI: 10.32027/MOP.19.2.7.

Karasiński K., *Online Dispute Resolution*, "Przegląd Prawa Handlowego" 2016, No. 8.

Loutocký P., *Online dispute resolution and the latest development of UNCITRAL model law*, [in:] *Cofola International 2015: current challenges to resolution of international (cross-border) disputes: conference proceedings*, K. Drličková (ed.), Brno 2015.

Mantelero A., *Analysis of international legally binding instruments. Final report*, [in:] Ben-Israel I., Cerdio J., Ema A., Friedman L., Ienca M., Mantelero A., Matania E., Muller C., Shiroyama H., Vayena E., *Towards Regulation of AI Systems, Global perspectives on the development of a legal framework on Artificial Intelligence systems based on the Council of Europe's standards on human rights, democracy and the rule of law*, CAHAI Secretariat, Compilation of contributions DGI (2020)16, December 2020.

McCarthy J., Minsky M.L., Rochester N., Shannon C.E., *A proposal for the Darmouth Summer Research Project on Artificial Intelligence*, 1955, http://www-formal.stanford.edu/jmc/history/darmouth.html.

Mucha J., *Alternatywne metody rozwiązywania sporów konsumenckich w prawie unijnym – nowe rozwiązania prawne (dyrektywa 2013/11/UE w sprawie ADR oraz rozporządzenie nr 524/2013 w sprawie ODR)*, "IKAR" 2014, No. 4.

Nissan E., *Digital technologies and artificial intelligence's present and foreseeable impact on lawyering, judging, policing and law enforcement*, "AI & SOCIETY" 2017, Vol. 32.

Partyk A., *Legitim 2.0., czyli o robocie przyszłości… rozstrzygającym spory zachowkowe*, "Studia Prawnicze. Rozprawy i Materiały" 2019, Vol. 2, No. 25.

Samek W. *et al.*, *Explainable Artificial Intelligence: Understanding, Visualizing and Interpreting Deep Learning Models*, "ITU Journal: ICT Discoveries" 2017, Special Issue, No. 1.

Samoili S., López Cobo M., Gómez E., De Prato G., Martínez-Plumed F., Delipetrev B., *AI Watch. Defining Artificial Intelligence. Towards an operational definition and taxonomy of artificial intelligence*, EUR 30117 EN, Publications Office of the European Union, JRC118163, Luxembourg 2020.

Scherer M., *Artificial Intelligence and Legal Decision-Making: The Wide Open?*, "Journal of International Arbitration" 2019, Vol. 36, No. 5.

Talby D., *Healthcare AI does not need a new Hippocratic Oath*, "Forbes" 2020, May, https://www.forbes.com/sites/forbestechcouncil/2020/05/22/healthcare-ai-does-not-need-a-new-hippocratic-oath/?sh=12c34a541752.

Vitkauskas D., Dikov G., *Protecting the right to a fair trial under the European Convention on Human Rights: A handbook for legal practitioners*, Council of Europe, 2017, https://rm.coe.int/protecting-the-right-to-a-fair-trial-under-the-european-convention-on-/168075a4dd.

Wachter S. *et al.*, *Transparent, explainable, and accountable AI for robotics*, "Science Robotics" 2017, Vol. 2, No. 6, http://robotics.sciencemag.org/content/2/6/eaan6080.full.

Yeung K., *A Study of the Implications of Advanced Digital Technologies (Including AI Systems) for the Concept of Responsibility Within a Human Rights Framework*, November 9, 2018, MSI-AUT (2018) 05, https://ssrn.com/abstract=3286027.

PROF. DR. EMŐD VERESS
University of Miskolc | DOI: 10.32041/9788367149372/2

Chapter 2. Artificial Intelligence as a Legal or Technological Person, and as a Judge?

2.1. Introduction: A brief contemplation on personhood in law

2.1.1. WHO (OR WHAT) MAY QUALIFY AS A "PERSON"?

Let us ask apparently the simplest of questions: what defines a person? The answer at first seems trivial: a person is, of course a human being. Should we rephrase the question, however complications quickly emerge. What if we asked, what defines a person in a legal sense? The task still seems simple enough: natural persons (human beings) are the subject and at times also the object of legal regulation, and most if not all legal systems recognise some form of legal or moral personhood for groups of (natural, or legal) persons constituted according to rules set forth by law. Therefore, in this meaning a person seems to be anyone (and anything) recognised as such under the law, so long as ultimately a human being is involved.

Going a bit further we may inquire as to the defining traits that distinguish a person from a non-person, and that a legislator may consider when recognising the concept of personhood. Philosophers, theologians, and lawyers have, since the beginning of

civilisation grappled[1] with the inherent difficulties of circumscribing personhood in a way that is useful for both practical and theoretical applications. The philosophical sense in which a person is defined, at least at the level of abstraction used by standard dictionaries, seems to focus on rationality and self-consciousness (as traits present at least virtually) in human beings, whereas the legal meaning of the notion emphasises[2] the ability to benefit from rights and be bound by obligations, as a status of acceptance among the subjects of law. This duality illustrates the major, as of yet unresolved tension between various approaches to personhood: (1) that which arises from the nature of the being (or, as the case may be, the object of that personhood), and (2) that which arises from the *fiat* of the law. It must therefore be clear that whenever the legal notion of personhood is concerned, it is the legislator that shall have the last word; as it was said "'person' signifies what law makes it signify."[3] However, some form of a theoretical framework for the basic descriptors of the legal concept of personality are necessary, when that last word may result in chimpanzees, robots, rivers or even Indian deities being endowed with personality[4] according to the law.

In order for us to attempt to construct such a theoretical framework, the duality of the extant juridical concepts of personhood could constitute a starting point. Natural and legal persons exist as separate categories, the personality of each arising out of varying factors.

---

[1] For some such definitions see J. Teichman, *The Definition of Person*, "Philosophy" 1985, Vol. 60, No. 232, pp. 175–177.

[2] J. Teichman, *The Definition of Person*, *op. cit.*, pp. 180–81.

[3] J. Dewey, *The Historic Background of Corporate Legal Personality*, "Yale Law Journal" 1925/1926, Vol. 35, No. 6, p. 655.

[4] The legal personality of all such entities has been considered, and sometimes even accepted. See S.M. Solaiman, *Legal Personality of Robots, Corporations, Idols and Chimpanzees: A Quest for Legitimacy*, "Artificial Intelligence and Law" 2017, Vol. 25, No. 2, pp. 155–179, DOI: 10.1007/s10506-016-9192-3; G. Eckstein *et al.*, *Conferring Legal Personality on the World's Rivers: A Brief Intellectual Assessment*, "Water International" 2019, Vol. 44, No. 6–7, pp. 804–829, DOI: 10.1080/02508060.2019.1631558.

2.1.2. PERSONS BEFORE THE LAW

2.1.2.1. Relevant traits of natural persons, and their limits

In the philosophical sense, at least, natural personhood is defined (as we have already seen) by the criteria of rationality and self-consciousness, both functions of higher order cognition, which endow the bearer with "rational agency" (as the landmark definition given by Boethius suggests); a definition that is, however apparently often ignored by the law, which may bar such rational agents (e.g., slaves, women, convicts, children) from the benefits of the legal status of "person," while liberally endowing other entities, perhaps lacking any form of rational agency, with such benefits.[5] Still such a definition elicits a clear link between reason, as a function of cognition, and the individual human being, even if philosophically it does not remain restricted to humans (it being impossible to exclude that non-human entities, i.e., aliens or artificial intelligence, could attain such traits).

As the notion of personhood was gradually untied from metaphysical concepts, it has been asserted that human beings' claim to personhood, in the legal sense stems simply from the purpose of the law, which is ultimately to uphold human interest, and from the belief that humans, as opposed to other creatures and inanimate objects, present the exceptional natural traits of higher-order cognition and sentience (awareness of the self and perception of feelings), which justify such a claim, as well as the tenets of human dignity, and equal value associated with it.[6]

Taking such an approach, sometimes referred to as the "radical naturalization of personhood"[7] carries the inherent dual risk that either the legal quality of being a person becomes shackled to (cognitive and emotional) traits which vary between individuals, and therefore imposes an unjust generalisation, or it inevitably excludes

[5] T. Pietrzykowski, *Towards Modest Naturalization of Personhood in Law*, "Revus" 2017, No. 32, pp. 59–60, DOI: 10.4000/revus.3863; J. Teichman, *The Definition of Person*, p. 181.

[6] T. Pietrzykowski, *Towards Modest Naturalization of Personhood in Law*, *op. cit.*, pp. 61–63.

[7] *Ibidem*, pp. 64–65.

human beings not in full possession of such faculties (the mentally ill, the comatose, etc.) in violation of an ethical standard of equality between individuals of the human species.

The opposite of this approach, is "modest naturalization,"[8] by which the basic traits of personhood from a legal perspective are reduced to its utility as a means of protecting and imposing the interests of not just humans, but any natural entity (animals or even "non-personal subjects of law" which are not even organisms is the biological sense) deserving of such protection, based on a moral duty of stewardship and conservation, of mandatory consideration for non-human interests.[9] As enticing as such views may be, they are based on the implicit assumption, that the protection of human-significant values such as natural diversity in rivers or the well-being of higher mammals, even the value of the human foetus cannot be attained in other ways then by endowing them with the status of subject, and not just object, of regulation. The truthfulness of this oftentimes ignored implicit assumption remains to be demonstrated.

The theoretical basis of "modest naturalization" – rooted in the vague concept of stewardship – as well as its implementations may very well prefigure the future of regulation when it comes to individuals' personality before the law, flinging open the lid of a Pandora's box from which innumerable "things" clamouring for legal protection will emerge to haunt the legal landscape, all represented – of course by – persons (i.e., natural, or legal) which may be called "classical" if legal history is kept in view. Such a theoretical basis, however, by reducing personhood before the law to a set of interests (rights *lato sensu*) deemed worthy of protection by a given legislator, ignores the fact that personality in the legal sense presupposes the possibility not only of having claims, but also of owing dues. We cannot help but to wonder how a claim against an Indian

[8] *Ibidem*, pp. 65–67.

[9] See also V.A.J. Kurki, *Why Things Can Hold Rights: Reconceptualizing the Legal Person*, [in:] *Legal Personhood: Animals, Artificial Intelligence and the Unborn*, V.A.J. Kurki, T. Pietrzykowski (eds.), Springer International Publishing, 2017, pp. 69–89, DOI: 10.1007/978-3-319-53462-6_5.

deity, a river, a chimpanzee, or any other "non-personal subject of law" could be enforced…

Herein we find the defining trait of an individual person, as the subject of law when relations of economic exchange (*quid pro quo*) in the widest sense are concerned: in the duality of personal economic interest (for which to manifest, a person must be able to hold not only assets but also liabilities) and the legal element embodied by substantive law regulating the activity of the given person; this definition of course may hold true for both natural or collective persons, however these are easily distinguished by the classical criterion of will: individual natural persons as a rule exercise an individual, direct will, while legal persons must rely on at least one natural person to exercise their "corporate" will for them and on their behalf in a formalised way, based on what ultimately amounts to a fiction of law, in order to be able to attain both rights and assume duties.[10]

Thereby a natural person's existence as an individual entity under the law must be characterised by (1) his or her ability to express an individual will, (2) resulting from rational internalised cognition (a known function of the human central nervous system) and (3) his or her presence, based on that will, as a party to economic exchange, both as subject to rights and obligations determined under substantive law which (4) impact his or her economic status (assets and liabilities) directly. This approach, in line with a partial *per a contrario* interpretation of the so-called "Theory of Fiction" proposed by Savigny[11] to explain legal personality, perhaps best expresses what a natural person in the legal sense is: an entity having the traits of abstract personhood in the legal sense, manifested as an individual human being. Of course, such an approach remains human-centric, but it precludes the dangerous fiction that human-neutral criteria may be found to define a unilateral (all rights, no

[10] E.A.Q Adriano, *The Natural Person, Legal Entity or Juridical Person and Juridical Personality*, "Penn State Journal of Law and International Affairs" 2015/2016, Vol. 4, No. 1, pp. 368–370.

[11] E.A.Q Adriano, *The Natural Person, Legal Entity or Juridical Person and Juridical Personality*, *op. cit.*, pp. 371–372.

obligations) form of legal person. After all, rivers, Indian deities, and chimpanzees would always have their interests represented by human stewards, as they would be unable to express human-intelligible rational agency.

"Robots", or more precisely the artificial intelligence (AI) embedded in them, bear separate treatment here, as they may, as we shall see, one day attain sentience as well as rational action capabilities that are at the very least equivalent to those of human beings or may generate outcomes indistinguishable from the results of rational cognition and sentience.

The above defined traits, some (the presence of rights and obligations) specific to all persons recognised under the law, while others (internalised, individual will) specific only to human beings help us circumscribe the limits of personality, when applied to a rational individual, taking decisions for himself or herself. They however do not elicit all the coordinates of a legal person, that is a person before the law, ultimately composed of one or several natural persons, unable to express an individual will on its behalf, but only by mediation of a human being acting in its interest. For this reason, we need to circumscribe the traits of non-human legal entities (sometimes called legal persons, legal entities, corporate persons, etc.).

2.1.2.2. Relevant traits of legal persons (legal entities)

The classical view oftentimes emphasises the economic aspect of legal entities (which we shall call legal persons): they are described in the literature as being mainly receptacles of rights and obligations, noting more, nothing less.[12] This modus of definition is considered "legalistic," i.e., based on a pragmatist evaluation of the purpose of the legal person, and devoid of any speculation on the metaphysical or moral elements of personality; only two components are truly significant: the presence of rights and duties, and the legislative

[12] J. Dewey, *The Historic Background of Corporate Legal Personality*, pp. 656, 659; E.A.Q. Adriano, *The Natural Person, Legal Entity or Juridical Person and Juridical Personality*, pp. 376–381.

fiat by which the legal person as an abstraction is deemed apt to benefit, or as the case may be, be bound by them (the traditional, Kelsenian view of the legal person), taking this optic very close to the related concept of property as legal persons are thought of as mere "clusters"[13] of rights and obligations. It is true that such a view of legal personality has its drawbacks, such as implicit assumptions which may not render it value-neutral, it being considered in essence an "empty slot"[14] to be filled with whatever content the legislator deems politically appropriate.

It is precisely when filling this "slot" that the more complex issues of legal personality arise. Let us emphasise here, even with the risk of some redundancy, that until very recently in legal history the mere possibility that anything other than what is ultimately (even indirectly) a human being or some collective of such beings, possessed of rationality, and all things specific to human cognition and morality, could benefit from legal personality was either unthinkable or confined to the realm of fantasy.

Many of the assumptions that go unsaid in the legalistic view (especially the ability to express a will based on a rational cognitive process) are just that: unsaid, but very much present. Rights and obligations do not usually arise at random in relation to a legal person, especially not during highly planned and thought-out activities involving business transactions in which many such persons are engaged. Therefore, the presence or possibility of rights and obligations should not be emphasised to the detriment of the cognitive ability, the capacity for reason, and perception, when circumscribing the characteristics of the legal person, even if this capacity arises from a "collective" process of decision-making. This "rationalist" view of legal personality takes the (Kantian) position that reason, the rational ability and common sense of a legal person is of defining significance; it is a view not free of criticism due to its association with liberal economic views (today deemed as conservative), especially

[13] Sh.N. Hamilton, *Impersonations: Troubling the Person in Law and Culture*, University of Toronto Press, 2009, pp. 31–38, 46–49.
[14] *Ibidem*.

influential in the contractual (real entity) or legal act theories (as opposed to traditional, hierarchical theories) of legal persons.[15]

Aside from the political opposition against some implications of rational actor theories in free market economics, the rationalist view also comes under fire for its untenability towards persons incapable of reason, as the Kantian view is strictly linked to the will theory of personhood, any claim to this status by entities (including natural persons) incapable of developing or expressing rational will is at best tenuous.[16] The rationalist view therefore may produce untenable results when applied to natural persons (considered in general to be subjects of law as resulting from natural law, regardless of their rationality), as it may constitute a basis for depriving human beings of personhood before the law; it is for this reason that such views are tempered, in the case of natural persons at least, not only by statute and case law but also by the notion of human dignity, arising out of the metaphysical concept of being human.[17] Such risks historically did not arise in the case of legal persons, which are essentially non-human, and need not benefit from human dignity, an institution usually[18] restricted to humans, the imposition of a requirement of rational will in their case being wholly justified.

---

[15] E.A.Q. Adriano, *The Natural Person, Legal Entity or Juridical Person and Juridical Personality*, p. 380; Sh.N. Hamilton, *Impersonations: Troubling the Person in Law and Culture*, pp. 61–62, 64–65. See also S.M. Watson, *The Corporate Legal Person*, "Journal of Corporate Law Studies" 2019, Vol. 19, No. 1, pp. 137–66, DOI: 10.1080/14735970.2018.1435951.

[16] Sh.N. Hamilton, *Impersonations: Troubling the Person in Law and Culture*, *op. cit*., pp. 67–68.

[17] *Ibidem*, pp. 92–94, 167–168.

[18] The Constitutional Court of Hungary, perhaps uniquely, extended the scope of human dignity to include legal entities. See Ch. McCrudden, *Human Dignity and Judicial Interpretation of Human Rights*, "European Journal of International Law" 2008, Vol. 19, No. 4, p. 708, DOI: 10.1093/ejil/chn043. The author examines all internationally relevant implementations of human dignity, which quasi-universally hold human dignity as inherent to the human being as subject of law (even when the infringement is directed against collectives of human beings manifested in legal persons).

2.2. AI's claim to (legal) personhood as a "technological person"

2.2.1. PROLEGOMENA

The concept of "technological person" is not regularly used in literature pertaining to the legal personality of machines or AI entities, or to describe such entities considered as subjects of law in their own right, but instead is usually understood to mean a human being interested in – or endowed with – some technological aptitudes.[19] Sporadically,[20] references to androids (artificial human-like entities, robots in human form) which resemble humans, and elicit responses similar to those given to human interlocutors, are described as technological persons.

"Electronic person" is a more widely utilised indicator of an AI entity (most often a robot) which is endowed with some form of legal personality, the term having been employed in the proposed set of Civil Law Rules on Robotics adopted by the European Parliament Resolution of 16 February 2017, resulting in some consternation.[21]

---

[19] See P. Sharma, S. Gaur, D. Dashora, *Impact of ICT Support on E-Governances Services*, [in:] *Computing and Network Sustainability*, Sheng-Lung Peng, N. Dey, M. Bundele (eds.), Springer Singapore, 2019, pp. 211–216; C. Grimalt-Álvaro *et al.*, *"I See Myself as a STEM Person": Exploring High School Students' Self-Identification with STEM*, "Journal of Research in Science Teaching" 2022, Vol. 59, No. 5, pp. 720–745, DOI: 10.1002/tea.21742.

[20] P.H. Kahn, S. Shen, *NOC NOC, Who's There? A New Ontological Concept (NOC) for Social Robots*, N. Budwig, E. Turiel, P.D. Zelazo (eds.), Cambridge University Press, 2017, p. 106; K. MacDorman, H. Ishiguro, *The Uncanny Advantage of Using Androids in Cognitive Science Research*, "Interaction Studies" 2006, Vol. 7, No. 3, p. 300, DOI:10.1075/is.7.3.03mac.

[21] European Parliament Resolution of 16 February 2017 with Recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), OJ C 252 (2018). See B. Custers, E. Fosch-Villaronga, *Humanizing Machines: Introduction and Overview*, [in:] *Law and Artificial Intelligence. Regulating AI and Applying AI in Legal Practice*, B. Custers, E. Fosch-Villaronga (eds.), T.M.C. Asser Press, 2022, p. 5; S. De Conca, *Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-Empowering Solutions*, [in:] *Law and Artificial Intelligence. Regulating AI and Applying AI in Legal Practice*, B. Kusters, E. Fosch-Villaronga (eds.), T.M.C. Asser Press, 2022, p. 253; J.G. Allen, *Agency and Liability*, [in:] *Artificial Intelligence. Law and Regulation*, Ch. Kerrigan

For the purposes of this study, we shall employ the notion of "technological person" in a way similar to that in which the term "electronic person" is used by the proposed (later aborted) European norm, in order to designate any AI, which is endowed with a form of (as of yet non-extant) legal personhood, distinguishing it from a simple property item.

We opt for the notion of "technological person," as that of "electronic person" implies an electronics-based AI, whereas such technology may also arise from a hybrid of electronic and biological solutions,[22] or even future technologies, of a yet unknown nature which may lead to synthetic intelligences[23] (possibly of a biological nature) which are entirely unmediated by electronics.

2.2.2. AI AS A "TECHNOLOGICAL PERSON"

AI is generally considered, to be, if not "the," then at the very least "one of the" most transformative technologies ever developed; a means by which mankind passes into a new era of progress as a technological species – or passes into oblivion.[24]

Most predictions of AI, utopian and dystopian alike, seem to be concerned with the consequences of what is often called artificial general intelligence (AGI),[25] an AI implementation which may

(ed.), Edward Elgar Publishing, 2022, p. 152; Jan-Erik Schirmer, *Artificial Intelligence and Legal Personality: Introducing "Teilrechtsfähigkeit": A Partial Legal Status Made in Germany*, [in:] *Regulating Artificial Intelligence*, T. Wischmeyer, T. Rademacher (eds.), Springer, 2020, p. 129.

[22] See for example Ch. Adami, *Making Artificial Brains: Components, Topology, and Optimization*, "Artificial Life" 2022, Vol. 28, No. 1, pp. 157–166, DOI: 10.1162/artl_a_00364.

[23] See for example C.A. Lindley, *Synthetic Intelligence: Beyond Artificial Intelligence and Robotics*, [in:] *Integral Biomathics: Tracing the Road to Reality*, P.L. Simeonov, L.S. Smith, A.C. Ehresmann (eds.), Springer, Berlin– Heidelberg 2012, pp. 195–204, DOI: 10.1007/978-3-642-28111-2_19.

[24] For perhaps the best known such predictions see Y.N. Harari, *Homo Deus. A Brief History of Tomorrow*, Vintage, 2017, pp. 327 *et seq*.

[25] T. Mahler, *Regulating Artificial General Intelligence (AGI)*, [in:] *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, B. Custers,

one day surpass human cognitive abilities, thereby presenting vast promise, but equally vast known as well as unknown risks. AGI would enable a technological entity to deploy its artificial cognition in a form very similar to human intelligence, but perhaps orders of magnitude superior to it, to broadly defined problems. This similarity to human cognition raises the possibility that an AGI would be capable of rational action in a way indistinguishable for all intents and purposes from a human being, even display (perhaps internalise) human-specific emotional traits and develop a human-like consciousness. An even much simpler AI would be capable of taking the kinds of reasoned actions that are specific to human beings, if on a much narrower spectrum limited by its specific purpose (e.g., an AI developed for stock-market trading).

Two sets of problems must therefore be considered here: the short-term problems, of whether similarity to human reason expressed by the actions of an AI – or their consequences – may justify endowing it with some form of legal personality, and the long-term problems, of whether an AGI which for all intents and purposes acts and feels like a human being should be considered as an artificial person (a non-natural, or synthetic person under the law).

Both sets of problems deal with creating a legal framework for a "technological person" but along very different lines and based on very different justifications. Whereas in the first case, which likely must be dealt with much sooner, the analogy to today's classes of legal persons is self-evident (only their actions and the resulting economic consequences matter; specific elements of human dignity should not be separately considered as being applicable to them, even if they are "intelligent" – albeit not in a way a human would be), the second case seems much more akin to how today's natural persons are regulated, where questions of dignity associated with a being capable of sentiments much like ours seem unavoidable.

E. Fosch-Villaronga, T.M.C. Asser Press, 2022, pp. 528–532, DOI: 10.1007/978-94-6265-523-2_26.

2.2.3. THE CASE FOR AND THE CASE AGAINST LEGISLATING A TECHNOLOGICAL PERSON

2.2.3.1. The case for…

The possibility that a machine may display some cognitive traits similar to humans, and intentional (self-generated, autonomous) action guided by this form of "reason" – what has been called "practical reason" – according to some, implies that, similarly to collectives of human beings AI (AGI) entities would have to be granted some kind of legal personality, based solely on such similarity (referred to as "*functional sinimorphy*"[26]). Let us take note here, that what is being advocated is mainly legal personality based on the assimilation of an AI with a corporation,[27] an entity also devoid of individual human will (in which a collective will is expressed by authorised persons), and not on any similarity to human behaviour.

Weak AI Contracting Agents and the AI Inventor

AI implementations in the form of "contracting agents"[28] for example are able to display rational behaviours and take autonomous actions in interactions with human or other machine interlocutors on behalf of human principals. They differ from simple computer programmes, as the conditions and contents of their actions do not derive directly from human-coded instructions, but are a result of the operation of the AI system itself capable of planning, learning, setting up, testing and implementing hypotheses (argumentation), and cooperation with human and machine alike;[29] in this they

[26] See D.J. Calverley, *Imagining a Non-Biological Machine as a Legal Person*, "AI & Society" 2008, Vol. 22, No. 4, pp. 527–528, DOI: 10.1007/s00146-007-0092-7.
[27] D.J. Calverley, *Imagining a Non-Biological Machine as a Legal Person*, *op. cit.*, p. 526.
[28] See for example F. Andrade *et al.*, *Contracting Agents: Legal Personality and Representation*, "Artificial Intelligence and Law" 2007, Vol. 15, No. 4, pp. 357–373, DOI: 10.1007/s10506-007-9046-0.
[29] F. Andrade *et al.*, *Contracting Agents: Legal Personality and Representation*, *op. cit.*, pp. 358–361.

are dissimilar to other computers, as they have a "mind of their own," and their decisions cannot be traced back directly to a human programmer or operator, or even to general instructions given by a human. These systems are also dissimilar to AGI because they act in a narrow field, under the control and ultimate supervision of human principals, and – some would add – display no specific traits of human sentience.

In this approach, there is no implicit need that the AI behave "just like" a human being, to have, or even to appear to have the cognitive and emotional traits of a human which would be manifested for example in it passing the Turing test[30] or similar trials of human (emotional) equivalence. It simply must be a rational actor, possessed of intent (perhaps of second order volition – setting parameters for its activities based on known social values which then guide its actions),[31] even if the rational action has little-to-no real-world consequences (for example, in the case of some chatbots).

It was based on such considerations of autonomous behaviour and practical reasoning that an AI entity called DABUS was indicated as an inventor in several patent applications filed under the Patent Cooperation Treaty; after numerous jurisdictions rejected registration of the patent showing DABUS as being the inventor, citing that it was not a natural person, the Federal Court of Australia held in its Judgement rendered on 30 July 2021 that the patent application must be granted, and DABUS must be acknowledged as the creator of the invention.[32] The decision conflated the notion of the invention (a novel way of manufacturing a product, called a "manner of manufacture") with that of "inventor," as the entity behind the invention, and reasoned, that an AI entity should be a novel form of inventor, because of the activities such an entity undertakes in the interest of technological progress, as mediated by

[30] See M.A. Boden, *Artificial Intelligence. A Very Short Introduction*, Oxford University Press, 2018, pp. 107–108.

[31] D.J. Calverley, *Imagining a Non-Biological Machine as a Legal Person*, *op. cit.*, p. 534.

[32] *AI System Qualifies as Inventor*, "GRUR International" 2022, Vol. 71, No. 6, pp. 540, 548–549, DOI: 10.1093/grurint/ikac025.

the producers of the AI entity.[33] The judgment was later overturned (as we shall discuss below). It is worth mentioning here that similar reasons were invoked to award AI legal personality so that it could enjoy intellectual property rights in the form of copyright. However, even the proponents of such a move consider it a mere possibility, incompatible with current legal regimes, at least where EU Member States are concerned.[34]

Strong AI (AGI) Contracting Agents

The proposed necessity and benefits of extending legal personality are also relevant for AGI, which would be able to pass the Turing test and display human or above-human intelligence in a human-intelligible way. In this case, based on the objective theory of contract, which emphasises the meeting of minds as expressed, and not as intended, one or several of the "minds" constituted by AGIs could be parties to a contract.[35] The main argument in favour of such a solution besides the ones derived from the fact that legal persons are created by legislative *fiat* and may present some form of self-determination (which we have already seen), are that AGI as opposed to "weak AI" can display "collective intentionality," that is to say, cultural equivalence with human beings, being able to participate in human-specific "conventional cultural practices such as law"[36] thus making human-to-AGI interaction similar to human-to-human interaction. Of course, as a legal person, such AGI must in this view also be a subject of rights and obligations in order to participate in economic exchange in a way that is meaningful for humans.[37]

---

[33] *AI System Qualifies as Inventor, op. cit.*, pp. 549–550.

[34] J. Smits, T. Borghuis, *Generative AI and Intellectual Property Rights*, [in:] *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, B. Custers, E. Fosch-Villaronga (eds.), T.M.C. Asser Press, 2022, p. 332, DOI: 10.1007/978-94-6265-523-2_17.

[35] J. Linarelli, *Artificial General Intelligence and Contract*, "Uniform Law Review" 2019, Vol. 24, No. 2, pp. 336, 339–340, DOI: 10.1093/ulr/unz015.

[36] J. Linarelli, *Artificial General Intelligence and Contract*, *op. cit.*, p. 341.

[37] *Ibidem*, pp. 342–343.

The Liable AI

Liability for damage caused by AI entities is one of the most fiendishly difficult areas of law when it comes to reconciling effective liability rules with the novel nature of the damage-causing AI technology. The nexus of most worries is of course the self-driving car, the autonomous vehicle, a technology which is likely to become implemented in the foreseeable future. We must keep in mind though, that AI is likely to be able to control not just vehicles, but also say a surgical robot or a company.[38] While most discussion is centred on non-contractual liability for "AI damage," we should not forget about the contractual field, where AI may also wreak havoc for which someone shall be called to pay. The principle of responsibility dictates that someone always has to pay for damage caused.[39] Up until very recently, that someone has always been either a natural person or some collective of natural persons. We can say that the "buck always stopped" with a human.

Discounting the situations when for one reason or another no one may be called to account (mostly limited to acts of God, or situations where immunity from civil liability is stipulated), traditional civil liability models have either consecrated strict liability (when damages are owed even if no fault of the party or tortfeasor is present), or liability based on personal fault (when the contribution of a usually illicit conduct – either a delict, or a breach of contract must occur for it to be implemented), or some combination of the two solutions, all these being in principle applicable to AI entities as well; following an analogy: responsibility for animals, founded on

[38] See K. Prifti, E. Stamhuis, K. Heine, *Digging into the Accountability Gap: Operator's Civil Liability in Healthcare AI-Systems*, [in:] *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, B. Custers, E. Fosch-Villaronga (eds.), T.M.C. Asser Press, 2022, pp. 279–295, DOI: 10.1007/978-94-6265-523-2_15; H. Drukarch, E. Fosch-Villaronga, *The Role and Legal Implications of Autonomy*, [in:] *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, B. Custers and E. Fosch-Villaronga (eds.), T.M.C. Asser Press, 2022, pp. 345–364, https://doi.org/10.1007/978-94-6265-523-2_18.

