*Lea Feuerbach**
*Doris Skaramuca***

DOI: 10.51204/Zbornik_UMKP_25144A
Originalni naučni rad

THE ROLE OF ARTIFICIAL INTELLIGENCE IN PREDICTING RECIDIVISM

Abstract: This paper explores the application of artificial intelligence (AI) in predicting recidivism among offenders, examining both the potential benefits and ethical concerns. We examine various machine learning models used for recidivism prediction. Each approach presents unique advantages and limitations in terms of accuracy, transparency, and real-world application. For instance, while some models may achieve high predictive accuracy, they often lack interpretability, making it difficult for judges and parole officers to fully trust the predictions. Conversely, more interpretable models might compromise some accuracy but offer clearer insights into how predictions are generated. A key focus of the paper is on the legislative frameworks guiding AI use in the criminal justice sphere. We compare the approaches taken in the United States and Europe, noting how differing legal and ethical standards shape the development of AI systems. In the U.S., AI tools have prompted significant debate regarding accountability and discrimination, especially given the history of bias within the system. In contrast, European nations often prioritize data protection and privacy, influencing their methodology for implementing predictive models. We also address the critical issue of bias within AI systems. Historical data used for training these models can perpetuate existing bias and potentially lead to disproportionate predictions for certain demographics. Finally, we discuss the need for interdisciplinary collaboration among technologists, legal experts, and ethicists in developing fair AI applications. This paper advocates for responsible deployment of AI tools in predicting recidivism, ensuring that they enhance, rather than undermine, the justice system.

* Ph.D. Candidate at the Ferenc Deák Doctoral School of the University of Miskolc and Scientific Researcher at the Central European Academy, Budapest, ORCID: ID orcid.org/0009-0004-3763-7889, lea.feuerbach@centraleuropeanacademy.hu.

** Ph.D. Candidate at the Ferenc Deák Doctoral School of the University of Miskolc and Scientific Researcher at the Central European Academy, Budapest, doris.skaramuca@centraleuropeanacademy.hu.

Keywords: recidivism prevention, artificial intelligence, explainable artificial intelligence, COMPAS, bias, transparency.

1. INTRODUCTION

The rapid advancement of Artificial Intelligence (AI) has had a profound impact on various sectors, including the criminal justice system. To fully understand its significance, it is essential to explore the origins, historical development, and definition of AI. The aim of this article is to explore the role of AI in the criminal justice system, with a particular focus on predicting recidivism. To achieve this, the legislative frameworks in the U.S. and Europe will be examined, highlighting the key differences between them. A brief overview of the general use of AI technology within the criminal justice system, particularly in risk assessment tools for predicting recidivism, will be provided. The potential benefits of using AI in risk assessment tools will be considered, while also addressing the key ethical and legal concerns associated with its application. Finally, it will consider future directions for the responsible and effective use of AI in reducing reoffending.

The origins of AI date back to the 1950s, with Alan Turing's work laying the conceptual foundation for thinking machines, however the formal establishment of the field followed in 1956, when John McCarthy introduced the term "Artificial Intelligence" at the Dartmouth Conference, marking the beginning of systematic research into machine-based reasoning and learning.[1] AI has made significant advancements in criminal law, providing new tools for law enforcement and the justice system. It broadly refers to systems that mimic human intelligence, performing tasks like pattern recognition, decision-making, and data analysis.[2] Moreover, includes the capability of a machine to sense and react to its environment on its own, carrying out tasks that would usually require human intelligence and decision-making, all without direct human involvement.[3] Pattern recognition is a key application of AI in criminal justice, where algorithms are used to replicate human abilities in identifying patterns, such as recognising faces, detecting anomalies in data, and predicting future events based on complex datasets.[4]

The historical development of AI can be divided into three distinct stages: rule-based systems, machine learning, and deep learning.[5] Initially, intelligent

1 S. L. Andresen, "John McCarthy: father of AI", *IEEE Intelligent Systems*, 5/2002, 84–85; S. Muggleton, "Alan Turing and the development of Artificial Intelligence", *AI Communications*, 1/2014, 3–10.

2 A. B. Simmons, S. G. Chappell, "Artificial Intelligence – Definition and Practice", *IEEE Journal of Oceanic Engineering*, 2/1988, 14–42; P. Gund, "Investigating Crime a Role of Artificial Intelligence in Criminal Justice Investigating Crime", *The Online Journal of Distance Education and e-Learning*, 2/2023, 1521.

3 C. Rigano, "Using Artificial Intelligence to Address Criminal Justice Needs", *NIJ Journal*, 280/2019, 2.

4 *Ibid.*, 3.

5 B. Dupont *et al.*, *Artificial Intelligence in the Context of Crime and Criminal Justice*, Montreal, 2018, 11.

systems were developed through rule-based approaches, where experts encoded specific knowledge into computers using predefined instructions.[6] These systems operated within clearly defined parameters but lacked flexibility in processing complex or novel data. With the emergence of machine learning, the focus shifted to enabling algorithms to learn from data rather than relying on manually programmed rules.[7] By analysing numerous labelled examples, these systems could identify patterns and make predictions without explicit human instruction.[8] Deep learning, a more advanced subset of machine learning, introduced the use of multilayered neural networks that mimic human cognitive structures.[9] Unlike traditional methods, deep learning algorithms automatically extract and process relevant features from raw data through hierarchical layers, significantly enhancing the system's capacity to handle large, unstructured datasets and complex tasks.[10] This progression reflects a significant shift from rigid programming to adaptive, data-driven intelligence.

While AI offers useful tools that can improve many areas of criminal law, at the same time, with the rise of AI, new types of crime have started to appear. Criminals use technologies like deepfakes, A/B optimisation or algorithmic profiling to commit offences in new ways.[11] This double impact of AI—creating new challenges but also offering helpful solutions—shows the need to study how these technologies affect both criminal behaviour and the justice system.

