Access to Justice in Eastern Europe
ISSN 2663-0575 (Print)
ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

Research Article

APPLICATION OF ARTIFICIAL INTELLIGENCE SYSTEMS IN CRIMINAL PROCEDURE: KEY AREAS, BASIC LEGAL PRINCIPLES AND PROBLEMS OF CORRELATION WITH FUNDAMENTAL HUMAN RIGHTS

Oksana Kaplina [1]***, Anush Tumanyants*** [2]***, Iryna Krytska*** [3] ***and Olena Verkhoglyad-Gerasymenko*** [4]

Submitted on 23 Mar 2023 / 1st Revision 11 Apr 2023 / 2nd Revision 5 May 2023
Approved **12 Jun 2023** / Published: **1 Aug 2023**

Summary: 1. Introduction. – 2. The main areas of possible application of AI systems in criminal proceedings. – 3. Basic principles of using AI systems in criminal justice and regulatory documents covering this issue. – 4. The problem of AI systems functioning in the context of fundamental human rights and freedoms. – 5. Conclusions.

Keywords: human rights; criminal proceedings; digital technologies; artificial intelligence; predictive justice, Ukraine.

1 Dr. Sc. (Law), Prof., Head of Department of Criminal Procedure, Yaroslav Mudryi National Law University, Kharkiv, Ukraine
o.v.kaplina@nlu.edu.ua
https://orcid.org/0000-0002-3654-673X
Corresponding author, responsible for writing and research. **Competing interests:** Any competing interests were included here by the authors. **Disclaimer:** Prof. Oksana Kaplina serves as a Member of the Scientific Advisory Board of the Supreme Court, but despite this, she does not represent any views of this body in this particular research, nor is she bound by their views. **Translation:** The content of this article was translated with the participation of third parties under the authors' responsibility. **Managing editor** – Dr. Serhii Kravtsov, Dr. Tetiana Tsuvina. **English Editor** – Julie Bold.

How to cite: O Kaplina, A Tumanyants, I Krytska, O Verkhoglyad-Gerasymenko 'Application of Artificial Intelligence Systems in Criminal Procedure: Key Areas, Basic Legal Principles and Problems of Correlation with Fundamental Human Rights' 2023 3 (20) Access to Justice in Eastern Europe 147-166. https://doi.org/10.33327/AJEE-18-6.3-a000314

2 Cand. of Science of Law (*Equiv. Ph.D.*), Docent, Associate Professor at the Department of Criminal Procedure, Yaroslav Mudryi National Law University, Kharkiv, Ukraine. **Co-author,** responsible for writing and research.
a.r.tumanyants@nlu.edu.ua https://orcid.org/0000-0001-6403-8436

3 Cand. of Science of Law (*Equiv. Ph.D.*), Senior Lecturer at the Department of Criminal Procedure, Yaroslav Mudryi National Law University, Kharkiv, Ukraine. **Co-author,** responsible for writing and research.
i.o.krytska@nlu.edu.ua https://orcid.org/0000-0003-3676-4582

4 Cand. of Science of Law (*Equiv. Ph.D.*), Assistant Professor at the Department of Criminal Procedure, Yaroslav Mudryi National Law University, Kharkiv, Ukraine. **Co-author,** responsible for writing and research.
o.v.verhoglyad-gerasymenko@nlu.edu.ua https://orcid.org/0000-0003-3461-0483

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

ABSTRACT

***Background**: Digital technologies are an important factor currently driving society' development in various areas, affecting not only traditional spheres, such as medicine, manufacturing, and education, but also legal relations, including criminal proceedings. This is not just about using technologies related to videoconferencing, automated distribution, digital evidence, etc. Development is constantly and rapidly moving forward, and we are now facing issues related to the use of artificial intelligence technologies in criminal proceedings. Such changes also entail new threats and challenges – we are referring to the challenges of respecting fundamental human rights and freedoms in the context of technological development. In addition, there is the matter of ensuring the implementation of basic legal principles, such as the presumption of innocence, non-discrimination and the protection of the right to privacy. This concern arises when applying artificial intelligence systems in the criminal justice system.*

***Methods:** The general philosophical framework of this research consisted of axiological and hermeneutic approaches, which allowed us to conduct a value analysis of fundamental human rights and changes in their perception in the context of the AI application, as well as apply in-depth study and interpretation of legal texts. While building up the system of the basic principles of using AI systems in criminal justice, we used the system-structural and logical methods of research. The study also relied on the comparative law method in terms of comparing legal regulation and law enforcement practice in different legal systems. The method of legal modelling was applied to emphasise the main areas of possible application of AI systems in criminal proceedings.*

***Results and Conclusions:** The article identifies the main possible vectors of the use of artificial intelligence systems in criminal proceedings and assesses the feasibility and prospects of their implementation. In addition, it is stated that only using AI systems for auxiliary purposes carries minimal risks of interference with human rights and freedoms. Instead, other areas of AI adoption may significantly infringe rights and freedoms, and therefore the use of AI for such purposes should be fully controlled, verified and only subsidiary, and in certain cases, prohibited altogether.*

1 INTRODUCTION

Digital technologies are gradually, year by year, progressively penetrating various areas of our lives. At the same time, the understanding of digitalisation in society is quite plural. Still, we should agree with its broad definition as 'legal, political, economic, cultural, social and political changes caused by the use of digital tools and technologies' .[5] In the mentioned definition, the key point is to indicate the changes caused by digital technologies when they are introduced into our lives. On the one hand, this can serve as the basis for its qualitative improvement and facilitation, while on the other hand, the disproportionate and unreasonable introduction of digital technologies can yield adverse effects. This includes, for example, unreasonable interference with human rights, the risk of unauthorised access to personal data due to insecurity, discriminatory decisions, and more .

These issues are particularly relevant in the context of the use of artificial intelligence (AI) systems, as evidenced, in particular, by the introduction to the White Paper on Artificial Intelligence: A European Approach to Excellence and Trust (the 'White Paper') published

5 Yulia Razmetaeva, Yurii Barabash and Dmytro Lukianov, 'The Concept of Human Rights in the Digital Era: Changes and Consequences for Judicial Practice' (2022) 5 (3) Access to Justice in Eastern Europe 44, doi: 10.33327/AJEE-18-5.3-a000327.

by the European Commission on 19 November 2020. It points to the rapid development of artificial intelligence and its positive impact on our lives, including improved healthcare through more accurate diagnostics, increased efficiency of production systems through predictive maintenance and ensuring a higher degree of security for Europeans . However, it also underscores the potential risks associated with AI , such as non-transparent decision-making, gender or other types of discrimination, invasion of privacy, or criminal misuse.[6]

The European community's anxiety about the rapid development of AI systems and their 'all-pervasive' nature has been reflected in the Regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain EU legislative acts.[7] Certain provisions of this European document warrant further analysis. Still, it is worth paying particular attention to the explanatory memorandum accompanying the proposal for the adoption of the AI Act which defines the reasons and objectives for the adoption of the relevant regulatory document. The memorandum outlines both the significant economic and social benefits in various industries and social activities, such as improved forecasting, optimisation of operations, and personalisation of services as well as the new risks and negative consequences for society and individuals associated with the introduction of AI systems. With this in mind, it is particularly noted that in light of 'the speed of technological change and the possible challenges of technological change and the possible challenges, the EU is committed to a balanced approach. It is in the Union's interest to maintain the EU's technological leadership and to ensure that Europeans can benefit from new technologies designed and operated in accordance with EU values, fundamental rights and principles'.[8]

An analysis of the prospective areas of introducing artificial intelligence technologies in the field of criminal justice and the associated challenges necessitates a fundamental understanding of the key categories used in the relevant subject area. The concept of artificial intelligence itself is pivotal . It is worth starting its analysis with John McCarthy, one of the influential proponents of AI systems, who considered it as 'the science and engineering of creating intelligent machines' , where 'intelligence' denotes 'a measurable part of the ability to achieve goals in the world' .[9] D. Castro and J. New further define AI as 'a branch of computer science devoted to the creation of computers and systems that perform operations similar to human learning and decision-making'. [10] These above definitions highlight AI primarily as a certain field of science or activity. [11]

However, the general approach to defining AI now focuses on the system's characteristics, the results of its activities, or the scope of its activities . For example, De Spiegeleire, Stephan, Matthijs Maas, and Tim Sweijs refer to AI as 'non-human intelligence measured by its ability to

6 White Paper on Artificial Intelligence: A European Approach to Excellence and Trust COM(2020) 65 final (*European Commission*, 19 February 2020) <https://commission.europa.eu/publications/white-paper-artificial-intelligence-european-approach-excellence-and-trust_en> accessed 18 March 2023.