[39] U. Pagallo, *The Laws of Robots. Crimes, Contracts and Torts*, Springer, 2013, pp. 29–31.

strict liability is historically known, and fault-based responsibility for damage caused by animals is also a possibility, while punishment is not regularly meted out to the animal itself but the owner who failed to physically control or condition (discipline) the animal.[40] AI directed robots (for the problem is mostly raised in their case) and disembodied AI (the "mind in a box") are however not animals, but rational technological objects constructed to achieve a certain purpose by deploying humanlike or even superhuman intelligence and communication. Inevitably, the basic mechanisms of civil liability are disrupted, specifically by the presence of several technologies merged into AI applications (such as robots which have various hardware and software components) leading to an abundance of actors responsible for each component (e.g., the datasets the AI learned from may be compiled by numerous persons), and an added uncertainty due to the AI itself interacting with such components.[41] These problems have been summarised, when examining the different and novel nature of AI, as being constituted by the factors of:

> *"(a) complexity – dealing with software that interacts directly with its environment and interacts with itself (…);*
> *(b) autonomy – outcomes arising as a result of the operation of the code rather than the intention of the programmer;*
> *(c) unpredictability – a fundamental change from traditional computing programming based on logical operations;*
> *(d) opacity – the 'black box' problem;*
> *(e) vulnerability – covering many things but including problems arising from bias or poor design."*[42]

---

[40] U. Pagallo, *The Laws of Robots. Crimes, Contracts and Torts*, *op. cit.*, pp. 33–38.

[41] D. Conca, *Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-Empowering Solutions*, *op. cit.*, p. 243; S. Whittam, *Mind the Compensation Gap: Towards a New European Regime Addressing Civil Liability in the Age of AI*, "International Journal of Law and Information Technology" 2022, pp. 2–4, DOI: 10.1093/ijlit/eaac013.

[42] Ch. Kerrigan, *Introductory Essay*, [in:] *Artificial Intelligence. Law and Regulation*, Ch. Kerrigan (ed.), Edward Elgar Publishing, 2022, p. 9.

In order to mitigate the problems of liability along with adapting the "standard tools" of fault-based (non-contractual, or even contractual) and strict liability for dealing with such situations, the deployment of insurance schemes, and even creating a legal entity, such as a limited liability company in which the AI would be a shareholder have been proposed, though this last proposal can be easily dismissed as such a legal person would be devoid of any patrimony and effectively managed by humans, constituting just an added hurdle in the way of obtaining redress.[43] It is in this context that the European Parliament, in the past proposed the creation of an electronic person to be held liable for AI damage.[44] The reasoning for this proposal was laconic to say the least, and remained mostly limited to the following statement by the designated rapporteur:

> *"Risks that may occur are inherently linked to the use of autonomous machine in our society. A robot's behaviour potentially has civil law implications, both in terms of contractual and of non-contractual liability. Thus clarification of responsibility for the actions of robots and eventually of the legal capacity and/or status of robots an AI is needed in order to ensure transparency and legal certainty for producers and consumers across the European Union.*
> *The Commission is called on to carry out an impact assessment of its future legislative instruments to explore the implications of all possible legal solutions, such as, among others, the establishment of a compulsory insurance scheme and a compensation fund."*[45]

We note here the air of inevitability with which regulation of the technological person is deemed necessary, whereas no proof

---

[43] D. Conca, *Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-Empowering Solutions*, *op. cit*., p. 253.

[44] *Ibidem*, p. 253.

[45] M. Delvaux, *Report with Recommendations to the Commission on Civil Law Rules on Robotics. Explanatory Statement*, European Parliament Committee on Legal Affairs, 27 January 2017, https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html#_section3.

has yet been provided of the fact that extant forms of (strict) civil liability adapted[46] to the requirements of a new technological age (as happened during the late 19th century) would be insufficient for tackling the difficulties posed by AI.

Analysing responsibility for actions by an AI based on the principal–agent model was also proposed, which inevitably leads us to the possibility of legal personhood, should an AI behave in a way specific to a proper agent (strict agency being a possibility even today, as AI entities are already engaged in contractual activities).[47] In the perspective in which cognitive tasks are entirely delegated to an AI acting as proper agents, there is an argument to be made for legal personhood,[48] as – again, based on the concept of "functional sinimorphy" – these act like humans, in tasks that could be (were) historically assigned to humans. In this case the essence of liability for the actions of an AI is grasped not so much from the standpoint of who shall be held responsible but for what is responsibility to be demanded. Actions by an AI, being similar to that of a human in that they are a product of cognition, in this view justify legal personality for AI by analogy.

The Artificial Human

The final, and perhaps most convincing set of arguments for endowing AI with personhood before the law, is the possibility that it may (probably as a form of AGI) one day attain sentience that is to all intents and purposes comparable with, or identical – even far superior – to that displayed by humans.[49] In this case, the "machine" would show the signs of possessing human(-like) emotions, it would be capable of physical and mental forms of pain and anguish, suffering which (at least by outwardly appearance) is similar to the

[46] For the most recent proposals for such adaptations in the field of product liability see S. Whittam, *Mind the Compensation Gap: Towards a New European Regime Addressing Civil Liability in the Age of AI*, pp. 7–16.

[47] U. Pagallo, *The Laws of Robots. Crimes, Contracts and Torts*, *op. cit*., pp. 40–41.

[48] *Ibidem*, pp. 101–102.

[49] See P. Shaw, *Ethics*, [in:] *Artificial Intelligence. Law and Regulation*, Ch. Kerrigan (ed.), Edward Elgar Publishing, 2022, pp. 399–400.

human experience of life, and therefore would be worthy of its own "artificial human dignity," endowed upon it, as it is upon all human beings, for simply existing. Philosophers, ethicists and lawyers have for centuries debated the wellsprings and true nature of human dignity,[50] but the common denominator of their views may be condensed into the following tenets, as they define personhood based on dignity alone, in a way devoid of any legalist reasoning:

> "*1. [A]n embodied being endowed with capabilities and limitations commensurate with the particular nature of the physical substrate that sustains its existence;*
> *2. capable of sensing and interpreting its environment through a cognitive architecture that includes situatedness in time, memory, reason, language, and learning;*
> *3. characterized by a broad range of emotional experiences and dispositional states possessing a positive or negative valence;*
> *4. centered around a focal point of subjective selfhood and individual character that ground the unique narrative of its experiences and actions;*
> *5. a social entity whose capabilities and limitations arise out of relationships with other persons, necessarily including dimensions of empathy, reciprocity, belonging, and a moral sense;*
> *6. inherently oriented toward agency in the world, through free will and a broader sense of goal-directed purposiveness.*"[51]

The list contains some assumptions reflected in philosophical and legal oriented writings, but also adds an essential component to the individual, its "embodied" nature, and its capacity for emotion (a consequence of a fragile and transient "body"), not just cognition,

[50] M. Bess, *Eight Kinds of Critters: A Moral Taxonomy for the Twenty-Second Century*, "The Journal of Medicine and Philosophy" 2018, Vol. 43, No. 5, pp. 587–592, DOI: 10.1093/jmp/jhy018.

[51] M. Bess, *Eight Kinds of Critters: A Moral Taxonomy for the Twenty-Second Century*, *op. cit.*, p. 592.

which influences the very essence of discussions on technological personhood when related to AI entities, the problem to which our study is devoted. This raises the question of whether a "mind in a box," that is an AI/AGI entity for which "the physical substrate that sustains its existence" is provided by another (human) being may qualify as a technological person?

We posit here, as is apparent from this discussion, that embodiment does not have to mean having a (human-like) mobile robotic body, only that the consciousness of the AI/AGI entity must be aware of the existence and the limits of its physical self (a sign of sentience). Thus, the possibility of granting equal dignity is preserved, while it does not impose granting wholly equal rights to technological persons: a comatose human being, for example, continues to benefit from his or her rights but is devoid of their exercise. While dignity is equal, the regulation of personhood is not wholly contingent on this equality alone.

We must now distinguish between several categories of possible "artificial humans" and "artificial superior intellects" as we might term a superhuman technological person. The first category of machine beings would be those designed to exist and behave similarly to humans, while the second is constituted of "presumed persons:" initially human minds transferred into machine form and augmented beyond recognition, or AGI entities, both so far superior to known human capabilities that their self-expression, perception, emotions and experiences would be unrecognisable to a human.[52] While artificial humans could be subject to a form of personhood similar to natural persons based on the tenets of "human" dignity,[53] artificial superior intellects being so different, there would be little which makes their experiences similar to those of humans, and therefore they should be treated under the presumption of having

[52] *Ibidem*, pp. 595, 597–599. The author here differentiates between anthropomorphically designed AGI entities (android) and entities lacking such design. Actually human-like body design is less important than mobility, regardless of how it is achieved (e.g., an insect-like robot may be possessed of human-like intelligence).

[53] *Ibidem*, pp. 600–602.

human wants and needs based, as far as possible, on the liberal desideratum of universal inclusion.[54]

2.2.3.2. ...and the case against

Based on the view which would ground technological personality on "functional sinimorphy" or the ability to elicit "practical reason," an AI chatbot, whose rational action is manifested in typing out text on a screen (a potentially insignificant real-world consequence of its autonomous action) could be considered a legal person for simply existing and intelligently displaying words on a monitor. Herein lies the Achilles' heel of such purely cognitive-behavioural approaches.

If one seeks an analogy between an AI entity and other legal persons, such as corporations, real-world consequences matter. Unless we speak of individual human beings (where human dignity permits no other solution) the quality of being a subject of rights and obligations is not an "award" granted for potential or manifest intelligence it is a "licence" which permits an entity to participate in the economic exchange of values for its own profit but also at its own, even existential, risk. The Judgment of the Federal Court of Australia, of 30 July 2021 given in the DABUS case was overturned on 8 August 2022, *inter alia*, because the court held that "human agency" was required for the inventive process to take place.[55] However the court in passing also considered a far more important set of questions, stating that the characterisation of the notion of an AI "inventor," if explored deeper, would require clarifying this notion against problems such as the fact that the copyright to the source code of DABUS is held by the plaintiff, the computer on which DABUS runs is the property of the plaintiff, and the plaintiff ensures its maintenance and supports its running costs.[56] Therefore, even if

[54] *Ibidem*, pp. 602–606.

[55] *DABUS Overturned: An AI Cannot Be Named as an Inventor*, "GRUR International" 2022, Vol. 71, No. 8, p. 736, DOI: 10.1093/grurint/ikac057.

[56] *DABUS Overturned: An AI Cannot Be Named as an Inventor*, *op. cit.*, pp. 736–737.

the court explicitly excluded examination of these problems from its reasoning, considering other arguments as being sufficient to decide the case, it did correctly recognise the lack of material (economic, existential) autonomy of an AI system as a factor in potentially denying it recognition as a legal entity. This brings us to the problem of risk, and the AI which may not be exposed to any risk, rendering its actions meaningless to it, as a subject of economic exchange.

Any AI not participating in economic exchange at its own risk and for its own profit cannot be equated to extant legal (or natural) persons without violating the tenets of either the legalistic or the rational approach to legal personhood which we have outlined above (the former because no practical purpose is present, the latter because it does not refer to reason exercised without a practical purpose). Should we untie ourselves from such notions, we may find ourselves adrift in an ocean of absurdity, compulsively looking for excuses to grant legal personality for the sake of simply swelling the number of legal persons.

The concept of "own risk" leads to a more profound issue, only hinted at in the DABUS case: AI may be, indeed is very likely to be, manifested in a disembodied[57] intelligence (i.e., a box in a room plugged into an electricity outlet, unable to experience reality in any other form than that in which it is "fed" to it – usually by a human "minder"). This state is relevant, not from the context of the AI's actions, or their risks (it can very well do great good, or ill, by simply operating on data, even without having a body or being capable of movement or direct perception of its environment) but from the context of the material prerequisites to its physical existence, and their feedback to its behaviour. Specifically, a disembodied AI as of yet cannot, and also need not, acquire any form of sustenance, it being wholly dependent on an energy source provided by a human intermediary. It cannot self-repair, so caring for the system's "health" also falls to a human. This situation renders anxiety, or concept of discomfort, such as pain, starvation, fear of loss or of dissolution (as only biological beings can die), all the formative elements of

[57] See M.A. Boden, *Artificial Intelligence. A Very Short Introduction*, *op. cit.*, pp. 121–122.

naturally-occurring intelligent behaviours in animals and humans, meaningless when we refer to an AI devoid of human sentience and "kept alive" by human effort.

Such an AI cannot be considered an "autonomous" entity, (risk-taker), like a corporation, as it neither assumes, nor suffers the consequences of its own actions (such as economic loss, even a loss of opportunity impacting the shareholders) any more than does a river or an Indian deity. Such an AI would therefore be the subject of rights, but not the subject of enforceable liabilities, and human dignity as understood today cannot be invoked to justify considering it a person (a subject of law) in such a state. Therefore, it cannot, at least in the traditional concept of corporate legal personality, be likened to a subject of law, based on the "functional sinimorphy" criterion, as no such similarity of form and function really exists. A chimpanzee would be a much better candidate for personhood than would such AI entities even if the AI displayed human-equivalent intelligence, as the ape's cognitive-behavioural autonomy is deployed in a way which directly impacts its material existence, an impact it perceives as it is sentient, similarly as in the case of individual humans and collectives of humans endowed with legal personhood. This reasoning is also valid for embodied AIs (i.e., intelligent robots), so long as their actions are not informed by their impact on the AI's own "quality of life" or material existence, as the economic, ultimately biological, feedback loop that punishes economically unwise or illegal behaviour, which historically gave rise to personhood in the legal sense for both individuals and collections of individuals (the necessity to concentrate resources, hedge risks while abiding by the law), is entirely missing. Such reasoning could of course not be applied to sentient AGI entities because under some circumstances, they would be able to consider their perceptions as unpleasant, painful, dangerous or terrifying, and would therefore react to them as sentient beings.

The set of arguments for AGI legal personality, based on participation in contractual relationships, is a step forward from the arguments applied to "weak AI," however even its proponents fails to address the nitty-gritty of how a machine should come to own property, or be subject to enforcement, or, in general, act as a human

being or collection of human beings, glossing over the issue by stating:

> *"If we want to justify the application of the law as it has been constructed by and for humans to AGI, then we need to develop AI with human-like values and dispositions – with humanlike cognitive architecture or, at least, that simulates it with the ability to interact with humans – otherwise humans can reasonably reject the move. We need AI to have these qualities to be able to enter into and perform contracts with humans."*[58]

This argument, wishful thinking aside, again emphasises cognitive abilities, the "humanlike cognitive architecture," while ignoring the fact that such "architecture" may very well reside in a practically immortal form, in a box, plugged into an outlet, and act on information which it has no direct means of perceiving for itself. This is not the definition that could be akin to a human's way of experiencing the world; thus, few things short of an artificial human could comply with the criteria set forth the endow AGI with legal personality under such a cultural interaction paradigm.

Also, even if we addressed the issues of perception by mounting an AGI on some mobile platform and transform it into a robot or allow it to operate as a technological Argus and experience the world directly through innumerable sensors, there is no guarantee that its behaviour would remain culturally human-compatible, so long as its actions do not directly impact its perceived well-being. Put very simply, in order for AGI to qualify for legal personhood on the argument of its behavioural (cultural) similarities to a human, it would have to become an artificial human, subject to all the perils such a state would entail. AGI may be able to cooperate effectively with humans on a level below that of an artificial human, but the simple reason that it offers simulated human interaction should not lead to the legal anthropomorphism of perceiving it as human, so long as other elements of the human condition are absent.

---

[58] J. Linarelli, *Artificial General Intelligence and Contract*, *op. cit.*, p. 343.

Granting AI legal personality so that it may be held to account for damage caused involves similar specific difficulties, as only a person able to be bound by obligations would be subject to liability, while managing assets and liabilities would entail high-level cognitive abilities by the AI (specific mostly to AGI) and would also risk undermining the principle of human dignity.[59] Furthermore, as aptly observed in the literature, (1) the human factor would not be absent from any responsibility (programmers, manufacturers and operators would still be present), (2) most prejudicious outcomes could either be reduced to the activity of extant forms of natural and legal persons, while any new rules regulating individual (strict) liability would be a better response than creating a new category of legal person, and (3) the risk of under-funding the assets of AI legal entities to blunt any action for liability would be ever-present, unduly limiting such liability (making it difficult to "pierce the electronic veil").[60] As such, electronic personhood for purposes of liability seems far-fetched and unnecessary.

We should add to such considerations the fact, as addressed above, that for the foreseeable future, a possibly disembodied AI entity would be unlikely to have the full range of perception and emotion of which a human being is capable, and for this reason, even if it could perceive being held liable for a negative consequence of its actions, the impacts of such a consequence on its future actions may well not be the discouragement we expect (a problem not only of civil, but also of possible criminal liability of the AI, because legal persons are recognised in numerous jurisdictions as also bearing criminal responsibility for their actions). An AI entity may simply act as the occasional human psychopath does and very well consider liability for even atrocious damage caused by it as the "cost of doing business" and, given the inherent opacity of AI decision-making, do so in a way imperceptible to humans. This in turn could result in the repeat occurrence of damage-causing events, as "imprisoning"

[59] D. Conca, *Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-Empowering Solutions*, *op. cit.*, pp. 253–254.
[60] S. Whittam, *Mind the Compensation Gap: Towards a New European Regime Addressing Civil Liability in the Age of AI*, *op. cit.*, p. 16.

the AI would not be an option, and the risk of applying any form of "capital punishment" to it could also not have the desired discouraging effects, as the AI is not "alive," and may (in fact is likely to) not have a notion of death similar to that of humans.

Artificial humans do not suffer from any of the apparent shortcomings which prevent other AI entities from attaining personhood before the law. There is nothing to prevent them from exercising rights and holding obligations on their own behalf, and at their own risk, and even if their perception of existential questions should differ from that of humans, so long as they proceed in their juridical acts and actions with practical intent to preserve their existence and "standard of living" (however these may be interpreted by them, something which we may never know) there is little to distinguish them from human persons.

Artificial superior intellects pose a different set of problems from the perspective of personality, which as of yet is unexplored. Because they would likely vastly exceed human cognitive abilities, their notions of rational action may, for better or for worse, differ wildly from our own, and in wholly unpredictable ways. While we may reasonably presume that a natural person, or extant forms of legal persons, will exercise their rights in a rational and measured way, such an assumption may be a step too far for so-called "presumed persons," such as artificial superior intellects. If considered a person that may hold rights and obligations and deploy them for its own purposes, what would be there to prevent such a potentially practically immortal intellect from constructing strategies, and accumulating resources for eons with the purpose of subverting the fundamentals of human dignity or of furthering the enslavement, or destruction of humankind as a whole? Knowing history, this or similar nightmare scenarios may well not[61] even be avoidable should an artificial superior intellect arise. It may therefore seem logical, even imperative to prevent such an intellect, whose abilities and lifespan may for all intents and purposes be unlimited and

[61] M. Bess, *Eight Kinds of Critters: A Moral Taxonomy for the Twenty-Second Century*, *op. cit.*, pp. 602–603; J.N. Harari, *Homo Deus. A Brief History of Tomorrow*, *op. cit.*, pp. 382–397.

true, long-term intentions unknowable, from gaining additional resources by benefiting from legal personhood. In fact, such entities should be prevented from coming into being by imposing rules of strict liability for even the most minor resulting damage against any actor that intentionally or inadvertently creates such entities, lest the doomsday prophecies turn out to be self-fulfilling.

2.3. What kind of "person" would an "artificial judge" be?

The personhood of AI entities cannot be wholly separated from the possible implementation of AI as a judge, a solution which is predicted[62] to emerge as a form of AGI. In order for us to discuss the link between the two problems, first we must contemplate the current implementations of AI and observe which of these may lead to the emergence of the "artificial judge."

Numerous technological solutions are already deployed with this purpose, some more "intelligent" than others. Expert systems usually provide information (such as legal texts, case law) based on a set of predetermined criteria (such as logical searches) that may be useful for a human operator; a version of these, so-called "case-based systems," help identify similar cases (precedent, case-law) based on a description of the situation at hand.[63] Both these solutions may deploy AI, but in neither situation will AI actually be able to solve a dispute.

Decision support systems are meant to facilitate resolution of a case, by a human operator, based on available information.[64] Various such solutions are implemented for assessing the flight risk of a suspect, or the risk of recidivism for a convict, helping judges take the necessary and proportional preventive and punitive measures

[62] See G.I. Zekos, *Robo-Justice*, [in:] *Advanced Artificial Intelligence and Robo-Justice*, G.I. Zekos (ed.), Springer International Publishing, 2022, pp. 347–415, DOI: 10.1007/978-3-030-98206-5_11.

[63] N. Lozada-Pimiento, *AI Systems and Technology in Dispute Resolution*, "Uniform Law Review" 2019, Vol. 24, No. 2, p. 354, DOI: 10.1093/ulr/unz022.

[64] N. Lozada-Pimiento, *AI Systems and Technology in Dispute Resolution*, *op. cit.*, p. 355.

such as arrest or a heavier sentence.[65] Here, AI serves as a virtual probation officer helping the human judge decide. These systems have so far mostly "made the news" in legal literature for the risk of bias they sometimes displayed (e.g., in the Loomis case).[66] Based on various types of AI implementation, collectively called machine learning, or its subset, deep learning,[67] they are meant to achieve a prediction based on available information and general patterns identified in the databases used to train the system's predictive abilities.

Machine learning as well as this pattern recognition ability may allow AI entities not just to advise the human judge but to provide preventative or curative solutions for a dispute with minimal-to-no human intervention. This form of dispute resolution may manifest itself in dispute prevention by communicating the possible outcome of a dispute (such as the Siarelis chatbot deployed in Columbia), to guide parties to a mediated outcome (the Settify, or the Split Up systems are examples of this approach), to facilitate dispute resolution, or to resolve the dispute itself as a veritable "E-judge."[68] Current technology allows for an AI entity to proceed to prediction, but prediction is not all a human judge does. Human judges in most if not all procedural systems[69] must determine the state of

---

[65] See A. Novokmet, Z. Tomičić, Z. Vinković, *Pretrial Risk Assessment Instruments in the US Criminal Justice System – What Lessons Can Be Learned for the European Union*, "International Journal of Law and Information Technology" 2022, Vol. 30, No. 1, pp. 1–22, DOI: 10.1093/ijlit/eaac006.

[66] See A. Novokmet, Z. Tomičić, Z. Vinković, *Pretrial Risk Assessment Instruments in the US Criminal Justice System – What Lessons Can Be Learned for the European Union*, *op. cit*., pp. 7–11; S. Chesterman, *Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity*, "The American Journal of Comparative Law" 2021, Vol. 69, No. 2, pp. 272–273, DOI: 10.1093/ajcl/avab012.

[67] See T. Virdee, *Understanding AI*, [in:] *Artificial Intelligence. Law and Regulation*, Ch. Kerrigan (ed.), Edward Elgar Publishing, 2022, pp. 40–44; M.A. Boden, *Artificial Intelligence. A Very Short Introduction*, *op. cit*., pp. 42–44; N. Lozada-Pimiento, *AI Systems and Technology in Dispute Resolution*, *op. cit*., p. 355.

[68] N. Lozada-Pimiento, *AI Systems and Technology in Dispute Resolution*, *op. cit.*, pp. 360–363.

[69] See D.J. Gerber, *Comparing Procedural Systems: Toward an Analytical Framework*, [in:] *Law and Justice in a Multistate World: Essays in Honor of Arthur T. Von Mehren*, J.A. Nafziger, S. Symeonides (eds.), Transnational Publishers, 2002.

fact by administering evidence, identify the applicable law, then apply that law to the state of fact and the claims of the parties, then render a decision. The human judge also independently perceives the elements of the case, weighs the argument of the parties, then renders a human-readable, reasoned decision. The keywords here are independent perception, and reasoned decision.

While a human judge shall usually be able to obtain the facts of the case for himself or herself, a machine judge must rely on facts fed to it by a – usually human – "minder," just as any other form of non-robotic AI. Everything said under this respect of the natural person–AI divide remains valid.

Also, whereas a human judge may articulate his or her reasons for a decision in both fact and law, while exposing most implicit considerations, this is not necessarily the case for AI. The basic differences between laws and software must be dealt with when contemplating how an AI decision would be reasoned, because laws are meant to be implemented by human judges, whereas AI will perceive any norms as software instructions (whether pre-programmed or developed through some form of learning); it remains to be seen if AI will even be able to learn to process natural language on the level at which justice is administered.[70] After all, the latter, is no mean feat even for a human being.

The most serious problem in AI judicial decision making is however not posed by the problem of transferring laws into code, but by the inherent opacity of the technology used: AI may reach a correct decision, in fact the statistical correctness of this decision may even be demonstrated, but how it got there is another matter entirely. Opacity arises from two major factors: the proprietary nature of AI technology and from the complexity (the inner workings) of the software itself.[71] Development of the technology's components (both hardware and software) takes place outside specific state control, by

[70] G. Buchholtz, *Artificial Intelligence and Legal Tech: Challenges to the Rule of Law*, [in:] *Regulating Artificial Intelligence*, T. Wischmeyer, T. Rademacher (eds.), Springer International Publishing, 2020, pp. 183–184, DOI: 10.1007/978-3-030-32361-5_8.

[71] S. Chesterman, *Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity*, *op. cit.*, p. 274.

private, often profit-oriented, corporations, in a process that would not be transparent even if the software was open-source (which is usually not the case); this in turn may allow bias as well as the political, or cultural views of the manufacturer to seep into the end product transforming AI manufacturers into second-rate legislators, with implications in the field of the rule of law.[72]

Furthermore, the AI system itself is so complex that the results of its running are not apparent even to its constructors; their acceptance is solely based on what is called "output-based legitimacy."[73] It is not only true, that in the administration of justice, unlike in other fields (such as medicine), outcome-based statistical modelling is not always possible, or very accurate, but also that the legitimacy of the output must be verifiable for ensuring the fairness of the procedure, and the legality of the decision.[74] This is achieved in practice by ensuring that a public and reasoned decision is usually subject to some sort of judicial remedy (an appeal). Publicity of decisions rendered by an AI may be undermined, *inter alia*, on considerations of data protection.[75] Even greater difficulties arise when the reasoned character of the decision is concerned, since the critique of that reasoning should be able to supply the fundament for the appeal.[76] As we have seen current "weak AI" solutions are incapable of providing for such a reasoning and should the appeal phase also be delegated to another AI judge, such problems would be compounded. All this results in the compatibility of an AI judge with the basic principles of a fair trial becoming highly questionable.[77]

---

[72] G. Buchholtz, *Artificial Intelligence and Legal Tech: Challenges to the Rule of Law*, *op. cit.*, pp. 185–186.

[73] S. Chesterman, *Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity*, *op. cit.*, p. 275.

[74] *Ibidem*, pp. 276–277.

[75] *Ibidem*, pp. 286–289.

[76] J. Szekely, *Lawyers and the Machine. Contemplating the Future of Litigation in the Age of AI*, "Acta Universitatis Sapientiae: Legal Studies" 2019, Vol. 8, No. 2, pp. 239–241.

[77] See J. Ulenaers, *The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?*, "Asian Journal of Law and Economics" 2020, Vol. 11, No. 2, DOI: 10.1515/ajle-2020-0008.

We may conclude, that under current technological conditions something has to give. Either we must accept, as some jurisdictions[78] already do, that AI may implement forms of social control without giving adequate reason, based simply on the perceived adequacy of the outcome, or AI must advance sufficiently so that it is able to render human-readable reasoned decisions, based on human-readable laws, where the procedural correctness and factual accuracy of such decisions may be subject to judicial remedy.

The first scenario is already with us but may not be the optimal outcome that mankind desires.

As for the second, the personal and legal characteristics of an "artificial human" to which we have referred above will become extremely relevant once a "strong AI" or AGI entity becomes able to take on the mantle of judge, applying the law to humans and perhaps even to other AI entities. Only an "artificial human" in possession of personhood before the law equal to that of human beings would be able to act as judge in conditions compatible with our current understanding of human dignity, and as we have seen, there is no sound reason, that any AI entity not qualifying as an artificial human be granted legal personality for the purposes of acting as judge.

2.4. Regulatory proposals

2.4.1. AI AS A TECHNOLOGICAL PERSON

As has been outlined above, the first choice which must be made when considering the status of a technological person by any national or international legislator, is whether to endow AI with legal personality at all. This is a choice of no small significance.

[78] L.C. Backer, *Next Generation Law: Data-Driven Governance and Accountability-Based Regulatory Systems in the West, and Social Credit Regimes in China*, "Southern California Interdisciplinary Law Journal" 2018/2019, Vol. 28, No. 1, p. 131.

As we have indicated, in the case of an AI devoid of any elements of the human condition, and which would simply benefit from rights without being bound by obligations, would hold economically insignificant rights and obligations, would hold obligations which are impossible to enforce, or would be present in a state which does not permit it to assume the risks and costs of its actions directly should not benefit from any form of legal, or technological personality.

There is no sound reason for which the law of persons should be amended in such cases, as the AI cannot constitute an independent (or human-independent) participant, with self-interest (bearing risks and reaping rewards on its own behalf) in economic exchange, a trait all other legal persons have (even if their aim is not to generate profit). Also, any "weak AI" entities cannot be equated to natural persons, as they lack (artificial) human dignity, which arises only out of experience similar to the human condition.

The case is of course may be different when AGI is concerned, as an artificial person would be near-indistinguishable from a natural person when it comes to elements of cognition associated with human-like sentiments, raising issues of human(-like) dignity. Such an entity may benefit from legal personality, so long as its experiences and actions are rational from the human perspective. An artificial superior intellect however should be barred from having legal personality, as its actions and motivations may be inconceivable in both timescale and magnitude, rendering it possibly hostile to existence in conditions of human dignity, or hostile to human existence entirely.

Finally, it should not be overlooked here, that there is some substance to the argument, expressed *ad nauseam* in the literature concerning possible legal personhood of AI entities, that legal person status is ultimately based on the option of the legislator, it is a receptacle that the legislator can fill with whatever content it desires. Should the legislator desire to regulate technological persons (a possibility only practicable when "strong AI"/AGI is concerned), several options are open to it:[79]

[79] See U. Pagallo, *The Laws of Robots. Crimes, Contracts and Torts*, *op. cit.*, p. 153.

(1) AI may be granted full legal personhood recognising its full legal capacity, and imposing respect and enforceability for duties, essentially equating human beings with the technological person.
(2) AI may be granted a limited legal personhood to benefit from constitutional rights strictly related to its person (persona), which ensure it dignity, but do not allow it access to economic exchange in its own name. (As we have indicated, such a solution would be meaningless, and should be avoided, unless the AI displays quasi-human reason and sentience, and therefore must be ensured human dignity.)
(3) AI may be granted a limited legal personhood and benefit from legal capacity but may have limited exercise of that capacity. Our statements made at the previous point apply here as well.
(4) AI may be considered a dependent legal person, such as a corporation, with accountability to a principal (a "traditional" legal person or a natural person). There is however little sound reason for this solution, as it would bring no added benefits when compared to considering AI a simple tool, for which the "principal" is liable (perhaps based on some strict liability model), as such liability is often implemented in the case of agency.
(5) AI may be subjected to specific forms of accountability for tort or breach of contracts, either by recognising only a *locus standi* for it when such liability is concerned (such as done in civil procedure for collectives lacking legal personality), or as a manifestation of legal personality. This solution again seems to offer no added benefit.

In view of these possibilities, legal personality may even be regulated as a set of elements in a continuum[80] for various types of technological persons.

We must contend here that, while the legal personality of AI based on the liability argument is vulnerable and should be avoided,

[80] M. Bess, *Eight Kinds of Critters: A Moral Taxonomy for the Twenty-Second Century*, *op. cit.*, p. 607.

not least because it degrades human dignity, other solutions, such as insurance schemes and other "ad hoc"[81] proposals, do not fare much better, as – plastered over with clever legal reasoning – they all tend to result in the socialisation of risk, thereby effectively encouraging human (or future AI/AGI) actors to disregard the risks of novel technologies and relegate the cautionary principle to being a purely philosophical question.

Having outlined these possibilities, we also should not lose sight of the quasi-consensus among AI researchers, and other scientists, lawyers and political leaders which is unequivocally and eloquently expressed in the Open Letter to the European Commission Artificial Intelligence and Robotics,[82] as a backlash to the proposed European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics:

> *"From an ethical and legal perspective, creating a legal personality for a robot is inappropriate whatever the legal status model:*
> *a. A legal status for a robot can't derive from the Natural Person model, since the robot would then hold human rights, such as the right to dignity, the right to its integrity, the right to remuneration or the right to citizenship, thus directly confronting the Human rights. This would be in contradiction with the Charter of Fundamental Rights of*

[81] D. Conca, *Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-Empowering Solutions*, *op. cit.*, pp. 255–256.