Despite AI extensive use across various facets of the criminal justice system—by both authorities and offenders—this article specifically examines its application in predicting recidivism. The focus on recidivism prediction is warranted due to its direct implications for judicial decision-making, public safety, and the resocialisation of offenders. Although there is no universally accepted definition of recidivism, and its scope varies depending on the disciplinary approach from which the phenomenon is examined,[12] for the purpose of this article, recidivism is broadly understand as a relapse into a previous pattern of behaviour, which, in the context of crime, refers to criminal conduct.[13] Recidivism

6 *Ibid*., 12.
7 *Ibid*., 13.
8 *Ibidem.*
9 *Ibid*., 15.
10 *Ibidem.*
11 B. H. M. Custers, "AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law", *Law and Artificial Intelligence* (eds. B. H. M. Custers, E. Fosch Villaronga), Springer, Heidelberg, 2022, 205–223.
12 More on the distinctions between criminological, normative, and penological perspectives in: A. M. Getoš Kalac, L. Feuerbach, "On (Measuring) Recidivism, Penal Populism and the Future of Recidivism Research", *Godišnjak Akademije pravnih znanosti Hrvatske*, 1/2023, 1–28.
13 L. Feuerbach, *Kriminološka analiza recidivizma: etiologija, fenomenologija i potencijalna prevencija*, Master thesis, University of Zagreb, Faculty of Law, 2022; L. Feuerbach, "Criminological Insights into Recidivism Trends in Croatia", *Proceedings of the International Scientific Conference "LIFE IN PRISON: Criminological, Penological, Psychological,*

serves as a key indicator of the effectiveness of prison and probation systems in their efforts to resocialise individuals and prepare them for independent living and lawful conduct following release.[14] At the same time, it poses a significant security concern for society. Research has consistently shown that a relatively small group of offenders is responsible for a disproportionately large share of criminal offences.[15] Given the widespread occurrence of recidivism and its significant implications for both the criminal justice system and societal safety, this article will focus on the growing use AI tools in predicting the likelihood of reoffending.

2. LEGISLATIVE FRAMEWORKS AND GENERAL APPLICATIONS OF AI IN CRIMINAL JUSTICE SYSTEM

The use of AI in law enforcement and criminal justice systems varies from country to country.[16] In the U.S. and the EU, both have established legislative frameworks to regulate AI and guide its practical application, though these frameworks are still evolving. While legislation exists in both regions, ongoing developments and updates are required to address the challenges posed by the rapidly advancing technology. These regions significant influence global AI policies and practices, with the U.S. being a leader in technological innovation and the EU known for its strong emphasis on data protection and human rights. This section will examine the role of AI in criminal justice in both regions, focusing on its influence on decision-making processes, data analysis, and the effectiveness of law enforcement practices. It will also explore the differences between the U.S. and the EU, particularly in how they balance the use of AI with the protection of personal data and human rights within the context of criminal justice.

2.1. United States Approach

The United States does not yet have a unified legal framework dedicated solely to AI, but several legislative and executive measures provide guidance on its development and use. The cornerstone is the National Artificial Intelligence Initiative Act of 2020, which coordinates federal efforts to advance AI research

Sociological, Legal, Security, and Medical Issues" (eds. M. Milićević, I. Stevanović, Lj. Ilijić), Institute of Criminological and Sociological Research, Belgrade, 2024, 219–235; A. M. Getoš Kalac, L. Feuerbach, *op. cit*.

14 A. M. Getoš Kalac, R. Bezić, P. Šprem, '"Ružno pače" hrvatskoga kaznenog pravosuđa – zatvorski sustav u svjetlu domaćih i europskih trendova', *Godišnjak Akademije pravnih znanosti Hrvatske*, 1/2021, 83–112.

15 A. R. Piquero, D. P. Farrington, A. Blumstein, "The Criminal Career Paradigm", *Crime and Justice*, 30/2003, 462.

16 M. Matić Bošković, *Implications of EU AI regulation for criminal justice*, Institute of Criminological and Sociological Research, Belgrade, 2024, 111.

and governance.[17] Complementing this, the *CHIPS and Science Act of 2022* allocates substantial funding for AI and related technologies.[18] Presidential directives, such as *Executive Order 13960* from 2020[19] and *Executive Order 14110* from 2023,[20] outline federal principles for trustworthy and secure AI. In addition, frameworks such as the NIST AI Risk Management Framework[21] and the Blueprint for an AI Bill of Rights[22] are non-binding documents that offer guidelines and best practices for the ethical and responsible use of AI. While these measures are not AI-specific criminal laws, they lay the groundwork for regulating AI applications, including in the justice and law enforcement sectors.

The United States is at the forefront of integrating artificial intelligence into criminal justice practices.[23] According to the U.S. Department of Justice, AI is being applied across a broad spectrum of areas to enhance the efficiency and effectiveness of the justice system.[24] The key areas of AI application include identification and surveillance, forensic analysis, predictive policing, and risk assessment. In the field of identification and surveillance, AI technologies are used to recognise faces, fingerprints, and other biometric identifiers, as well as to track licence plates and detect gunshots.[25] In forensic analysis, AI improves both the speed and accuracy of investigations, aiding in DNA comparison, the tracing of

17 United States Congress, National Artificial Intelligence Initiative Act of 2020, H.R. 6216, 116th Congress, *https://www.congress.gov/bill/116th-congress/house-bill/6216/text*, 15 April 2025.

18 United States Congress, CHIPS and Science Act of 2022, Public Law No: 117–167, *https://www.congress.gov/bill/117th-congress/house-bill/4346/text*, 15 April 2025.

19 United States President, Executive Order 13960: Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, 85 FR 78939, 2020, *https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government*, 15. April 2025.

20 United States President, Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 88 FR 75191, 2023, *https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence*, 15. April 2025.

21 National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), AI 100-1, 2023, *https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf*, 15. April 2025.