7 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 (*European Commission*, 21 April 2021) <https://artificialintelligenceact.eu/the-act> accessed 18 March 2023.

8 ibid.

9 John McCarthy, 'What is Artificial Intelligence? Basic Questions' in *English++ : English for Computer Science Students: Complementary Course Book open book* (English++ project, Jagiellonian Language Center Jagiellonian University 2008) 141 <https://englishplusplus.jcj.uj.edu.pl/listenings/what-is-artificial-intelligence/fulltext/index.html> accessed 18 March 2023.

10 Daniel Castro and Joshua New, *The Promise of Artificial Intelligence* (Center for Data Innovation 2016) 36 <https://datainnovation.org/2016/10/the-promise-of-artificial-intelligence> accessed 18 March 2023.

11 Goda Strikaitė-Latušinskaja, 'The Rule of Law and Technology in the Public Sector' (2023) 6 (1) Access to Justice in Eastern Europe 28, doi: 10.33327/AJEE-18-6.1-a00010.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

reproduce human mental skills, such as pattern recognition, natural language understanding, adaptive learning from experience, and developing strategies and rationales for others'.[12]

The official definition of 'AI system' is provided in the aforementioned AI Act, developed by the European Parliament and the European Council of the EU - thus, paragraph 1 of Article 3 of this document defines it as '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' .[13]

Nevertheless, even an analysis of the above definitions and a review of some of the literature on this relevant topic, it becomes apparent that fully encompassing everything within this field with a single term is, as D. Castro and J. New assert, an unattainable goal. In fact, according to researchers, t his attempt to define the entirety of this domain, led to the emergence of, for example, the concepts of 'weak' and 'strong' AI, which, however, did not add to the ease of understanding . In particular, ' weak' artificial intelligence has come to be understood as technologies that are limited to one function and one task,[14] and 'strong' AI, often also called artificial general intelligence, is 'a hypothetical type of AI that can match or exceed human-level intelligence and apply this ability to solve problems and any type of task, just as the human brain can easily learn to drive a car, cook a meal, and write code' .[15]

Some terms that may also be mentioned in the relevant scientific papers within this field or in certain regulatory documents are machine learning and deep learning. Machine learning is proposed to be understood as ' a part of AI that focuses on giving algorithms the ability to learn to perform tasks without explicit instructions, allowing them to adapt in the presence of new data' ,[16] while deep learning is considered 'an advanced type of machine learning that can detect abstract or complex patterns in data using a multi-layer artificial neural network' .[17] D. Castro and J. New explain the use of the term ' deep' by the fact that this type of AI has more ' layers' than simple machine learning approaches, which allows for more complex processing.[18] Specialised literature also distinguishes certain areas of AI functioning or its branches, for example, natural language processing,[19] computer vision,[20] and cognitive computing. [21]

12 Stephan De Spiegeleire, Matthijs Maas and Tim Sweijs, *Artificial Intelligence and the Future of Defense: Strategic Implications for Small- and Medium-Sized Force Providers* (Hague Centre for Strategic Studies 2017) ch 2, 27.

13 Artificial intelligence Act (n 7).

14 Castro and New (n 10) 36.

15 Irving Wladawksy-Berger, " Soft' Artificial Intelligence Is Suddenly Everywhere' *The Wall Street Journal* (16 January 2016).

16 'What Is Machine Learning' (*TechTarget*, updated in March 2021) <http://whatis.techtarget.com/definition/machine-learning> accessed 18 March 2023.

17 Amit Karp, 'Deep Learning Will Be Huge – and Here's Who Will Dominate It' (*Venture Beat*, 2 April 2016) <http://venturebeat.com/2016/04/02/deep-learning-will-be-huge-and-heres-who-will-dominate-it> accessed 18 March 2023.

18 Castro and New (n 10) 37.

19 Matt Kiser, 'Introduction to Natural Language Processing (NLP)' (*Search Medium*, 29 April 2016) <https://medium.com/@mattkiser/an-introduction-to-natural-language-processing-e0e4d7fa2c1d> accessed 18 March 2023.

20 TS Huang, 'Computer Vision: Evolution and Promise' in CE Vandoni (ed), *19th School of Computing: Egmond aan Zee, The Netherlands, 8-21 Septemder 1996* (CERN 1996) 21, doi: 10.5170/CERN-1996-008.21.

21 Bernard Marr, 'What Everyone Should Know About Cognitive Computing' (*Forbes*, 23 March 2016) <http://www.forbes.com/sites/bernardmarr/2016/03/23/what-everyone-should-know-about-cognitive-computing/#a64bc145d6e7> accessed 19 March 2023; John E Kelly III, 'Computing, Cognition, and the Future of Knowing: How Humans and Machines are Forging a New Age of Understanding' (2016) 28 (8) Computing Research News <https://cra.org/crn/2016/09/computing-cognition-future-knowing-humans-machines-forging-new-age-understanding> accessed 19 March 2023.

At this point, having a general vision of what AI is, its functional purpose and what tasks it can perform, we will try to identify the main areas of possible use of AI systems in criminal proceedings. In this part of our work, we plan to demonstrate both existing examples of using such technologies in certain countries and promising areas of AI use. However, in the next part, we will not evaluate the compliance of such vectors of AI use with the principles that can be derived from the relevant regulatory framework - these issues will be addressed in the final part of this article.

2 THE MAIN AREAS OF POSSIBLE APPLICATION OF AI SYSTEMS IN CRIMINAL PROCEEDINGS

Let us try to identify the main possible areas of AI use that are in some way related to criminal proceedings. To facilitate the perception of information and its systematic presentation, we will distinguish certain groups based on their functional focus and the main purpose of AI application.

(1) *Related to the collection and processing of evidence*. We propose to include in this group: a) recognition of images, such as people and objects in video and photo images; b) DNA analysis; c) identification of weapons and other objects.

(2) *Related to the so-called 'predictable' decision-making*. In this aspect, certain areas of such decisions can be highlighted: a) pre-trial release of a person from custody; b) selection of the most appropriate type and measure of punishment, including probation.

(3) Related to the performance of auxiliary tasks arising in criminal proceedings, which may be embodied in the following: a) automatic preparation of forms of certain procedural documents, including summonses, applications, petitions, and complaints ; b) generalisation and systematisation of evidence; c) search of relevant case law; d) forecasting of judicial prospects; e) automated preparation of court transcripts using natural language recognition technologies; f) provision of legal advice using chatbots, for example .

We will now focus on some of the above areas in more detail. As for the recognition of images, such as people and objects , in video and photo images, it should be noted that the specialised literature on this issue primarily refers to crime prevention, i.e. the use of these technologies to identify potential offenders. However, in our opinion, such AI functions may well be used to collect and analyse potential evidentiary information, for example, by comparing images of a person from CCTV cameras from or near the scene of an incident with information from various databases, as well as with open data from social networks. The same may apply to the comparison of not only images of people but also images of particular material objects.