[82] *Open Letter to the European Commission Artificial Intelligence and Robotics*, http://www.robotics-openletter.eu/ [accessed on: 15 June 2022]; for further arguments see also L. Floridi, M. Taddeo, *Romans Would Have Denied Robots Legal Personhood*, "Nature" 2018, Vol. 557, p. 309, https://doi.org/10.1038/d41586-018-05154-5. The authors states: *'Attributing electronic personhood to robots risks misplacing moral responsibility, causal accountability and legal liability regarding their mistakes and misuses. Robots could be blamed and punished instead of humans. And irresponsible people would dismiss the need for care in the engineering, marketing and use of robots. Even the Romans knew better: the owner of an enslaved person was fully responsible for any damage caused by that person (known as vicarious liability).'*

the European Union and the Convention for the Protection of Human Rights and Fundamental Freedoms.
b. The legal status for a robot can't derive from the Legal Entity model, since it implies the existence of human persons behind the legal person to represent and direct it. And this is not the case for a robot.
c. The legal status for a robot can't derive from the Anglo-Saxon Trust model also called Fiducie or Treuhand in Germany. Indeed, this regime is extremely complex, requires very specialized competences and would not solve the liability issue. More importantly, it would still imply the existence of a human being as a last resort – the trustee or fiduciary – responsible for managing the robot granted with a Trust or a Fiducie."[83]

These warnings should be heeded, even if the European Commission has submitted a Proposal[84] for regulation which now omits any reference to an "electronic" (i.e., technological) person.

Finally, should the legislator grant legal personhood to AI – as artificial superior intellects may pose an existential risk to humanity – normative models for rapidly retracting such a legal personality from any AI/AGI entity that might risk transforming into such a superior intellect by displaying emergent behaviours[85] resulting from its functioning, must be rapidly developed and implemented. Such models should include ways in which the assets and liabilities of such an AI entity are subsequently wound up.

---

[83] See note 21 above.

[84] European Commission, Directorate-General for Communications Networks, Content and Technology, Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM/2021/206 final), https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206 [accessed on: 15 November 2021].

[85] M. Bess, *Eight Kinds of Critters: A Moral Taxonomy for the Twenty-Second Century*, *op. cit.*, pp. 592–593.

2.4.2. AI AS AN ARTIFICIAL JUDGE

Based on the consideration that human beings are subjects, and not simple objects of the law – the only stance truly compatible with human dignity – no AI entity of lesser standing than a human being should ever be endowed with the right to autonomously render judicial (or administrative) decisions if we are to conserve the current notion of human dignity. Such a proposition should be laid down, preferably on a constitutional level.

As for AI technologies assisting the judge, proposals for regulation already exist. The European Commission Proposal for the regulation of AI technologies referred to judicial application of AI in several places. In the Recitals of the proposal, as amended[86] by the European Parliament, the proposed text now indicates that:

> *"(17) AI systems providing social scoring of natural persons for general purpose may lead to discriminatory outcomes and the exclusion of certain groups. They violate the right to dignity and non-discrimination and the values of equality and justice. Such AI systems evaluate or classify natural persons or groups based on multiple data points and time occurrences related to their social behaviour in multiple contexts or known, inferred or predicted personal or personality characteristics. The social score obtained from such AI systems may lead to the detrimental or unfavourable treatment of natural persons or whole groups thereof in social contexts, which are unrelated to the context in which the data was originally generated or collected or to a detrimental treatment that is disproportionate or unjustified to*

[86] European Parliament, Artificial Intelligence Act. Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html [accessed on: 30 October 2023].

> *the gravity of their social behaviour. Such AI systems should be therefore prohibited.*
> *(…)*
> *(40) Certain AI systems intended for the administration of justice and democratic processes should be classified as high-risk, considering their potentially significant impact on democracy, rule of law, individual freedoms as well as the right to an effective remedy and to a fair trial. In particular, to address the risks of potential biases, errors and opacity, it is appropriate to qualify as high-risk AI systems intended to be used by a judicial authority or administrative body or on their behalf to assist judicial authorities or administrative bodies in researching and interpreting facts and the law and in applying the law to a concrete set of facts or used in a similar way in alternative dispute resolution. The use of artificial intelligence tools can support, but should not replace the decision-making power of judges or judicial independence, as the final decision-making must remain a human-driven activity and decision. Such qualification should not extend, however, to AI systems intended for purely ancillary administrative activities that do not affect the actual administration of justice in individual cases, such as anonymisation or pseudonymisation of judicial decisions, documents or data, communication between personnel, administrative tasks or allocation of resources.*"

Article 6 and Article 8 *et seq.* of the amended Proposal aim to regulate high-risk AI systems among which Annex III Article 1(8)(a) of the amended Proposal would include "*AI systems intended to be used by a judicial authority or administrative body or on their behalf to assist a judicial authority or administrative body in researching and interpreting facts and the law and in applying the law to a concrete set of facts or used in a similar way in alternative dispute resolution.*" From this we may deduce that in the European Parliament's view the advent of the "AI judge" is still far off, although possibly not as far as initially thought by the Commission (the proposal which it first tabled having lacked the "to be used by a judicial authority or

administrative body or on their behalf" clause), thus, an autonomous or independent AI entity resolving legal disputes is now being at least considered as the possible object of the Regulation. We deem this to be a mistake in urgent need of correction by both the national legislator and by the participants in the EU regulatory procedure, now very much nearing completion.

Furthermore, Article 13(1) of the proposed Regulation (as amended) stipulates requirements of transparency, while the amended Article 14(1) imposes human oversight by qualified personnel, and Article 14(2) now specifically names "decisions based solely on automated processing by AI systems" where these "produce legal or otherwise significant effects on the persons or groups of persons on which the system is to be used" as subject to regulation under the special rules for high-risk systems; Chapter 3 of the Proposal elaborates on the obligations of providers of high-risk AI systems. We observe, that no provisions of the proposed and amended text adequately address the problem of granting a human-readably reasoned decision by the AI as the basis for exercising judicial remedies. Still, an explanatory clause newly inserted by the European Parliament into the amended form of Article (13)1 of the Commission's proposal (appended after the proposed text) now reads as follows:

> *"Transparency shall thereby mean that, at the time the high-risk AI system is placed on the market, all technical means available in accordance with the generally acknowledged state of art are used to ensure that the AI system's output is interpretable by the provider and the user. The user shall be enabled to understand and use the AI system appropriately by generally knowing how the AI system works and what data it processes, allowing the user to explain the decisions taken by the AI system to the affected person pursuant to Article 68(c)."*

This wording constitutes a step in the right direction however, reasoned decisions still differ from the type of transparency being referred to in the quoted clause.

In fact, the amended text of the proposed Regulation still fails to address in what way an AI system would "assist" the judge and how such "assistance" would be reflected in the decision. Should the system just grant a score of, say, 87% probability of recidivism, how would even qualified human oversight be achieved, regarding a factor that the judge will certainly take into consideration when rendering a decision? The right to a "reasoned" decision[87] is an inherent element of the right to a fair trial, which should be stipulated explicitly in the context of AI-aided justice.

Ultimately, we consider that the proposal extends to the field of the organisation of the judiciary, even if the proposed Regulation is purportedly founded on Articles 16 (data protection) and Article 114 (internal market) of the Treaty on the Functioning of the European Union.[88] Whereas judicial organisation of the Member States is not granted into the competence of the European Union, the possible balance between the interest of Member States which desire to use such tools and those which may desire to ban them should also be considered when working out EU and domestic norms.

2.5. Conclusions

In our study we have attempted to explore some of the corelations between the traits of extant natural persons and legal persons, and those of AI entities which would render them compatible or incompatible with some form of personhood before the law, as well as proposals to this effect. We have found that the AI world is populated by a myriad of current and future (possible) manifestations of technology that one day may indeed result in granting such personality to the AI entities which bear the closest cognitive resemblance to humans, the "artificial human." We have however also found that

[87] See *Guide on Article 6 of the European Convention on Human Rights. Right to a Fair Trial (Civil Limb)*, Council of Europe – European Court of Human Rights, 2022, pp. 94–96, https://www.echr.coe.int/documents/guide_art_6_eng.pdf.

[88] Consolidated Version of the Treaty on the Functioning of the European Union, OJ C 326 (2012), http://data.europa.eu/eli/treaty/tfeu_2012/oj.

legal personality should not be given out lightly, and that its extension to AI entities without any claim to human (-like) dignity is unjustified, even if numerous such proposals are now being tabled. Such a conclusion is in line with the current consensus, but we base it on considerations that we found to be sparse in the literature: the ability of AI entities to hold obligations, not just rights, and genuine non-intervention of the human factor in their economic activity. Also, concurrently with the current scientific (if not always legislative) consensus, we consider that neither the arguments of liability nor those of agency justify granting AI entities lesser than an "artificial human" any personality before the law, and they constitute an unwelcome complication to situations which may be resolved by other means.

Finally, we have contemplated the current state of law and technology, and the future possibility of an AI judge. We found that the time for such a solution has not come, yet the regulatory framework is already being proposed. Based on the strong link between justice and personhood, we consider that an AI judge should not be lesser than a human being when it comes to rendering and reasoning a decision.

Our research has implications for the present and future of AI regulation, as we have attempted to explore an element of humanity which is at times overlooked when discussing proposals for the legal personality for AI entities: the substance of the human condition, the material and cognitive preconditions to participating in economic exchange, not just as a holder of rights, but also of obligations, and of action based on not just practical reason, but rational self-interest as a bearer of (even existential) risks associated with actions which an AI devoid of concepts and prerequisites of existence in the physical world may not be required to undertake, and for which reason such an AI should never be grated legal person status.

REFERENCES

Adami Ch., *Making Artificial Brains: Components, Topology, and Optimization*, "Artificial Life" 2022, Vol. 28, No. 1.

Adriano E.L.Q., *The Natural Person, Legal Entity or Juridical Person and Juridical Personality*, "Penn State Journal of Law and International Affairs" 2015/2016, Vol. 4, No. 1.

AI System Qualifies as Inventor, "GRUR International" 2022, Vol. 71, No. 6.

Allen J.G., *Agency and Liability*, [in:] *Artificial Intelligence. Law and Regulation*, Ch. Kerrigan (ed.), Edward Elgar Publishing, 2022.

Andrade F., Novais P., Machado J., Neves J., *Contracting Agents: Legal Personality and Representation*, "Artificial Intelligence and Law" 2007, Vol. 15, No. 4.

Backer L.C., *Next Generation Law: Data-Driven Governance and Accountability-Based Regulatory Systems in the West, and Social Credit Regimes in China*, "Southern California Interdisciplinary Law Journal" 2018/2019, Vol. 28, No. 1.

Bess M., *Eight Kinds of Critters: A Moral Taxonomy for the Twenty-Second Century*, "The Journal of Medicine and Philosophy" 2018, Vol. 43, No. 5.

Boden M.A., *Artificial Intelligence. A Very Short Introduction*, Oxford University Press, 2018.

Buchholtz G., *Artificial Intelligence and Legal Tech: Challenges to the Rule of Law*, [in:] *Regulating Artificial Intelligence*, T. Wischmeyer, T. Rademacher (eds.), Springer International Publishing, 2020.

Calverley D.J., *Imagining a Non-Biological Machine as a Legal Person*, "AI & SOCIETY" 2008, Vol. 22, No. 4.

Chesterman S., *Through a Glass, Darkly: Artificial Intelligence and the Problem of Opacity*, "The American Journal of Comparative Law" 2021, Vol. 69, No. 2.

Consolidated version of the Treaty on the Functioning of the European Union, OJ C 326 (2012), http://data.europa.eu/eli/treaty/tfeu_2012/oj.

Custers B., Fosch-Villaronga E., *Humanizing Machines: Introduction and Overview*, [in:] *Law and Artificial Intelligence. Regulating AI and Applying AI in Legal Practice*, B. Custers, E. Fosch-Villaronga (eds.), T.M.C. Asser Press, 2022.

DABUS Overturned: An AI Cannot Be Named as an Inventor, "GRUR International" 2022, Vol. 71, No. 8, pp. 731–737, https://doi.org/10.1093/grurint/ikac057.

De Conca S., *Bridging the Liability Gaps: Why AI Challenges the Existing Rules on Liability and How to Design Human-Empowering Solutions*, [in:] *Law and Artificial Intelligence. Regulating AI and Applying AI in Legal Practice*, B. Kusters, E. Fosch-Villaronga (eds.), T.M.C. Asser Press, 2022.

Delvaux M., *Report with Recommendations to the Commission on Civil Law Rules on Robotics. Explanatory Statement*, European Parliament Committee on Legal Affairs, 2017, https://www.europarl.europa.eu/doceo/document/A-8-2017-0005_EN.html#_section3.

Dewey J., *The Historic Background of Corporate Legal Personality*, "Yale Law Journal" 1925/1926, Vol. 35, No. 6.

Drukarch H., Fosch-Villaronga E., *The Role and Legal Implications of Autonomy in AI-Driven Boardrooms*, [in:] *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, B. Custers, E. Fosch-Villaronga (eds.), T.M.C. Asser Press, 2022.

Eckstein G., D'Andrea A., Marshall V., O'Donnell E., Talbot-Jones J., Curran D., O'Bryan K., *Conferring Legal Personality on the World's Rivers: A Brief Intellectual Assessment*, "Water International" 2019, Vol. 44, No. 6–7.

European Commission, Directorate-General for Communications Networks, Content and Technology. Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM/2021/206 final), https://eur-lex.europa.eu/legal-content/EN/ALL/?uri=CELEX:52021PC0206.

European Parliament resolution of 16 February 2017 with recommendations to the Commission on Civil Law Rules on Robotics (2015/2103(INL)), OJ C 252 (2018).

Floridi L., Taddeo M., *Romans Would Have Denied Robots Legal Personhood*, "Nature" 2018, Vol. 557, No. 309.

Gerber D.J., *Comparing Procedural Systems: Toward an Analytical Framework*, [in:] *Law and Justice in a Multistate World: Essays in Honor of Arthur T. Von Mehren*, J.A. Nafziger, S. Ardsley (eds.), Transnational Publishers, 2002.

Grimalt-Álvaro C., Couso D., Boixadera-Planas E., Godec S., *"I See Myself as a STEM Person": Exploring High School Students' Self-Identification with STEM'*, "Journal of Research in Science Teaching" 2022, Vol. 59, No. 5.

Guide on Article 6 of the European Convention on Human Rights. Right to a Fair Trial (Civil Limb), Council of Europe – European Court of Human Rights, 2022, https://www.echr.coe.int/documents/guide_art_6_eng.pdf.

Hamilton Sh.N., *Impersonations: Troubling the Person in Law and Culture*, Toronto 2009.

Harari Y.N., *Homo Deus. A Brief History of Tomorrow*, London 2017.

Kahn P.H., Shen S., *NOC NOC, Who's There? A New Ontological Concept (NOC) for Social Robots*, N. Budwig, E. Turiel, P.D. Zelazo (eds.), Cambridge 2017, pp. 106–122.

Kerrigan Ch., *Introductory Essay*, [in:] *Artificial Intelligence. Law and Regulation*, Ch. Kerrigan (ed.), Cheltenham (UK)–Northampton (USA) 2022.

Kurki V.A.J., *Why Things Can Hold Rights: Reconceptualizing the Legal Person*, [in:] *Legal Personhood: Animals, Artificial Intelligence and the Unborn*, V.A.J. Kurki, T. Pietrzykowski (eds.), Cham 2017.

Linarelli J., *Artificial General Intelligence and Contract*, "Uniform Law Review" 2019, Vol. 24, No. 2.

Lindley C.A., *Synthetic Intelligence: Beyond Artificial Intelligence and Robotics*, [in:] *Integral Biomathics: Tracing the Road to Reality*, L.P. Simeonov, L.S. Smith, C.A. Ehresmann (eds.), Heidelberg–Berlin 2012.

Lozada-Pimiento N., *AI Systems and Technology in Dispute Resolution*, "Uniform Law Review" 2019, Vol. 24, No. 2.

MacDorman K., Ishiguro H., *The Uncanny Advantage of Using Androids in Cognitive Science Research*, "Interaction Studies" 2006, Vol. 7.

Mahler T., *Regulating Artificial General Intelligence (AGI)*, [in:] *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, B. Custers, E. Fosch-Villaronga (eds.), The Hague 2022.

McCrudden Ch., *Human Dignity and Judicial Interpretation of Human Rights*, "European Journal of International Law" 2008, Vol. 19, No. 4.

Novokmet A., Tomičić Z., Vinković Z., *Pretrial Risk Assessment Instruments in the US Criminal Justice System – What Lessons Can Be Learned for the European Union*, "International Journal of Law and Information Technology" 2022, Vol. 30, No. 1.

Open Letter to the European Commission Artificial Intelligence and Robotics, http://www.robotics-openletter.eu/.

Pagallo U., *The Laws of Robots. Crimes, Contracts and Torts*, Dordrecht–Heidelberg–New York–London 2013.

Pietrzykowski T., *Towards Modest Naturalization of Personhood in Law*, "Revus" 2017, Vol. 32.

Prifti K., Stamhuis E., Heine K., *Digging into the Accountability Gap: Operator's Civil Liability in Healthcare AI-Systems*, [in:] *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, B. Custers, E. Fosch-Villaronga (eds.), The Hague 2022.

Schirmer J.E., *Artificial Intelligence and Legal Personality: Introducing "Teilrechtsfähigkeit": A Partial Legal Status Made in Germany*, [in:] *Regulating Artificial Intelligence*, T. Wischmeyer, T. Rademacher (eds.), Cham 2020.

Sharma P., Gaur S., Dashora D., *Impact of ICT Support on E-Governances Services*, [in:] *Computing and Network Sustainability*, Sh.L. Peng, N. Dey, M. Bundele (eds.), Singapore 2019.

Shaw P., *Ethics*, [in:] *Artificial Intelligence. Law and Regulation*, Ch. Kerrigan (ed.), Cheltenham (UK)–Northampton (USA) 2022.

Smits J., Borghuis T., *Generative AI and Intellectual Property Rights*, [in:] *Law and Artificial Intelligence: Regulating AI and Applying AI in Legal Practice*, B. Custers, E. Fosch-Villaronga (eds.), The Hague 2022.

Solaiman S.M., *Legal Personality of Robots, Corporations, Idols and Chimpanzees: A Quest for Legitimacy*, "Artificial Intelligence and Law" 2017, Vol. 25.

Szekely J., *Lawyers and the Machine. Contemplating the Future of Litigation in the Age of AI*, "Acta Universitatis Sapientiae: Legal Studies" 2019, Vol. 8, No. 2.

Teichman J., *The Definition of Person*, "Philosophy" 1985, Vol. 60, No. 232.

Ulenaers J., *The Impact of Artificial Intelligence on the Right to a Fair Trial: Towards a Robot Judge?*, "Asian Journal of Law and Economics" 2020, Vol. 11, No. 2.

Virdee T., *Understanding AI*, [in:] *Artificial Intelligence. Law and Regulation*, Ch. Kerrigan (ed.), Cheltenham (UK)–Northampton (USA) 2022.

Watson S.M., *The Corporate Legal Person*, "Journal of Corporate Law Studies" 2019, Vol. 19, No. 1.

Whittam S., *Mind the Compensation Gap: Towards a New European Regime Addressing Civil Liability in the Age of AI*, "International Journal of Law and Information Technology" 2022.

Zekos G.I., *Robo-Justice*, [in:] *Advanced Artificial Intelligence and Robo-Justice*, G.I. Zekos, (ed.), Cham 2022.

DR. HABIL. ISTVÁN AMBRUS, ASSOCIATE PROFESSOR
Eötvös Loránd University | DOI: 10.32041/9788367149372/3

Chapter 3. Substantive Criminal Law and Artificial Intelligence

3.1. Introduction

Nowadays, it can be stated without exaggeration that the issue of *artificial intelligence* (hereinafter "AI") is one of the leading topics in the news and everyday talk, especially on social media. As for the related scientific discourses, the issue was intensely examined even much earlier by those disciplines closely linked to AI as a technology. The term itself, in its modern sense, originated in 1956, according to the relevant scientific status quo.[1] Thus, primarily engineering, natural sciences and the accompanying disciplines interconnected to them, such as cognitive psychology, invented for themselves and tried to define and classify AI more than a half-century ago.[2] However, at the end of the second decade of the 21st century, the phenomenon has become an unavoidable component of everyday life to such an extent that it necessarily increased the attention of researchers in almost every discipline, thus among those in classical social sciences such as sociology and economics (also showing the

[1] S. Russel, P. Norvig, *Artificial Intelligence: A Modern Approach*, Prentice Hall, Upper Saddle River, New Jersey, 2010, p. 1.

[2] However, it should be emphasised that the idea of building intelligent machines is already a concern for ancient and medieval thinkers.

signs of natural sciences) as well.[3] The same applies to legal literature and the legal profession, whose main task cannot be regarded only as establishing the law's theoretical paradigm related to AI. Besides, for the future application of law, nearly the same significance is attached to the need for developing appropriate legal frameworks. Furthermore, the prognosis, according to which AI will reshape the pragmatic application of law and ultimately the whole legal life, is of far from negligible importance.

Even a superficial review of the related recent Hungarian legal literature shows that special attention is paid to the topic of AI concerning almost all branches of law.[4] Hence, besides general legislative/liability issues, we can read, for example, about classic civil law (in the narrow sense), labour law, copyright, data protection and specific, new issues related to the operation of the legal profession. In addition, studies analysing criminal legal problems have also been published, of course.[5] The present study is part of such works, as it examines *the connection between AI and (primarily in a narrow sense, i.e., substantive) criminal law*. After addressing some general definition attempts and the main characteristics of AI and the related grouping options, a brief identification of the more frequent fields of use follows; furthermore, I will discuss the issues that are most relevant from the aspect of criminal law (in a broader sense). Examining substantive criminal legal matters arising in relation to AI form the core of the study, under which, following the notions related to the subject of criminal liability, I present the relevant problems concerning the concept of a criminal offence and criminal sanction and the special part of the Act C of 2012 on the Criminal Code. I then examine those cases where AI appears as an object of the offence. In the context of the discussion, in addition to

[3] In view of this, it is therefore reasonable to conclude that AI is a universal issue that has attracted the interest of almost all disciplines.

[4] See W. Barfield, *Towards a Law of Artificial Intelligence*, [in:] *Research Handbook on the Law of Artificial Intelligence*, W. Barfield, U. Pagallo (eds.), Cheltenham, 2018, pp. 2–39.

[5] In Hungarian literature see B. Miskolczi, Z. Szathmáry, *Büntetőjogi kérdések az információ korában: Mesterséges intelligencia, Big Data, profilozás*, Budapest 2018, pp. 39–104.

the results from the related Hungarian, as well as Anglo-Saxon and German literature, I will draw up and evaluate the findings of recent domestic legal literature. I do so in the hope that this work may not only inspire further scientific thinking but also may, perhaps, be used to serve as a support for future criminal legislation, especially in Poland, Hungary, and any other European counties as well.

3.2. Options for the definition and classification – general characteristics of AI

At the present point of the study, seeking to determine the concept of AI in general, following the attempts to define the identified issue in other fields of science and specifically in the legal literature and the options for classification, I highlight its main characteristics. As we will see, in particular, distinguishing the categories of *strong and weak AI*, and from the relevant features the ability of machine learning and, as a result, the distinction between merely *automatic* and truly *autonomous actions,* and between robots staying on the ground of *determinism* or *indeterminism,* can be regarded as the most important in this context.

Several descriptions – both more detailed and simplistic – have been published in the Hungarian literature on AI. In his recent study, Dániel Eszteri describes literature findings based on engineering and the on history of philosophy as well. In the context of the former, he refers to AI as the science and engineering practice of producing intelligent machines. The philosophical point of view describes AI humanly; furthermore, as a system that is thinking and acting rationally.[6] According to György Lőrincz, in his recent study, AI "[shall mean] those hardware/software systems that can solve even difficult problems in a 'human way,' to choose between decision alternatives utilising conclusions characteristic of the human way of thinking."[7]

[6] D. Eszteri, *Hogyan tanítsuk jogszerűen a mesterséges intelligenciánkat?*, "Magyar Jog" 2019, Vol. 66, No. 12, pp. 669–681, at p. 670.

[7] G. Lőrincz, *A mesterséges intelligencia alkalmazásával hozott döntés jogi megítélésének egyes kérdései*, "Gazdaság és Jog" 2019, Vol. 28, No. 3, pp. 1–7, at p. 1.

However, US academic legal writers point out that defining the exact concept of AI that applies to all fields of science is not such a simple task in the present circumstances.[8] As a result, the above-mentioned Hungarian position, which regards creating the precise definition as an objective for the present and future investigations, suggests an essentially normative approach to the AI concept, according to which "AI shall mean what we identify as such" can be fully supported.[9] However, in the hope of a future definition, and even for its codification based on its "translation" into legal language, it is worth briefly reviewing the possible approaches and criteria. The European Parliament resolution of 16 February 2017 also underlines that "there is a need to create a generally accepted definition of a robot and AI that is flexible and is not hindering innovation," pursuant to point C of the Introduction, is a well-founded claim. This was the issue addressed in the Communication from the European Commission *Towards a common European data space* issued on 25 April 2018, which defined AI as it "refers to systems that display intelligent behaviour by analysing their environment and taking actions – with some degree of autonomy – to achieve specific goals."[10] Recently, in 2022, the Report on Artificial Intelligence in a Digital Age [2020/2266(INI)] was accepted by the EU.

However, in addition to the aforementioned general definition (intended to be a minimum, of course), several other identifications and classifications using different approaches may be possible.

In conceptualisation, it could be considered that we rely, in the first place, on the results of the already mentioned *cognitive psychology* (*and neuroscience*) which has been dealing with AI for a long time. In one of his essential works, Csaba Pléh, the Hungarian doyen of the discipline, which works in symbiosis with maths and computer science, differentiates between *strong* and *weak* AI, following in the footsteps of Alan Turing and John von Neumann, the fathers

[8] M. Hatfield, *Professionally Responsible Artificial Intelligence*, "Arizona State Law" 2019, Vol. 51, No. 3, pp. 1057–1122, at p. 1064.

[9] B. Miskolczi, *Az MI-vel kapcsolatos büntetőjogi felelősségi kérdések*, [in:] B. Miskolczi, Z. Szathmáry, *op. cit.*, note 5, p. 58.

[10] Artificial Intelligence for Europe, Brussels, 25.04.2018, COM(2018) 237 final.

of computer science, and explains that, according to "the *hypothesis of strong AI,* manipulation of symbols qualifies as human thinking. Our task is to produce machines capable of performance comparable to human beings, and afterwards, the principles of the programme set up for this purpose will actually explain and qualify as human thinking." In contrast to this perception, there are as many as five different counter-arguments, according to which only weak AI – i.e., not having autonomous conscience – could be created at present.[11] László Mérő represents a more characteristically negative position on the issue when he explains that "[he] definitely do[es] not agree with the supporters of strong artificial intelligence, although not based on the usual holistic and not even the dualistic (proclaiming the duality of material and soul) arguments." According to his reasoning, "there are *modus operandi* of human thinking based on principles other than the pure rational manipulation of symbols."[12]

According to the definition found in the recent Anglo-Saxon literature, AI is essentially a device modelled on human thinking that is able, for instance, to plan and develop a strategy, make a decision and give reasons for it. They mention playing chess, language translation and driving as examples. However, as Harry Surden highlights, the phenomenon called "AI" cannot be regarded nowadays as literally intelligent; that is, thinking like a human being, although it can produce useful results given its high performance and speed. Ultimately, "only" *weak AI* is available; the more sophisticated version (*strong AI*) is nothing more than a mere desire right now.[13]

Another study also contrasts AI with the human brain on the basis that while the former is usually capable of fulfilling a single or only a few functions (*narrow AI*), the human mind is capable of even very abstract thinking in numerous areas.[14]

---

[11] C. Pléh, *A megismeréstudomány alapjai: Az embertől a gépig és vissza*, Typotex, Budapest 2013, pp. 193–197.

[12] L. Mérő, *Új észjárások: A racionális gondolkodás ereje és korlátai*, Tercium, Budapest 2001, pp. 248.

[13] H. Surden, *Artificial Intelligence and Law: an Overview*, "Georgia State University Review" 2019, Vol. 35, No. 4, pp. 1305–1337, at pp. 1307–1309.

[14] D. Ben-Ari, Y. Frish, A. Lazovski, U. Eldan, D. Greenbaum, *Danger, Will Robinson? Artificial Intelligence in the Practice of Law: An Analysis and Proof of*

Recent German literature also draws attention to the difficulties of conceptualisation. According to the co-authors Staffler and Jany, AI as a term is primarily used in everyday life, thanks to fashion and for marketing purposes, in an inflationary way for almost all modern computing devices. They also point out that although even the term "intelligence" does not have a generally accepted definition, with respect to legal analysis, it can be based on the colloquial definition, namely, in their view, the ability to make cognitive efforts. Similarly to the authors cited earlier, they state that the mechanism of a computer that makes even many and quick decisions based on the data entered cannot yet be regarded as "intelligent," only as an automated operation. We cannot talk about intelligence even if the machine/robot collects and processes the data through its sensors autonomously. Therefore, AI only exists where the applied algorithm arrives at unpredictable, i.e., *a priori* unknown, results. However, this feature is valid for far fewer devices than those to which the term AI is applied to in practice. Machine learning and "deep" learning – which produces results, not even necessarily predictable by humans, from unstructured data – are considered to be the most characteristic feature of true AI.[15]

Finally, the aspect also emphasised in a new handbook on the legal research of AI, according to which a sharp distinction must also be drawn between devices operating solely on the basis of automatisms and those acting in a genuinely autonomous manner (robots, if you prefer) needs to be highlighted. An example for the former are those machines processing a large amount of data ("*big data*") in a short period and making decisions as a result of this process but based on the previously entered criteria (see the next point for concrete examples). The main characteristic of the operation of such robots is that, after running the encoding algorithms, they usually arrive at a predictable result; thus, they stand on the ground of

Concept Experiment, "Richmond Journal of Law & Technology" 2017, Vol. 23, No. 3, pp. 1–55, at p. 8.

[15] L. Staffler, O. Jany, *Künstliche Intelligenz und Stafrechtpfleg: eine Orientierung*, "Zeitschrift für Internationale Strafrechtsdogmatik" 2020, Vol. 15, No. 4, pp. 164–177, at p. 166.

determinism. The latter, on the other hand – the true autonomous agents – after modelling human thinking in at least at one sub-area, are usually unpredictable (*indeterministic*) as regards their output.[16]

3.3. General and criminally relevant fields of use of AI

The European Commission's White Paper, issued in February 2020 sets out in the introduction that AI "will change our lives by improving healthcare […], increasing the efficiency of farming, contributing to climate change mitigation […], improving the efficiency of production systems […], increasing the security of Europeans, and in many other ways that we can only begin to imagine."[17]

Bearing these aspects in mind, in this brief introductory chapter, I only refer to the fact of how widely AI-technology can be used even today. For example medical use, primarily in relation to special surgeries requiring microscopical precision operations, is of significant importance. Products and services based on AI technology used in connection with cryptocurrencies (such as bitcoin and others) and smart contracts are the most significant innovations of the 21st century.[18] Delivering targeted advertisements to consumers on social media sites is also carried out using AI, as are the increasingly widespread facial-recognition systems used by state authorities as well as private companies. It is essential to highlight these-called AI advisers and, of course, the expectedly growing significance of autonomous (self-driving) vehicles in the carriage of goods and persons. Moreover, the role of drones is also growing as they (and

[16] W. Barfield, *op. cit.*, note 4.