22 United States Office of Science and Technology Policy, Blueprint for an AI Bill of Rights. The White House, 2022, *https://marketingstorageragrs.blob.core.windows.net/webfiles/Blueprint-for-an-AI-Bill-of-Rights.pdf*, 15. April 2025.

23 T. Sushina, A. Sobenin, "Artificial Intelligence in the Criminal Justice System: Leading Trends and Possibilities", *Proceedings of the 6th International Conference on Social, economic, and academic leadership (ICSEAL-6-2019)*, Atlantis Press, Paris, 2020, p. 433; 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*, 1/2022, 2.

24 U.S. Department of Justice, Final report on Artificial Intelligence and Criminal Justice, 2024, pp. 9–11. *https://www.justice.gov/olp/media/1381796/dl.*

25 *Ibidem.*

seized drugs, and the prioritisation of electronic evidence.[26] Predictive policing is another area where AI plays a significant role, with law enforcement agencies utilising historical data to identify locations where crimes are likely to occur and individuals who may be at higher risk of involvement in criminal activity—either as offenders or victims.[27] The idea of predicting the time and location of crime is not a novel concept. As early as the 19th century, the cartographic (or geographic) school of criminology sought to identify high-risk areas for crime by analysing environmental and social factors.[28] This early approach laid the groundwork for modern predictive policing tools such as the PredPol algorithm, which similarly aims to forecast the times and locations where specific crimes are most likely to occur, enabling targeted police patrols to prevent them, however it does not assess who is likely to commit a crime.[29] Additionally, AI-powered risk assessment tools are employed to estimate the likelihood of specific outcomes within the justice process, such as reoffending or failing to appear in court, thereby informing judicial decisions and resource allocation.[30]

2.2. European Union Approach

The European Union is actively working to establish a legal framework for the use of AI in the judiciary. The Council of Europe took an early step toward addressing the intersection of artificial intelligence and criminal law through its European Committee on Crime Problems (CDPC), which conducted a feasibility study exploring the potential for a dedicated legal instrument in this field.[31] Key initiatives include the European Parliament Resolution 2020/2016 (INI), which addresses AI in criminal law and its application by police and judicial authorities[32], and the Proposal for a Regulation on harmonised rules for AI, aimed at ensuring consistency across EU Member States.[33] Additionally, in 2018, the European Commission for the Efficiency of Justice (CEPEJ) adopted the *Ethical*

26 *Ibidem.*

27 *Ibidem.*

28 E. A. Fattah, *Criminology: Past, Present and Future*, Palgrave Macmillan UK, London, 1997, 208–214; M. Friendly, "The life and works of André-Michel Guerry, revisited", *Sociological Spectrum*, 4–6/2022, 1–34.

29 B. Dupont *et al.*, *op. cit*., 88

30 U.S. Department of Justice, *op. cit.,* 9–11.

31 European Committee on Crime Problems, Feasibility Study on a Future Council of Europe Instrument on Artificial Intelligence and Criminal Law, Strasbourg, 2020. *www.coe.int/cdpc*. 15 April 2025.

32 European Parliament, Resolution 2020/2016 (INI) on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters, Official Journal of the European Union, 2020, *https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html*, 15 April 2025.

33 European Commission, Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending certain

Charter on the Use of Artificial Intelligence in the Judiciary, establishing key principles to guide the responsible and ethical implementation of AI technologies within judicial systems.[34] Furthermore, the High-Level Expert Group on AI presented the *Ethics Guidelines for Trustworthy Artificial Intelligence*, which set out requirements for AI to be lawful, ethical, and robust.[35] These efforts aim to balance innovation with the protection of fundamental rights and the rule of law. While the mentioned EU legal framework directly addresses AI in the judiciary, broader EU legislation indirectly limits its application through the protection of personal data and human rights. For example, the General Data Protection Regulation (GDPR) imposes strict rules on data processing, including the use of AI in profiling and decision-making, ensuring that AI technologies respect individuals' privacy rights.[36] Similarly, the Charter of Fundamental Rights of the European Union[37] guarantees the right to a fair trial and protection against discrimination, which also influences how AI systems are deployed within judicial contexts to safeguard these fundamental freedoms.

AI is widely used in the EU's criminal justice systems. Driven by the exponential growth of digital data and technological advancement, AI offers a suite of tools capable of enhancing the effectiveness, precision, and responsiveness of criminal justice systems.[38] From predictive analytics and digital forensics to biometric identification and generative technologies, AI has introduced new dimensions to crime prevention, investigation, and operational planning, redefining traditional law enforcement paradigms. Key uses include data analytics, where AI systems process large datasets to detect crime patterns, correlate external factors and predict resource needs.[39] Predictive policing employs statistical modelling to forecast where crimes

---

Union legislative acts (COM/2021/206 final), 2021, *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2021%3A206%3AFIN*, 15 April 2025.

34 European Commission for the Efficiency of Justice, European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment, Strasbourg, 2018. *https://www.europarl.europa.eu/cmsdata/196205/COUNCIL%20OF%20EUROPE%20-%20European%20Ethical%20Charter%20on%20the%20use%20of%20AI%20in%20judicial%20systems.pdf*, 15 April 2025.

35 High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy artificial intelligence. European Commission, 2019, https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419, 15 April 2025.

36 European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2016, *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679,* 15 April 2025.

37 European Union, Charter of Fundamental Rights of the European Union, Official Journal of the European Union, C 326/391, 2012, *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A12012P*, 15 April 2025.

38 Europol Innovation Lab, AI and Policing the Benefits and Challenges of Artificial Intelligence for Law Enforcement, Luxemburg, 2023.