Considering this vector of AI use for criminal justice purposes, Ch. Rigano emphasises that the analysis of video and photo materials using AI creates opportunities for obtaining information about people, objects and actions that can significantly assist in criminal investigations. At the same time, the researcher also acknowledges that these processes are not fully automated at the current stage, leading to certain difficulties connected with significant investments in personnel with relevant knowledge and experience. Ch. Rigano also draws attention to the substantial potential of this area of using AI systems, extending beyond tasks like face matching and object identification. He points out the potential for AI to detect complex events such as accidents and crimes, both in real- time and after they occur , and, most importantly, without the requirement of human intervention at all.[22]

22 Christopher Rigano, 'Using Artificial Intelligence to Address Criminal Justice Needs' (2019) 280 NIJ Journal 38-9 <https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs> accessed 19 March 2023.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

A research report prepared in 2018 by the Australian National University of Cybercrime Surveillance provides some examples of the use of so-called 'computer vision' technologies for pre-trial investigation of crimes. Particularly, the analysis is focused on the 2004 Morris programme , which allows obtaining information from images and related multidimensional data through an automated understanding of photos and videos. Several areas of application for these technologies were identified, namely object recognition, motion detection and evaluation, scene reconstruction and event detection.[23]

However, as the analysis of certain ECtHR decisions shows, such as *Gaughran v. the United Kingdom* (Application no. 45245/15) and *S. and Marper v. the United Kingdom* (Applications nos. 30562/04 and 30566/04), the collection of identification data about a person (in particular, photographs, DNA and fingerprints) after they have been released from criminal liability or served a sentence, might violate Article 8 of the ECHR, provided that the relevant restrictions are disproportionate.[24]

As part of a brief analysis of the second vector of AI systems' use, which is related to assistance in making certain court decisions, i.e. the so-called 'predictive justice' , we will also provide some examples of the current use of these technologies in criminal proceedings. These examples share two characteristics. Firstly, they rely on an approach founded on the analysis of potential risks based on data on a person's previous behaviour, criminal record, living conditions, and more. Secondly, the main areas of application - first of all, for deciding on releasing a person from custody or applying alternative preventive measures to them that are not related to their isolation, as well as rendering final judgments regarding sentencing or assessing the possibility of probation for an individual .

These systems include the following examples that have already been implemented or are currently being developed and tested on the European continent:

- OASys (Offender Assessment System) and OGRS (Offender Group Reconviction Scale) are based on risk assessment and prediction of reoffending. The first is used in pre-sentencing reports to inform sentencing and probation decisions. In the case of a conviction, if the likely sentence for the offence is less than two years, the results of the OASys assessment might influence the choice between imprisonment and community service. The OGRS is used in post-sentence reports to predict the possibility of reoffending (England and Wales);
- RisCanvi is a system aimed at assessing risks at the post-trial stage (Spain);
- Cassandra is a risk assessment programme designed to automate the process of providing pre-trial reports by probation officers and can also be used by the court to assess whether a person should be detained at the trial stage. Among the potential directions of use of this system is the development of standards for the accounting of court decisions in order to 'identify unfair judicial practice' through the analysis of court decisions (Ukraine);
- HART (Harm Assessment Risk Tool) is used to profile suspects and predict their risk of reoffending. The resulting data is used to make an alternative decision on whether to charge the suspect or work with them on an out-of-court settlement

23 Roderic Broadhurst and others, *Artificial Intelligence and Crime: Report of the ANU Cybercrime Observatory for the Korean Institute of Criminology* (KIC, ANU Cybercrime Observatory, College of Asia and the Pacific 2019) 16, 24, doi: 10.2139/ssrn.3407779.

24 *Gaughran v the United Kingdom* App no 45245/15 (ECtHR, 13 February 2020) <https://hudoc.echr.coe.int/eng?i=001-200817> accessed 20 March 2023; *S and Marper v the United Kingdom* App nos 30562/04 and 30566/04 (ECtHR, 4 December 2008) <https://hudoc.echr.coe.int/eng?i=001-90051> accessed 20 March 2023.

under the 'Checkpoint' programme, which, if successful, allows the person not to be prosecuted (England).[25]

The United States is a mong the Western countries using AI tools to predict and assess risks, assisting judges in making decisions . For instance, in certain states, such as Alabama, Virginia and New Jersey, AI is used in criminal proceedings when considering the potential release of a person on bail or other non-isolation measures, as well as during the sentencing stage. The most well-known AI systems used in the US are PSA (Public Security Analyse), COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), and PTRA (Pre-trial Risk Analyse).

It is important to note that systems like PSA is primarily designed to offer judges recommendations regarding pre-trial detention or release. Interestingly, judges often tend to trust AI technologies.[26] However, t he importance of using AI systems as auxiliary tools is also emphasised in the report prepared by Fair Trial, which notes that they cannot completely replace human decision-making, but should only help inform and assist the judge.[27]

The final area we intend to consider in this part of our research covers various segments related to the performance of auxiliary tasks that arise in criminal proceedings. Within this realm, there are several noteworthy examples of AI technology applications, including :

1) Voice and speech recognition, which in criminal proceedings could be used not only for verification of persons or for speech recognition, for example, in a video, but also for automated typing of procedural decisions (Russell & Norvig, TextAloud, Acapela, Amazon Ivona, Grandview Research);[28]

2) Prediction of possible court decisions and assessment of judicial prospects. It is about determining the likelihood of success of a case based on decisions that have already been made and predicting possible options for resolving a dispute (generating representative decisions) by sorting court decisions and analysing relevant statistics;[29]

3) The usage of chatbots based on several AI subfields (natural language processing, machine learning, and big data analysis) to obtain legal advice, select information on a specific topic, and more;[30]

4) Measuring the accuracy and quality of court decisions. [31]

Now we have outlined the main possible areas of AI use in the interests of criminal proceedings and their participants; it is crucial to establish clear regulations that comply with fundamental legal principles, particularly considering the potential risks associated with the use of AI in certain legal contexts. For this reason, in the next part of our work, we will focus on analysing regulatory frameworks, primarily international and regional, in terms of the use of AI systems in criminal proceedings.

25 'Automating Injustice: The Use of Artificial Intelligence & Automated Decision-Making Systems in Criminal Justice in Europe' (*Fair Trial*, 9 September 2021) 24-6 <https://www.fairtrials.org/articles/publications/automating-injustice> accessed 20 March 2023.

26 Benoit Dupont and others, *Artificial Intelligence in the Context of Crime and Criminal Justice: A Report for the Korean Institute of Criminology* (KIC, ICCC 2018) 116-9, 126-7, doi: 10.2139/ssrn.3857367.

27 Automating Injustice (n 25) 24.

28 Broadhurst and others (n 23) 16.

29 Dupont and others (n 26) 116.

30 Broadhurst and others (n 23) 16.

31 Dupont and others (n 26) 133-4.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

3 BASIC PRINCIPLES OF USING AI SYSTEMS IN CRIMINAL JUSTICE AND REGULATORY DOCUMENTS COVERING THIS ISSUE

It is important to note that the execution of justice, primarily criminal justice, is inextricably linked to interference with fundamental human rights and freedoms and making decisions that are often determinative of a person's fate. That is why the use of AI technologies in this area should be clearly regulated, controlled and comply with certain standards, as it carries risks of discrimination, restriction of the right to a fair trial, and the right to privacy, among others . Considering that Ukraine is already on the path of integration into the European Union, we propose to focus on certain regulatory documents that play a key role in regulating the basic principles, approaches and recommendations for the use of AI in the field of criminal justice for European states.