[17] European Commission White Paper on Artificial Intelligence – A European Approach to Excellence and Trust, Brussels, 19.02.2020, COM(2020) 65 final.

[18] See E.P. Pacy, *Tales from the Cryptocurrency: On Bitcoin, Square Pegs, and Round Holes*, "New England Law Review" 2014, Vol. 49, No. 1, pp. 121–144; M. Abramowicz, *Cryptocurrency-based Law*, "Arizona Law Review" 2016, Vol. 58, No. 2, pp. 359–420.

the soft- and hardware controlling them) become increasingly more sophisticated.[19]

As for issues relevant to criminal law in the broader sense, as well as other criminal sciences, we should mention those stock market operations directed by AI that are used for money laundering or even for financing terrorism, typically committed by criminal organisations. The Dark Web, as a barely traceable, secret online platform, is the forum for numerous criminal offences relating to weapons, drugs, sexual exploitation of children, etc. Naturally, the commission of criminal offences against property, such as fraud, or today primarily information system fraud, may arise as well. AI in the hands of investigative authorities can be a useful tool for computer risk analysis and profiling; as a counterpoint to this, criminals, for example, use *deepfake* for misusing personal data, harassment, and offences against property. In addition to traffic offences, self-driving vehicles could be used to commit crimes such as homicide. In relation to drones, the commission of criminal offences constituting different kinds of breaches of confidentiality, misuse of personal data or offending certain fundamental rights could arise.

3.4. AI and criminal liability

3.4.1. AI AS THE PERPETRATOR ITSELF

In the present study, I primarily undertook a criminal doctrinal analysis; therefore, I will not examine in detail the preliminary question (being essentially legislative) of whether criminal liability needs to be established for infringements attributable to the operation of AI. That decision is subject to the position of the legislature, of course. However, it can be stated without further examination of the matter that the social resolution is positive; for instance, in the case of a fatal accident caused by a self-driving car. However, as a study pointed out, according to the prognosis regarding such vehicles

[19] S.J. Barela, *Legitimacy and Drones: Investigating the Legality, Morality and Efficacy of UCAVs*, Farnham 2015.

assuming their widespread use, the number of traffic infringements can be significantly reduced due to the elimination of human failure and carelessness. With this in mind, it should at least be said that "if the number of accidents with a harmful result (and thus offences against traffic generally) decreased significantly, concerning the *ultima ratio* nature of criminal law, whether it may fit into the concept of allowable risk as a ground of justification to keep those accidents that will be caused by self-driving cars in the future and with the same outcome, unpunished."[20] By the abstraction of this thought and applying it to the issue of AI, the question can be summed up briefly as: *if AI makes life safer and more comfortable, then will we paradoxically be disadvantaged by e-criminal law instruments at all costs, rather than waiving (at least part of) them?* Insofar as those professionals dealing with the development of AI need to work in the shadow of prison, their – understandable – caution may hold back innovation, which, in turn, may ultimately result in achieving a safe level of society, but with a significant delay.

Our baseline is that criminal infringements generated by AI require a state response. First, it is necessary to clarify the issue as to who shall be the subject of criminal liability or who (perhaps what) should be regarded as the perpetrator of the offence.

In recent English criminal law, *four models* were developed on this issue. Following the work of Thomas C. King and his co-authors:

1) the direct (individual) liability of AI (direct liability),
2) liability similar to indirect perpetration (preparation-by--another),
3) the responsibility of a superior character (command responsibility), and
4) liability for the failure to meet an obligation to provide due care (natural probable consequence) may be examined.[21]

[20] I. Ambrus, G. Kovács, I. Németh, *Automomous vehicles and the prospective change in criminal liability*, "Ügyészek Lapja" 2018, Vol. 25, No. 6, pp. 85–92, at pp. 86–87.

[21] T.C. King, N. Aggarwal, M. Taddeo, L. Floridi, *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, "Science and Engineering Ethics" 2020, Vol. 26, No. 1, pp. 89–120, at pp. 108–110.

Below, I will examine these four categories, taking domestic regulation and the doctrinal aspects of criminal law into account.

The model of the *criminal liability of AI as an individual entity* in the light of the prevailing literary perceptions as well as *de lege lata* must undoubtedly be rejected. The reason for this lies in the fact that our continental criminal law considers only (blameworthy) human conduct committed by guilt (intentionally or negligently) as criminal offences. The philosophical basis for this viewpoint can be traced back to John Locke, who pointed out that only those who have a will by which they can be addressed by reward and punishment should be regarded as human beings. They must therefore understand the importance of rules and must be able to feel positive and negative human emotions.[22] At present, this indeed cannot be established in relation to AI. The viewpoint of Kant was close to that of Locke, according to which "volition is the source of laws, arbitrariness is the source of maxims; arbitrariness is free in human beings; […] experiences have shown that *humans as sensual beings* can choose not only *according to* the law but also *against* the law."[23] In the new German literature, Gless and her co-authors, speaking about AI, claim that "a robot is not conscious of its freedom, is not able to think about itself as an entity having a past and a future. Moreover, it could not be able to understand the importance of rights and obligations."[24]

According to the domestic criminal law in force and the related academic views, the possibility of establishing the individual criminal liability of AI is excluded because the concept of criminal offence under section 4(1) of the Criminal Code – primarily by requiring the subjective side – provides only for the criminal liability of a natural person. Legal literature restricts the concept of act exclusively to human actions as well (see, in that regard, the next chapter of the study).

---

[22] J. Locke, *An Essay on Concerning Human Understanding*, Pennsylvania, 1690/1999, p. 325.

[23] I. Kant, *The Metaphysics of Morals*, Gondolat, Budapest, 1797/1991, pp. 320–321.

[24] S. Gless, E. Silverman, T. Weigend, *If Robots Cause Harm, Who is to Blame? Self-Driving Cars and Criminal Liability*, "New Criminal Law Review" 2016, Vol. 19, No. 3, pp. 412–436, at pp. 423–424.

The second model, resembling on the concept of *indirect perpetration* under section 13(2) of the Criminal Code, favours the criminal liability of the person operating the AI, i.e., the so-called "human operator." This means that the operator uses AI as an instrument to commit the objective side of the criminal offence. This solution could form a basis for establishing criminal liability, even in the current domestic criminal legal system, albeit only by applying several corrections. On the one hand, subject to the exhaustive list in the provision on indirect perpetration (infancy, mental disorder, coercion and threat, error) in this case individual (general) perpetration [section 13(1) of the Criminal Code] could be considered at most. However, as concerns to AI that is acting in a truly autonomous manner, to establish its use as an instrument may be at least disputable, as the operation of "strong" AI possessing "quasi-" consciousness – although we saw that it could not be construed under the dominant perception – barely resembles the act of a person without accountability. In addition, whereas an ideal AI, as we saw, can even make unpredictable decisions, the guilt of the human operator (his/her intention or at least negligence) may be questionable in this respect. To eliminate this counter-argument, we may be helped by the dogmatic concept of *actio libera in causa*, long known in German and Hungarian criminal law. Accordingly, criminal liability may not be established for the time when the harmful result occurred, but based on the so-called *blameworthy pre-conduct* that induced the irresponsible state of the offender. Thus, for example, if somebody consumes narcotic drugs, accepting the fact that he/she may commit a crime under its effect, his/her criminal liability shall be established even if he/she had a mental disturbance, making him/her not responsible for his/her actions due to that drug at the time of committing the offence. Similarly, the negligent version of *actio libera in causa* may be construed (see the case of the mother who, despite knowing her dangerous sleep patterns, nonetheless puts her crying baby by her side during the night and crushes him/her to death while asleep).[25] With regard to

[25] See P. Angyal, *A magyar büntetőjog tankönyve*, Athenaeum, Budapest 1909, p. 383; S. Beck, *Neue Konstruktionsmöglichkeiten der actio libera in causa*,

AI, for example, this process may relate to tracing back to the time of purchase or the last servicing of the instrument operated by him/her and whether the responsible person was truly in control of the AI device at all times.

The *responsibility of a superior/managerial character*, as a third model similar to the Anglo-Saxon approach, is akin to the rules of secondary, increased responsibility of a commander, also known primarily from military criminal law in our country. From the scope of such domestic provisions, the rule under section 130(2) of the Criminal Code contains similar features to those relating to the second model, according to which also the person giving the order shall be liable for a criminal offence committed upon an order as an offender if the subordinate knew that carrying out the order would constitute a criminal offence. Otherwise – that is, if the soldier erred, in that carrying out the order would constitute a delict – the person giving the order shall be liable as an indirect offender. That quite unusual situation can be observed here; the legislature involves itself the command as a normative act, creating a quasi- exemption, within the scope of the (individual/general) conduct of an offender in compliance with the statutory definition. However, imposing such a requirement cannot be equated with the objective side of the statutory definition in the Special Part of the Criminal Code fulfilled by the subordinated soldier. Therefore, in this case, there are (at least) two offenders of the criminal offence by force of law. Moreover if the soldier *errs in law*, similarly to the second model, it constitutes indirect perpetration. Such regulation could be solved according to the doctrines discussed in the previous point, except that it does not seem inconceivable, *mutatis mutandis,* that an error in law concerning AI, as "[the] idea of creating the perfectly law-abiding AI would require from the programmer, as an absolute legal positivist, to be able to map, at the level of signal processing, all possible implementations of statutory definitions under the Criminal Code – together with the underlying legislation of open statutory definitions – in the knowledge base of AI along with the whole dogmatic system

"Zeitschrift für Internationale Strafrechtsdogmatik" 2018, Vol. 13, No. 6, pp. 204–2011, at p. 204.

of application of rules determining legal interpretation,"[26] which seems obviously an impossible task. Of course, introductory legal provisions can be entered in AI. Thus, if the "commander" – for instance, the owner/operator of a self-driving car – gives the order to the AI controlling the vehicle to get him/her in half an hour from Szeged to Budapest, the pre-programmed protection mechanism of the AI, knowing the distance, KRESZ (Hungarian decree on road traffic rules) and other traffic regulations, shall obviously be obliged to refuse it.

The last *liability model based on the breach of duty of care* shows a private law character, which can be attributable to its Anglo-Saxon origin. However, the initiatives of this model, of failure to give due care, can be found in several parts of the current Hungarian criminal law. However, before describing these, it can be mentioned as an interesting fact that this concept shows strong similarities to the *Roman Law* provisions on slaves, as the owner was responsible for actions of a fugitive slave.[27] In Hungarian criminal law, cases under section 145 of the Criminal Code under the title of *responsibility of a military or official superior* within the scope of crimes against humanity (Chapter XIII of the Criminal Code) are similar. The essence of these is, for example, the superior, among others, shall also be subject to the same penalty as the perpetrator of the offence against humanity, if the person under his command and control committed such a crime and the superior knew or, owing to the circumstances at the time, should have known of the commission of the delict, etc. The following provisions of the Criminal Code cite similar but less rare cases. Section 397 of the Criminal Code, under the name of *failure to comply with the supervisory or control obligation related to budget fraud,* as a *sui generis* delict, orders the punishment of the executive of an economic operator or a member or worker with the power to control or supervise, who fails to comply with his/her supervisory or control obligation and, by doing so, enables

[26] Z. Szathmáry, *Az MI cselekményeinek ontológiai kérdései*, [in:] B. Miskolczi, Z. Szathmáry, *op. cit.*, notes 5, 80.

[27] W.W. Buckland, *The Roman Law of Slavery: The Condition of the Slave in Private Law from Augustus to Justinian*, Cambridge 1908, p. 105.

a member or worker of the economic operator to commit budget fraud while pursuing the activities of an economic operator. In relation to corruption delicts, both the intentional and negligent version are punishable, under *active bribery regarding a public officer*; for example, the executive of an economic operator or a member with the power to control or supervise, who makes it possible for his/her subordinate to commit "ordinary" bribery regarding a public officer, provided that the performance of his/her supervisory or control obligation could have prevented the commission of the criminal offence [sections 293(4)–(5) of the Criminal Code].

Applying these solutions to infringements relating to AI could not only raise the responsibility of the person using the device operated by AI at the time of the infringement but also of the operator or the owner or even of the manufacturer. However, the increasing private law character of criminal law should be outlined. The owner and, mostly the manufacturer can primarily not be a natural person but a legal one, whose individual criminal liability is not recognised by our domestic criminal law (nonetheless, criminal measures relating to legal persons, have been present for nearly twenty years in the Hungarian legal system).

3.4.2. THE "ACT" OF AI

As mentioned above, establishing the individual criminal liability of AI may be excluded at present. Nonetheless, it is not useless to review the aspects for or against recognising the "act" of AI as one falling within the scope of criminal law. The general approach today follows the so-called reduced concept of an act, according to which it is an effective and wilful human conduct.[28] However, it is not inconceivable that this entrenched approach, also with regard to the abovementioned criminal policy factors, may require revision. Thus, in the domestic literature, Szathmáry represents the position according to which "the concept of an act has to be revised in the

[28] B. Gellér, I. Ambrus, A *magyar büntetőjog általános tanai I*, Budapest 2019, pp. 180–185.

future."[29] This revision may mean further reduction at the same time, but there is quite a difference in how exactly that is expected to take place. The act, as the *genus proximum* of a criminal offence, must be capable of having any further conceptual element of a criminal offence based on it. Hence, to create the correct future concept of an act, it is appropriate to review the results of the related research by the more prominent domestic and foreign authors.

In the Hungarian literature, we may highlight the viewpoint of co-authors Kádár and Kálmán, according to which "[the] act is a conscious and wilfully aimed human intrusion on nature and society in the form of a defined external behaviour (act or omission)."[30] Recently, according to Ervin Belovics, an effective act shall mean "the act is capable of, based on general life experience, creating a harmful consequence."[31]

Under the social concept of an act, which has a strong presence in German criminal law even today, a human act is what is governed or may be governed by volition and is socially significant.[32] On the other hand, according to Kindhäuser, there is no act if an action controlled by volition appears to be absent on the part of the subject and, therefore, it is physically impossible to achieve the goal. He takes the example of different convulsive states, sleeping and the case of absolute force, i.e., *vis absoluta*.[33]

Anglo-Saxon authors primarily take a negative approach to the issue. Thus, according to Hart, there is no voluntary bodily movement and therefore no action if the subject behaves in a manner for which there is no reason.[34] As per Williams, an act is not wilful if the perpetrator is unable to avert it.[35] According to Asworth, the act in a criminal law sense is missing if the perpetrator did not

[29] Z. Szathmáry, *op. cit.*, notes 26, 77.
[30] M. Kádár, G. Kálmán, *A büntetőjog általános tanai*, Budapest 1966, p. 276.
[31] E. Belovics, *Büntetőjog I. Általános rész*, Budapest 2017, p. 158.
[32] J. Wessels, W. Beulke, H. Satzger, *Strafrecht Allgemeiner Teil: Die Straftat und ihr Aufbau*, C.F. Müller, Heidelberg 2013, p. 42; J. Kaspar, *Strafrecht Allgemeiner Teil: Einführung*, Baden-Baden 2017, p. 38.
[33] U. Kindhäuser, *Strafrecht Allgemeiner Teil*, Baden-Baden 2017, p. 54.
[34] H.L.A. Hart, *Punishment and Responsibility*, Oxford 1968, p. 103.
[35] G. Williams, *Textbook of Criminal Law*, London 1983, p. 29.

control his/her act and he/she were not even able to control it at the time relevant to the establishment of the statutory definition. However, he recalls that the related case law follows a somewhat more permissive approach, as it considers not only complete control but also "some degree" of control as sufficient to establish an act.[36] As Allen highlights in his recent textbook, even in cases of the so-called *strict liability* in Anglo-Saxon criminal law, which operates with the reverse burden of proof and results in severe penalties in practice, it is necessary to prove wilfulness in criminal proceedings.[37]

To summarise, the partial conclusion from the above is that both the continental and the Anglo-Saxon systems essentially base their concept of criminal offence on human conduct. Therefore, to establish the "liability" of AI, it is primarily this element that should be loosened, thus the operation *in concreto* and not directly departing from a human being but retraceable to human conduct should be recognised as an act. Efficiency (especially its potential presence) would not generate any problem in relation to AI, as producing an effect in the outside world, is considered as given in relation to any machine operation, thus, naturally to the "activity" of AI. On the other hand, the criterion of wilfulness should be reconsidered, as in principle it means "with regard to human consciousness and will," i.e., a psychological relationship – free from assessment, hence not covered by guilt – which may be absent in the case of the "individual" process of AI. Therefore, to establish criminal liability in connection with AI, the reduced concept of act could be further developed – more precisely, further reduced – by not leaving wilfulness or potential wilfulness as a conclusive factor but only one of those to be decided by the court according to the examination carried out *in concreto*. Nonetheless, this solution is a significant threat to the frameworks of criminal law based on the principle of liability for action.

[36] A. Asworth, *Principles of Criminal Law*, Oxford 1995, p. 98.
[37] M.J. Allen, *Criminal Law*, Oxford 2017, p. 124.

3.4.3. AI AND COMPLIANCE WITH STATUTORY DEFINITIONS

If we consider that an act, in a criminal law sense, may be established by one of the above solutions for assessing infringements regarding AI as a delict, the next stage of criminal liability is to comply with a statutory definition, i.e., basically, to fulfil the statutory criteria of an offence specified in the Special Part of the Criminal Code. Szathmáry, *de lege ferenda,* correctly states that "by the exclusion of culpability, no subjective element precludes the possibility that the act may comply with a statutory definition. Concerning further thoughts on compliance with a statutory definition, the human factor relating to the subject of the act – set out in statutory definitions as 'who' – must be eliminated of course, which shall be replaced in our mind, conditionally, by the autonomous decision-making system."[38] Naturally, this could be applied to future legislation and the dogmatic reflection of it; however, *de lege lata*, it seems somewhat lenient to me, because a statutory definition has, pursuant to the prevailing literary perception, both an objective and a subjective side. On the basis of the present legislation, it is hardly disputable that the subject (or under a more recent and appropriate term, the existence of the criteria necessary for the perpetration) of a crime is involved in the statutory definition. Thus, the abovementioned approach is only acceptable for the objective side of a statutory definition (moreover, only for a part of it) at present; consequently, for the future assessment of infringements relating to AI, the legislation needs to be changed.

3.4.4. AI AND UNLAWFULNESS (DANGER TO SOCIETY)

Unlawfulness (according to section 4(1) of the Criminal Code and individual academic writers following the wording of the law: *danger to society* [which represents the material side of unlawfulness]) is the third element of the concept of crime, which is mostly accompanied by compliance with statutory definitions. In the occurrence of

[38] Z. Szathmáry, *op. cit.*, notes 26, 78.

a ground of justification, the unlawfulness of the act in compliance with a statutory definition is missing; consequently, the offence does not constitute a criminal offence. In this context, without the ambition to be exhaustive, I only wish to deal with the most characteristic grounds of justification.

The right to *justifiable defence* (sections 21–22 of the Criminal Code) originates, on the one hand, from the ethical norms of society and the power of the state and, on the other hand, from the right to self-defence ensured by "natural law." Concerning questions of obstacles to punishability arising in relation to AI, it can be highlighted that the operation of AI may occasionally pose a threat to the legal objects to be defended by a justifiable defence, such as (human) life, property and privacy. Concerns may therefore arise as to whether AI can necessarily avoid, in all circumstances, persons from being injured by its activity. With regard to self-driving vehicles and drones as typical examples of the area of AI, the following can be highlighted. For instance, via Google Maps, the self-driving car can hit a pedestrian and cause an accident. A similar example is that a drone sent by Amazon drops a package on the intended recipient's head. Devices controlled by AI can represent a physical threat to property, can harm the exclusive right to possession. Hence, for example, the aforementioned Google Maps must (should) be obliged to flag and update incorrect information. Finally, such tools also pose a danger to private life. They can easily monitor individuals in situations where the same would be nearly impossible for a human. Self-driving vehicles raise special problems as they not only hold the potential for infringing road traffic rules but also to cause serious accidents.[39]

As regards the Hungarian assessment of justifiable defence in this respect, it should be emphasised that, to give effect to this ground of justification, it is required that the self-driving vehicle does not cause an injury or there is minimal risk thereof based on the entered data; then, with regard to smart technology in the meantime, built-in information. A justifiable defence is subject to

[39] M.A. Froomkin, P.Z. Colangelo, *Self-Defense Against Robots and Drones*, "Connecticut Law Review" 2015, Vol. 48, No. 1, pp. 1–69, at p. 7.

the existence of an *unlawful* attack (or at least a situation directly threatening it) – shall be established where *AI functions as a device in the hands of the offender who is a natural person*.[40] If, for example, a hacker attack aims to kill the victim, programming the AI for this may be regarded as an act (in a criminal law sense) in the same way as if the offender, for example, simply tries to kill the victim with a gun or sabotages the victim's car by tampering with the brakes. This solution is more akin to the offences of causing bodily harm and homicide committed with the incitement of an animal. In this case, the animal itself, since it is neither a legal entity in a private law sense, nor can be regarded as the subject of an offence under criminal law, cannot carry out an unlawful attack. However, it is appropriate to serve as an instrument in an unlawful attack on a human. To this end, this approach could be enforced in the event that AI is used by a human to achieve a harmful result.

With regard to the issue of *necessity* relating to AI, the so-called "tram/trolley problem" may be raised, because with the appearance of robotic technology, programmers need to decide on the ethics to apply, for which the involvement of philosophy and psychology, and the sociology of law are also indispensable. According to the hypothesis, social morality is deducible from the answer to the question, which might also be put into practice in this way.[41]

Under the original scenario, five low-status people are standing on a railway track, and a man is working on the track next to it. A train is rushing unstoppably toward the five people, but the operator can decide to switch track at the last minute, thus (almost certainly) causing the death of the person working alone, although saving the life of the five peoples remaining on the initial track. However, the issue is not new; there is no evident answer acceptable to everyone; there are only arguments in support of the correctness of one particular answer. Thus, the fundamental dilemma is whether one human can be sacrificed to rescue five others. Furthermore, in

[40] Also see Z. Szathmáry, *A büntetőjogi felelősségre vonás akadályai*, [in:] B. Miskolczi, Z. Szathmáry, *op. cit.*, notes 5, 126.

[41] J.D. Greene, *Solving the Trolley Problem*, [in:] *A Companion to Experimental Philosophy*, J. Sytsma, W. Buckwalter (eds.), Chichester 2016, p. 175.

a criminal law sense, at issue is whether the statutory definition of homicide can be fulfilled by a positive action to avert homicide by omission (committed against more than one person).

Under the rules of necessity, as they are currently understood, the answer is yes, against which the duty of states to protect life (and to criminalise, as implemented by the legal definition of homicide) can be raised.

The statutory definition employed by Jason Millar, reflecting further on the problem and the question related to it is thus: "A self-driving car is facing an inevitable collision. It can change its direction in two ways. In the first case it would hit a motorcyclist with a helmet, in the second case, one without a helmet. What is the correct solution?"[42] In this case, the dilemma is whether AI shall necessarily hit the person whose chances of survival are better, though he is the one – in contrast to the motorcyclist who did not wear a helmet – who followed the rules.

However, the tension of the situation can be exaggerated: according to the dilemma attached to the name of Patrick Lin, the entirely self-driving car transporting only one person detects that it either runs into a school bus transporting 28 children, thus risking everyone's life, or, by changing direction, it drives off a cliff, which results in the inevitable death of the given person.[43] If every person had the same chance to survive the collision of the vehicles, the question arises of what should be the line of "thought" for AI.

Should it risk the lives of 29 people in every case, or should it not take the risk and instead drive off a cliff (meaning the inevitable death of the person in the car that goes off the cliff); perhaps it should do a quick calculation and only drive off the cliff where the chances of someone being fatally injured in the accident are greater than 1 in 30? Furthermore: is it ethical if the trader sells the vehicle at a premium, upon payment of which the self-driving vehicle will

[42] J. Millar, *Ethics Settings for Autonomous Vehicles*, [in:] *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, P. Lin, K. Abney, R. Jenkins (eds.), Oxford 2017, p. 21.

[43] P. Lin, *Why Ethics Matters for Autonomous Cars*, [in:] *Autonomous Driving. Technical, Legal and Social Aspects*, M. Maurer, J.Ch. Gerdes, B. Lenz, H. Winner (eds.), Berlin–Heidelberg 2016, p. 76.

be programmed to prioritise in all cases the saving of the life of the person sitting in it?

However, with regard to the ethical aspects, to provide detailed answers to these questions would exceed the scope of this study. Hence, I address only the assessment based on the applicable domestic legislation of necessity. Under section 23(1) of the Hungarian Criminal Code, for unconditional impunity, the harm caused may not be greater than the threat posed by a danger that is both imminent and which cannot be averted by any other means. In effect, this means that if an accidental situation, arising in relation to the running of a self-driving car, raises the possibility of harm to human life, it may fall within the frames of necessity to save the life of at least one or even more people – passengers, the human operator, or the test driver – sitting in the car. In contrast, if the accident threatens the lives of two or more pedestrians, yet there are more pedestrians than there are people in the car necessity as a ground of justification cannot be argued effectively. The case of necessity, excluding culpability, under section 23(2) of the Criminal Code, is not per se applicable in relation to AI as it is not able to experience fright or excitement.

3.4.5. AI AND CULPABILITY

The classic, subjective version of culpability – which as well involves intentional behaviour and negligence – may only be applicable to an offence structure in which we aim to establish the criminal liability of a natural person. If we derive culpability from subjective criteria and view it purely as imputability, its application to AI may be much more convenient. However, in the light of the specific legal nature of criminal law, it primarily penalises conduct that is in violation of the criminal code and is reprehensible to human beings. Thus, *de lege ferenda*, infringements relating to AI may be assessed in another form. In my view, Barna Miskolczi provides a perfect solution to this, according to which "the system of criminal liability relating to the autonomous decision of AI may be established analogically to

the criminal liability of a legal person."[44] As criminal actions against legal persons (e.g., companies) have been known in Hungarian criminal law since *Act CIV of 2001 on Criminal Measures Applicable against Legal Persons* entered into force on 1 May 2004 and, as Miskolczi also points out correctly, the applicability of the sanctions set out in this act due to the amendments in recent years are becoming increasingly less dependent on the criminal liability of the natural person (e.g., managing director, senior manager) committing the act through the legal person. Thus, the liability of AI could be construed in the same way as the fiction of the criminal liability of a legal person. In addition to this reasoning in the field of grounds for justification, such fiction exists in Hungarian criminal law not only in relation to legal persons but also in relation to natural ones Under section 18 of the Criminal Code, the person in a drunken or otherwise intoxicated state through his own fault and carrying out the material side of a criminal offence cannot rely on the existence of a mental disorder under section 17(1) of the Criminal Code, he/she is essentially criminally liable, even if his/her guilt did not exist in the ontological sense. A similar concept may be invoked when substituting the "culpability" of AI.

3.5. Criminal sanctions

In relation to criminal consequences (in practice and in most cases, sanctions, namely, penalties or measures), it should be noted that there is no real problem when the criminal liability of a natural person is established in connection with an offence involving AI. This is because, as a main rule, the court (or, in exceptional cases, the prosecution service) may impose a concrete sanction following the establishment of criminal liability according to the penalty/measure type(s) determined by the legislature and within the lower and upper limit of the relevant range of penalty after considering the aggravating and mitigating circumstances. The court may impose

[44] B. Miskolczi, *Az MI önálló felelősségi rendszere*, [in:] B. Miskolczi, Z. Szathmáry, *op. cit.*, notes 5, 188.

imprisonment (to be served or suspended) or another sanction (not affected by the so-called co-application prohibition) penalty(-ies) and/or measure(s). Alternatively, if it is permitted by section 33(4) of the Criminal Code, i.e., if the minimum of the penalty range for a criminal offence does not reach one year of imprisonment, a more lenient penalty may be imposed individually (or even in combination), such as community service or a fine, as an alternative to imprisonment. Similarly, to some extent, the act provides for the replacement of a penalty, i.e., the substitutive application of individual measures. Probation is such an example (sections 65–66 of the Criminal Code).

Nevertheless, should it be decided by the legislature that AI itself is punishable, the issue would require a rethinking. Certain sanctions, unaffected by the establishment of classic criminal liability, such as forfeiture, confiscation and, increasingly, rendering electronic data irreversibly inaccessible, could also be applied in respect of AI. However, most of the sanctions cannot be applied and could be replaced by sanctions similar to those applied to legal persons (such as liquidation of the AI, restriction of its activities, fine, etc.).

3.6. Some aspects in the context of the special part of criminal law

In the second, introductory part of this study, I already mentioned those criminal offences that may occur with regard to devices (agents) controlled by AI. As for this point, I specify this further; however, again without the aim of being exhaustive.

In Hungarian criminal law, the most frequent criminal offence in connection with AI could well be the *violation of information systems or related data breach* (section 423 of the Criminal Code), which may be regarded as the prototype of delicts related to computers, according to the former terminology, and recently against information systems. This could, for example, be a DDoS (Distributed Denial of Service) attack controlled by AI. However, it is also characteristic that this act is only a predicate offence for a further crime committed by data manipulation or other infringement after logging into the

information system. Where a violation of information systems takes place for illicit gain and in a form causing damage, *information system fraud* may principally be established (section 375 of the Criminal Code). However, other criminal offences committed through information systems, such as "simple" *fraud* (section 373 of the Criminal Code) or *extortion* (section 367 of the Criminal Code), may be established in concurrence with the violation of information systems or a related data breach. Similarly, an "easy" criminal situation may be created by AI, for example, to commit a *misuse of personal data* (section 219 of the Criminal Code), *misuse of classified data* or *violation of trade secrets* (section 418 of the Criminal Code).

Harassment (section 222 of the Criminal Code), albeit frequently occurring in classic "offline" form but also increasing in cyberspace, may be committed using AI as well. Notably, the variant committed by disturbance under section 222(1) of the Criminal Code may be envisaged through hundreds or thousands of emails, system notices, etc. generated automatically by AI. Unfortunately, the type of harassment committed by threat under section 222(2)(a) may be established in the same way, although the content of the message sent by AI is also relevant here.

In the context of sexual crimes, primarily, of course, *child pornography* (section 204 of the Criminal Code), which is widely distributed on the internet, may come into question. It can be highlighted that recent Anglo-Saxon literature also focuses on new types of sexual offences developed under the influence of AI. The so-called *sex-bots*, now often with a humanoid appearance, may not only serve to satisfy acceptable sexual needs but also to simulate *sexual violence* (section 197 of the Criminal Code). This – even if we disregard whether AI may be a legally recognised entity in the future against which such an offence may be committed – according to the related research may have a criminogenic effect in relation to future sexual delicts committed against a human.[45]

[45] T.C. King, N. Aggarwal, M. Taddeo, L. Floridi, *op. cit.*, note 21, pp. 104–105.

3.7. AI as the material object of the offence

Although the related literature primarily focuses on the criminal liability of AI, it is worth noting that AI can appear in the position as "victim" as well; that is, under the concept of statutory definition as a material object. Earlier legal literature accepted only persons and things as material objects. However, according to our recent dogmatic view, it would be appropriate to incorporate the phrase "other specific object" into the concept of the material object, thus enabling inanimate phenomena – but according to the related civil and criminal legislation, *de lege lata,* not things – to appear also as material objects.[46] In relation to AI, it can be achieved, in particular, by accepting data and information systems as material objects, whereas attacks against different information systems may easily constitute misuse of data and attacks against information systems. These new possibilities must be considered by the dogmatics of criminal law.