39 *Ibid.*, 12–14.

are likely to occur, enabling proactive deployment.[40] They have been implemented in various European jurisdictions, notably in the Netherlands, Germany, Austria, France, Estonia, and Romania.[41] Concurrently, other Member States, including Luxembourg, Portugal, and Spain, are engaged in assessing the feasibility and potential benefits of adopting such technologies within their respective law enforcement frameworks.[42] Furthermore, digital forensics uses AI to recover and analyse digital evidence, while computer vision and biometrics facilitate video surveillance, facial recognition, and behavioural analysis for identification and categorisation.[43] Natural Language Processing (NLP) allows law enforcement to interpret and extract key information from text or audio without compromising data protection.[44] Moreover, generative AI opens new possibilities by creating synthetic content, and AI tools overall contribute to more effective resource allocation and strategic planning in policing.[45] Additionally, in the EU countries, AI-supported risk assessment tools are used to inform judicial decisions and predict the likelihood of recidivism. For example, the United Kingdom uses the Offender Assessment System (OASys) to evaluate reoffending risk and guide rehabilitation strategies;[46] Germany applies tools such as SAPROF to assess both risk and protective factors;[47] and the Netherlands employs RISc to support probation services in structured risk evaluations.[48]

2.3. The EU Artificial Intelligence Act

Artificial Intelligence Act[49] (AI Act) is the first comprehensive law on AI in the world which came into force on August 1, 2024. It is a risk-based regulation directly applicable in the EU Member States. According to the Recital 59 of the AI Act, AI systems used by law enforcement can significantly impact criminal proceedings and positions of the parties in question.[50] If these systems are discriminatory or

40 *Ibid*., 14–17.

41 *European Crime Prevention Network, Artificial intelligence and predictive policing: risks and challenges*, Brussels, 2022, 3.

42 See more details in: *Ibidem.*

43 Europol Innovation Lab, *op. cit*., 20–28.

44 *Ibid*., 17–20.

45 *Ibid*., 28–30.

46 E. Tiarks, "Report on Artificial Intelligence and the Administration of Justice in the United Kingdom Predictive Justice", *e-Revue Internationale de Droit Pénal*, 2023, 1–12.

47 D. Yoon, A. Spehr, P. Briken, "Structured assessment of protective factors: a German pilot study in sex offenders", *Journal of Forensic Psychiatry & Psychology*, 6/2011, 834–844.

48 L. M. van der Knaap *et al*., "Reevaluating Interrater Reliability in Offender Risk Assessment", *Crime & Delinquency*, 1/2012, 147–163.

49 Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act).

50 *Ibid.,* Recital 59, *https://artificialintelligenceact.eu/recital/59/,* 15 April 2025.

inaccurate, it can undermine fundamental rights like presumption of innocence and the right to a fair trial, especially if the AI lacks transparency and explainability.[51] Therefore, certain AI systems intended for law enforcement, particularly those assessing crime risks or evaluating evidence, should be classified as high-risk to maintain public trust and accountability.[52] High-risk AI systems, as outlined in Annex 3 of the EU AI Act, include applications that significantly impact individuals' rights and freedoms, such as AI systems for assessing the risk of a natural person offending or re-offending not solely on the basis of the profiling of natural persons as referred to in Article 3(4) of Directive (EU) 2016/680, or to assess personality traits and characteristics or past criminal behaviour of natural persons or groups.[53] These systems are subject to rigid requirements for risk assessment and mitigation to ensure they operate safely and transparently. Furthermore, developers of high-risk AI systems must implement rigorous governance frameworks to maintain compliance with safety and ethical standards throughout the technology's lifecycle. It's crucial to ensure that AI tools in law enforcement do not create inequality or hinder defendants' rights.

2.4. Where is the difference?

In the United States, AI use in criminal matters is more widespread and varies significantly due to the lack of a unified legislative framework, relying instead on voluntary guidelines, executive orders, and existing laws. Conversely, in the European Union, the deployment of AI in criminal justice is characterised by a more cautious and uniform approach, underpinned by stringent regulatory frameworks. The emphasis on mandatory safeguards, while potentially slowing adoption, serves to mitigate risks such as bias and misuse, ensuring a more rights-respecting implementation of AI technologies. A key difference between the two regions lies in their approach to the protection of confidential data and intellectual property.[54] In the U.S., legal professionals often remain cautious about fully recognising the right to access such data, typically placing greater emphasis on protecting private interests like intellectual property. In contrast, EU countries prioritise the right to information—particularly regarding the logic behind algorithmic decisions—as reflected in the GDPR.

3. RISK ASSESMENT TOOLS FOR PREDICTING RECIDIVISM

As outlined in the introduction, recidivism presents a significant concern in the criminal justice system, impacting both community safety and the effectiveness of resocialisation efforts. Although the applications of AI technologies in criminal

51 *Ibidem*

52 *Ibidem.*

53 *Ibid.,* Annex III, Art. 6. d. *https://artificialintelligenceact.eu/annex/3/,* 15 April 2025.

54 T. Sushina, A. Sobenin, *op. cit.*, 434.

law are broad, this article has a limited scope and will focus on the use of AI-driven risk assessment tools in predicting recidivism. Prior to the development of algorithmic predictive tools, recidivism prediction relied on the judgment and intuition of criminal justice professionals or was based on statistical calculations.[55] However, due to human biases and the need for reform in the criminal justice system, there was a push for more effective methods of assessing recidivism risks.[56]

Risk assessment tools are not a recent development; actuarial instruments have been utilised within the United States justice system since the 1960s.[57] One of the earlier predictive models was the Salient Factor Score, which was employed from early 1970s.[58] However, the tool had a limited scope, as its assessment was based on just seven factors.[59] The Salient Factor Score assessed recidivism risk based on the number of prior convictions, number of prior commitments, age at the time of first offence, nature of the commitment offence, parole or probation history, history of drug dependence, and employment or education status in the two years prior to incarceration.[60] While there are several types of risk assessment tools in practice[61], the focus of this article will be on COMPAS (Correctional Offender Management Profiling for Alternative Sanctions), which is frequently examined in academic research due to its prevalence. This tool gained significant attention after the non-profit organization ProPublica challenged its effectiveness in predicting criminal behaviour and also pointed out its discriminatory impact on black defendants.[62] Since COMPAS development in 1998 by Northpointe, (which later rebranded to "Equivant" in January 2017), has been used to assess over 1 million offenders.[63] The recidivism prediction component, known as the recidivism risk scale, has been an integral part of the tool since 2000.[64] This privately developed algorithm, which judges in certain U.S. federal states are required to use, evaluates 137 factors based on either the defendant's responses or information from criminal records—ranging from household conditions and financial stability to family and criminal history—and assigns a risk score from 1 (low) to 10 (high) as a tool to assist judicial decision-making, without being the sole determinant in

55 *Ibidem.*

56 M. M. Farayola *et al.*, "Ethics and Trustworthiness of AI for Predicting the Risk of Recidivism: A Systematic Literature Review", *Information*, 8/2023, 426.