In our opinion, the starting point of our analysis should be Conclusion No. 14 (2011) 'Judiciary and Information Technology' ,[32] prepared by the Consultative Council of European Judges in 2011. This document has not yet mentioned AI but focuses on the benefits of using information technology in court proceedings, particularly on the possible improvement of the process of consideration and movement of court cases, raising the level of court activity in general, simplifying access to information on the progress and results of court proceedings for parties to the proceedings and third parties. We would like to emphasise a crucial aspect that this conclusion , regarding an important fundamental provision concerning the special role of a person in the use of digital technologies in the administration of justice. It emphasised the mandatory participation of a judge in making all decisions on the use and development of IT in the judicial system, as well as the impossibility of replacing the powers of a judge to examine and evaluate evidence with information technology. Otherwise, this would violate the rule of law and the guarantees provided for in Article 6 of the European Convention on Human Rights. It will become evident later in the paper that this principle is a 'red thread' in further European regulations on this aspect. Therefore, although the mentioned Conclusion has not yet directly regulated the issues related to the use of artificial intelligence, it has generally defined the guidelines for the further use of this type of digital technology.

A significant milestone in establishing guidelines for the utilisation of artificial intelligence technologies in the justice system was the adoption of the European Charter on the Ethical Use of AI in Judicial Systems and their Environment in 2018. This charter, issued by the European Commission for the Efficiency of Justice (CEPEJ) under the Council of Europe, holds great importance.[33] This document not only identifies priority areas for the use of AI in judicial systems but also accounts for differences in the scope of its application in decision-making in civil, commercial and administrative cases, on the one hand, and criminal proceedings, on the other. In addition, this charter underscores the paramount importance of upholding the fundamental human rights proclaimed by the European Convention on Human Rights and the Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data. These fundamental rights must be taken into account when employing AI and generally serve as a guide for its use.

Nevertheless, the Charter's special significance lies in the fact that, firstly, it contains definitions of the category of 'artificial intelligence' , as well as concepts basic to the IT sector,

32 Opinion No (2011) 14 Judiciary and Information Technology (IT) (adopted CCJE, 9 November 2011) <https://rm.coe.int/168074816b> accessed 20 March 2023.

33 CEPEJ, *European ethical Charter on the use of Artificial Intelligence in judicial systems and their environment: Adopted at the 31st plenary meeting of the CEPEJ (Strasbourg, 3-4 December 2018)* (Council of Europe 2019) <https://rm.coe.int/ethical-charter-en-for-publication-4-december-2018/16808f699c>accessed 17 March 2023.

in particular: 'data' , 'meta-data' , 'database' , 'algorithm' , ' chatbot' , ' data mining' , 'machine learning' , and more . Secondly, it sets out five fundamental principles for using artificial intelligence technologies in judicial systems and their environment.

These are the following principles: (1) *respect for fundamental human rights*; (2) *non-discrimination* (the essence of this principle is based on the basics of artificial intelligence algorithms that can be used at the development stage and is based on the idea that it is inadmissible to use data that is a priori biased against certain groups of people - by race, age, gender); (3) *quality and security* (refers to the requirements to use certified sources for the machine learning mechanism and to ensure the requirements for a secure environment that apply to the storage and use of the created models and algorithms); (4) *transparency, impartiality and fairness* (this principle is based on approaches that provide for the use of accessible and understandable methods of data processing, as well as external verification); (5) *'under user control'* (the user must have access to review court decisions and data used by AI at any time, as well as the ability to make the necessary adjustments to the decision or cancel it).

Based on the Charter and the European Convention on Human Rights, as well as relevant secondary EU legislation (General Data Protection Regulation (GDPR), Regulation (EU) 2018/172521 (EUDPR), the Electronic Privacy Directive and the Law Enforcement Agencies Directive (LED)) and practice, the High Level Expert Group on AI published Ethical Guidelines for Trustworthy AI on 8 April 2019.[34] The recommendations define seven key requirements that AI systems must satisfy: 1) human involvement and supervision; 2) technical reliability and security; 3) privacy and data management; 4) transparency; 5) non-discrimination and fairness; 6) societal and environmental welfare; and 7) accountability and responsibility.

The next step on the way to the controlled introduction of artificial intelligence in the administration of justice can be recognised as the 'White Paper on Artificial Intelligence. A European Approach to Excellence and Trust' ,[35] published on 19 February 2020. This document is not directly devoted to the problems of using relevant algorithms in criminal proceedings. Still, it can be considered as a certain 'roadmap' of the necessary regulatory adjustments since the White Paper contains proposals and recommendations for possible changes to European legislation that could contribute to the reliable and safe development of artificial intelligence in Europe with full respect for the interests and rights of EU citizens, in particular the right to a fair trial. This European document defines specific areas of improvement of the EU legal framework to address possible risks and situations: a) effective application and enforcement of current EU and national legislation; b) limitation of the scope of current EU legislation; c) change in the functionality of AI systems . At the same time, the focus should be on high-risk applications of artificial intelligence, i.e., those that may most likely interfere with fundamental human rights, particularly the right to privacy and respect for human dignity.

The requirements for them, according to the White Paper, should include certain characteristics: 1) the quality of the initial training data underlying the learning algorithms; 2) controlled monitoring and storage of data and records, including those related to algorithm performance and data processing problems; 3) proactive provision of necessary information to users regarding the use of high-risk AI systems, including to promote responsible use of AI, build trust and facilitate compensation, if necessary; 4) reliability and accuracy of high-risk AI software - such systems need to be developed responsibly and with due consideration

34 High-Level Expert Group on Artificial Intelligence, 'Ethics Guidelines for Trustworthy AI' (*European Commission*, 8 April 2019) <https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai> accessed 17 March 2023.

35 White Paper (n 6).

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

of the risks they might pose (including requirements that ensure reproducibility of results, ensure that AI systems can adequately handle errors or inconsistencies at all stages of the life cycle); 5) mandatory human supervision (AI should not undermine human autonomy, as reliable, ethical and human-centred use of AI can only be achieved by ensuring appropriate human involvement in high-risk AI programs).

Furthermore, on 6 October 2021, the European Parliament adopted a Resolution on AI in criminal law and its use by police and judicial authorities in criminal cases. The Resolution confirms the following key aspects: (a) all AI solutions for law enforcement and the judiciary must fully respect the principles of human dignity, non-discrimination, freedom of movement, the presumption of innocence and the right to defence - otherwise, the use of AI technologies should be prohibited; (b) the security of AI systems, which must be reliable and sustainable to prevent potentially catastrophic consequences of malicious attacks on them; (c) special human control before launching certain critical AI apps; (d) the use of only those AI systems that comply with the principle of confidentiality and data protection (without allowing creeping operation). To implement this r esolution, in December 2021, the European Commission adopted two proposals: first, a Regulation laying down rules for digital connection in judicial communication procedures in civil, commercial and criminal matters, and second, a Directive aligning the existing rules of communication with the rules of the proposed Regulation.[36]

Based on the standards set forth in the international documents analysed above, as well as considering the recommendations offered by various non-governmental organisations in their reports and opinions, including those on improving the relevant legal acts, we aim to formulate our own list of basic principles for the utilisation of AI systems for the purposes and tasks of criminal justice:

- Priority of human rights. The essence of this principle is obvious and follows from the fact that the introduction of any innovations, including AI technologies, should be based on a human-centred approach. New technologies should 'adjust' to the fundamental human rights and freedoms enshrined in numerous international documents over the decades and not change them. Undoubtedly, the vast majority of rules contained in conventions and other documents are interpreted based on the concept of a 'living tree' , i.e. taking into account their dynamic nature and current circumstances. However, this does not mean that adopting AI systems can justify the appropriateness of even a slight derogation from fundamental human rights and freedoms or narrow their scope in any way.