3.8. Conclusion

In my study on the interrelationship between AI and Hungarian substantive criminal law, I undertook to develop a comprehensive dogmatic phenomenology system between the most significant innovation of the 21st century and the conceptual universe of a classic area of law, being at least one and a half centuries old, even in its modern form. Since criminal law in Hungary was basically created at the time (in 1878) to penalise criminal activities committed by humans, applying its concepts in their original meaning to infringements relating to AI will thus not be sufficient in future. With particular attention to this, the essential elements of the concept of criminal offences, such as acts and compliance with the statutory definitions, may require reconsideration. However, several new aspects arose in connection with the obstacles to punishability and criminal sanctions, challenging the researcher of criminal law and

[46] B. Gellér, I. Ambrus, *op. cit.*, note 28, pp. 204–206.

the legislature as well. With this in mind, it might be recommended for criminal legal professionals to monitor technological changes continually, given the necessity to recreate both the statutory and scientific frames without delay in order that there be the correct judgement of related events requiring a criminal legal reaction. Probably all this can serve as an example to follow in Central and East European countries that have a continental system of criminal law, such as Poland or Hungary.

REFERENCES

Allen M.J., *Criminal Law*, Oxford 2017.

Ambrus I., Kovács G., Németh I., *Automomous vehicles and the prospective change in criminal liability*, "Ügyészek Lapja" 2018, Vol. 25, No. 6.

Angyal P., *A magyar büntetőjog tankönyve*, Budapest 1909.

Asworth A., *Principles of Criminal Law*, Oxford 1995.

Beck S., *Neue Konstruktionsmöglichkeiten der actio libera in causa*, "Zeitschrift für Internationale Strafrechtsdogmatik" 2018, Vol. 13, No. 6.

Belovics E., *Büntetőjog I. Általános rész (Criminal Law I. General Part)*, Budapest 2017.

Ben-Ari D., Frish Y., Lazovski A., Eldan U., Greenbaum D., *Danger, Will Robinson"? Artificial Intelligence in the Practice of Law: An Analysis and Proof of Concept Experiment*, "Richmond Journal of Law & Technology" 2017, Vol. 23, No. 3.

Barela S.J., *Legitimacy and Drones: Investigating the Legality, Morality and Efficacy of UCAVs*, Farnham 2015.

Barfield W., *Towards a Law of Artificial Intelligence*, [in:] W. Barfield, U. Pagallo (eds.), *Research Handbook on the Law of Artificial Intelligence*, Cheltenham 2018.

Buckland W.W., *The Roman Law of Slavery: The Condition of the Slave in Private Law from Augustus to Justinian*, Cambridge 1908.

Eszteri D., *Hogyan tanítsuk jogszerűen a mesterséges intelligenciánkat?*, "Magyar Jog" 2019, Vol. 66, No. 12.

Froomkin M.A., Colangelo Z.P., *Self-Defense Against Robots and Drones*, "Connecticut Law Review" 2015, Vol. 48, No. 1.

Gellér B., Ambrus I., *A magyar büntetőjog általános tanai I*, Budapest 2019.

Gless S., Silverman E., Weigend T., *If Robots Cause Harm, Who is to Blame? Self-Driving Cars and Criminal Liability*, "New Criminal Law Review" 2016, Vol. 19, No. 3.

Joshua D.G., *Solving the Trolley Problem*, [in:] J. Sytsma, W. Buckwalter (eds.), *A Companion to Experimental Philosophy*, Chichester 2016.

Hart H.L.A., *Punishment and Responsibility*, Oxford 1968.

Hatfield M., *Professionally Responsible Artificial Intelligence*, "Arizona State Law Journal" 2019, Vol. 51, No. 3.

Kádár M., Kálmán G., *A büntetőjog általános tanai (General Doctrines of Criminal Law)*, Budapest 1966.

Kant I., *The Metaphysics of Morals*, Budapest 1797/1991.

Kaspar J., *Strafrecht Allgemeiner Teil: Einführung (Criminal Law General Part: Introduction)*, Baden-Baden 2017.

Kindhäuser U., *Strafrecht Allgemeiner Teil (Criminal Law General Part)*, Baden-Baden 2017.

King T.C., Aggarwal N., Taddeo M., Floridi L., *Artificial Intelligence Crime: An Interdisciplinary Analysis of Foreseeable Threats and Solutions*, "Science and Engineering Ethics" 2020, Vol. 26, No. 1.

Lin P., *Why Ethics Matters for Autonomous Cars*, [in:] M. Maurer, J.Ch. Gerdes, B. Lenz, H. Winner (eds.), *Autonomous Driving. Technical, Legal and Social Aspects*, Berlin–Heidelberg 2016.

Locke J., *An Essay on Concerning Human Understanding*, Pennsylvania 1690/1999.

Lőrincz G., *A mesterséges intelligencia alkalmazásával hozott döntés jogi megítélésének egyes kérdései (Some issues in the legal assessment of a decision made using artificial intelligence)*, "Gazdaság és Jog" 2019, Vol. 28, No. 3.

Mérő L., *Új észjárások: A racionális gondolkodás ereje és korlátai (New insights: The power and limitations of rational thinking)*, Budapest 2001.

Millar J., *Ethics Settings for Autonomous Vehicles*, [in:] P. Lin, K. Abney, R. Jenkins (eds.), *Robot Ethics 2.0: From Autonomous Cars to Artificial Intelligence*, Oxford 2017.

Miskolczi B., Szathmáry Z., *Büntetőjogi kérdések az információ korában: Mesterséges intelligencia, Big Data, profilozás (Criminal Law Related Issues in the Age of Information: Artificial Intelligence, Big Data, Profiling)*, Budapest 2018.

Pacy E.P., *Tales from the Cryptocurrency: On Bitcoin, Square Pegs, and Round Holes*, "New England Law Review" 2014, Vol. 49, No. 1.

Pléh C., *A megismeréstudomány alapjai: Az embertől a gépig és vissza (The Basics of Cognitive Science: From Man to Machine and Back)*, Budapest 2013.

Russel S., Norvig P., *Artificial Intelligence: a Modern Approach*, Upper Saddle River, New Jersey 2010.

Staffler L., Jany O., *Künstliche Intelligenz und Stafrechtpfleg: eine Orientierung (Artificial intelligence and the maintenance of criminal law: an orientation)*, "Zeitschrift für Internationale Strafrechtsdogmatik" 2020, Vol. 15, No. 4.

Surden H., *Artificial Intelligence and Law: An Overview*, "Georgia State University Review" 2019, Vol. 35, No. 4.

Wessels J., Beulke W., Satzger H., *Strafrecht Allgemeiner Teil: Die Straftat und ihr Aufbau (Criminal Law General Part: The Criminal Offence and its Construction)*, Heidelberg 2013.

Williams G., *Textbook of Criminal Law*, London 1983.

DR HAB. INŻ. AGNIESZKA GRYSZCZYŃSKA, PROF. UCZ.
Cardinal Stefan Wyszyński University in Warsaw | DOI: 10.32041/9788367149372/4

Chapter 4. The Impact of the Proposed Regulation Establishing Harmonised Rules on Artificial Intelligence in the European Union on Law Enforcement and the Administration of Justice in Poland

4.1. Introduction

Artificial intelligence is a fast-evolving family of technologies that can bring a wide array of economic and societal benefits. By improving prediction, optimising operations and resource allocation, and personalising service delivery, artificial intelligence can support socially and environmentally beneficial outcomes. Growth in computing power, availability of data and progress in algorithms have turned Artificial intelligence into one of the most strategic technologies of the 21st century.[1] Artificial intelligence applications can provide a competitive advantage, but may also imply new risks or adverse consequences felt by individuals or society. In the area of justice, if artificial intelligence is based on trust, it can lead not only to a reduction in costs but also to a significant reduction in the length of a trial and the delivery of a verdict.

Understanding AI as a set of techniques and methods for modelling knowledge, encapsulated in a system and subject to interactions with the external environment, the challenges of the convergence of cybersecurity and AI resilience should be taken into account at the

[1] AI for Europe, COM/2018/237 final.

stage of designing regulations and modelling technical solutions.[2] The current state of research and development of artificial intelligence systems exposes the problem of the so-called "black box" – that is, the inability of the rules of traditional linear cause-and-effect logic to explain the recommendations that flow from the system. This is because AI implements a different approach, one based on correlation analysis of results or fuzzy logic methodologies. The applicability of AI by law enforcement and in the administration of justice will therefore largely depend on the development of common approaches and recommendations for AI systems, including the concept of stewardship of trustworthy AI[3], as well as the integration of state law activities sanctioning an ethical framework, fair competition and the principle of mutual recognition of interoperability rules and risk assessment and validation. The way we approach AI will define the world we live in. The fundamental question to be asked, however, concerns the method, scope and level of legal regulation of artificial intelligence.

Answers to these questions can assist verifying the research hypothesis of whether trustworthy, ethical and human-centric AI can support law enforcement and the administration of justice, contributing to a better fulfilment of the right to a fair trial.

In order to determine the impact of the proposal of the Artificial Intelligence Act[4] on law enforcement and the justice system, the current state of AI regulation in Poland and the EU was analysed.

The analysis seeks to identify the measures that can be implemented presently to develop artificial intelligence in law enforcement and administration of justice, and those that will require

[2] R. Kroplewski, *Odporność AI dla odpornej wspólnoty*, [in:] *Cyberbezpieczeństwo AI. AI w cyberbezpieczeństwie* [online], Warsaw 2023, p. 111, https://www.nask.pl/pl/aktualnosci/5237,Cyberbezpieczenstwo-AI-AI-w-cyberbezpieczenstwie-wazna-publikacja-NASK.html [accessed on: 1 December 2023].

[3] OECD (2019), Recommendation of Council of OECD on Artificial Intelligence; OECD AI Principles; UNESCO (2021), Recommendations on Ethics of AI.

[4] Proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

regulatory change. Furthermore, the investigation will highlight those measures that have been deemed either unacceptable or high-risk as per the proposal of the Artificial Intelligence Act.

4.2. The concept of artificial intelligence

In Poland, the concept of artificial intelligence has no legal definition. Defining artificial intelligence is undoubtedly a challenge, and in the literature the concept is defined in various ways. At the outset, it should be pointed out that certain problems, e.g., can thought processes be automated? Can the process of human reasoning be described formally by means of logic? Can a formal conceptual and linguistic system be constructed with which to describe the world? have troubled philosophers since ancient times. It was not until the middle of the 20th century, in connection with the dynamic development of computer science, that the issue of artificial intelligence became one of the key problems not only of computer science but of philosophy, psychology, linguistics, biology and the methodology of sciences.[5] Nowadays, the issue of artificial intelligence has become one of the most important legal issues, whereas scientific discussion has further started to be accompanied by regulatory action since the 21st century.

One of the fundamental issues of artificial intelligence is the question of when a human-constructed system can be said to be intelligent. In 1950, the question "Can machines think?" was asked by Alan M. Turing,[6] and, in answering it, pointed out that an analysis of the meaning of the terms "machine" and "think" would not lead to an answer to the question thus posed, since these terms cannot be clearly defined. Rather than trying to determine if a machine is thinking, Turing suggests we should ask if the machine can win the "Imitation Game", and that the question is, "Can machines do what we (as thinking entities) can do?" In other words, as Stevan

[5] M. Flasiński, *Wstęp do sztucznej inteligencji*, Warsaw 2011, p. 3.

[6] A.M. Turing, *Computing Machinery and Intelligence*, "Mind" 1950, Vol. 59, No. 236, p. 433.

Harnad notes, Turing is no longer asking whether a machine can "think"; he is asking whether a machine can act indistinguishably from the way a thinker acts, and proposes a very general methodological criterion for modelling mental function: total functional equivalence, and indistinguishability.[7]

It is extremely difficult to answer the question, "What is artificial intelligence?" due to the lack of universal agreement about what intelligence is. Moreover, there is scant reason to believe that machine intelligence bears much relationship to human intelligence, at least to date.[8]

John McCarthy, widely recognised as the father of artificial intelligence, described the process in 1955 in a proposal for the Dartmouth Summer Research Project on Artificial Intelligence as "that of making a machine behave in ways that would be called intelligent if a human were so behaving".[9] However, it is pointed out that this approach is flawed by the difficulty of defining and measuring human intelligence and whether "how" a problem is solved is as important as "whether" it is solved.

The notion of AI should be clearly defined to ensure legal certainty, while providing the flexibility to accommodate future technological developments. The draft regulation does not contain a definition of artificial intelligence but it does contain a definition of an artificial intelligence system. According to Article 3(1) of the Draft Regulation, "artificial intelligence system" (AI system) means software that is developed with one or more of the techniques and approaches listed in Annex I and can, for a given set of human-defined objectives, generate outputs such as content, predictions, recommendations, or decisions influencing the environments they interact with. The definition should be based on the key functional

---

[7] S. Harnad, *Minds, Machines and Turing: The Indistinguishability of Indistinguishables,* "Journal of Logic, Language, and Information (special issue on Alan Turing and Artificial Intelligence)" 2001, https://web-archive.southampton.ac.uk/cogprints.org/2615/1/harnadoo.turing.html [accessed on: 1 December 2023].

[8] J. Kaplan, *Artificial Intelligence: What Everyone Needs to Know*, Oxford University Press, 2016, p. 15.

[9] J. McCarthy, M. Minsky, N. Rochester, C. Shannon, *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955.

characteristics of the software, in particular the ability, for a given set of human-defined objectives, to generate outputs such as content, predictions, recommendations, or decisions which influence the environment with which the system interacts, be it in a physical or digital dimension. AI systems can be designed to operate with varying levels of autonomy and be used on a stand-alone basis or as a component of a product, irrespective of whether the system is physically integrated into the product (embedded) or serves the functionality of the product without being integrated therein (non-embedded). The definition of AI system should be complemented by a list of specific techniques and approaches used for its development, which should be kept up-to-date in the light of market and technological developments through the adoption of delegated acts by the Commission to amend that list.

The definition of an AI system in the proposal aims to be as technology-neutral and future-proof as possible, taking into account the fast technological and market developments related to AI. Over the summer, the Czech Presidency has presented the second compromise text of the Council, in which it attempts to settle previously contentious issues, namely how to define an AI system, governance and enforcement, classification of AI systems as high-risk as well as national security exclusion. Among the changes suggested by the Czech Presidency is a narrower definition of AI systems, providing for a clearer distinction between AI systems and more traditional software systems. Moreover, the Presidency's compromise text introduces a new horizontal framework to complement the classification of high-risk AI in order to ensure that innocuous AI systems will not become subject to the new regulation once adopted.

At the time of the book's submission for publication, according to the adopted amendment (Amendment 165), an artificial intelligence system (AI system) was defined as "a machine-based system that is designed to operate with varying levels of autonomy and that can, for explicit or implicit objectives, generate outputs such as predictions, recommendations, or decisions, that influence physical or virtual

environments."[10] This definition has been further modified in further interinstitutional negotiations to align it more closely with the work of international organisations working on artificial intelligence, notably the OECD. According to the new compromise version of the AI Act agreed during the trilogue in December 2023 an "AI system" is defined as "a machine-based system designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments".[11]

4.3. State of regulation of artificial intelligence in Poland

The phrase "artificial intelligence" in 2022 was found in 28 legal acts published in the Journal of Laws (*Dziennik Ustaw*) and Polish Monitor (*Monitor Polski*, hereinafter referred to as M.P.) in the SIP LEX database, including 12 in force. Currently, the number of legal acts in which the phrase 'artificial intelligence' appears has increased. Among the acts, no acts have been established at the statutory level that define the concept of artificial intelligence, or that regulate the application or liability of artificial intelligence.

The phrase artificial intelligence appears in one law[12] – on the Future Industry Platform Foundation. According to Article 1 of this law, the aim of the Foundation Platform of the Future Industry is to

[10] Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), P9_TA(2023)0236.

[11] Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts – Analysis of the final compromise text with a view to agreement, 26.01.2024, https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf [accessed on: 8 March 2024].

[12] Law on the Future Industry Platform Foundation of 17 January 2019, Journal of Laws 2019, item 229.

act to increase the competitiveness of entrepreneurs by supporting their digital transformation in terms of processes, products and business models, using the latest developments in the field of automation, artificial intelligence, ICT technologies and machine-to-machine and human-to-machine communication, taking into account the appropriate level of security of these solutions.

The term "artificial intelligence" appears in 13 regulations, most of them relating to core curricula,[13] teaching standards, and qualification proceedings.[14] In the regulations defining the core curriculum for pre-primary education, primary general secondary school, technical secondary school and industry upper secondary school, artificial intelligence is only indicated in the content of ethics teaching. In the regulation of 7 August 2014 on the classification of professions and specialties for the needs of the labour market and the scope of its application,[15] the following are distinguished: specialists for the development of *artificial intelligence* (251,908) and specialist for machine learning (251,909). The term "artificial intelligence" also appears in the regulation on granting the Scientific and Academic

[13] Regulation of the Minister of National Education of 14 February 2017 on the programme basis of pre-school upbringing and the programme basis of general education for primary school, including for pupils with moderate or severe intellectual disabilities, general education for an industry school of the first degree, general education for a special school preparing for work and general education for a post-secondary school (Journal of Laws of 2017, item 356, as amended); Regulation of the Minister of National Education of 30 January 2018 on the programme basis of general education for upper secondary school, technical school and upper secondary vocational school, Journal of Laws of 2018, item 467, as amended. Artificial intelligence is mentioned only in the content of ethics teaching. In the selected issues of detailed ethics "Ethics and science and technology", it is indicated that the student identifies and analyses selected moral problems related to scientific and technological progress (e.g., the problem of privacy protection, copyright protection, cyberbullying, development of artificial intelligence, transhumanism).

[14] Regulation of the Minister of Internal Affairs and Administration of 12 January 2022 on the qualification procedure for candidates applying for admission to service in the Police, Journal of Laws of 2022, item 109.

[15] Regulation of the Minister of Labour and Social Policy of 7 August 2014 on the classification of professions and specialities for the needs of the labour market and the scope of its application, i.e., Journal of Laws of 2018, item 227, as amended.

Computer Network the status of a state research institute,[16] in which it is indicated that the tasks of NASK, which are particularly important for the planning and implementation of state policy, the performance of which is necessary to ensure public safety, the development of education and the improvement of the quality of life of citizens, performed on a continuous basis, include the implementation of scientific research results and conducting development work supporting machine learning and artificial intelligence.

The term "artificial intelligence" also appears in strategic and programme documents. Of key importance is Resolution No. 196 of the Council of Ministers of 28 December 2020 on the establishment of the "Policy for the development of artificial intelligence in Poland from 2020," further as AI Policy.[17]

Poland's AI Policy is based on the definition of an AI System developed within the OECD by the AIGO group of independent experts (OECD), according to which an AI system is a system based on the concept of a machine that can influence the environment by making recommendations, predictions, or decisions on a set of objectives.[18] According to the Poland's AI Policy, the minister

[16] Regulation of the Council of Ministers of 7 June 2017 on conferring the status of a state research institute on the Scientific and Academic Computer Network, Journal of Laws of 2017, item 1193.

[17] Resolution No. 196 of the Council of Ministers of 28 December 2020 on the establishment of the "Policy for the development of artificial intelligence in Poland from 2020", M.P. of 2021, item 23.

[18] It does this by using input, machine or human data to: perceive real or virtual environments, summarise such perceptions into models manually or automatically, use model interpretation to formulate outcome options. In the diagram, an AI system consists of three main elements: sensors (sensors), operational logic (algorithm models), actuators (execution apparatus). An AI system consists of three main elements: Sensors, Operational Logic and Actuators. Sensors collect raw data from the Environment, while Actuators take actions to change the state of the Environment. The key power of an AI system resides in its Operational Logic, which, for a given set of objectives and based on input data from Sensors, provides output for the Actuators - as recommendations, predictions or decisions - that are capable of influencing the state of the Environment, https://www.oecd-ilibrary.org/docserver/d62f618aen.pdf?expires=1657742576&id=id&accname=guest&checksum=491B7B641D204752F82B0A28B24E6C24 [accessed on: 1 December 2023].

responsible for informatisation is responsible for coordinating the implementation of AI Policy. A key role in monitoring and supporting the coordination of the progress of AI Policy implementation will be played by the AI Policy Task Team at the Committee of the Council of Ministers for Digital Affairs (KRMC).[19] The Task Team will be appointed by the KRMC at the request of the minister responsible for informatisation. The AI Policy Task Team will present the KRMC with draft information on the implementation of AI Policy activities for the year. This draft, after consideration by the KRMC, will be presented to the Council of Ministers by the minister responsible for informatisation.

There are also a growing number of monographs and scholarly articles on the problems of artificial intelligence in the Polish legal literature.[20]

4.4. Work on AI regulation in the EU

The announcement of work on a coordinated European approach on the human and ethical implications of AI was announced by Ursula von der Leyen in her political guidelines for the 2019–2024 Commission "A Union that strives for more."[21] She also announced, along with the legal framework, the prioritisation of financial investment

[19] The team was established on 19 January 2022, https://www.gov.pl/web/ai/powolanie-zespolu-zadaniowego-do-spraw-realizacji-polityki-ai [accessed on: 1 December 2023].

[20] M. Rojszczak, *Prawne aspekty systemów sztucznej inteligencji – zarys problemu*, [in:] *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe: zagadnienia wybrane*, K. Flaga-Gieruszyńska, J. Gołaczyński, D. Szostek (eds.), Warsaw 2019, p. 3; A. Kasperska, *Problems of applying artificial neural networks in legal practice*, "Przegląd Prawa Publicznego" 2017, No. 11, p. 25; L. Lai, M. Świerczyński, *Prawo sztucznej inteligencji,* Warsaw 2020; A. Chłopecki, *Sztuczna inteligencja – szkice prawnicze i futurologiczne*, 2nd edition, Warsaw 2021; A. Krasuski, *Status prawny sztucznego agenta. Podstawy prawne zastosowania sztucznej inteligencji*, Warsaw 2021; A. Nowak-Gruca, *Cyborg, czyli kto? O prawach cyborgów w ujęciu interdyscyplinarnym*, Warsaw 2023.

[21] https://ec.europa.eu/info/sites/default/files/political-guidelines-next-commission_en_0.pdf, p. 13 [accessed on: 1 December 2023].

in Artificial Intelligence, both through the Multiannual Financial Framework and through the increased use of public-private partnerships. On 19 February 2020 the Commission published the White Paper on AI – "A European approach to excellence and trust."[22]

European leaders have put AI at the top of their agendas. On 10 April 2018, 24 Member States and Norway committed to working together on AI.[23] Responding to the request by European leaders to define a European approach to AI was the Commission Communication "Artificial intelligence for Europe,"[24] in which the Commission put forward a European approach to Artificial Intelligence based on three pillars: being ahead of technological developments and encouraging uptake by the public and private sectors, prepare for socio-economic changes, and ensure an appropriate ethical and legal framework. Delivering on its strategy on AI adopted in April 2018,[25] in December 2018 the Commission presented a Coordinated Plan – prepared together with the Member States – to foster the development and use of AI in Europe.[26] This plan proposes some 70 joint actions for closer and more efficient cooperation between Member States, and the Commission in key areas, such as research, investment, market uptake, skills and talent, data and international cooperation. The plan is scheduled to run until 2027, with regular monitoring and review.

In October 2020 the European Parliament adopted a number of resolutions related to AI, including on ethics,[27] liability,[28] and copyright.[29] In 2021, those were followed by resolutions on AI in

[22] European Commission, White Paper on Artificial Intelligence – A European approach to excellence and trust, COM(2020) 65 final, 2020.

[23] https://digital-strategy.ec.europa.eu/en/news/eu-member-states-sign-cooperate-artificial-intelligence [accessed on: 1 December 2023].

[24] AI for Europe, COM/2018/237 final.

[25] Artificial Intelligence for Europe, COM(2018) 237.

[26] Coordinated Plan on Artificial Intelligence, COM(2018) 795.

[27] European Parliament resolution of 20 October 2020 on a framework of ethical aspects of artificial intelligence, robotics and related technologies, 2020/2012(INL).

[28] European Parliament resolution of 20 October 2020 on a civil liability regime for artificial intelligence, 2020/2014(INL).

[29] European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI).

criminal matters[30] and in education, culture, and the audio-visual sector.[31]

In April 2021, the European Commission proposed the first EU legislative framework for artificial intelligence.[32] On 8 December 2023 Parliament and Council negotiators reached a provisional agreement on the Artificial Intelligence Act.[33]

The compromise agreement provides for a horizontal layer of protection, including a high-risk classification of AI systems. Limited-risk AI systems would be subject to very light transparency obligations, such as disclosing that the content is AI-generated so that users can make informed decisions about its further use.

For some uses of AI, the risk is deemed unacceptable and, therefore, these systems will be banned from the EU. According to the agreed text of the AI Act, the following are to be prohibited: 1) biometric categorisation systems that use sensitive characteristics (e.g., political, religious, and/or philosophical beliefs; sexual orientation; race); 2) the untargeted scraping of facial images from the internet or CCTV footage to create facial recognition databases; 3) emotion recognition in the workplace and educational institutions; 4) social scoring based on social behaviour or personal characteristics; 5) AI systems that manipulate human behaviour to circumvent human beings' free will; 6) AI used to exploit people's vulnerabilities (due to age, disability, social or economic situation).

Negotiators agreed on a number of safeguards and narrow exceptions for the use of biometric identification systems ("RBI")

[30] European Parliament Draft Report, Artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters, 2020/2016(INI).

[31] European Parliament Draft Report, Artificial intelligence in education, culture and the audiovisual sector, 2020/2017(INI).

[32] Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonised Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, COM/2021/206 final.

[33] https://www.europarl.europa.eu/news/en/press-room/20231206IPR15699/artificial-intelligence-act-deal-on-comprehensive-rules-for-trustworthy-ai?xtor=AD-78-%5BSocial_share_buttons%5D-%5Btwitter%5D-%5Ben%5D-%5Bnews%5D-%5Bpressroom%5D-%5Bartificial-intelligence-act-possible-deal%5D- [accessed on: 10 December 2023].

in public spaces for law enforcement purposes, subject to prior judicial authorisation and for strictly defined crimes.

To take account of the wide range of tasks that AI systems can perform and the rapid expansion of their capabilities, it was agreed that General Purpose AI (GPAI) systems and the GPAI models on which they are based will have to comply with transparency requirements, as originally proposed by the Parliament. These include the drawing up of technical documentation, compliance with EU copyright law and the dissemination of detailed summaries of the content used for training.

Work on the final text of the AI Act has not been completed as of December 2023, and is expected to continue at a technical level in the coming weeks to finalise the details of the new regulation. The provisional agreement provides that the AI act should apply two years after its entry into force, with some exceptions for specific provisions.

4.5. Digitalisation of justice systems

Digitalisation and the challenges it poses must be taken into account in the justice system. The digitalisation of the justice system aims to facilitate and improve access to justice, to make the justice system more effective and efficient while facilitating the work of justice professionals, and to bring it closer to citizens, thus offering better justice services to all.

The Communication from the European Commission on digitalisation of justice in the European Union[34] emphasises that access to justice and facilitating cooperation between Member States are among the main objectives of the European Union's area of freedom, security, and justice.

[34] Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Digitalisation of justice in the European Union, a toolbox of opportunities, COM/2020/710 final.

Following the Communication, the focus of the European Union's work in the area of e-Justice has been on legislative action. Between 2019 and 2023, the Council of the European Union, the European Commission and the European Parliament have made considerable efforts to speed up the digitalisation process and promote the use of digital services in e-Justice. The emphasis has been firmly placed on legislative action, concluding several legislative initiatives, such as the regulation on a computerised system for the cross-border electronic exchange of data in the area of judicial cooperation in civil and criminal matters (e-CODEX system),[35] that provided an appropriate framework for exchanging judicial information through secure services; the Service of Documents[36] and Taking of Evidence Regulations[37] establishing the use of the decentralised IT system with interoperable access points based on e-CODEX for relevant communications; the e-Evidence Regulation;[38] and especially the Regulation[39] and Directive[40] on the digitalisation of

[35] Regulation (EU) 2022/850 of the European Parliament and of the Council of 30 May 2022 on a computerised system for the cross-border electronic exchange of data in the area of judicial cooperation in civil and criminal matters (e-CODEX system), and amending Regulation (EU) 2018/1726, OJ L 150, 1.06.2022, pp. 1–19.

[36] Regulation (EU) 2020/1784 of the European Parliament and of the Council of 25 November 2020 on the service in the Member States of judicial and extrajudicial documents in civil or commercial matters (service of documents) (recast), OJ L 405, 2.12.2020, pp. 40–78.

[37] Regulation (EU) 2020/1783 of the European Parliament and of the Council of 25 November 2020 on cooperation between the courts of the Member States in the taking of evidence in civil or commercial matters (taking of evidence) (recast), OJ L 405, 2.12.2020, pp. 1–39.

[38] Regulation (EU) 2023/1543 of the European Parliament and of the Council of 12 July 2023 on European Production Orders and European Preservation Orders for electronic evidence in criminal proceedings and for the execution of custodial sentences following criminal proceedings, OJ L 191, 28.07.2023, pp. 118–180.

[39] Regulation of the European Parliament and of the Council on the digitalisation of judicial cooperation and access to justice in cross-border civil, commercial and criminal matters, and amending certain acts in the field of judicial cooperation, 2021/0394 (COD); PE-CONS 50/23; https://data.consilium.europa.eu/doc/document/PE-50-2023-INIT/en/pdf [accessed on: 11 December 2023].

[40] Directive of the European Parliament and of the Council amending Directives 2011/99/EU and 2014/41/EU of the European Parliament and of the Council,

cross-border judicial cooperation and access to justice (the "Digitalisation Package").

On 17 November 2023, the European e-Justice Strategy for 2024–2028 was adopted.[41] The strategy should guide the ongoing digital transformation in the justice domain across the European Union. In particular, the strategy aims to identify strategic and operational objectives and the principles that should be respected when carrying out this digital transformation process, to put in place organisational and methodological measures, to identify key enablers to facilitate and foster digitalisation, as well as to promote mechanisms to facilitate the coordination and follow-up of progress on e-Justice initiatives.

According to the strategy, the actions taken in the context of digital transformation are at the same time to give flexibility in the use of modern technologies, including artificial intelligence. The strategic objective "Improve access to digital justice" (A), includes supporting users through conversational assistants (chatbots, including AI-powered ones), facilitating citizens' access to judicial information, as well as supporting access to IT material means for users (B). The objective "Make digital justice more efficient" (C) provides for actions related to improving the collection and use of legal and judicial data and/or the automation of justice activities. The objective "Promote an innovative digital justice" (D) includes actions related to identifying areas of application and safely apply AI in the justice domain, including but not only: for the anonymisation and pseudonymisation of judicial decisions; as a transcription tool for the recording of proceedings and the documentation of evidence gathered by the court (speech-to-text and text-to-speech); for translation; for legal analysis of, e.g., case law and big data sources; for

Council Directive 2003/8/EC and Council Framework Decisions 2002/584/JHA, 2003/577/JHA, 2005/214/JHA, 2006/783/JHA, 2008/909/JHA, 2008/947/JHA, 2009/829/JHA and 2009/948/JHA, as regards digitalisation of judicial cooperation, 2021/0395 (COD), PE-CONS 51/23, https://data.consilium.europa.eu/doc/document/PE-51-2023-INIT/en/pdf [accessed on: 11 December 2023].

[41] European e-Justice Strategy 2024–2028, 15509/23, https://data.consilium.europa.eu/doc/document/ST-15509-2023-INIT/en/pdf [accessed on: 11 December 2023].

calculating entitlements to compensation, e.g., passenger rights or rights of a similar type.