57 K. Schwerzmann, "Abolish! Against the Use of Risk Assessment Algorithms at Sentencing in the US Criminal Justice System", *Philosophy & Technology*, 4/2021, 1888.

58 P. B. Hoffman, S. Adelberg, "The Salient Factor Score: A Nontechnical Overview", *Federal Probation*, 1/1980, 44.

59 *Ibid.*, 49.

60 *Ibidem.*

61 The most common used are the PSA, the Federal PTRA and COMPAS. For more details see: A. Novokmet, Z. Tomičić, Z. Vinković, *op. cit.*, 29.

62 European Commission for the Efficiency of Justice, *op. cit.*; T. Sushina, A. Sobenin, *op. cit.*, p. 435; European Crime Prevention Network, *op. cit.,* p. 11.

63 J. Dressel, H. Farid, "The accuracy, fairness, and limits of predicting recidivism", *Science Advances*, 1/2018, 1.

64 *Ibidem.*

sentencing.[65] One of the major concerns surrounding COMPAS lies in the fact that Northpointe has not disclosed the internal logic or methodology behind its recidivism prediction algorithm.[66] Lack of transparency concerns were central to *Loomis v. Wisconsin*[67] *case*, where the defendant argued that being sentenced based on a proprietary algorithm—without access to the methodology behind its risk score—violated his right to a fair and informed sentencing process.[68] However, the Wisconsin Supreme Court dismissed the defendant's claims, stating that when used appropriately, COMPAS could support the judge's assessment by complementing other sentencing evidence and contributing to the development of a sentencing plan tailored to the individual.[69] Moreover, COMPAS has received significant criticism due to concerns about its fairness, transparency, and potential bias, particularly following investigations that revealed racial and gender disparities in its predictions.[70] Further criticism emerged from a study indicating that individuals recruited through a popular online crowdsourcing platform—presumably with minimal or no expertise in criminal justice—were just as accurate and fair as COMPAS in predicting recidivism.[71] Given all these concerns, it is not surprising that the use of the COMPAS tool is primarily focused in the United States. While the EU countries has explored the use of risk assessment tools in criminal justice, COMPAS has not been implemented. Instead, some EU countries have developed or are exploring their own risk assessment tools, which are typically tailored to their specific legal systems and ethical standards.[72] The use of such tools in Europe is closely scrutinised in light of the EU's strong data protection laws, such as the GDPR, which raises concerns about privacy and transparency, especially regarding the use of algorithms in judicial decision-making.

4. ARE THERE ANY POTENTIAL BENEFITS OF AI IN RECIDIVISM PREVENTION?

The incorporation of cutting-edge technologies, especially AI, into the rehabilitation and oversight of offenders represents a significant evolution in correctional approaches, as it aims to improve the efficacy of rehabilitation

65 *Ibidem*; European Commission for the Efficiency of Justice, *op. cit*., 52; C. Wang *et al.,* "In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction", *Journal of Quantitative Criminology*, 2/2023, 556; A. Alqatawna, "Utilizing Artificial Intelligence (AI) in Criminal Justice and Policing", *Comparative Law Review*, 30/2024, 24.

66 J. Dressel, H. Farid, *op. cit*.; G. van Dijck, "Predicting Recidivism Risk Meets AI Act", European Journal on Criminal Policy and Research, 3/2022, 409.

67 Loomis v. Wisconsin, 881 N.W.2d 749 (Wis. 2016) 137 S. Ct. 2290 (2017).

68 L. H. Noel, "The Use of Artificial Intelligence in Gauging the Risk of Recidivism", *Judges' Journal*, 58/2019, 39; G. van Dijck, *op. cit*. 409; C. Wang *et al*., *op. cit*., 524–525.

69 van Dijck, *op. cit*., 409.

70 M. M. Farayola *et al*., *op. cit*.

71 J. Dressel, H. Farid, *op. cit*.

72 As mentioned previously in section 2.2.

initiatives while also providing ongoing support for individuals both throughout their time in confinement and as they reintegrate into society after release.[73] To create a more successful transition and reduce recidivism rates, correctional systems can tailor programs to meet the individual needs of each offender, all while monitoring their progress in real-time and offering personalized interventions.[74] AI has the potential to enhance rehabilitation by assessing individual criminogenic factors and developing tailored plans that improve behaviours, skills, and access to education and employment.[75] States like Virginia have seen lower recidivism rates, highlighting the effectiveness of targeted rehabilitative programs in reducing reoffending.[76] We could argue that this innovative approach also contributes to a more humane and effective justice system, where the focus shifts from mere punishment to actual support for individuals seeking a second chance. However, as previously mentioned, this paper focuses on AI risk assessment tools rather than those tools that could influence the improvement of operations within correctional institutions or the rehabilitation of offenders. Therefore, we will consider the following criteria: accuracy, fairness and bias.