This priority of human rights is highlighted in a 2021 report by Fair Trial, which concludes that in order to properly protect individuals and their fundamental rights, including the right to a fair trial and the presumption of innocence, the use of AI systems in criminal proceedings for prediction, profiling and risk assessment should be prohibited. The authors of the report argue that a significant number of AI systems detect certain correlations between inputs rather than causal relationships between a person's characteristics and their likely behaviour, and therefore, as noted in the report, these types of high- impact, fact-based criminal judgements should not be delegated to automated processes.[37]

- Usefulness. Integrating any new technologies into criminal proceedings, as well as into any other sphere of human life, should always be aimed at bringing the level of relevant activities to a qualitatively new and improved level. Such improvement

36 Directorate-General for Justice and Consumers, 'Digitalisation of Cross-Border Judicial Cooperation' (*European Commission*, 1 December 2021) <https://commission.europa.eu/publications/digitalisation-cross-border-judicial-cooperation_en> accessed 19 March 2023.

37 Automating Injustice (n 25) 32.

should not only meet the requirements of promptness and economy but also benefit those for whom any innovations are introduced - that is, an individual or society as a whole. The point is that the adoption of AI systems, in particular, should be carried out only if and only in those areas where the benefits will be far greater than the possible risks it might cause.

In this context, an example of using the well-known and aforementioned COMPAS system seems appropriate. Specifically, in 2018, J. Dressel and H. Farid tried to assess the accuracy of this system, which led to quite striking conclusions - the system was accurate on average only in 65% of cases. In other words, as the researchers summarised, 'the recidivism predictions made by COMPAS were no more accurate than those made by people with little or no criminal justice experience or by simple statistical analysis based on two characteristics'.[38]

B. Dupont, Y. Stevens and others, in our opinion, quite appropriately expressed the principle under consideration, noting in their study that AI might remain ' an attractive tool that can satisfy the intellectual curiosity of policymakers, as well as computer and data scientists, but this technological intervention and judicial prodding can use unreliable data to harm individuals who come into contact with the criminal justice system, without demonstrating why such algorithms are needed in the first place' . At the same time, as rightly noted, such tools ' must not only be justified but also promote due process (or procedural fairness)'.[39]

- Impartiality and prohibition of discrimination. It seems obvious that one of the goals of integrating AI systems into the judiciary, especially when it comes to so-called 'predictive' justice, is not only to automate processes to speed up and simplify them but also, above all, to reduce the influence of the subjective, 'human' factor. However, many negative examples, which we will mention in the last part of this article, indicate that the use of AI tools sometimes increases the risk of biased, discriminatory treatment of a person. First of all, this is due to the fact that the results of AI systems' analyses are based on pre-entered input data, which is, naturally, constantly updated and modified, but may be biased at the stage of their input - for example, skin colour, area of residence, and social origin.

As noted above, many AI tools operate on the basis of identifying correlations between input data and an individual. In light of this, it is fair to say that ' law enforcement actions or court decisions that are either influenced by racial or ethnic profiling or targeted at less economically advantaged individuals may lead to biased data about certain groups in society'.[40]

The requirement of impartiality and non-discrimination is also emphasised in the Fair Trial recommendations to the EU Commission's AI Act in August 2021. For example, it may be noted that this document 'contains unclear and unspecific requirements for 'bias', none of which prevent discrimination and bias. Artificial intelligence systems used in law enforcement and criminal justice should be subject to mandatory independent bias testing, but the possibility of such testing depends on the availability of criminal justice data, which is sorely lacking in the EU' .41 Fully supporting the view that the prohibition of

38 Julia Dressel and Hany Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4 (1) Science Advances doi:10.1126/sciadv.aao5580.

39 Dupont and others (n 26) 134, 137.

40 Serena Oosterloo and Gerwin van Schie, 'The Politics and Biases of the 'Crime Anticipation Systems' of the Dutch Police' in J Bates and others (eds), *BIAS 2018 : Bias in Information, Algorithms, and Systems: Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems co-located with 13th International Conference on Transforming Digital Worlds (iConference 2018), Sheffield, United Kingdom, 25 March 2018* (Springer 2018) 30 <https://ceur-ws.org/Vol-2103> accessed 21 March 2023.

41 'EU Commission 'AI Act' Consultation: Fair Trials' Response' (*Fair Trials*, 13 August 2021) 2 <https://www.fairtrials.org/articles/publications/eu-commission-ai-act-consultation-fair-trials-response> accessed 21 March 2023.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

discrimination and bias in the use of AI systems should be of primary importance, and the relevant risks should be absolutely minimised by ensuring transparency of the relevant tools and their ongoing testing, we will now analyse the following principles, which are completely correlated to the principle under consideration.

Fully supporting the view that the prohibition of discrimination and bias in the use of AI systems should be of primary importance, and the relevant risks should be absolutely minimised by prioritising transparency in the utilisation of these relevant tools and ensuring their ongoing testing. With this in mind, we will now analyse the following principles, which are intrinsically linked to the principle under consideration.

- Transparency. In order to immediately identify the most ' edge' aspects that necessitate the highlighting of this principle, let us quote the thesis that, for example, 'the American criminal justice system ... is one of the most privatised in the world, with an entire industry developing and selling a wide range of products and services to meet its growing needs' .42 There are no significant differences in other countries of the world, where private entities create various AI tools and can later be used by public authorities. This leads to a certain conflict of interest - on the one hand, the developers' intention to protect their trade secrets, and on the other hand, the objective need for transparency in procedural decision-making and openness of AI systems used in criminal proceedings. However, the vast majority of international documents, as well as authors of numerous reports on these issues, are unanimous in this regard - the priority should be given to the public interest, and therefore the algorithms of AI systems and the results of their work should be open.

Notably, the aforementioned Fair Trial report emphasises the crucial role of transparency in AI systems used within the realm of criminal justice. It highlights the need for system processes to be open source and not subject to legal protections such as trade secrets or intellectual property claims. At the same time, individuals should be notified of all cases where an AI system has been applied and has influenced, or could have influenced, a decision in criminal proceedings against them.43 Thus, the special significance of the principle of transparency of AI systems is emphasised not only in terms of ensuring transparency for users of AI systems but also for individuals who are affected by AI or decisions made with its assistance .44 It is also noted that the system's transparency is a prerequisite for its verification and testing by independent auditors (organisations or individuals), which in the long run, makes it possible to avoid bias.[45]

- Multi-stage verification. Numerous analyses and recommendations on the problems of integrating AI systems into the criminal justice system focus on the need for a thorough testing regime, which is the minimum necessary guarantee to reduce the risk of discrimination and ensure equality before the law. That is, the use of AI systems in real-life situations should take place after lengthy 'trials' , not during them. It is underlined that if the relevant tests have not been carried out and/or if it cannot be proved that the AI system is not discriminatory, ' it should be legally prohibited from being used' .46 In this context, it is appropriate to pay attention to some of the specifics of such a framework - namely, that it should be systematic and multi-stage and be conducted by independent stakeholders and not, for example, by the developers themselves.

42 Dupont and others (n 26) 134.
43 Automating Injustice (n 25) 36.
44 EU Commission 'AI Act' Consultation (n 41) 2
45 Broadhurst and others (n 23) 24.
46 Automating Injustice (n 25) 36.