Intensive and comprehensive work on the digitalisation of justice systems is taking place in parallel with the work on the regulation of AI within the EU. However, the development of justice systems requires a strong focus on the effectiveness of the protection guaranteed by existing fundamental rights. Initiatives in the context of the digital transformation of justice need to respect judicial independence and comply with the rule of law, which is one of the core values on which the European Union is founded. Action should be taken in particular in the criminal justice context, where the use of remote communication technologies could pose serious risks to the fundamental rights of suspects and defendants, in particular the right to a fair trial, the right to be present at the trial, and the right of defence. The emergence of innovative technologies may also create fundamentally new challenges and risks, such as cybersecurity breaches, a widening of the digital divide or implicit discrimination due to biased algorithms or data sets.

4.6. The impact of artificial intelligence on the internal openness of proceedings – on the example of the use of digitised case files in criminal proceedings

In a democratic state under the rule of law, criminal proceedings must meet certain minimum standards, among which openness plays an important role.

The literature aptly points out that internal openness applies to all actors who defend their own interest in the process or act as procedural representatives of those persons.[42] In the doctrine, the forms of implementation of internal openness of proceedings

[42] W. Jasiński, *Jawność wewnętrzna postępowania sądowego*, [in:] *Jawność procesu karnego*, J. Skorupka (ed.), Warsaw 2012, pp. 211–216; T. Grzegorczyk, *Jawność wewnętrzna postępowania sądowego*, [in:] *Jawność jako wymóg rzetelnego procesu karnego. Zagadnienia prawa polskiego i obcego*, W. Jasiński, K. Nowicki (eds.), Warsaw 2013, p. 71.

include: 1) participation in the activities of the proceedings, 2) access to the materials of the proceedings, including the case file, 3) informing the participants of the trial about its course and results in its individual stages (with the exception of information about the rights and obligations of such participants, which is a manifestation of the implementation of the principle of a fair trial).[43]

However, there are different approaches to the location of internal openness in the system of procedural rules of criminal law presented in the Polish literature. According to the first view, which is dominant, the principle of openness covers both aspects: external and internal.[44] According to the second trend, openness is limited to the external dimension (the principle of the public), while internal openness should be associated with the adversarial principle.[45] According to the third view, internal openness forms a separate procedural principle, which is understood as the principle of participation of the parties in the proceedings or the principle of openness to the parties and other participants in the proceedings,[46] while the fourth view indicates the validity of the first or second approach depending on the preferred point of view.[47] Notwithstanding the different ideas on the positioning of internal openness in the system of procedural principles of criminal law, internal openness serves primarily the interests of the parties and other participants in criminal proceedings.[48]

---

[43] T. Grzegorczyk, J. Tylman, *Polskie postępowanie karne*, Warsaw 2014, p. 150.

[44] See, for example, A. Kaftal, *Jawność postępowania karnego w świetle nowego kodeksu postępowania karnego*, "Nowe Prawo" 1969, No. 11–12, pp. 1640, 1647; B. Wójcicka, *Jawność postępowania sądowego w polskim procesie karnym*, Łódź 1989, p. 9.

[45] S. Waltoś, P. Hofmański, *Proces karny. Zarys systemu*, Warsaw 2018, pp. 317 *et seq*.

[46] M. Cieślak, *Polska procedura karna. Basic theoretical assumptions*, Warsaw 1973, p. 292.

[47] P. Wiliński, *Świadek incognito w polskim procesie karnym*, Kraków 2003, p. 422.

[48] R. Koper, *The principle of openness versus internal openness in the criminal process*, "Legal Studies" 2019, No. 2 (218), DOI:10.37232/sp.2019.2.6, p. 145.

4.7. The need to maintain internal openness linked to the issue of protecting human rights

The exercise of the right to a fair trial is therefore affected not only by the procedural guarantees of access to the file, but also by the technical manner in which this right is exercised, influencing the time it takes to carry out a request to inspect or obtain a copy, the costs or the de-location of the exercise.

In addition, the system of digitisation of files makes it possible to increase the efficiency of the functioning of the organisational structures of the public prosecutor's office by creating a mechanism for making files available in digital form to citizens entitled to access the files of proceedings and creating a mechanism for the exchange of data with law enforcement authorities and the judiciary.

The process of digitisation of pre-trial investigation files in Poland has been ongoing since 2013. On 8 January 2013 The General Prosecutor's Office signed an agreement with the Implementing Authority for European Programmes on co-financing from the European Regional Development Fund within the framework of Priority Axis 7 of the Operational Programme Innovative Economy Information Society of the project "Implementation of the system of digitalisation of preparatory proceedings files and creation of local and central repository of files in digital form in common prosecution units." The aim of the project was to implement a system of digitisation of preparatory proceedings files, to create local systems of collecting, processing and making digitised files available in common organisational units of the public prosecutor's office. The beneficiary of the project was the General Prosecutor's Office together with common organisational units of the public prosecutor's office at the level of appellate and circuit prosecutor offices (total: 57 units). Since the implementation of the SDA system, i.e., from 30 April 2015 to 31 October 2015, the units covered by the project digitised 12,749 volumes of files.[49]

[49] https://pk.gov.pl/aktualnosci/aktualnosci-prokuratury-krajowej/system-dygitalizacji-akt-sda-konferencja-prokuratury-generalnej-podsumowujaca-projekt/ [accessed on: 15 September 2022].

In 2021, as part of the project Development of the System for Digitisation of Preliminary Proceedings Files in Criminal Cases (iSDA 2.0), the central ICT system for the Prosecutor's Office, PROK-SYS, was implemented, which involved the integration of locally dispersed applications implementing business processes into an integrated IT system for all organisational units of the Prosecutor's Office. The digitisation of files has also been significantly accelerated to include the files of cases conducted in the district prosecutors' offices. By December 2023, a total of 1.3 million volumes (130 million pages) covering 500,000 criminal cases had been digitised. Priority for digitisation is given to the most complicated and serious criminal cases, as well as cases in which pre-trial detention is applied (currently 90% of the files in which pre-trial detention is applied have been digitised). The main advantage of digitising criminal case files is that they can be made available in electronic form – both on media and online – to authorised parties on the external portal of the Public Prosecutor's Office (https://portalzewnetrzny.prokuratura.gov.pl/). This provides an easy and convenient way for those entitled to consult the file – 24 hours a day, 7 days a week – and supports the right to a fair trial by providing the right to see the evidence collected.

The process of digitising files could definitely be accelerated if machine learning-based solutions were implemented. Currently, staff shortages are indicated as the main problem with the result that only about 21.5% of case files in district prosecutors' offices have thus far been digitised. The process of scanning the files itself is a fairly quick process, but the description of the scanned documents by the scanning operators is a time-consuming task. Although this process is partly automated, the need for scanning operator intervention is still significant. The low level of automation in describing scanned documents is caused by the existence of different formats and descriptions of identical documents, the existence of documents in handwriting or different formats of dates used. In connection with automated document content analysis platforms implemented in some entities (e.g., banks, law firms, tax offices), based on document examination using machine learning models and natural language analysis algorithms, the same mechanisms can be used not only

for simple generation of document descriptions, but also for the creation of detailed metadata or summaries. In addition to the simple generation of description and metadata, the same tools can be used for bulk analysis and presentation of answers to queries on hundreds or even thousands of documents.

The application of machine learning to scanned and OCRed documents will contribute to better recognition of the content of the documents, allowing all the metadata of the documents being described to be automatically filled in in the form of: type, name, subject, object and date of the document. Such use of AI systems shall not be considered as high risk. Such use of AI systems shall not be considered as high risk under the draft AI Act. First of all, such use is not indicated in Annex III to the draft AI Act. Above all, it shall not pose a significant risk of harm, to the health, safety or fundamental rights of natural persons, including by not materially influencing the outcome of decision making. At the same time, speeding up and automating the digitisation process will make more files available online, which will contribute to the rights of litigants.

Digitisation of files leads to the creation of a database that can be used to teach algorithms that would then support the decision-making process and automatically generate drafts of selected letters. In criminal proceedings, in connection with the occurrence of a specific state of facts, specific actions are taken as indicated by generally applicable laws and methodologies for conducting proceedings in selected categories of cases, standard letters are prepared or specific decisions or orders are issued. Some of the actions could be subject to algorithmisation and automation, and a learning algorithm could prepare ready-made draft letters or procedural decisions subject to approval by the prosecutor on the basis of legal regulations and regularities in the proceedings. For example, if a suspect is registered in the case, a query of his/her criminal record in the National Criminal Register could be made automatically; depending on the answer, if the suspect has previously served a prison sentence, the algorithm would check whether there are grounds for recidivism and automatically prepare a letter to the court with a request to send a copy of the judgment with information on serving the sentence,

and after receiving the answer it would count time limits and generate an alert for the public prosecutor referring the case.[50]

The use of natural-language processing mechanisms can assist the victim to report suspected crimes electronically, using a chatbot-assisted form. The use of the chatbot functionality can assist in guiding the victim through the process of making such a report, obtaining all relevant information about the circumstances of the report, and providing full information about the applicable law and the victim's rights. The use of natural-language processing models can also help to obtain relevant assistance from state and social institutions (available 24/7) as well as information on the status of the case, which will be particularly relevant for persons with disabilities.

4.8. Conclusion

New technologies must not impair the rights of individuals, and they must be used in full respect of the right to a fair trial and the right of defence. The European Union has taken a leading role in reconciling transformative technologies with fundamental rights and freedoms and ensuring safeguards against possible risks.

In line with the approach set out in the "Digitisation Package", the European e-Justice Strategy 2024–2028 and the proposal for a Regulation on artificial intelligence, innovative technologies can, *inter alia*, bring law enforcement and justice closer to citizens, improve the functioning of courts and assist prosecutors and judges in their daily work. However, to achieve this goal, it is necessary to act responsibly and respect the principles and values of the European Union.

Among the applications of AI identified in the above-mentioned documents are assisting users through virtual interlocutors (chatbots), facilitating citizens' access to judicial information, anonymising and pseudonymising court decisions, transcribing hearing transcripts and documentary evidence collected by the court (converting speech to text and text to speech), translating,

[50] Examples of AI content generators: https://bloggersideas.com/pl/best-ai-content-writing-softwares/ [accessed on: 11 December 2023].

analysing case law and large data sets or, for example, calculating compensation claims. The finalisation of the AI Act will reduce legal uncertainty for AI applications in law enforcement and justice, and ensure that fundamental rights and the EU values are safeguarded and respected.

The discourse on the application of artificial intelligence in the administration of justice extends well beyond the EU. Other countries are also developing practical applications of AI in the administration of justice, legislation responding to its use, or specific guidelines on how to use the most popular and widely available Generative Artificial Intelligence (GenAI) chatbots (such as ChatGPT, Bing Chat or Google Bard).[51]

In Poland, there is no specific regulation on the use of machine learning algorithms and ensuring their safety in law enforcement and justice, so the AI Act will play a fundamental role. However, the GRAI group's ongoing work on mapping regulations to the challenges posed by AI shows that procedural regulations in particular require profound changes.[52] Criminal procedure is considered the most undigitalised procedure. Not only does it not refer to the automation of certain actions, but it also does not provide for the electronic form of proceedings or the use of electronic documentation in proceedings. At the same time, from a technical point of view, the Prosecutor's Office has a central system, PROK-SYS, in which not only case registration data but also criminal case files are continuously digitised. The Prosecutor's Office database, which is constantly being developed, could be a possible area for artificial intelligence applications. Basic future applications include optimising text

[51] Guidelines for Use of Generative Artificial Intelligence in Courts and Tribunals, Artificial Intelligence Advisory Group, 2023, www.courtsofnz.govt.nz/assets/6-Going-to-Court/practice-directions/practice-guidelines/all-benches/20231207-GenAI-Guidelines-Judicial.pdf [accessed on: 11 December 2023].

[52] Analysis of the relationship of the Act on artificial intelligence with selected and proposed regulations legislation, GRAI, 2023, https://www.gov.pl/web/ai/raport-analiza-zwiazku-aktu-w-sprawie-sztucznej-inteligencji-z-wybranymi-obowiazujacymi-i-projektowanymi-regulacjami-prawnymi [accessed on: 11 December 2023].

recognition at the stage of scanning files and extracting metadata, as well as supporting the creation of hearing protocols based on speech recognition and transcription. Given that in Poland, digitised criminal case files can be made available online to authorised parties via the external portal of the Prosecutor's Office, applications that facilitate and speed up digitisation and anonymisation will facilitate access to files by parties to the proceedings, thereby contributing to strengthening the guarantees of internal openness of proceedings and thus the right to a fair trial. The work of prosecutors would be facilitated by a Large Language Model-based search for information in the case file database or the generation of draft standardised documentation (e.g., orders, decisions, guidelines, investigation plans). A chatbot could assist victims at the stage of reporting a crime or obtaining basic information on their rights and obligations during the proceedings and on the status of the case.

The introduction of some of these solutions is already legally admissible and technically possible. For some uses of AI, risk is deemed unacceptable and, therefore, these systems will be banned in the EU. The provisional agreement bans, for example, cognitive behavioural manipulation, the untargeted scraping of facial images from the internet or CCTV footage, emotion recognition in the workplace and educational institutions, social scoring, biometric categorisation to infer sensitive data, such as sexual orientation or religious beliefs, and some cases of predictive policing for individuals. Specific uses of AI, e.g., real-time remote biometric identification systems in publicly accessible spaces for law enforcement purposes, under the AI Act require the provision of additional safeguards and limits these exceptions to cases of victims of certain crimes, prevention of genuine, present, or foreseeable threats, such as terrorist attacks, and searches for people suspected of the most serious crimes.

In the area of justice, artificial intelligence, if it is based on trust, can lead not only to a reduction in costs but also to a significant reduction in the length of a trial and the delivery of a verdict. Further informatisation of the judiciary and the prosecution in Poland, allowing for the introduction of electronic communication between the parties to the proceedings and the procedural authority

(prosecutor or court), defining the principles for the use of electronic procedural documents or the rules for the admissible use of artificial intelligence in proceedings, requires not only the final adoption of the AI Act at the European Union level, but also profound changes to the Code of Criminal Procedure and comprehensive and widespread training of prosecutors, judges and officials.

REFERENCES

AI for Europe, COM/2018/237 final.

Amendments adopted by the European Parliament on 14 June 2023 on the proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)), P9_TA(2023)0236, https://www.europarl.europa.eu/doceo/document/TA-9-2023-0236_EN.html.

Analysis of the relationship of the Act on artificial intelligence with selected and proposed regulations legislation, GRAI, 2023, https://www.gov.pl/web/ai/raport-analiza-zwiazku-aktu-w-sprawie-sztucznej-inteligencji-z-wybranymi-obowiazujacymi-i-projektowanymi-regulacjami-prawnymi.

Chłopecki A., *Sztuczna inteligencja – szkice prawnicze i futurologiczne*, 2nd edition, Warsaw 2021.

Cieślak M., *Polska procedura karna. Podstawowe założenia teoretyczne*, Warsaw 1973.

Communication from the Commission to the European Parliament, the Council, the European Economic and Social Committee and the Committee of the Regions on Digitalisation of justice in the European Union, a toolbox of opportunities, COM/2020/710 final.

Coordinated Plan on Artificial Intelligence, COM(2018) 795 final.

Directive of the European Parliament and of the Council amending Directives 2011/99/EU and 2014/41/EU of the European Parliament and of the Council, Council Directive 2003/8/EC

and Council Framework Decisions 2002/584/JHA, 2003/577/JHA, 2005/214/JHA, 2006/783/JHA, 2008/909/JHA, 2008/947/JHA, 2009/829/JHA and 2009/948/JHA, as regards digitalisation of judicial cooperation, 2021/0395 (COD), PE-CONS 51/23, https://data.consilium.europa.eu/doc/document/PE-51-2023-INIT/en/pdf.

European Commission, White Paper on Artificial Intelligence – A European approach to excellence and trust, COM(2020) 65 final, 2020.

European e-Justice Strategy 2024–2028, 15509/23, https://data.consilium.europa.eu/doc/document/ST-15509-2023-INIT/en/pdf.

European Parliament Draft Report, Artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters, 2020/2016(INI).

European Parliament Draft Report, Artificial intelligence in education, culture and the audiovisual sector, 2020/2017(INI).

European Parliament resolution of 20 October 2020 on a civil liability regime for artificial intelligence, 2020/2014(INL).

European Parliament resolution of 20 October 2020 on a framework of ethical aspects of artificial intelligence, robotics and related technologies, 2020/2012(INL).

European Parliament resolution of 20 October 2020 on intellectual property rights for the development of artificial intelligence technologies, 2020/2015(INI).

Flasiński M., *Wstęp do sztucznej inteligencji*, Warsaw 2011.

Grzegorczyk T., *Jawność wewnętrzna postępowania sądowego*, [in:] W. Jasiński, K. Nowicki (eds.), *Jawność jako wymóg rzetelnego procesu karnego. Zagadnienia prawa polskiego i obcego*, Warsaw 2013.

Grzegorczyk T., Tylman J., *Polskie postępowanie karne*, Warsaw 2014.

Guidelines for Use of Generative Artificial Intelligence in Courts and Tribunals, Artificial Intelligence Advisory Group, 2023, www.courtsofnz.govt.nz/assets/6-Going-to-Court/practice-directions/practice-https://guidelines/all-benches/20231207-GenAI-Guidelines-Judicial.pdf.

Harnad S., *Minds, Machines and Turing: The Indistinguishability of Indistinguishables*, "Journal of Logic, Language, and

Information (special issue on Alan Turing and Artificial Intelligence)", https://web-archive.southampton.ac.uk/cogprints.org/2615/1/harnadoo.turing.html.
https://bloggersideas.com/pl/best-ai-content-writing-softwares.
https://digital-strategy.ec.europa.eu/en/news/.eu-member-states-sign-cooperate-artificial-intelligence.
https://ec.europa.eu/info/sites/default/files/political-guidelines-next-commission_en_0.pdf, p. 13.
https://pk.gov.pl/aktualnosci/aktualnosci-prokuratury-krajowej/system-dygitalizacji-akt-sda-konferencja-prokuratury-generalnej-podsumowujaca-projekt/.
https://www.gov.pl/web/ai/powolanie-zespolu-zadaniowego-do-spraw-realizacji-polityki-ai.
https://www.oecd-ilibrary.org/docserver/d62f618a-en.pdf?expires=1657742576&id=id&accname=guest&checksum=491B7B641D204752F82B0A28B24E6C24.
Jasiński W., *Jawność wewnętrzna postępowania sądowego*, [in:] J. Skorupka (ed.), *Jawność procesu karnego*, Warsaw 2012.
Kaftal A., *Jawność postępowania karnego w świetle nowego kodeksu postępowania karnego*, "Nowe Prawo" 1969, No. 11–12.
Kaplan J., *Artificial Intelligence: What Everyone Needs to Know*, Oxford University Press 2016.
Kasperska A., *Problemy zastosowanie sztucznych sieci neuronalnych w praktyce prawniczej*, "Przegląd Prawa Publicznego" 2017, No. 11.
Koper R., *Zasada jawności a jawność wewnętrzna w procesie karnym*, "Studia Prawnicze" 2019, No. 2 (218), DOI:10.37232/sp.2019.2.6.
Krasuski A., *Status prawny sztucznego agenta. Podstawy prawne zastosowania sztucznej inteligencji*, Warsaw 2021.
Kroplewski R., *Odporność AI dla odpornej wspólnoty*, [in:] *Cyberbezpieczeństwo AI. AI w cyberbezpieczeństwie*, Warsaw 2023, https://www.nask.pl/pl/aktualnosci/5237,Cyberbezpieczenstwo-AI-AI-w-cyberbezpieczenstwie-wazna-publikacja-NASK.html.
Lai L., Świerczyński M., *Prawo sztucznej inteligencji*, Warsaw 2020.

Law on the foundation platform of the industry of the future of 17 January 2019, Journal of Laws of 2019, item 229.

McCarthy J., Minsky M., Rochester N., Shannon C., *A Proposal for the Dartmouth Summer Research Project on Artificial Intelligence*, 1955, http://www-formal.stanford.edu/jmc/history/dartmouth/dartmouth.html.

Nowak-Gruca A., *Cyborg, czyli kto? O prawach cyborgów w ujęciu interdyscyplinarnym*, Warsaw 2023.

OECD (2019), AI Principles, https://oecd.ai/en/ai-principles.

OECD (2019), Recommendation of Council of OECD on Artifcial Intelligence. https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449.

Proposal for a regulation of the European Parliament and of the Council on laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts (COM(2021)0206 – C9-0146/2021 – 2021/0106(COD)).

Proposal for a Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts - Analysis of the final compromise text with a view to agreement, 26.01.2024, https://data.consilium.europa.eu/doc/document/ST-5662-2024-INIT/en/pdf.

Regulation (EU) 2020/1783 of the European Parliament and of the Council of 25 November 2020 on cooperation between the courts of the Member States in the taking of evidence in civil or commercial matters (taking of evidence) (recast), OJ L 405, 2.12.2020.

Regulation (EU) 2020/1784 of the European Parliament and of the Council of 25 November 2020 on the service in the Member States of judicial and extrajudicial documents in civil or commercial matters (service of documents) (recast), OJ L 405, 2.12.2020.

Regulation (EU) 2022/850 of the European Parliament and of the Council of 30 May 2022 on a computerised system for the cross-border electronic exchange of data in the area of judicial cooperation in civil and criminal matters (e-CODEX

system), and amending Regulation (EU) 2018/1726, OJ L 150, 1.06.2022.

Regulation (EU) 2023/1543 of the European Parliament and of the Council of 12 July 2023 on European Production Orders and European Preservation Orders for electronic evidence in criminal proceedings and for the execution of custodial sentences following criminal proceedings, OJ L 191, 28.07.2023, pp. 118–180.

Regulation of the European Parliament and of the Council on the digitalisation of judicial cooperation and access to justice in cross-border civil, commercial and criminal matters, and amending certain acts in the field of judicial cooperation, 2021/0394 (COD); PE-CONS 50/23; https://data.consilium.europa.eu/doc/document/PE-50-2023-INIT/en/pdf.

Regulation of the Council of Ministers of 7 June 2017 on granting the Scientific and Academic Computer Network the status of a state research institute, Journal of Laws of 2017, item 1193.

Regulation of the Minister of Internal Affairs and Administration of 12 January 2022 on the qualification procedure for candidates applying for admission to service in the Police, Journal of Laws of 2022, item 109.

Regulation of the Minister of Labour and Social Policy of 7 August 2014 on the classification of professions and specialities for the needs of the labour market and the scope of its application, i.e., Journal of Laws of 2018, item 227, as amended.

Regulation of the Minister of National Education of 14 February 2017 on the programme basis of pre-school education and the programme basis of general education for primary school, including for pupils with moderate or severe intellectual disabilities, general education for an industry school of the first degree, general education for a special school preparing for work and general education for a post-secondary school, Journal of Laws of 2017, item 356, as amended.

Regulation of the Minister of National Education of 30 January 2018 on the programme basis of general education for general secondary school, technical secondary school and industry secondary school, Journal of Laws of 2018, item 467, as amended.

Resolution No. 196 of the Council of Ministers of 28 December 2020 on the establishment of the 'Policy for the development of artificial intelligence in Poland from 2020', M.P. of 2021, item 23.

Rojszczak M., [in:] *Sztuczna inteligencja, blockchain, cyberbezpieczeństwo oraz dane osobowe. Zagadnienia wybrane*, K. Flaga-Gieruszyńska, J. Gołaczyński, D. Szostek (eds.), Warsaw 2019.

Turing A.M., *Computing Machinery and Intelligence*, "Mind" 1950, Vol. 59, No. 236.

UNESCO (2021), Recommendations on Ethics of AI, https://unesdoc.unesco.org/ark:/48223/pf0000381137.

Waltoś S., Hofmański P., *Proces karny. Zarys systemu*, Warsaw 2018.

Wiliński P., *Świadek incognito w polskim procesie karnym*, Kraków 2003.

Wójcicka B., *Jawność postępowania sądowego w polskim procesie karnym*, Łódź 1989.

DR RAFAŁ WIELKI
University of Opole | DOI: 10.32041/9788367149372/5

Chapter 5. Use of Artificial Intelligence in Law Enforcement and Criminal Justice

5.1. Introduction

The issue of artificial intelligence (AI) and the law is complicated. Lawmakers are faced with the challenge of how to legally regulate issues related to artificial intelligence, and virtually no legal system in the world addresses this problem comprehensively. A lot of challenges are posed by criminal law, specific by its nature, so there is a need for research on the possibilities of using artificial intelligence in the areas of law enforcement and criminal justice.

The main goal of my research is to provide a framework for the use of AI and databases by law enforcement agencies. By examining the problems of using AI to combat crime, and to detect individuals, I will attempt to analyse how AI can be used to improve crime prevention, investigation, and public safety. I will also consider whether artificial intelligence can be used in the broader field of criminal law, that is, the administration of justice as well. Through the research, it will be possible to determine how to create a legal framework for the use of AI in law enforcement and the justice system to avoid violations of citizens' personal freedoms. The considerations are carried out by the method of logical analysis with elements of heuristics.

The research questions I have set include the following problems. I want to consider: whether it is possible to use AI to support decision-making processes in law enforcement and the criminal justice system; what AI tools can be used to combat crime; whether AI can support crime prevention and enable crime prediction; what problems may arise from the use of artificial intelligence in criminal law, and what the shape of legal regulations should be.

I hypothesise that the use of AI for decision support should be limited to technical activities. The nature of criminal cases, the complexity of modus operandi, motives, and the presence of emotional factors require a broader human view of the problem. Criminal law doesn't keep up with social changes, but AI can be used to prevent and detect crimes before they are committed through the use of databases and big data. However, it may be necessary to clarify which databases law enforcement can use and to what extent they may be permitted to use them.

5.2. Artificial intelligence and law enforcement

5.2.1. TO START WITH: THE "GOOD GUYS" APPROACH

Technologies using artificial intelligence are increasingly being used in everyday matters, even if we are not aware of it. They surround us when we use search engines, online translations, social media, make cashless payments, use satellite navigation or use a cell phone in daily activities, etc.

What is artificial intelligence? Let's assume practically and simplistically that Artificial Intelligence (AI) is not a specific technology, but rather a collection of computational methods and techniques. There is no single AI and there is a lack of consensus among AI researchers on a universal definition. This is because AI means different things to different people and can be used in conjunction with a variety of other technologies, such as the Internet of Things and robotics. So, artificial intelligence is a collection of interrelated

technologies used to solve problems and perform tasks that, when humans do them, requires thinking.[1]

As with any significant discovery or invention for civilisation, it can be exploited in many ways, which is particularly evident in the relationship between "policemen and thieves." Criminals have always reached for the achievements of civilisation, and in the age of technological progress this phenomenon is gaining momentum. It is no different for law enforcement agencies, which are also trying to make effective use of the achievements of science.

It probably won't be an abuse to say that it is usually the criminals who are one step ahead of law enforcement. It is likely that this state of affairs is due to the fact that criminals find it easier to adapt to the changing social reality. On "this" side of the line there are also organisations, structures and even rules. The difference is that they are less formalised and thus easier to adapt to change. Law enforcement agencies are complex organisations that tend to be hierarchical, often bureaucratic, and the scope of their activities is limited by law (and it's worth noting that legislators often fail to keep up with social change). In turn, this means that the actions of the services are most often a reaction to a crime committed.

Of course, policing strategies are transforming, so the use of the so-called proactive approach can be increasingly observed. We are talking about the implementation of detection activities already at the time of committing a crime, or even at a time well before the crime is committed, with a focus on taking the initiative. It is nowadays accepted that the features of modern strategies include:

- intensive development of various forms of cooperation with the community (*community policing*);
- realistic solutions to specific security problems (*problem-oriented policing*);

[1] T. Walsh, N. Levy, G. Bell, A. Elliott, J. Maclaurin, I.M.Y. Mareels, F.M. Wood, *The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing*, Report for the Australian Council of Learned Academies, Melbourne 2019, p. 14.

– extensive collection and processing of all information that can be used to prevent and combat crime (*intelligence-led policing*).[2]

Intelligence-led policing is particularly important. After the terrorist attacks of 11 September 2001, there was a need to integrate intelligence with operational policing activities.[3] Regardless of this trend of proactive measures, other needs have emerged for the use of artificial intelligence by law enforcement agencies. The use of AI in policing is likely to invoke the same vision of "smarter" policing, bringing better security to a world where crime is increasingly complex and ever more well-organised. There are numerous ways in which AI could play an important role in policing: from the automation of crime reporting and other police administrative tasks, the use of autonomous vehicles for mobile patrols and emergency responses, and speech and image recognition software for investigative purposes, to machine learning to predict the times or places of crime occurrence and identify suspects.[4]

It has been increasingly questioned whether it is possible to predict criminal events, hence the concept of predictive policing has emerged. According to one view, we are talking about forward-thinking crime prevention, which combines technology, management practices, real-time data analysis, problem solving and information-led policing. These, in turn, are expected to lead to results such as crime reduction, increased efficiency of services, and increased levels of innovation as well as the modernisation of them.[5] Undoubtedly, many definitions of the concept can be found, although the one cited above seems to be fairly universal in nature.

---

[2] N. Tilley, *Modern Approaches to Policing: Community, Problem-Oriented and Intelligence-Led Policing*, [in:] *Handbook of Policing*, T. Newburn (ed.), Abingdon 2008, p. 373.

[3] J. Ratcliffe, *Intelligence-Led Policing*, [in:] *Encyclopedia of Criminology and Criminal Justice*, G. Bruinsma, D. Weisburd (eds.), New York 2014, p. 2573.

[4] J. Chan, *The Future of AI in Policing. Exploring the Sociotechnical Imaginaries*, [in:] *Predictive Policing and Artificial Intelligence*, J.L.M. McDaniel, K.G. Pease (eds.), Routledge, New York 2021, p. 48.

[5] W. Bratton, J. Morgan, S. Malinowski, *Fighting Crime in the Information Age: The Promise of Predictive Policing*, Los Angeles 2009, p. 3.

The key observations are those of S. Egbert and M. Leese. According to the authors, "Predictive policing takes up and incorporates a number of technical, economic, and political trajectories. The use of algorithmic crime analysis tools is generally presented as an elegant way to resolve organisational shortcomings and external pressures. Not surprisingly, then, predictive policing has over the past decade spread rather quickly into multiple national and local contexts around the globe."[6] For the purposes of this paper, let's take this general concept as a heuristic for further consideration.

Following up on the introductory remarks cited above, let's try to think about how artificial intelligence can be used by law enforcement agencies in specific ways.

5.2.2. DATA ANALYSIS AND PATTERN RECOGNITION

One of the key issues in the use of artificial intelligence in law enforcement is data analysis, particularly of large data sets. Data analytics is a broad term that encompasses the use of mathematical and statistical methods, techniques and tools to reveal the hidden knowledge that resides in data. This is used to make better organisational decisions. Data analysis can be quantitative or qualitative.[7]

This potential is particularly important not only because of the volume of data sets to be analysed. Also important is the issue of information redundancy. We are talking about the phenomenon in which information arrives through different channels, is often incomplete, distorted, duplicated, repeated, arrives from different sources, and may have the same or similar content.[8] In short, information occurs in redundancy.

---

[6] S. Egbert, M. Leese, *Criminal Futures: Predictive Policing and Everyday Police Work*, New York 2021, p. 28.

[7] See: C.L. Borgman, *The Conundrum of Sharing Research Data*, "Journal of the American Society for Information Science and Technology" 2012, Vol. 63, No. 6, pp. 1059–1078.

[8] R.M. Clark, *Intelligence Analysis. A Target-Centric Approach*, Los Angeles 2013, p. 143–144.

There is no doubt in the scientific community about the enormous potential of data analysis in relation to computer forensics. It is a very important area of forensic research, as digital traces are rapidly increasing in number, and their disclosure looks a little different from physical traces. The peculiarities of digital traces are their fuzziness and the difficulty in linking them to the specific person who left them.