4.1. Accuracy

AI tools are capable of searching vast databases and logically connecting them much faster and more systematically than humans. Therefore, they are able to boost efficiency in important decision-making processes. But can we say they are more accurate than human judges when predicting recidivism? According to Hunter, Bagaric and Strobbs, there are three primary methodologies utilized to assess an offender's risk of reoffending.[77] The first method involves unstructured clinical assessments, which rely on subjective judgment and lack empirical validation, rendering them unreliable for algorithm-based systems.[78] The second, known as actuarial assessment, employs empirical data to forecast the likelihood of future offenses through statistical algorithms, though these tools are relatively recent and may be met with caution.[79] The third method, risk and needs assessments, not only estimates the likelihood of

73 D. D. Lee, "AI Detective: Solving Crimes with Artificial Intelligence", *SkyCuration*, 2024.

74 *Ibidem.*

75 A. Patterson, "Can AI really help predict recidivism and help with rehabilitation efforts?", *The Criminal Law Practitioner*, 2024.

76 *Ibidem.*

77 D. Hunter, M. Bagaric, N. Stobbs, "A Framework for the Efficient and Ethical Use of Artificial Intelligence in the Criminal Justice System", *Florida State University Law Review*, 47/2020, 774–778.

78 *Ibidem.*

79 *Ibidem.*

reoffending but also identifies specific interventions that could mitigate that risk.[80] Although these tools are not infallible, studies indicate that well-administered risk and needs assessments can accurately predict recidivism in approximately 70% of cases.[81] Higher accuracy can definitely be considered a benefit, although AI tools are not infallible. Dressel and Farid's 2018 study found that widely used commercial software for predicting recidivism is as accurate as predictions made by individuals with minimal criminal justice knowledge, both reaching about 65% accuracy.[82] But, a closer analysis of the Dressel and Farid study raised concerns about how human predictions were gathered, potentially misrepresenting true human judgment in predicting reoffending. Their study emphasized key risk factors in a way that may have enhanced accuracy artificially. In 2020, a group of researchers in California discovered that algorithms are considerably more precise than humans in predicting which defendants will be arrested for new crimes. In their experiments, they examined how streamlined versus enriched information affects prediction accuracy, finding that enriched details might lead to better statistical predictions than human judgments.[83] Additionally, they explored the role of feedback on accuracy and how base rates of reoffending influence predictions, noting that statistical algorithms consistently outperform humans in adjusting to base rates.[84] Experiments done in 2024 yielded some concerning findings regarding the inclusion of race information in prompts, which appears to have a more significant effect than previously suggested, particularly for the Hispanic group.[85] In contrast to the conclusion by Dressel and Farid, which stated there was insufficient evidence to indicate that race inclusion affects overall accuracy or fairness, this research showed that race information notably influenced human decision-making for the Hispanic group by about 6%.[86] It is important to note that the prediction of recidivism can lead to two types of errors: false positives, predicting an offender will re-offend when they won't, and false negatives, predicting they won't re-offend when they will.[87] Ethically assessing these errors requires considering the criminal justice system's response, as the consequences of mispredictions can vary significantly.[88] Therefore, evaluating the performance of risk

80 *Ibidem.*
81 *Ibidem.*
82 J. Dressel, H. Farid, *op. cit.* 3.
83 Z. Lin, J. Jung, S. Goel, J. Skeem, "The limits of human predictions of recidivism", *Science Advances*, 2020, 5.
84 *Ibidem.*
85 K. Mallari, J. Adebayo, K. Inkpen, M. T. Wells, A. Gordo, S. Tan, "Generative Models, Humans, Predictive Models: Who Is Worse at High-Stakes Decision Making?", *arXiv*, 2024 (last revised February 14, 2025), 9.
86 *Ibidem.*
87 J. Ryberg, "Artificial intelligence at sentencing: when do algorithms perform well enough to replace humans?", *AI Ethics*, 2024, 3–4.
88 *Ibidem.*

assessment algorithms versus human predictions should focus not only on accuracy but also on the implications of their error profiles.[89]

4.2. Fairness and Bias

One of the primary justifications for implementing artificial intelligence is its potential to eliminate human bias, thereby facilitating more equitable decision-making processes. However, it was later concluded that bias and discrimination contribute to unfairness in AI systems, particularly in predicting recidivism, often originating from the training datasets used.[90] Discrimination can be classified into direct discrimination, which involves deliberate unfair treatment based on protected attributes, and indirect discrimination, resulting from decisions influenced by unprotected attributes like zip codes.[91] Discrimination can also be categorized as explainable, where the outputs can be justified, or in-explainable, where reasons for the outcomes are lacking.[92] Bias arises from harmful characteristics in the data or from preconceived notions, leading to decisions that deviate from actual values.[93] The primary sources of bias include data bias, model bias, and evaluation bias, all of which can harm an organization's reputation and erode customer trust.[94] One of the biggest concerns about AI tools in risk assessment is their potential to make biased decisions or even reinforce existing biases. However, humans are prone to bias in decision making too. Many types of cognitive bias exist, and studies support that these biases may influence judges' decisions. In a systematic review of 23 studies across several countries, most focused on identifying biases rather than debiasing techniques, with a majority revealing significant effects of biases on decision-making.[95] Whether AI is used solely as an auxiliary tool to assist the judge in decision-making or completely replaces it, it is essential to emphasize that algorithms can be developed in a way that reduces the presence of bias and achieves fair results. Future efforts should focus on integrating fairness techniques across all development phases, improving data quality, and diversifying datasets sourced from various jurisdictions. In this way, the optimization of AI

89 *Ibidem.*

90 M. M. Farayola, I. Tal, B. Malika, T. Saber, and R. Connolly, "Fairness of AI in Predicting the Risk of Recidivism: Review and Phase Mapping of AI Fairness Techniques" *Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES '23)*, Association for Computing Machinery, 2023, 2–3.

91 *Ibidem.*

92 *Ibidem.*

93 *Ibidem.*

94 *Ibidem.*

95 T. M. S. Neal, P. Lienert, E. Denne, J. P. Singh, "A general model of cognitive bias in human judgment and systematic review specific to forensic mental health", *Law and Human Behavior*, 2/2022, 99–120.

risk assessment tools would be achieved, current potential problems would be circumvented, and the ultimate goal of reducing human bias in decisions that can change human lives would be reached.