Thus, the reports propose that the requirement for independent testing of AI systems by an independent body before and after deployment in criminal justice systems should be incorporated into legislation. Otherwise, the operational use of AI tools should be prohibited.[47] Based on the above, the relevant verification should occur at the development stage, as well as before the direct implementation, and continuously after the implementation of AI systems in real-life criminal proceedings.[48]

The significance of such verification might be demonstrated by the following example: regular testing of the initial classifiers on which the AI system is based by measuring the quality of their predictions makes it possible to verify the reliability of the relevant indicators in predicting and assessing risks and to evaluate the suitability of their use. The received outcomes, for example, regarding a significant number of false-positive or false-negative results in this case, will highlight the need to remove the relevant indicator from the so-called 'decision-making tree' and the expediency of reviewing and re-analysing previous decisions based on this indicator.[49]

In summarising the analysis of the multi-stage verification principle, it is evident that it is intricately linked to the other principles we have already described. Particularly, implementing this principle necessitates adherence to the requirement of transparency in AI systems and their operational results. Transparency is crucial not only for independent auditors and also for individuals affected by the use of these AI tools . Moreover, a multi-stage independent audit guarantees non-discrimination and safeguards against bias, as it allows identifying not only inefficient inputs but also discriminatory ones. In addition, the principle we are considering is also pivotal for ensuring compliance with other principles that will be further analysed in this article, such as explainability, controllability, and contestability, as their implementation seems impossible without ensuring permanent open independent verification at various stages of AI systems' operation.

- Subsidiarity and controllability. The essence of this principle can be explained by the fact that AI systems cannot entirely replace human involvement in making relevant procedural decisions. That is, they can be used either as auxiliary tools, for instance, to systematise and structure information, automate the filling of certain forms and their mailing, select relevant legislation and practice, and more. Additionally, they can provide additional information of a reference nature that a human judge can consider, particularly when deciding the possible release of an individual from custody pending trial. At the same time, the predictions and risk assessment made by the AI system should not be binding and should be assessed by the judge on par with other information available in criminal proceedings. At the same time, the use of the results of AI tools should be fully controlled by a human prior to its implementation and not after the fact during the subsequent inspection.

Nevertheless, while such standards are in place in most cases , the risk that law enforcement will over-rely on AI as something 'objective' and 'absolutely free from error' is still too high. In particular, it is noted that ' the mere requirement of having a human decision-maker 'in the know' or authorised to review or verify the automated decision is not sufficient, as it may lead to an overestimation of the ability or willingness of human decision-makers to doubt and overturn automated decisions. The sole requirement that an automated decision be reviewed by a human may turn the review into a mere stamping exercise, which in practice

47 'Briefing Paper on the Communication on Digitalisation of Justice in the European Union' (*Fair Trials*, 12 January 2021) <https://www.fairtrials.org/articles/publications/digitalisation-of-justice-in-the-european-union> accessed 22 March 2023.

48 EU Commission 'AI Act' Consultation (n 41) 10.

49 Broadhurst and others (n 23) 16.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

would not constitute any oversight or control' .50 In this regard, a special role will be played by high-quality and proper training of all participants in criminal proceedings who may be involved in the operation of AI tools, including investigators, inquirers, prosecutors, judges, attorneys, and representatives of probation authorities, which should cover not only the explanation of the operation principles of the relevant systems but also the risks associated with the operation of AI and the key principles of its use in criminal proceedings.

- Explainability and comprehensibility. This principle is an obvious extension of several of the previous ones we have identified, particularly transparency, verifiability, and controllability, since its implementation is possible only if algorithms are open, subject to multi-stage verification, and if AI systems used are under human control. We see the essence of explainability and comprehensibility in the nature of AI algorithms' actions should be logical and predictable, and the results of the relevant tools should consistently follow the source data on which the relevant system is based. At the same time, such results should be of the same type for similar factual circumstances and not situational and random. In relation to accessibility, it should be added that this principle not only extends to specialists in the corresponding field, such as independent experts conducting testing but also, and above all, to law enforcement officers who utilise the results of AI systems in their activities. Additionally, accessibility applies as well as to the persons against whom AI tools have been applied.

In support of our above statement, we would also like to present recommendations from several reports and guidelines formulated in different parts of the world in the context of integrating AI systems into the criminal justice system. Specifically, the importance of making any AI-influenced decisions in the criminal justice system understandable to a layperson is emphasised, i.e., explaining the relevant decisions should not require technical knowledge. It is equally important to notify individuals that they have been the subject of an automated decision made by an AI system, which will create preconditions for a possible appeal of the decision.[51]

In Article 32 of the Toronto Declaration titled 'Protecting the right to equality and non-discrimination in machine learning systems,' several duties are imposed on states to guarantee responsibility and the greatest possible transparency in the use of machine learning systems within the public sector. These duties serve as the preconditions for explaining and understanding the use of these technologies, thereby enabling effective verification and, if necessary, the right to appeal and prosecution. These specific duties include (a) publicly disclosing where machine learning systems are used in the public sphere; (b) providing information that clearly and easily explains how automated decision-making using machine learning is performed; (c) documenting actions taken to identify, record and mitigate discriminatory or other impacts that violate human rights.[52] The above standards appear universal for most national legal systems that face the new challenge of introducing AI systems into the judiciary.

Still, this principle is not always feasible in practice due to the peculiarities of the functioning of certain AI systems, particularly machine learning systems that use neural networks. These systems are built in such a way that it is impossible to decipher the decision-making process

50 Automating Injustice (n 25) 32.

51 Briefing Paper (n 47) 10.

52 'The Toronto Declaration: Protecting the right to equality and non-discrimination in machine learning systems' (*Amnesty International*, 17 May 2018) <https://www.amnesty.org/en/documents/pol30/8447/2018/en> accessed 22 March 2023.

and the results they produce.[53] As some reports on the subject state, 'some machine learning algorithms are simply too complex to be understood with a reasonable degree of accuracy; this is particularly true when AI systems include 'neural networks'. Decision-making processes of this kind are referred to as 'intuitive' because they do not follow a specific logical method, making it impossible to analyse the exact process by which a decision is made'.[54]

There is also an opinion that some AI systems are incomprehensible to humans since 'the machine learning algorithms that support them are able to identify and rely on geometric relationships that humans cannot visualise... that is, it is beyond human capabilities' .[55] In our opinion, the way out of this situation should be that the results of such systems should be exclusively advisory and fully controlled by humans. At the same time, law enforcement officers should be separately informed about the relevant features of the functioning of such AI tools.

- Appealability. It seems generally logical that ensuring transparency, explainability and comprehensibility of the course and results of the operation of AI systems in the criminal justice sector goes beyond simply providing individuals with the knowledge and relevant facts. First and foremost, they should have a real opportunity to review the relevant procedural decision made, particularly using AI systems. In this case, the grounds for appeal should include both the procedural decision itself in terms of its illegality, unmotivated and/or unreasonable nature, as well as the procedure for applying AI tools, the prerequisites for such application, conclusions drawn based on the functioning of AI systems, and the initial data used in processing.

The conclusion of the recommendations on the AI Act also underlines the crucial importance of the guarantee of appealability - in particular, it notes that if AI can have a significant impact on people when it is used in law enforcement and criminal proceedings, it is essential that there are effective ways to review not only AI decisions but also the system itself. Meanwhile, the authors of the recommendations emphasise that the relevant Act, unfortunately, does not simplify the regulation of this issue and does not provide clear ways to appeal or compensate for damage to persons who try to appeal against AI systems or their decisions[56]. Therefore, there is currently a need to improve the regulatory framework in this area.

- Security and respect for privacy. The essence of this initial provision can be revealed in the context of the European Charter for the Ethical Use of AI in Judicial Systems and Their Environments, which, among other things, enshrines a similar principle of quality and security. This document states that the use of AI systems in the processing of court decisions and source data should be based on certified sources and using models developed on the basis of an interdisciplinary approach in a secure technological environment. It is further specified that 'the data entered into the software implementing the machine learning algorithm must come from certified sources and must not be changed until it is actually used by the learning mechanism' . This means that the entire process must be monitored to ensure that no tampering or modifications have been made. In addition, the models and algorithms created must also be able to be stored and executed in a secure environment.