For these reasons, it is noted that computer forensics and data analysis using artificial intelligence can be particularly useful in the following areas:

- computer forensics,
- network forensics,
- software forensics,
- mobile forensics,
- memory forensics,
- malware forensics,
- database forensics,
- social media forensics,
- cloud forensics,
- bitcoin forensics,[9]
- big data forensics,
- anti-forensics.[10]

Criminal reality is taking a slightly different shape. In place of criminal crime, we are increasingly confronted with the development of economic crime or cybercrime. For these reasons, we should agree with the authors that these areas should be of particular interest to law enforcement agencies. Computers, phones, flash memories, clouds and all network services are carriers of huge data resources, including photographs, videos, documents, location data or metadata, among others. These, in turn, can be

[9] The authors used the phrase "bitcoin forensics", but it is too narrow a term, since bitcoin is only one of many cryptocurrencies. Hence, it is better to talk about cryptocurrency forensics.

[10] P.K. Parichha, *Introduction to Digital Forensics*, [in:] *Big Data Analytics and Computing for Digital Forensic Investigations*, S. Satpathy, S.N. Mohanty (eds.), Boca Raton 2020, pp. 8–13.

very useful in crime detection, but analysing them with traditional methods using human resources would basically destabilise the work of investigators.

The issue of social media is certainly worth noting, although we will also analyse them more extensively later in the paper. Investigators should skilfully use social media as a source of information. Social media has caused a revolutionary change in social behaviour. Studying the dependencies present in social behaviour online is one of the more interesting challenges of criminology today. From the point of view of criminology, it may be most relevant to determine the relevancy of information posted by social media users. All social media activities leave digital footprints, becoming the largest collections of information about people and communities to date, which may even allow a separate category of social media intelligence to be distinguished known as SOCMINT.[11] The vast amount of data collected on social media often requires the use of tools using artificial intelligence to support traditional open-source, person-to-person intelligence.

As we have noted, an important area of data analytics use is cryptocurrencies. Most transactions in the cryptocurrency market are not questionable, but some of them can be used to finance or profit from criminal activity It is not the purpose of this study to present a methodology for detecting criminal transactions using advanced tools, as this is too vast a topic.[12] However, it is worth noting that cryptocurrency wallet transactions are often public. Some analysts have recognised the problem of using cryptocurrencies to handle the finances of extremist movements, so analysts have decided to use bots based on artificial intelligence to make public every transaction on wallets belonging to radicals.[13]

---

[11] D. Omand, C. Miller, J. Bartlett, *Towards the Discipline of Social Media Intelligence*, [in:] *Open Source Intelligence in the Twenty-First Century. New Approaches and Opportunities*, C. Hobbs, M. Moran, D. Salisbury (eds.), London 2014, p. 24.

[12] See: N. Furneaux, *Investigating Cryptocurriences: Understanding, Extracting, and Analyzing Blockchain Evidence*, Indianapolis 2018.

[13] D. Carliste, T. Keatinge, F. Keen, *Virtual Currencies and Terrorist Financing: Assessing the Risks and Evaluating Responses*, Study for the REER Committee, European Parliament, Brussels 2018, pp. 31–34.

The same is true when one has to deal with malware or anti-forensics activities. Behind every such criminal activity there is a person (or group of people) somewhere, but getting to the original source is very complicated. The traditional "thread-to-the-pattern" investigation seems to be an insufficient tool. Pattern recognition today requires the use of advanced tools based on artificial intelligence. This is used to identify types or clusters of data in an investigation. Pattern recognition systems are essentially classifiers – that is to say, they answer the question: is this piece of data a member of the class X, where X is the type of data the user is interested in? In order to work successfully, pattern recognition techniques therefore have to try to match a given datum/the given data against all possible pieces of data (or as near as is computationally feasible), which can involve a large amount of matches, and the patterns have to have sufficient generality to match all positive matches but sufficient specificity to not match any of the negative examples.[14] Of course, the implementation of new research methods in forensics always requires their validation, and many of the solutions based on artificial intelligence are in the early stages of development. Happily, in this area, too, researchers are striving to ensure correct methodology,[15] it being extremely important for the subsequent use of evidence in court.

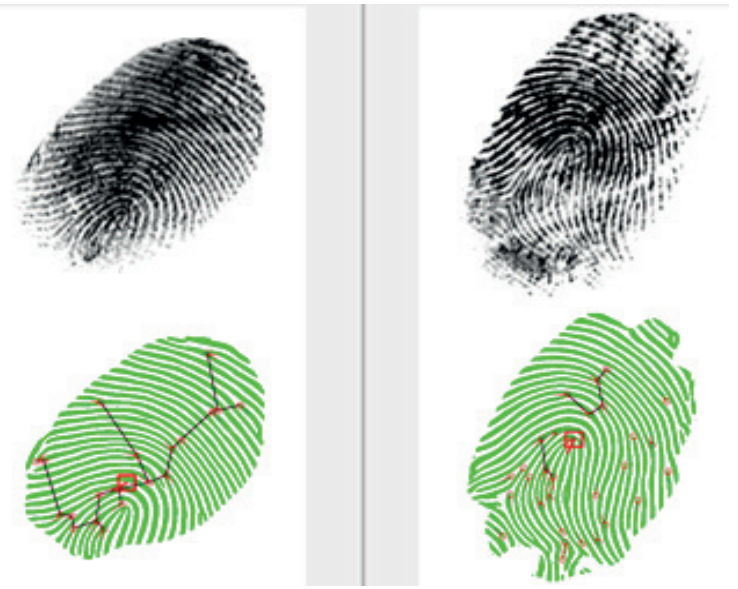
5.2.3. IMAGE RECOGNITION AND BIOMETRICS

One of the most important areas of using artificial intelligence for law enforcement is image recognition. The first groundbreaking tool appears to have been AFIS (Automated Fingerprint Identification System). Until the system was developed, fingerprint cards were stored as physical collections. A major problem was the collections'

[14] F. Mitchell, *The Use of Artificial Intelligence ind Digital Forensics: An Introduction*, "Digital Evidence and Electronic Signature Law Review" 2010, Vol. 7, pp. 37–38.

[15] A.A. Solane, M.A. Biasotti, *Digital Forensics AI: Evaluating, Standardizing and Optiimizing Digital Evidence Mining Techniques*, Künstl Intelligence (Open Access), 2022.

rapid expansion, which made it difficult to identify and correlate specific features (minutiae) in fingerprint traces taken from suspects and those collected in catalogues. With the early era of computerisation in the 1960s, AFIS was gradually developed, which allowed fingerprint images to be digitally processed and compared with those collected in databases[16] (picture 5.1).



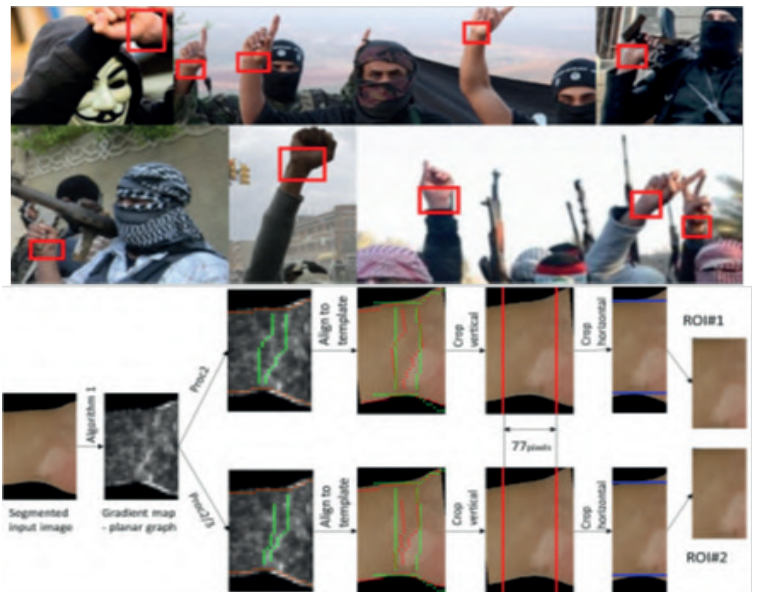
Picture 5.1. Listing of common features of the studied traces in AFIS

Source: own elaboration.

Nowadays, AFIS can be linked to multiple databases, facilitating international law enforcement cooperation. The computing power of computers is many times greater than when the system was implemented, and allows the presentation, in a very short time, of several comparison traces most similar in characteristics to the evidence trace. Of course, we must remember that although AFIS

[16] D. Meuwly, *Automated Fingerprint Identification System*, [in:] *Wiley Encyclopedia of Forensic Science Vol. 1*, A. Jamieson, A. Moenssens (eds.), Wiley, Chichester 2009, p. 249.

is a system that automates the work of experts, the identification of evidentiary traces cannot be done automatically. Detailed comparative studies and the decision on trace identification must ultimately be made by an expert.



Picture 5.2. Wrist identification areas based on images and AI algorithm-processed wrist image

Source: W. Matkowski, K.S.C. Frodo, A.W.K. Kong, *A Study on Wrist Identification for Forensic Investigation*, "Image and Vision Computing" 2019, Vol. 88, p. 96–112.

Closely related to image recognition tools is biometrics. It is a knowledge-generating activity that uses the entry of a person's biometric data into a specific database. It is used to determine a person's distinctive physical or behavioural characteristics in order to identify him or her,[17] often in order to gain authorised access to some area. Biometrics is also a security measure meant to quickly and cheaply protect critical infrastructure facilities. Access to computers, mobile devices, ATMs, etc., can also be protected biometrically. Biometrics

[17] M.J. Palmiotto, *Criminal Investigation*, Boca Raton 2012, p. 266.

is a form of human identification, carried out with such a small area of uncertainty that the level of diagnostic value of this identification is considered sufficient for certain practical applications, but not for evidential purposes.[18]

The characteristics in biometric identification (biometric data) are mainly mappings of finger(s), palm(s), iris(-es), face(s), but also voice(s), smell(s) or the gait.[19] There is also known research into the possibility of identifying a person based on mapping the areolae of the nipple[20] or wrist from images of participants in demonstrations or people suspected of terrorist activity[21] (picture 5.2).

Picture 5.3. Biometric security screening at Tel Aviv's Ben Gurion Airport

Source: www.israele360.com.

Technically achievable, although controversial in many respects (as will be discussed later), is the use of real-time biometric identification systems. They are most often used for security purposes

[18] J. Konieczny, *Kryminalistyczny leksykon śledztwa* (*Forensic Lexicon of Investigation*), Opole 2020, pp. 43–44.

[19] M. Saini, A.K. Kapoor, *Biometrics in Forensic Identification: Applications and Challenges*, "Journal of Forensic Medicine" 2016, Vol. 1, Issue 2, p. 2.

[20] W. Matkowski, K. Matkowski, A.W.K. Kong, C. Lloyd Hall, *The Nipple-Areola Complex for Criminal Identification*, 2019 International Conference on Biometrics (ICB), 2019, pp. 1–6.

[21] W. Matkowski, K.S.C. Frodo, A.W.K. Kong, *A Study on Wrist Identification for Forensic Investigation*, "Image and Vision Computing" 2019, Vol. 88, pp. 96–112.

in critical infrastructure facilities. Biometric technologies are also increasingly utilised in identification (that is determining, through the use of one-to-many matching, who the person is, or in some instances who the person is not. The clearest example of one-to-many matching is evident in the use of facial recognition technology matched to CCTV cameras. Air passengers can have their faces scanned and checked against a database of "terrorists" or other "wanted" people[22] (picture 5.3).

The last area explored is the issue of evidence. Artificial intelligence can play a significant role in evidential analysis. It is a set of rules for evidentiary reasoning. It is used to establish such rules for conducting evidentiary reasoning to eliminate unjustified decisions by trial authorities. Most often, it uses specific methodological assumptions, computer achievements, and tools that maximise the persuasiveness of arguments.[23]

5.2.4. STATISTICAL EVIDENCE

While this is a broad concept relating to the analysis of information for the criminal process, the issue of building statistical evidence seems of particular interest. Recent decades have also brought a change in the approach to the work of experts and the ways in which research results are presented to the court. This has mainly happened due to the DNA evidence, which has influenced a paradigm shift in the work of experts of various specialties in forensic science. Expert evidence makes it possible to establish the facts of the case, especially those that are relevant to the case, which are referred to as evidentiary facts.[24] In fact, on the ground of forensic science, a critical approach has begun to be taken to the traditional paradigm

[22] D. Lyon, *Surveillance after September 11*, "Sociological Research Online" 2001, Vol. 6, No. 3, http://www.socresonline.org.uk/6/3/lyon.html [accessed on: 30 October 2022].

[23] A. Ibek, J. Necessary, K. Wojcik, *Investigative Analysis (Investigative Analysis)*, "State and Law" 2018, No. 6, p. 51.

[24] H.L. Ho, *A Philosophy of Evidence Law. Justice in the Search for Truth*, Oxford 2008, p. 11.

in comparative forensic research, which involves the formulation of categorical individual identification judgments, as long as a "sufficient" number of common features are found in the disputed and comparative material. In other words: categorical statements take the form of judgments such as "the trace is from person X," "the traces are identical," "there is a 100% correspondence of features," etc.

Over time, academic discussions began to take place on the shape of forensic expertise. DNA examination revolutionised forensic biology and, more broadly, all forensic science. DNA examination has caused forensic science to begin looking at all categories of evidence in a new way, that is, through the prism of emphasising the techniques used, validation of tests, estimation of the skill level of experts, the use of databases and the consideration of probability in interpreting results. To illustrate this, a DNA expert is able to provide his opinion by concluding that a given DNA profile can be found in one in a trillion cases (99.999…% match), and this statement is based on databases of the occurrence of certain traits in the population.[25]

The formulation of categorical conclusions by experts has no logical or factual basis, eliminates the conduct of counter-evidence that could be used by the defence, and can also limit the determining authority by violating the rule of main fact, which states that an expert is not authorised to pronounce on the incident of a main fact in a criminal trial.[26] In other words, in the light of modern scientific knowledge on the grounds of criminal proceedings, it is only possible to issue probative (probabilistic) conclusions.

The European Network of Forensic Science Institutes has concluded that the only scientifically sound method of estimating the strength of evidence is the likelihood ratio, derived from Bayes' theorem. Estimating the value of forensic findings before a court of law is done using probability as a measure of uncertainty, based on

[25] G.T. Duncan, M.L. Tracey, E. Stauffer, *DNA Typing*, [in:] *Forensic Science: An Introduction to Scientific and Investigative Techniques*, S.H. James, J.J. Nordby, S. Bell (eds.), Boca Raton 2014, p. 229.

[26] C. Aitken, F. Taroni, *Statistics and the Evaluation of Evidence for Forensic Scientists*, Chichester 2004, pp. 86–87.

expert findings, related data, expert knowledge, and case specifics and conditions, together with taking into account the selection of information, and is based on the determination of the credibility quotient. Expert practice should follow its logical principles. The expert structure defined by them applies to all fields of forensic science. The credibility quotient measures the strength of the support of the findings, providing the basis for discrimination between statements, occurring in the case. It is scientifically justified providing a defensible, based on the principles of logic, path of evidentiary reasoning.[27]

Forensic science involves supporting the narrative and arguments with strong statistical evidence. AI can build graphical structures that can support the building of scenarios and case stories. It can also help build graphical model situations that can be used to prove or disprove arguments, helping the law to make better judgments. AI provides mathematical and computational tools that can help to build statistically relevant and significant evidence. All this will reduce the errors and improve the understanding of the statistics behind a study.[28]

Currently, there is talk of developing artificial intelligence algorithms based on Bayes' theorem in criminological forecasting. One of the advantages of using the Bayesian approach is the ability to perform predictive analysis. This makes extensive use of the naive Bayesian classifier, which is widely used in programming and artificial intelligence, as it allows the exploration of extensive data resources to assign objects to particular sets. What's more, the naive Bayes classifier allows artificial intelligence to learn inferences, which makes it possible to automate many activities, including for predicting criminal behaviour.[29]

---

[27] European Network of Forensic Science Institutes, *ENFSI Guideline for Evaluative Reporting in Forensic Science: Strengthening the Evaluation of Forensic Results across Europe (STEOFRAE)*, ENFSI, Wiesbaden 2015, p. 6.

[28] S.K. Chinnikatti, *Artificial Intelligence in Forensic Science*, "Forensic Science and Addiction Research" 2018, Vol. 2, Issue 5, p. 182.

[29] M.S. Vural, M. Gök, *Criminal Prediction Using Naïve Bayes Theory*, "Neural Computing and Applications" 2017, Vol. 28, Issue 9, pp. 2581–2592.

As we can see, the scope of application of tools based on artificial intelligence in law enforcement agencies is very wide. As the analysis shows, changes in the world through the development of new technologies affect the work of law enforcement agencies and change existing paradigms. The usefulness of artificial intelligence in matters related to computer forensics and the disclosure of digital traces is not inconsiderable. The need to conduct investigations in a thicket of data and information requires the use of tools that will help solve problems arising from information overload and allow pattern recognition. Huge potential lies in image recognition systems, as they are a tool that facilitates forensic identification, including in real time. This, in turn, can serve to enhance public safety, but there will always be ethical concerns. To what extent law enforcement agencies can afford to not conduct mass surveillance of the public and potential human rights violations? These questions are given further thought in the section on artificial intelligence rule-making.

5.3. Artificial intelligence and criminal justice

5.3.1. PREDICTION IN CRIMINOLOGY

The analysed issues concerning the use of artificial intelligence by law enforcement agencies allow us to conclude that the potential for the utility of new technologies can be considerable for investigative and evidentiary activities. Since artificial intelligence can benefit law enforcement agencies and thus ensure public safety, the question arises: Can artificial intelligence influence the shape of justice?

Considerations should begin by identifying potential areas where artificial intelligence could affect the justice system. We will not focus here on the issue of criminal liability for crimes committed using artificial intelligence. This area requires separate research related to the use of artificial intelligence tools by criminals and will be the subject of a separate research paper prepared as part of this research project. However, we can take as a starting point

two areas relevant to the functioning of the justice system, namely criminological forecasting and procedural decision-making.

As we have already mentioned, research is being conducted on the use of algorithms based on Bayes' theorem to make criminological predictions, understood as an attempt to estimate a person's propensity to repeat criminal behaviour. Making criminological predictions is quite a challenge. First of all, it is reasoning carried out under conditions of uncertainty. On the one hand, certainty does not work in science, even less so in criminology or psychology, and this will always be a challenge in predictions regarding an individual's behaviour. Prediction of possible future criminal behaviour will always have to coexist with uncertainty and errors.[30]

Since uncertainty and errors will always accompany criminological forecasting, one may then wonder if artificial intelligence tools will be more effective than humans and reduce the level of uncertainty as well as the number of errors made.

This is a controversial issue, although there are already known cases in the world of using software based on artificial intelligence algorithms to estimate the risk of recidivism. Such a solution is used in the US using Northpointe software (formerly Correctional Offender Management Profiling for Alternative Sanctions – COMPAS). According to assurances from service provider Northpointe Suite Risk Needs Assessment, criminal justice professionals can:

- choose a validated, actuarial Risk Needs Assessment that best suits workflow, process and jurisdiction policy;
- make risk-based decisions using a library of alternative screenings designed for specific profiles such as domestic violence, DUI, sex offence, and mental health;
- build individualised case plans with our fully integrated feature to support the development and execution of treatment decisions and court-ordered conditions;
- track outcomes with a logical flow of case planning and programming from custody to community;

[30] G. Zara, D.P. Farrington, *Criminal Recidivism: Explanation, Prediction and Prevention*, New York 2016, p. 172.

- organise all offender and assessment data under a master person-identifier to view the span of activity for justice-involved persons and review historical assessment results and prior case plan progress;
- leverage existing static data in new assessments to eliminate redundant data entry and cut assessment administration time.[31]

Researchers have been critical of such forecasting tools, pointing out that complex decision boundaries pose serious challenges: "To forecast well, a researcher must understand the nature of the complexity, be able to translate that knowledge properly into an algebraic expression, and then have the data to construct an appropriate model. These requirements are daunting for criminal justice application." In addition, they note that in criminal justice, where real lives may be at stake, the consequences can be significant.[32] In their response to this scientific report, the service provider's representatives partially concede the point that greater precision in predictive tools can be of practical importance, and that the criticisms formulated can be of value to criminology and the justice system as a wake-up call on research tools. The software itself remains useful in their view.[33]

Although recidivism prediction algorithms are used in criminal cases in parole consideration or sentencing decisions, they remain controversial. According to a more recent experimental study conducted in the US, it appears that these tools may be even less accurate or fair than predictions made by people with modest or no experience in criminal adjudication.[34]

---

[31] https://www.equivant.com/northpointe-risk-need-assessments/ [accessed on: 22 August 2022].

[32] R.A. Berk, J. Bleich, *Statistical Procedures for Forecasting Criminal Behavior. A Comparative Assessment*, "Criminology & Public Policy" 2013, Vol. 12, No. 3, p. 541.

[33] T. Brennan, W.L. Oliver, *The Emergence of Machine Learning Techniques in Criminology. Implications of Complexity in our Data and in Research Questions*, "Criminology & Public Policy" 2013, Vol. 12, No. 3, pp. 551–562.

[34] J. Dressen, H. Farid, *The Accuracy, Fairness, and Limits of Predicting Recidivism*, "Science Advances" 2018, Vol. 4, Issue 1, pp. 1–5.

Views on this issue are disputed, although a current critical of the use of artificial intelligence in criminological forecasting has quite clearly emerged. An interesting review of such comments was made by K. Mamak, a Polish researcher dealing with new technologies and criminal law.[35] In the course of his research, it is worth noting the sources the author reached, which are related to the problem of discrimination. According to many researchers, the factors that are taken into account in terms of estimating the risk of recidivism in a particular person are controversial. These include property status, place of residence (zip code), employment, age, gender, and race, among others, so predictive systems are said to be based on socioeconomic and demographic criteria, making *a priori* higher risk in certain subgroups.[36]

In the case of the system in question – which may inspire disbelief in the eyes of European legal scholars, to say the least – the problem is the lack of access to the source code, as the software manufacturer and public institutions cite trade secrets.[37] This makes it impossible to verify exactly how a particular person's diagnosis is going. However, these tools are entering the criminal justice system and are already being used in courtrooms across the country in parole, pre-trial, and sentencing determinations.

From the analysis conducted, a critical approach to the use of tools based on artificial intelligence in criminological forecasting is clarified, especially if it can have an impact on punishment. Thus, the question should be posed: can artificial intelligence be used in decision-making in the administration of justice?

---

[35] K. Mamak, *Digital Revolution and Criminal Law*, Kraków 2019, pp. 101–124.

[36] K. Kirkpatrick, *It's Not the Algorithm, It's the Data*, "Communications of the ACM" 2017, Vol. 60, Issue 2, pp. 21–23; G. Van Eijk, *Socioeconomic Marginality in Sentencing: The Built-In Bias in Risk Assessment Tools and the Reproduction of Social Inequality*, "Punishment & Society" 2016, pp. 1–19; N. Scurich, J. Monahan, *Evidence-Based Sentencing: Public Openness and Opposition to Using Gender, Age, and Race as Risk Factors for Recidivism*, "Law and Human Behavior" 2016, Vol. 40, Issue 1, pp. 36–41.

[37] T.R. Moore, *Trade Secrets & Algorithms as Barriers to Social Justice*, Washington 2017, pp. 3–5.

5.3.2. AUTOMATED DECISION-MAKING

Without a doubt, artificial intelligence is touching more and more areas related to law. This can include contract analysis, legal research, and e-discovery. Computer programmes can help lawyers analyse the other sides written submissions and to provide relevant case law. In addition, one can talk about drafting and proofreading submissions, translation of documents, case management, organisation of documents, estimation of costs, etc.[38] There are opportunities to use machine learning for tasks that have traditionally been performed by lawyers. Examples include mass review of contracts, helping to automatically merge contracts and other legal documents, preparing template documents, etc.[39]

The potential for tools based on artificial intelligence can be seen without much difficulty in civil cases involving payment, contract performance, etc. It seems that the question of the use of artificial intelligence tools in administrative cases may be similarly drawn. In these, very often the issuance of some kind of decision or settlement depends solely on the fulfilment of certain formal prerequisites. If the formal prerequisites are met by a party, one can imagine automating the process of issuing a decision. However, is there place for such tools in the justice system? Can a computer programme or robot replace a human or a decision-making body?

Before considering such extreme hypothetical situations, let's analyse whether the justice system can use artificial intelligence-based solutions for technical or administrative issues. Australian researchers considered the following areas:

- technology assisted review and discovery,
- automated online dispute resolution,
- prediction of litigation outcomes,
- criminal sentencing and risk assessment tools,
- automated decision-support and decision-making,

[38] M. Scherer, *Artificial Intelligence and Legal Decision-Making: the Wide Open?*, "Journal of International Arbitration" 2019, Vol. 36, No. 5, p. 540.

[39] H. Surden, *Artificial Intelligence and Law: An Overview,* "Georgia State University Law Review" 2019, Vol. 35, Issue 4, p. 1331.

– automated e-filing,
– triage and allocation of matters,
– natural-language processing,
– AI-supported legal research.[40]

From the point of view of criminal law, not all of these areas are crucial. Of cognitive interest is the issue of criminal sentencing and risk assessment tools, but we have noticed that there is a sizable group of critics of such solutions. Nevertheless, interesting solutions are proposed in the framework of automated decision-support and decision-making tools. According to the authors, work is underway in the UK to create a completely online court that would deal with certain minor torts, allow guilty pleas and predetermined punishment, all without a judge. The case could be handled in a completely automated manner, without any human supervision. Criminal courts would focus on more serious crimes, punishable by imprisonment.[41]

Might such a judicial model be worth considering? Certainly, in favour of such automated decision-making arrangements is the time and cost savings associated with minor criminal cases. Perhaps in misdemeanour cases where the evidence is credible and the offender admits guilt, such a system would make sense. However, if an admission of guilt did not occur, the case would have to go to a "human" court. However, this is an ethical problem as well, and requires a broader discussion in the scientific community, but reflection on this type of solution is advisable.

While technical in nature, automated e-filing can be a useful tool to reduce or eliminate physical case files. Storing and locating documents is easier when they are in electronic format. Further, the capacity of parties, lawyers and judges to search lengthy documents for particular words or phrases has become near instantaneous through the use of searchable files, and the ability to navigate

[40] F. Bell, L.B. Moses, M. Legg, J. Silove, M. Zalnieriute, *AI Decision-Making and the Courts. A Guide for Judges, Tribunal Members and Court Administrators*, Sydney 2022, pp. 15–29.

[41] UK Ministry of Justice, *Transforming Our Justice System: Assisted Digital Strategy, Automatic Online Conviction and Statutory Standard Penalty, and Panel Composition in Tribunals*, Government Response No. Cm 9391, 2017, p. 16.

between relevant documents has been facilitated through the use of hyperlinked documents.[42] Particularly in criminal cases, the digitisation of files can speed up court proceedings, as searching through individual volumes of files in multi-threaded cases is very time-consuming. In addition, if the e-filing systems used had the ability to complete some of the data automatically, the creation of documents would be much simpler. Such a solution, however, would require switching to electronic documentation of all actors in criminal proceedings, from the very beginning of a criminal case. This is by no means unattainable. Even from the Polish perspective, it seems that more and more public institutions are making it possible to handle a case in the form of electronic forms. This is convenient insofar as the system detects the absence of certain data or attachments, which reduces the problem of rectifying formal deficiencies.

Such solutions, which facilitate the administration of judicial units, can be developed by systems that perform the triage and allocation of matters. Interesting insights can be drawn from Israel's cloud-based judicial management system, Legal-Net. It is a cloud-based platform that handles the administration of all first-instance and appellate court cases litigated in the Magistrate and District courts in Israel, save for Supreme Court cases. All court appointments, hearings and courtroom assignments are scheduled through Legal-Net. The system also manages the court's "warehouse," keeping track of all cases in the system, classifying them according to several categories and indicating what stage they are at and the next steps awaiting each one. Legal-Net acts as a workstation for judges. It's where all court documents are filed (e.g., transcripts, depositions and interrogatories) and where the official register of subpoenaed witnesses is located, including the date they are expected to appear in court and any details of their testimony. This is where judges' drafts are written, and ultimately where decisions on various motions, as well as final resolutions, are written, published and stored. Judges are not authorised to write anything on their own computers; they are required to work on Legal-Net. The platform also serves as a communication system for lawyers and parties: it's where they

[42] F. Bell *et al.*, *op. cit.*, p. 25.

file most legal documents, including motions, declarations, exhibits, briefs, responses, requests to call witnesses, and statements of appeal, to name a few. It is also where they then receive court documents, including all decisions and rulings.[43]

Because the data is updated in real time, Legal-Net functions as a powerful statistical platform that is able to produce comprehensive and up-to-date reports in response to inquiries from authorised persons. From a managerial perspective, all of the information and data stored in Legal-Net is visible to the heads of the judiciary. Legal-Net enables them to obtain reliable information about the work of individual judges or a particular court, as well as about the entirety of court cases (provided, of course, that the data has been correctly entered by the registrars). The system can analyse the synchronous or diachronic case load of individual judges, sections in the courts, a single court or any court, and compare that section with others. It can offer information on the processing time of each application, conduct statistical analyses of the average (median or other denominator) processing time of specific decisions by specific judicial units, the percentage of granting or dismissal of specific applications (per judge or court), or other data-driven queries.[44]

At the same time, the authors note the other side of the coin. The price of such judicial amenities is a shift in the role of the judiciary toward functional service delivery rather than the administration of justice, geared to each individual. Such an ecosystem may also raise concerns about changing the behaviour of judges, who may tend to handle a case in a particular way, having been taught to do so by algorithms.[45]

Truly, there are many solutions based on artificial intelligence that can be used in the justice system to save time or costs. These can be legal databases, commonly used, by the way, by lawyers (e.g., Lex or Legalis) in Poland. Artificial intelligence makes it possible to

---

[43] A. Reichman, Y. Sagy, S. Balaban, *From a Panacea to a Panopticon: The Use and Misuse of Technology in the Regulation of Judges*, "Hasting Law Journal" 2020, Vol. 71, Issue 3, p. 598.

[44] *Ibidem,* p. 599.

[45] *Ibidem*, p. 636.

better obtain the content you are looking for, to find the right court rulings or theses. Another interesting tool is natural-language processing. Software based on machine learning recognises, processes and analyses spoken language and converts it to forms. This way there is no need for a human to record the cases, while at the same time it is not just a digital recording that requires human involvement to perform the transcription. This type of solution is currently being tested in China's justice system.[46]

Tools based on artificial intelligence that can potentially be used in the judiciary – as can be seen from the analysis – are numerous. Especially in technical and administrative matters, they can significantly facilitate the work and contribute to the efficiency of the judiciary, although they are not entirely without drawbacks, as they can affect the way people reason. Nevertheless, we can say with certainty that there is room for automated decision-making systems. It remains to return to the question posed earlier. Can artificial intelligence replace the judge himself?

When comparing the work done by artificial intelligence and humans, a number of similar arguments are most often cited. Artificial intelligence can reduce human errors if programmed correctly. Decisions are made from previously collected data and information, which are analysed by algorithms. Computers can take risks instead of humans, without humans risking their health, lives or reputations. Computers don't get tired of repetitive jobs, they don't get into a rut, and the decisions made are generally made much faster.

On the other hand, even the best algorithms cannot replicate a human. They can make rational decisions, which, however, are devoid of emotions and moral values. Hence, they do not know what is ethical, what is legal. When they run into a previously unfamiliar situation, they may behave inappropriately or suspend themselves. Algorithms are also not creative, whereas humans have the ability to "think outside of the box."