5. ETHICAL AND LEGAL RISKS

The Thomson Reuters Institute and the National Centre for State Courts released a guide for legal practitioners outlining AI risks in the legal field.[96] Key concerns include overreliance on AI outputs, the possibility of privacy breaches, and the potential for AI to perpetuate biases.[97] Given the fact we have discussed the latter in the previous chapter, in this chapter we will deal with opaqueness of AI tools instead.

Overreliance on AI occurs when users trust incorrect AI recommendations, often due to a lack of understanding of the AI's capabilities and performance. Although using AI can sometimes lead to better outcomes than working alone, AI systems can still make unpredictable mistakes, which calls for caution. This highlights the need for effective human oversight, as users may struggle to address AI shortcomings when they overly depend on these systems. Users tend to overrely on AI due to automation bias, favouring its recommendations even when its performance fluctuates, which can lead to inconsistent trust levels.[98] Additionally, confirmation bias causes users to trust AI more when its suggestions align with their beliefs, reinforcing their existing assumptions about the system's reliability.[99] Developers must navigate two key regulatory areas: collective self-regulation by professional groups and oversight by larger third-party bodies, like government regulators.[100] An example of effective self-regulation can be seen with OpenAI's transparent approach to GPT-3, where the company engaged experts for public beta testing and welcomed feedback to prioritize ethical considerations.[101] However, despite developers' best efforts, issues may not be apparent until after an application is released, prompting the need for post-release oversight from corporations and regulatory bodies, similar to organizations like the FAA or FDA.[102] While some self-regulation within the AI industry could alleviate the challenges

96 Thomson Reuters Institute/National Center for State Courts AI Policy Consortium for Law and Courts, "Principles and Practices for Using AI Responsibly and Effectively in Courts: A Guide for Court Administrators, Judges, and Legal Professionals", 2025, 6.

97 *Ibidem.*

98 S. Passi, M. Vorvoreanu, "Overreliance on AI: Literature review", *AETHER AI Ethics and Effects in Engineering and Research*, 2022, 6, *https://www.microsoft.com/en-us/research/wp-content/uploads/2022/06/Aether-Overreliance-on-AI-Review-Final-6.21.22.pdf?msockid=2040a001794f606c38a6b50478e66134,* 28 April, 2025.

99 *Ibid.,* 7.

100 K. LaGrandeur, "How safe is our reliance on AI, and should we regulate it?", *AI Ethics* 1/2021, 97–98.

101 *Ibidem*

102 *Ibid*., 98.

of government oversight, third-party regulation remains essential due to the potential misuse of AI technologies.[103]

Privacy and data protection are crucial for stakeholders in criminal justice systems, particularly regarding the use of offenders' personal information in risk assessment tools for predicting recidivism. The implementation of open-source AI models for predicting recidivism risk raises significant privacy concerns, as these models rely on sensitive data that must be ethically managed.[104] Research addressing privacy and data governance in this context is lacking, which calls for clearer accountability and the protection of offenders' rights.[105] Ensuring the security of these systems against breaches is essential to safeguarding both data privacy and the integrity of risk assessments.[106] It is paramount for courts to prioritize the protection of sensitive private information. By adopting encryption methods, ensuring secure storage practices, and setting up strict access controls, they can greatly lower the chances of unauthorized access and data breaches, helping to safeguard privacy and maintain trust and security in the legal process. The GDPR applies to AI systems that handle personal data, mandating organizations to evaluate and address risks linked to data processing, which is particularly important for AI projects.

The use of algorithms in decision-making can be problematic due to a lack of transparency, making it difficult to dispute or appeal decisions impacting people's lives.[107] The reason is that AI algorithms inner workings remain largely opaque, leaving users and even some developers confused about how decisions, predictions, and conclusions are derived. Deep learning networks, in particular, exhibit a "black box problem," where the processes behind decisions are opaque, raising ethical concerns like algorithmic bias.[108] "Closed-box" or "black box" systems refer to algorithms that humans cannot fully understand, even though they can excel at pattern detection and reasoning.[109] Opacity can vary based on the interests and expertise of stakeholders, as critical elements may be unknown to them.[110] Machine learning algorithms can be opaque either because their decision-making mechanisms are inaccessible or because key inputs are not available to programmers or observers.[111] Unlike simpler algorithms that may be obscure due to proprietary reasons, deep learning's complexity makes transparency nearly

103 *Ibidem*.
104 M.M. Farayola *et al*., *op. cit.*, 18.
105 *Loc. cit.*
106 *Ibidem*.
107 W. J. von Eschenbach, "Transparency and the Black Box Problem: Why We Do Not Trust AI.", Springer, *Philosophy and Technology*, 34/2021, 1612–1613.
108 *Ibidem*.
109 A. A. Solanke, "Explainable digital forensics AI: Towards mitigating distrust in AI-based digital forensics analysis using interpretable models", *ScienceDirect, Forensic Science International: Digital Investigation*, 301403, 42/2022, 3.
110 W. J. von Eschenbach, *op. cit*., 1613.
111 *Ibidem*.

unattainable, even for specialists.[112] While these systems can perform effectively, their lack of clear decision-making explanations reduces credibility, particularly in high-stakes fields like law.

6. FUTURE PERSPECTIVE AND EXPECTATIONS

It would be ideal if AI were used to achieve efficiency and speed, but without potential negative effects. There are several ideas that, if implemented, could bring about the best of both worlds.