53 EU Commission 'AI Act' Consultation (n 41) 7.

54 ibid 8.

55 Yavar Bathaee, 'The Artificial Intelligence Black Box and the Failure of Intent and Causation' (2018) 31 (2) Harvard Journal of Law & Technology 889.

56 EU Commission ' AI Act' Consultation (n 41) 2.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

A review of some of the existing ethical norms in the field of AI by Benoît Dupont, Yuan Stevens, Hannes Westermann, and Michael Joyce examined five such frameworks: CNIL (French Data Protection Authority) in France (2017), Japanese Society for Artificial Intelligence Ethical Guideline in Japan (2017), the House of Lords – Select Committee on Artificial Intelligence in the UK (2018), Ethically Aligned Design V. 2 by the IEEE on an international scale (2018), and the Montreal Declaration for a Responsible Development of Artificial Intelligence in Canada (2018). Based on this, it is stated that all these documents are united by a basic set of principles, including not only transparency, usefulness for people and society, and accountability, which were reflected in our work earlier but also respect for privacy. In this context, we note the possible emergence of new risks associated with the use of open data by criminals to achieve their illegal goals.[57] Therefore, the public sector, in the case of the adoption of AI systems, should instead guarantee mechanisms to prevent such risks.

As mentioned earlier, the effectiveness of prediction and operation of AI systems in risk assessment directly depends on the amount of data it can receive and process. However, a significant quantity of data on individuals also has the opposite side, namely, the risk of illegal access to it and its use for purposes unrelated to criminal proceedings. Thus, the state's task , if it wants to use AI tools in the field of criminal justice, is to guarantee the security and confidentiality of the relevant data. This can be ensured both through the appropriate digital infrastructure and the use of cloud storage technologies.

4 THE PROBLEM OF AI SYSTEMS FUNCTIONING IN THE CONTEXT OF FUNDAMENTAL HUMAN RIGHTS AND FREEDOMS

The aforesaid analysis has allowed us to establish that the use of AI systems in criminal proceedings may, in one way or another, affect the provision of certain fundamental human and civil rights and freedoms provided for, in particular, by such documents as the ECHR and the ICCPR. First of all, we refer to the prohibition of any form of discrimination enshrined in Article 14 of the ECHR and Article 4 of the ICCPR; the right to a fair trial in terms of guaranteeing the presumption of innocence (Article 6(2) of the ECHR and Article 14(2) of the ICCPR) and the decision of the case by an impartial, unbiased and independent court (Article 6(1) of the ECHR and Article 14(1) of the ICCPR). The use of AI tools can also lead to interference with the right to privacy guaranteed, for example, by Article 8 of the ECHR, but the problems associated with this have already been outlined in the previous part of our work. In this regard, our analysis will focus on the potential challenges that may arise from the juxtaposition of the desire to introduce innovations in criminal proceedings and the need to respect and uphold fundamental human rights.

The first problem that requires thorough consideration by human rights organisations in the context of integrating AI tools into criminal procedure is the impossibility of completely overcoming discriminatory practices in the operation of AI systems. The factors that cause the existence of this phenomenon seem quite obvious - any AI based on machine learning systems operates based on relevant indicators and input data, which to one degree or another, will be potentially discriminatory towards certain groups of people. This negative impact can only be ' overcome' by removing all input data with signs of 'discrimination' . At the same time, such removal will eventually lead to the removal of almost all classification indicators from AI systems, making the relevant AI tool ineffective or unsuitable for operation.

57 Dupont and others (n 26) 143-7.

To illustrate the signs of discrimination in AI systems, we will consider some examples. For instance, the NDAS system uses data obtained during detentions and searches, which in the police activity of England and Wales targets black people almost ten times more than white ones. The Delia, Sensing Project, HART, and RADAR-iTE systems use ethnicity data, and the RisCanvi system includes ethnicity data in its risk assessment. However, even less obvious classification indicators, such as home addresses, telephone codes, and so on, can potentially be discriminatory, as they overlook the fact that many European countries have pronounced ethnic segregation of the population. In practice, this increases the likelihood that AI systems will inadvertently establish a correlation between ethnicity and predicted risk.[58]

Regrettably , we have to acknowledge the widespread belief that certain factors directly influencing prejudice and discrimination against certain groups of people can be difficult to completely eradicate . Nevertheless, there is a glimmer of hope in the efforts of certain companies, such as IBM, who are developing tools to help organisations put the principles of openness and transparency of code, data and variables into practice. The company's AI OpenScale technology, launched in 2018, can automate bias detection and mitigation for a wide range of machine learning systems. By providing explanations for how decisions are made, this technology seeks to instil confidence in their outcomes, according to its authors.[59] Another potential tool that could help enhance the interpretability and reliability of machine learning methods in the future is the Explainable AI programme developed by the US defence research agency DARPA.[60] These new programmes could be particularly beneficial , especially in the field of criminal justice.

The issue at the heart of the functioning of AI systems in the context of the presumption of innocence is that the use of relevant tools is essentially carried out through the so-called 'predictive justice' . This means risk assessment, prediction of possible illegal activity of a suspect or accused in case of their release from custody or, for example, probation, creation of profiles of relevant persons without sufficient factual data and observance of the standards of proof typical for the relevant stages and stages of criminal proceedings. This can potentially lead to prejudicial treatment or prosecution of individuals in violation of the presumption of innocence.

An example that illustrates the potential violation of the presumption of innoncence is the functioning of certain AI systems. In particular, ProKid uses data on the crimes committed by other individuals in close contact with the child, against the child, as well as the child's own victimisation and even the victimisation of other persons in the child's environment as indicators of the likelihood of the child committing offences in the future. Similarly, the NDAS system uses data based on a person's environment to profile them and assess their likelihood of committing a crime in the future. W hile RisCanvi uses information on the criminal history of a person's family or parental criminal history and their 'criminal or antisocial friends' . According to the authors of the relevant report, this constitutes 'criminalisation by association, without any actual evidence or establishment of guilt'.[61]

Thus, on the one hand, it is usually not a matter of actually convicting a person based on the analysis using AI tools. Still, it is noted that relevant predictions and risk assessments can lead to unjustified police surveillance, harassment and arrests, as well as influence decisions on bringing a person to criminal liability, applying non-isolation measures and

58 Automating Injustice (n 25) 29-30.

59 David Kenny, 'How AI OpenScale Overcomes AI's Biggest Challenges' *(IBM, 2023)* <https://newsroom.ibm.com/IBM-watson?item=30695&mhsrc=ibmsearch_a&mhq=AI%20OpenScale>accessed 22 March 2023

60 Matt Turek, 'Explainable Artificial Intelligence (XAI)' (*DARPA*, 2018) <https://www.darpa.mil/program/explainable-artificial-intelligence>accessed 22 March 2023.

61 Automating Injustice (n 25) 30-1.

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

probation.[62] Obviously, this can hardly be considered acceptable in view of the requirement that the issue of guilt or innocence of a person can only be resolved in an impartial and comprehensive trial by a court decision. This makes it crucial to ensure that the principles of subsidiarity and controllability, as well as appealability described earlier, are implemented in law and practically realised. One approach that can enhance transparency and accountability is to subject relevant reports generated by AI systems to open court review. This requires presenting the reports to the prosecution and defence within a reasonable period of time, allowing them to be able to cross-examine them before the judge.