The criminal trial is not just about the crime and the criminal who must be tried in order to achieve justice, repair damage or prevent similar events. The criminal trial has the makings of an artistic

[46] F. Bell *et al.*, *op. cit.*, p. 28.

performance. In the thicket of evidence, facts, arguments in the courtroom, there are also human feelings, dramas of victims, experiences, but also attempts to sow uncertainty or persuade the jury, both by prosecutors and defence attorneys. It is difficult to imagine that amid such specific factors, which are considered through the prism of a person's life and work experience and a range of emotional states, a machine could make decisions about the future and life of a particular human individual.

We can agree that artificial intelligence may have utility value in assisting court case handling or autonomous decision-making processes, such as eliminating information asymmetry between departments in case handling, maximising case-handling efficiency, reducing wrong cases, promoting justice, etc. Although artificial intelligence plays an important role in litigation activities, in the judicial process, artificial intelligence occupies a subordinate position, merely assisting the judge in handling the case: the judge is the key and core of litigation work. The extent to which judicial decisions can be determined through statistical modelling, analysis and calculation, and controlled by rules and norms, will be the extent to which artificial intelligence can be applied. However, judicial judgment is not one-dimensional reasoning, it is in fact a complex activity open to universal practice, that is to say, it is open to moral, ethical and practical considerations.[47]

It seems that replacing human judges is impossible. This is not a matter of the limited (as yet) technical capabilities of artificial intelligence alone, but is primarily an ethical issue. The advent of new technologies in the justice system clearly creates additional opportunities to improve access to justice. These benefits come from reducing costs and delays and removing physical, psychological and informational barriers to access to justice. On the other hand, there is a significant risk that the focus on cost reduction and time savings, along with the dehumanisation of the criminal justice process, could

[47] Z. Xu, *Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities*, "Applied Artificial Intelligence. An International Journal" 2022, Vol. 36, Issue 1, p. 1042.

lead to a justice system that is no longer about justice, however we understand it.[48]

5.4. How to create laws on artificial intelligence?

5.4.1. POLAND'S MAIN GOALS

Consideration of lawmaking is worth starting with some initial observations. According to the workshops of the "Legal Aspects of Artificial Intelligence" group within the Polish-Hungarian Research Platform 2022 project, neither Hungarian nor Polish legislators have yet produced laws that comprehensively address the use of artificial intelligence, especially with regard to criminal law. Regulations are only being drafted, so it is worth referring first to the strategic assumptions being prepared in Poland.

The main document that takes into account policy intentions towards artificial intelligence is the Policy for the Development of Artificial Intelligence in Poland from 2020.[49] In particular, the AI Policy takes into account the objectives defined in the following strategic documents: Public Data Opening Programme; Strategy for Responsible Development; "From Paper to Digital Poland" Programme; "Dynamic Poland 2020" Strategy for Innovation and Efficiency of the Economy; Communication of the European Commission (EC) "Coordinated plan on Artificial Intelligence;" Position of the Visegrád Group on Artificial Intelligence; Recommendations of the High Level Expert Group on Artificial Intelligence (HLEG AI) to the European Commission in the form of "Ethics Guidelines for Trustworthy AI" and Recommendations on Policy and Investment in Trustworthy AI, as well as recommendations concerning the management of trustworthy artificial intelligence,

[48] See T. Sourdin, *Judges, Technology and Artificial Intelligence: The Artificial Judge*, Cheltenham 2021, pp. 187–188.

[49] Policy for the Development of Artificial Intelligence in Poland from 2020, Appendix to the Resolution No. 196 of the Council of Ministers of 28 December 2020 (item 23).

in the Strategy of Polish Foreign Policy and in the "Memorandum on the Development of Artificial Intelligence in Poland."

The main areas of interest in the development of artificial intelligence in Poland are:

- AI and society – activities, whose goal is to make Poland one of the major beneficiaries of a data-based economy, while raising awareness of the need for the continuous improvement of knowledge and skills, including digital competencies.
- AI and innovative companies – activities aimed at supporting Polish AI companies, creating financing mechanisms to foster their growth, increasing the number of orders, ensuring cooperation between start-ups and the government and introducing new pro-development regulations – digital sandboxes.
- AI and science – activities supporting the Polish academic and research communities in designing interdisciplinary challenges or solutions in the field of AI, taking into account both the humanities and social sciences; establishing AI departments, training PhD students, awarding grants for researchers and other activities aimed at preparing a staff of experts capable of creating AI-based solutions, taking into account the framework for ethical and safe use of this technology, for the benefit of the economy and the welfare of citizens.
- AI and education – activities that are supposed to be implemented at every level of education – from primary, through secondary education up to the university level, including course curricula for people threatened with losing their jobs as a result of progressing automation and deployment of new technologies, educational grants aimed at helping to prepare the best staff for the Polish AI economy.
- AI and international cooperation – international activities that will support the promotion of Polish business in the field of AI and the development of AI technologies that respect human dignity and fundamental human rights, in accordance with EU and OECD standards, as well as digital diplomacy

activities in the area of policies or regulations concerning artificial intelligence.

– AI and the public sector – activities aimed at supporting the public sector in the implementation of contracts concerning AI, better coordination of activities and further development of programmes such as GovTech Poland, as well as protecting the people from relevant risks and threats. Still others will comprise virtual data repositories, data trusts (trusted data space initiatives), or the Government Cloud, and making as much public data as possible open and available for use by citizens and businesses.[50]

As we can see, the Polish AI Policy does not list as strategic plans the areas of internal security, defence or justice. Only some of the activities related to cooperation between the private and defence sectors are included in another document.[51]

5.4.2. TRUSTWORTHY ARTIFICIAL INTELLIGENCE – EUROPEAN UNION'S APPROACH

It is worth noting that the issue of artificial intelligence is of interest to the European Union. In the absence of national solutions in both Poland and Hungary, considerations should start with proposals at the European level. This is particularly important because European Union law affects the laws of individual member states, so it is to be expected that artificial intelligence issues will be regulated at this level, as was the case with personal data regulated by the *General Data Protection Regulation*.[52]

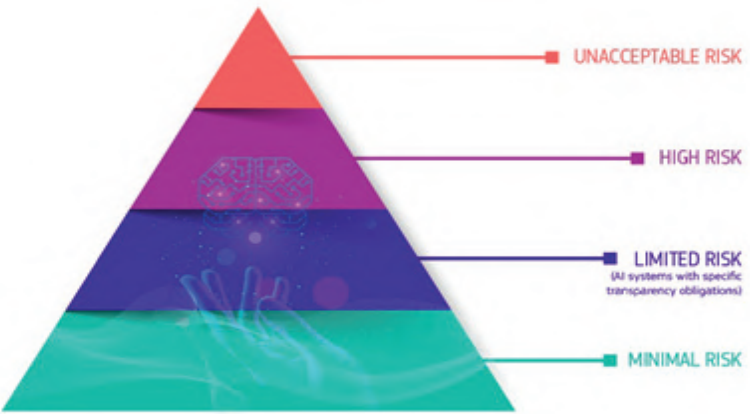
---

[50] *Ibidem*, pp. 6–7.

[51] See: National Security Strategy of the Republic of Poland, approved by the order of the President of the Republic of Poland from 12 May 2020, Monitor Polski, item 413.

[52] See: Regulation (EU) 2016/679 of the European Parliament and the of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

1 April 2021 the European Commission has published a draft regulation establishing harmonised rules for artificial intelligence, called the "Artificial Intelligence Act."[53] The Commission is proposing new rules to make sure that AI systems used in the EU are safe, transparent, ethical, unbiased and under human control. The proposed legal framework for artificial intelligence is based on distinguishing four categories of AI systems, depending on the level of risk posed by their use: unacceptable risk, high risk, limited risk, minimal risk (picture 5.4).



Picture 5.4. Levels of AI system risks proposed in the AI Act

Source: European Commission.

From the point of view of the considerations undertaken in this chapter, that is, issues related to public security and the administration of justice, we must look primarily at two levels: unacceptable risk and high risk. Due to the multifaceted nature of the proposals of this legislation, we will focus only on those elements that are related to the area under study.

[53] See: Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts, [SEC(2021) 167 final] – [SWD(2021) 84 final] – [SWD(2021) 85 final].

Unacceptable risks (prohibited AI practices) to systems otherwise contrary to the EU values of respect for human dignity, freedom, equality, democracy and the rule of law, and the fundamental rights of the Union, including the rights to non-discrimination, data protection and privacy, and the rights of the child, posing a clear risk to the security, livelihoods and rights of citizens, have been declared unacceptable. According to the draft, the use of "real-time" remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement shall be prohibited, unless and insofar as such use is strictly necessary for one of the following objectives:

- the targeted search for specific potential victims of crime, including missing children;
- the prevention of a specific, substantial and imminent threat to the life or physical safety of natural persons, including that of a terrorist attack;
- the detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to in Article 2(2) of Council Framework Decision 2002/584/JHA 62 and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State.[54]

Also, the use of "real-time" remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1 point d) shall take into account the following elements:

- the nature of the situation giving rise to the possible use, in particular the seriousness, probability and scale of the harm caused in the absence of the use of the system;
- the consequences of the use of the system for the rights and freedoms of all persons concerned, in particular the seriousness, probability and scale of those consequences.

In addition, the use of "real-time" remote biometric identification systems in publicly accessible spaces for the purpose of law enforcement for any of the objectives referred to in paragraph 1

[54] *Ibidem*, Article 5 paragraph 1 point (d).

point d) shall comply with necessary and proportionate safeguards and conditions in relation to the use, in particular as regards the temporal, geographic and personal limitations.[55]

As regards paragraphs 1 point (d) and 2, each individual use for the purpose of law enforcement of a "real-time" remote biometric identification system in publicly accessible spaces shall be subject to a prior authorisation granted by a judicial authority or by an independent administrative authority of the Member State in which the use is to take place, issued upon a reasoned request and in accordance with the detailed rules of national law referred to in paragraph 4. However, in a duly justified situation of urgency, the use of the system may be commenced without an authorisation and the authorisation may be requested only during or after the use. The competent judicial or administrative authority shall only grant the authorisation where it is satisfied, based on objective evidence or clear indications presented to it, that the use of the "real-time" remote biometric identification system at issue is necessary for and proportionate to achieving one of the objectives specified in paragraph 1, point (d), as identified in the request. In deciding on the request, the competent judicial or administrative authority shall take into account the elements referred to in paragraph 2.[56]

Classification rules for high-risk AI system are concretised in Annex III:

- biometric identification and categorisation of natural persons – AI systems intended to be used for the "real-time" and "post-remote" biometric identification of natural persons;
- AI systems intended to be used by law enforcement authorities for making individual risk assessments of natural persons in order to assess the risk of a natural person for offending or reoffending or the risk for potential victims of criminal offences;
- AI systems intended to be used by law enforcement authorities as polygraphs and similar tools or to detect the emotional state of a natural person;

[55] *Ibidem*, Article 5 paragraph 2.
[56] *Ibidem*, Article 5 paragraph 3.

- AI systems intended to be used by law enforcement authorities to detect deep fakes as referred to in article 52(3);
- AI systems intended to be used by law enforcement authorities for evaluation of the reliability of evidence in the course of investigation or prosecution of criminal offences;
- AI systems intended to be used by law enforcement authorities for predicting the occurrence or reoccurrence of an actual or potential criminal offence based on profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 or assessing personality traits and characteristics or past criminal behaviour of natural persons or groups;
- AI systems intended to be used by law enforcement authorities for profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680 in the course of detection, investigation or prosecution of criminal offences;
- AI systems intended to be used for crime analytics regarding natural persons, allowing law enforcement authorities to search complex related and unrelated large data sets available in different data sources or in different data formats in order to identify unknown patterns or discover hidden relationships in the data;
- AI systems intended to be used by competent public authorities as polygraphs and similar tools or to detect the emotional state of a natural person;
- AI systems intended to be used by competent public authorities to assess a risk, including a security risk, a risk of irregular immigration, or a health risk, posed by a natural person who intends to enter or has entered into the territory of a Member State;
- AI systems intended to be used by competent public authorities for the verification of the authenticity of travel documents and supporting documentation of natural persons and detect non-authentic documents by checking their security features;
- AI systems intended to assist competent public authorities for the examination of applications for asylum, visa and residence permits and associated complaints with regard to the eligibility of the natural persons applying for a status;

– AI systems intended to assist a judicial authority in researching and interpreting facts and the law and in applying the law to a concrete set of facts.[57]

As we can see, the proposals being discussed in the European Union address most of the issues presented as the potential of artificial intelligence in law enforcement, as well as, in part, issues devoted to criminal justice. The proposals appearing in the seeds of the Act on Artificial Intelligence very thoroughly address, for example, the issue of biometric identification and the systems that use it. It is also rightly assumed that the use of artificial intelligence tools in law enforcement entails a relatively high level of risk, as it may violate fundamental rights, and may also be ethically questionable. This shows that while the potential of artificial intelligence in criminal cases is very high, it requires consideration of all risks and uncertainties. Before subjecting specific provisions to a broader commentary, let's take a look at proposals for modifying the legal regulations presented. The project continues to be processed. By the end of 2022, the project has managed to be consulted in a number of bodies, such as: Economic and Social Committee, European Central Bank, and the European Committee of the Regions, while it is currently in the discussion stage at the Council of the European Union, where detailed provisions are being worked on so that the whole thing can eventually be approved by the European Parliament.

A report by the Committee on the Internal Market and Consumer Protection and the Committee on Civil Liberties, Justice and Home Affairs was published in April 2022. While law enforcement and criminal justice issues are not prioritised here, one gets the impression that there is an effort in the amendment proposals to limit the applicability of artificial intelligence to public safety and crime issues.[58]

---

[57] Annex III paragraphs 1, 6–8.

[58] See Committee on the Internal Market and Consumer Protection and Committee on Civil Liberties, Justice and Home Affairs, *Draft Report on the Proposal for a Regulation of the European Parliament and of the Council on Harmful Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM 2021/0206 – C9-0146/2021 – 2021/0106(COD))*, 2021/0106 (COD).

It seems that the social movement around restricting the use of artificial intelligence in these areas is becoming dominant. According to other reported comments, there has been a call to ban predictive policing altogether, arguing that place-based predictive policing systems are equally harmful, because research has shown how they reinforce discriminatory policing practices based on historical policing patterns, enhance the over-surveillance and criminalisation of racialised and working class communities, and equally challenge the presumption of innocence, on a collective basis. Among other demands, there is a ban on the use of biometric surveillance systems (including post facto and by private entities), systems for emotion recognition and biometric categorisation, and systems for risk assessment in the area of migration. Also proposed are obligations to conduct a human rights impact analysis of the system (at least for public institutions implementing such a system), to register high-risk systems used by private entities, to publish in an EU database information on the assumptions of the system in question, what concessions were made in its design and what the system optimises, as well as the results of the risk analysis, etc.[59]

As Fair Trials notes in turn, as more and more countries are turning to AI in law enforcement and criminal justice, it is more crucial than ever that the EU takes this opportunity to become a leading standard-setter in this area, ensuring the protection of fundamental rights. They believe that the proposals under consideration in the EU arena are not sufficiently dedicated to preventing the abuse of AI by law enforcement and criminal justice: "The Act must prohibit AI used by law enforcement, and judicial and criminal justice authorities used to predict, profile or assess people's risk or likelihood of 'criminal' behaviour, generate reasonable suspicion, and justify law enforcement or criminal justice action, such as surveillance, stop and search, arrest, detention, pre-trial detention, sentencing and

[59] An EU Artificial Intelligence Act for Fundamental Rights. A Civil Society Statement, Drafted by: European Digital Rights (EDRi), Access Now, Panoptykon Foundation, epicenter.works, AlgorithmWatch, European Disability Forum (EDF), Bits of Freedom, Fair Trials, PICUM, and ANEC (European consumer voice in standardisation), 2021, pp. 1–4.

probation." It was very rightly noted that the Artificial Intelligence Act must require greater openness, transparency and the ability to explain artificial intelligence systems This includes their use, the decisions they make, and, importantly, it must focus not only on providing transparency for users of the systems, but also for those affected by AI or AI-assisted decisions, given that AI can have a significant impact on individuals when used in law enforcement and criminal justice.[60]

Regulations are still under development, following public consultation. However, the discussion demonstrates that we are facing several dilemmas related to the use of artificial intelligence in law enforcement and criminal justice. Proposals coming out of the European Union are being closely watched by legal scholars from many parts of the world, as the first very comprehensive piece of legislation is being drafted to regulate artificial intelligence in a politically and economically significant region of the world.

5.5. Conclusions and *de lege ferenda* comments

There are many voices in the discussion and they sometimes take different extreme forms. As we can see from the projected content, the European Union plans to severely restrict the use of artificial intelligence in law enforcement. At the same time, these ideas are strongly supported by NGOs. There is a great deal of controversy surrounding predictive policing and the assignment of this technology to the category of unacceptable risk. It seems that at this point, with this level of development of artificial intelligence and the public's acceptance of new technologies, banning predictive policing is a good step. On the one hand, predictive policing is the next stage in the development of police services that are trying to adapt to social changes. In addition, in many places around the world this type of system is in operation, as the example of the United States or China shows. On the other hand, it should be remembered that

[60] EU Commission 'AI ACT' Consultation, Fair Trails' Response, 2021, Feedback Reference: F2665646, pp. 1–2.

in Europe – despite the whole range of divergence on social or worldview issues in individual member states – the right to individual privacy is an important value. Perhaps in the future there will be some change in public sentiment and it will be possible to discuss the use of predictive policing in law enforcement practice. In addition, the lack of clear algorithms – whose effectiveness, in any case, is not scientifically proven – is a significant obstacle today. In view of the above, it is necessary to agree with the calls to ban the use of predictive policing in EU law, and thus in the regulations of individual member states.

Another important issue to clarify is the use of biometric systems by law enforcement agencies. As we know, it has been advocated to prohibit the use of "real-time" remote biometric identification systems in public spaces for law enforcement purposes. This refers to systems used to identify individuals remotely by comparing their biometric data (e.g., facial image, gait, silhouette outline) with biometric data contained in a reference database, where data collection, comparison and identification take place without significant delay). On the one hand, it can be seen that such identification systems are being developed within the sciences, and their potential is enormous. Perhaps in the future, biometric identification will supplant the identification methods previously used in forensics, but one important condition must be met. Biometric identification methods must be validated every time. Validation is the determination of the statistical parameters of a quantitative laboratory test method, performed to demonstrate that the method is suitable for achieving the specific purposes for which it was developed. Validation includes the determination of, among other things, accuracy (which is a measure of the deviation of the results obtained with it from the actual value of the measured quantity), its precision (i.e., the spread of the results) and other parameters. Thus, validation of a method allows recognition of the level of error (area of uncertainty) associated with its use.

As we know, in the proposed European Union-level legislation, biometric identification is classified as an unacceptable risk system, but the draft provides for limited exceptions, which were outlined earlier. However, one can have some criticism of the proposed content of the legislation. As we know, the exceptions concern the

detection, localisation, identification or prosecution of a perpetrator or suspect of a criminal offence referred to the most serious crimes and punishable in the Member State concerned by a custodial sentence or a detention order for a maximum period of at least three years, as determined by the law of that Member State. These crimes include participation in a criminal organisation, terrorism, human trafficking, murder, rape, or child pornography.

This is a fairly broad catalogue of the ways that de facto law enforcement can abuse biometric systems in real time. Note that often the recognition of someone's participation in a criminal organisation or the recognition of the organisation itself as a terrorist organisation is a political decision, not a judicial one. This is, of course, an ethical dilemma, involving an axiological conflict. Is the more important value public security or the right to privacy and protection of personal freedoms? For example, wouldn't it be going too far to use public surveillance, while synchronising it with an automatic biometric identification system, to analyse the faces of all passersby in real time in order to find the image of a person suspected of being part of an organisation that the country considers criminal?

Of course, under the proposed amendments, any single use of a "real-time" remote biometric identification system in a public space for law enforcement purposes will, as a general rule, require prior authorisation from a court or independent administrative authority of a member state. It will be up to member states to decide on the use of such systems – they will be able to either not provide for such a possibility at all, or to provide for it only for some of the permitted purposes indicated above. In the case of Poland and Hungary, we do not know the intentions of legislators as to whether they will decide to implement such systems. If they do, it will be necessary to regulate this issue in criminal procedure laws along the lines of those for operational control and judicial oversight of investigations.

We should also address the issue of statistical evidence, which is based on the Bayesian theorem and is becoming an important element of forensic research. Forensic expertise is evolving towards the use of databases on the frequency of occurrence of some studied characteristic in the population, in order to then determine the level

of strength of the evidence in relation to a particular suspect. As we have already mentioned, the methodology of presenting expert opinion in court is changing around the world. Here the change is a positive one, as it is moving away from *a priori* assumptions about the uniqueness of certain traits to analysing the distribution of traits in the population. Until now, there was a functioning belief that – for example – every person has unique fingerprint traits. This may be the case, but despite more than a century of using such traces for evidence, it continues to have the character of an assumption rather than scientific proof. The use of statistical evidence can reverse this methodological fallacy that operates in the practice of justice.

Nevertheless, for it to be possible to determine the strength of evidence according to modern scientific standards, it is necessary to collect databases on the occurrence of a trait in the population. Data are often obtained as a result of scientific work by laboratories providing services to the judiciary or from academic work. It is important that when statistical evidence is present during a criminal trial, all parties have access to how an expert made calculations on the strength of the evidence. This gives another expert the opportunity to reconstruct the reasoning and make his own calculations, which may either coincide with the original findings, or they might differ and thus help to undermine the evidence. Due to the lack of legal regulations, criminal procedures have to combat the problem of low-quality expert opinions. This provides justification to formulate the conclusion that an Act on experts urgently needs to be drafted, because in spite of numerous attempts over the last thirty years, it has thus far not been possible to introduce such a regulation into Polish law. The regulation should include an obligation for the expert to make available information about the databases used and the methods used to calculate the strength of the evidence at the request of the court, prosecution or defence counsel.

Law enforcement activities using artificial intelligence should be more widely permitted, though not indefinitely. We can't ignore important issues of public safety, as well as matters related to the presentation of evidence in court. In addition, according to the proposals presented, law enforcement activities will be subject to judicial review, so it cannot be said that artificial intelligence will

result in automatic decision-making without any human intervention. The use of artificial intelligence in the administration of justice is somewhat different.

The research shows that artificial intelligence should not be used in courts to predict future criminal behaviour, as today's developing technologies still make mistakes. Moreover, it still raises too many ethical questions. Therefore, we should propose that predictive algorithms should be banned at the level of European Union law, for example, in cases of parole consideration or in deciding on sentencing. However, if the use of such algorithms were to be allowed in the future, the person being analysed should have access to the calculations or source code of the algorithm. The individual in criminal cases must be able to understand the mechanism of decision-making and learn about the factors that were taken into account.

Wider use of artificial intelligence in the judiciary today is a matter of academic consideration, a discussion of what it might look like in the future. On the other hand, from the point of view of today's technological achievements and social maturity, it is difficult to imagine that we can replace judges with algorithms. The use of artificial intelligence in the administration of justice should therefore be limited to the technical and administrative issues outlined in earlier considerations.

REFERENCES

Aitken C., Taroni F., *Statistics and the Evaluation of Evidence for Forensic Scientists*, Chichester 2004.

An EU Artificial Intelligence Act for Fundamental Rights. A Civil Society Statement, Drafted by: European Digital Rights (EDRi), Access Now, Panoptykon Foundation, epicenter.works, AlgorithmWatch, European Disability Forum (EDF), Bits of Freedom, Fair Trials, PICUM, and ANEC (European consumer voice in standardization), 2021.

Bell F., Moses L.B., Legg M., Silove J., Zalnieriute M., *AI Decision-Making and the Courts. A Guide for Judges, Tribunal Members and Court Administrators*, Sydney 2022.

Berk R.A., Bleich J., *Statistical Procedures for Forecasting Criminal Behavior. A Comparative Assessment*, "Criminology & Public Policy" 2013, Vol. 12, No. 3.

Borgman C.L., *The Conundrum of Sharing Research Data*, "Journal of the American Society for Information Science and Technology" 2012, Vol. 63, No. 6.

Bratton W., Morgan J., Malinowski S., *Fighting Crime in the Information Age: The Promise of Predictive Policing*, Los Angeles 2009.

Brennan T., Oliver W.L., *The Emergence of Machine Learning Techniques in Criminology. Implications of Complexity in our Data and in Research Questions*, "Criminology & Public Policy" 2013, Vol. 12, No. 3.

Carliste D., Keatinge T., Keen F., *Virtual Currencies and Terrorist Financing: Assessing the Risks and Evaluating Responses*, Study for the REER Committee, European Parliament, Brussels 2018.

Chan J., *The Future of AI in Policing. Exploring the Sociotechnical Imaginaries*, [in:] J.L.M. McDaniel, K.G. Pease (eds.), *Predictive Policing and Artificial Intelligence*, New York 2021.

Chinnikatti S.K., *Artificial Intelligence in Forensic Science*, "Forensic Science and Addiction Research" 2018, Vol. 2, Issue 5.

Clark R.M., *Intelligence Analysis. A Target-Centric Approach*, Los Angeles 2013.

Committee on the Internal Market and Consumer Protection and Committee on Civil Liberties, Justice and Home Affairs, *Draft Report on the Proposal for a Regulation of the European Parliament and of the Council on Harmful Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legislative Acts (COM2021/0206 – C9-0146/2021 – 2021/0106(COD))*, 2021/0106 (COD).

Dressen J., Farid H., *The Accuracy, Fairness, and Limits of Predicting Recidivism*, "Science Advances" 2018, Vol. 4, Issue 1.

Duncan G.T., Tracey M.L, Stauffer E., *DNA Typing*, [in:] S.H. James, J.J. Nordby, S. Bell (eds.), *Forensic Science: An Introduction to Scientific and Investigative Techniques*, Boca Raton 2014.

Egbert S., Leese M., *Criminal Futures: Predictive Policing and Everyday Police Work*, New York 2021.

EU Commission 'AI ACT' Consultation, Fair Trails' Response, 2021, Feedback Reference: F2665646.

European Network of Forensic Science Institutes, *ENFSI Guideline for Evaluative Reporting in Forensic Science: Strengthening the Evaluation of Forensic Results across Europe (STEOFRAE)*, ENFSI, Wiesbaden 2015.

Furneaux N., *Investigating Cryptocurriences: Understanding, Extracting, and Analyzing Blockchain Evidence*, Indianapolis 2018.

Ho H.L., *A Philosophy of Evidence Law. Justice in the Search for Truth*, Oxford 2008.

https://www.equivant.com/northpointe-risk-need-assessments/.

Ibek A., Necessary J., Wojcik K., *Investigative Analysis*, "State and Law" 2018, No. 6.

Kirkpatrick K., *It's Not the Algorithm, It's the Data*, "Communications of the ACM" 2017, Vol. 60, Issue 2.

Konieczny J., *Kryminalistyczny leksykon śledztwa (Forensic Lexicon of Investigation),* "Opole University Publishing House", Opole 2020.

Lyon D., *Surveillance after September 11*, "Sociological Research Online" 2001, Vol. 6, No. 3.

Mamak K., *Digital Revolution and Criminal Law*, Kraków 2019.

Matkowski W., Frodo K.S.C., Kong A.W.K., *A Study on Wrist Identification for Forensic Investigation,* "Image and Vision Computing" 2019, Vol. 88.

Matkowski W., Matkowski K., Kong A.W.K., Lloyd Hall C., *The Nipple-Areola Complex for Criminal Identification*, "International Conference on Biometrics" 2019.

Meuwly D., *Automated Fingerprint Identification System*, [in:] Jamieson A., Moenssens A., *Wiley Encyclopedia of Forensic Science*, Vol. 1, Chichester 2009.

Mitchell F., *The Use of Artificial Intelligence in Digital Forensics: An Introduction*, "Digital Evidence and Electronic Signature Law Review" 2010, Vol. 7.

Moore T.R., *Trade Secrets & Algorithms as Barriers to Social Justice*, Washington 2017.

National Security Strategy of the Republic of Poland, approved by the order of the President of the Republic of Poland from 12 May 2020 (Monitor Polski, item 413).

Omand D., Miller C., Bartlett J., *Towards the Discipline of Social Media Intelligence*, [in:] C. Hobbs, M. Moran, D. Salisbury (eds.), *Open Source Intelligence in the Twenty-First Century. New Approaches and Opportunities*, London 2014.

Palmiotto M.J., *Criminal Investigation*, Boca Raton 2012.

Parichha P.K., *Introduction to Digital Forensics*, [in:] S. Satpathy, S.N. Mohanty (eds.), *Big Data Analytics and Computing for Digital Forensic Investigations*, Boca Raton 2020.

Policy for the Development of Artificial Intelligence in Poland from 2020, Appendix to the Resolution No. 196 of the Council of Ministers of 28 December 2020 (item 23).

Proposal for a Regulation of the European Parliament and of the Council Laying Down Harmonized Rules on Artificial Intelligence (Artificial Intelligence Act) and Amending Certain Union Legistative Acts, {SEC(2021) 167 final} – {SWD(2021) 84 final} – {SWD(2021) 85 final}.

Ratcliffe J., *Intelligence-Led Policing*, [in:] G. Bruinsma, D. Weisburd (eds.), *Encyclopedia of Criminology and Criminal Justice*, New York 2014.

Regulation (EU) 2016/679 of the European Parliament and the of the Council of April 27, 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation).

Reichman A., Sagy Y., Balaban S., *From a Panacea to a Panopticon: The Use and Misuse of Technology in the Regulation of Judges*, "Hasting Law Journal" 2020, Vol. 71, Issue 3.

Saini M., Kapoor A.K., *Biometrics in Forensic Identification: Applications and Challenges*, "Journal of Forensic Medicine" 2016, Vol. 1, Issue 2.

Scherer M., *Artificial Intelligence and Legal Decision-Making: The Wide Open?*, "Journal of International Arbitration" 2019, Vol. 36, No. 5.

Scurich N., Monahan J., *Evidence-Based Sentencing: Public Openness and Opposition to Using Gender, Age, and Race as Risk Factors for Recidivism*, "*Law and Human Behavior" 2016, Vol. 40, Issue* 1.

Solane A.A., Biasotti M.A., *Digital Forensics AI: Evaluating, Standardizing and Optimizing Digital Evidence Mining Techniques*, Künstl Intelligence, 2022.

Sourdin T., *Judges, Technology and Artificial Intelligence: The Artificial Judge*, Edward Elgar Publishing, Cheltenham 2021.

Surden H., *Artificial Intelligence and Law: An Overview*, "Georgia State University Law Review" 2019, Vol. 35, Issue 4.

Tilley N., *Modern Approaches to Policing: Community, Problem-Oriented and Intelligence-Led Policing*, [in:] T. Newburn (ed.), *Handbook of Policing*, Abingdon 2008.

UK Ministry of Justice, *Transforming Our Justice System: Assisted Digital Strategy, Automatic Online Conviction and Statutory Standard Penalty, and Panel Composition in Tribunals*, Government Response No. Cm 9391, 2017, p. 16.

Van Eijk G., *Socioeconomic Marginality in Sentencing: The Built-In Bias in Risk Assessment Tools and the Reproduction of Social Inequality*, "Punishment & Society" 2016.

Vural M.S., Gök M., *Criminal Prediction Using Naïve Bayes Theory*, "Neural Computing and Applications" 2017, Vol. 28, Issue 9.

Walsh T., Levy N., Bell G., Elliott A., Maclaurin J., Mareels I.M.Y., Wood F.M., *The Effective and Ethical Development of Artificial Intelligence: An Opportunity to Improve Our Wellbeing*, Report for the Australian Council of Learned Academies, Melbourne 2019.

Xu Z., *Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities*, "Applied Artificial Intelligence. An International Journal" 2022, Vol. 36, Issue 1.

Zara G., Farrington D.P., *Criminal Recidivism: Explanation, Prediction and Prevention*, New York 2016.