Analysts also note that most regulations and acts that define the sphere and standards of use of AI systems in criminal proceedings do not apply to Europol and other international organisations working in the field of law enforcement and judicial cooperation. At the same time, for example, Europol currently collects and stores a significant amount of sensitive personal data in its databases and information systems.[63]

5 CONCLUSIONS

Thus, this research has identified and analysed the key existing or potential areas of use of AI systems in criminal proceedings, which can be divided into certain groups, namely: (1) related to the collection and processing of evidence, such as recognition of patterns in video and photo images, DNA analysis, identification of weapons and other objects, and so on (2) related to the 'predictive' decision-making; (3) related to the performance of auxiliary tasks arising in criminal proceedings, particularly, automatic preparation of forms of certain procedural documents, generalisation and systematisation of evidence, selection of relevant case law, forecasting of judicial prospects, automated preparation of court transcripts using natural language recognition technologies, and so on .

The following study has shown that only the third group of AI systems vectors pose minimal risks in terms of disproportionate and uncontrolled interference with human rights and freedoms. However, it is important to note that the first two groups may encroach on the rights and freedoms protected , for example, by Articles 6, 8 and 14 of the ECHR. Therefore, the use of AI tools for the relevant purpose in these areas should be fully controlled, verified and only employed as subsidiary measures, and in certain cases - prohibited altogether.

Furthermore, the authors have formulated and characterised the basic principles of using AI systems, in particular: the priority of human rights, usefulness, impartiality and prohibition of discrimination, transparency, multi-stage verification, subsidiarity and controllability, explainability and comprehensibility, appealability, security and respect for privacy. It seems that it is the consideration of these principles in law-making and law enforcement activities in the context of using AI tools in criminal proceedings that will ensure the necessary balance between the benefits of modern technologies and the fundamental rights and freedoms of humans and citizens.

62 European Union Agency for Fundamental Rights, *EU-MIDIS II Second European Union Minorities and Discrimination Survey: Main results* (FRA 2017) <http://fra.europa.eu/en/publication/2017/second-european-union-minorities-and-discrimination-survey-main-results>accessed 22 March 2023.

63 EU Commission 'AI Act' Consultation (n 41) 14.

REFERENCES

1. Bathaee Ya, 'The Artificial Intelligence Black Box and the Failure of Intent and Causation' (2018) 31 (2) Harvard Journal of Law & Technology 889.
2. Broadhurst R and others, *Artificial Intelligence and Crime: Report of the ANU Cybercrime Observatory for the Korean Institute of Criminology* (KIC, ANU Cybercrime Observatory, College of Asia and the Pacific 2019) doi: 10.2139/ssrn.3407779.
3. Burns E, 'Machine Learning' (*TechTarget*, March 2021) <http://whatis.techtarget.com/definition/machine-learning> accessed 18 March 2023.
4. Castro D and New J, *The Promise of Artificial Intelligence* (Center for Data Innovation 2016) <https://datainnovation.org/2016/10/the-promise-of-artificial-intelligence> accessed 18 March 2023.
5. David Kenny, 'How AI OpenScale Overcomes AI's Biggest Challenges' *(IBM, 2023)* <https://newsroom.ibm.com/IBM-watson?item=30695&mhsrc=ibmsearch_a&mhq=AI%20OpenScale>accessed 22 March 2023
6. De Spiegeleire S, Maas M and Sweijs T, *Artificial Intelligence and the Future of Defense: Strategic Implications for Small- and Medium-Sized Force Providers* (Hague Centre for Strategic Studies 2017).
7. Dupont B and others, *Artificial Intelligence in the Context of Crime and Criminal Justice: A Report for the Korean Institute of Criminology* (KIC, ICCC 2018) doi: 10.2139/ssrn.3857367.
8. Huang TS, 'Computer Vision: Evolution and Promise' in CE Vandoni (ed), *19th School of Computing: Egmond aan Zee, The Netherlands, 8-21 Septemb er 1996* (CERN 1996) 21, doi: 10.5170/CERN-1996-008.21.
9. Julia Dressel and Hany Farid, 'The Accuracy, Fairness, and Limits of Predicting Recidivism' (2018) 4 (1) Science Advances doi:10.1126/sciadv.aao5580.
10. Karp A, 'Deep Learning Will Be Huge – and Here's Who Will Dominate It' (*Venture Beat*, 2 April 2016) <http://venturebeat.com/2016/04/02/deep-learning-will-be-huge-and-heres-who-will-dominate-it> accessed 18 March 2023.
11. Kelly III JE, 'Computing, Cognition, and the Future of Knowing: How Humans and Machines are Forging a New Age of Understanding' (2016) 28 (8) Computing Research News <https://cra.org/crn/2016/09/computing-cognition-future-knowing-humans-machines-forging-new-age-understanding> accessed 19 March 2023.
12. Kenny D, 'How AI OpenScale Overcomes AI's Biggest Challenges' *(IBM, 2023)* <https://newsroom.ibm.com/IBM-watson?item=30695&mhsrc=ibmsearch_a&mhq=AI%20OpenScale>accessed 22 March 2023.
13. Kiser M, 'Introduction to Natural Language Processing (NLP)' (*Search Medium*, 29 April 2016) <https://medium.com/@mattkiser/an-introduction-to-natural-language-processing-e0e4d7fa2c1d > accessed 18 March 2023.
14. Marr B, 'What Everyone Should Know About Cognitive Computing' (*Forbes*, 23 March 2016) <http://www.forbes.com/sites/bernardmarr/2016/03/23/what-everyone-should-know-about-cognitive-computing/#a64bc145d6e7> accessed 19 March 2023.
15. McCarthy J, 'What is Artificial Intelligence? Basic Questions' in *English++ : English for Computer Science Students: Complementary Course Book open book* (English++ project, Jagiellonian Language Center Jagiellonian University 2008) 141 <https://englishplusplus.jcj.uj.edu.pl/listenings/what-is-artificial-intelligence/fulltext/index.html> accessed 18 March 2023.
16. Oosterloo S and Schie G van, 'The Politics and Biases of the ' Crime Anticipation Systems' of the Dutch Police' in Bates J and others (eds), *BIAS 2018 : Bias in Information, Algorithms, and Systems: Proceedings of the International Workshop on Bias in Information, Algorithms, and Systems co-*

Access to Justice in Eastern Europe
ISSN 2663-0575 (Print) ISSN 2663-0583 (Online)
Journal homepage http://ajee-journal.com

located with 13th International Conference on Transforming Digital Worlds (iConference 2018), Sheffield, United Kingdom, 25 March 2018 (Springer 2018) 30 <https://ceur-ws.org/Vol-2103> accessed 21 March 2023.

17. Razmetaeva Yu, Barabash Yu and Lukianov D, 'The Concept of Human Rights in the Digital Era: Changes and Consequences for Judicial Practice' (2022) 5 (3) Access to Justice in Eastern Europe 41, doi: 10.33327/AJEE-18-5.3-a000327.
18. Rigano Ch, 'Using Artificial Intelligence to Address Criminal Justice Needs' (2019) 280 NIJ Journal 36 <https://nij.ojp.gov/topics/articles/using-artificial-intelligence-address-criminal-justice-needs> accessed 19 March 2023.
19. Strikaitė-Latušinskaja G, 'The Rule of Law and Technology in the Public Sector' (2023) 6 (1) Access to Justice in Eastern Europe 28, doi: 10.33327/AJEE-18-6.1-a00010.
20. Turek M, 'Explainable Artificial Intelligence (XAI)' (*DARPA*, 2018) <https://www.darpa.mil/program/explainable-artificial-intelligence>accessed 22 March 2023
21. Wladawksy-Berger I, " Soft' Artificial Intelligence Is Suddenly Everywhere' *The Wall Street Journal* (16 January 2016).