Steps are being taken to address this issue through the creation of models that clarify the complex and non-linear decision-making processes, making them more comprehensible to humans. This type of model, known as explainable artificial intelligence (XAI), shows potential in tackling the black box problem in AI; however, their current state limits their effectiveness in rendering these processes more transparent to a majority of observers.[113] XAI faces significant challenges in government decision-making, often seen purely as a technical issue when it is inherently complex.[114] The public's lack of expertise and the politicized nature of these decisions can breed distrust, complicating the explanation of algorithms.[115] To address this, XAI should be viewed as a socio-technical challenge, focusing on building trust and making the decision-making process transparent, while strategies should include shifting from merely explaining algorithms to explaining decisions, negotiating algorithms, and emphasizing value-sensitive approaches.[116] The new 2025 study presents the RCN method, which integrates deep learning, clustering techniques, and explainable AI to improve predictions of recidivism and offender profiling and proves development of explainable AI models that are equally or more accurate than the opaque ones is possible and an alternate way to go.[117] By optimizing machine learning models with Keras and employing clustering methods like k-means and t-SNE, the model achieved nearly 75% accuracy and identified 10,661 recidivists, though it struggled with 4,038 false positives and 3,262 false negatives.[118] The inclusion of SHAP values enhanced model interpretability, making AI-driven decisions more transparent for criminal justice stakeholders, while future work should focus on broadening the feature set and exploring

112 *Ibidem.*

113 *Loc. cit.*, 1608.

114 H. de Brujin, M. Warnier, M. Janssen, "The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making", *Government Information Quarterly* 39/2022, 6–7.

115 *Ibidem.*

116 *Ibidem.*

117 M. Cavus *et al.*, "Transparent and bias-resilient AI framework for recidivism prediction using deep learning and clustering techniques in criminal justice", *Applied Soft Computing*, 2025.

118 *Ibidem.*

advanced machine learning techniques to improve predictive accuracy.[119] Therefore, XAI is one of the possible options for developing a transparent AI that we can trust, whose decisions we understand and can scrutinize; however, there are also other methods of control and protection that we can utilize.

In an integrative literature review, Ejjami presents a framework for ethically and efficiently integrating AI into the legal field by proposing roles such as AI Legal Oversight Officer, AI Legal Compliance Officer and AI Legal Quality Assurance Officer, which are essential for maintaining the integrity of legal processes while improving operational efficiency and decision-making accuracy through AI technologies.[120] These, as well as the general raising of awareness among citizens and the work on educating both citizens and professionals, is the path towards a better understanding and protection of human rights and towards a fairer and more efficient judicial system.

7. CONCLUSION

In conclusion, using AI to predict recidivism is a double-edged sword that brings both exciting opportunities and serious ethical dilemmas to the table in our criminal justice system. On one hand, AI has the potential to improve how accurately and efficiently we assess risks, but these benefits cannot overshadow the important ethical considerations and the real effects these technologies can have on people's lives. Different machine learning models each come with their own strengths and weaknesses when it comes to making reliable predictions. One major challenge we face is balancing the desire for accurate predictions with the need for these predictions to be understandable. For those involved in critical decisions—like judges and parole boards—trusting the AI's recommendations while being able to grasp how they were generated is crucial. Adding another layer of complexity are the varying laws and cultural attitudes towards technology in places like the United States and Europe. In the U.S., discussions often revolve around accountability and the urgency to confront the historical biases, particularly regarding how they affect minority groups in the justice process. On the other hand, European countries tend to prioritize rigid legislation and strict data protection laws and privacy rights, which influences how they develop and apply these predictive technologies. To navigate these challenges, it's vital for policymakers to step up and create clear guidelines for the ethical use of AI in predicting recidivism. By prioritizing transparency, explainability, and accountability in AI systems, we can help build public confidence in these technologies. The goal should be to ensure that AI supports and safeguards the rights of all individuals involved in the justice system. Ultimately, the responsible deployment of AI in

119 *Ibidem.*

120 R. Ejjami, "AI-Driven Justice: Evaluating the Impact of Artificial Intelligence on Legal Systems", *International Journal for Multidisciplinary Research (IJFMR)*, 3/2024, 24.

predicting recidivism requires a concerted effort to address ethical concerns, incorporate diverse perspectives, and challenge existing biases.

BIBLIOGRAPHY

Alqatawna A., "Utilizing Artificial Intelligence (AI) in Criminal Justice and Policing", *Comparative Law Review*, 30/2024.

Andresen S. L., "John McCarthy: father of AI", *IEEE Intelligent Systems*, 5/2002.

Cavus M., Benli M.N., Altuntas U. *et al*., "Transparent and bias-resilient AI framework for recidivism prediction using deep learning and clustering techniques in criminal justice", *Applied Soft Computing*, 2025.

Custers B. H. M., "AI in Criminal Law: An Overview of AI Applications in Substantive and Procedural Criminal Law", *Law and Artificial Intelligence* (eds. B. H. M. Custers, E. Fosch Villaronga), Springer, Heidelberg, 2022.

de Brujin H., Warnier, M., Janssen, M., "The perils and pitfalls of explainable AI: Strategies for explaining algorithmic decision-making", *Government Information Quarterly* 39/2022.

Dressel J., Farid, H., "The accuracy, fairness, and limits of predicting recidivism", *Science Advances*, 1/2018.

Dupont B. *et al.*, *Artificial Intelligence in the Context of Crime and Criminal Justice*. Montreal, 2018.

Ejjami, R., „AI-Driven Justice: Evaluating the Impact of Artificial Intelligence on Legal Systems", *International Journal for Multidisciplinary Research (IJFMR)*, 3/2024.

European Committee on Crime Problems, *Feasibility Study on a Future Council of Europe Instrument on Artificial Intelligence and Criminal Law*, Strasbourg, 2020. www.coe.int/cdpc.

European Crime Prevention Network, *Artificial intelligence and predictive policing: risks and challenges*, Brussels, 2022.

Europol Innovation Lab, *AI and Policing the Benefits and Challenges of Artificial Intelligence for Law Enforcement*, Luxemburg, 2023. *https://doi.org/10.2813/0321023*.

Farayola M. M. *et al.*, "Ethics and Trustworthiness of AI for Predicting the Risk of Recidivism: A Systematic Literature Review", *Information*, 8/2023.

Farayola M. M. *et al.*, "Fairness of AI in Predicting the Risk of Recidivism: Review and Phase Mapping of AI Fairness Techniques", *Proceedings of the 18th International Conference on Availability, Reliability and Security (ARES '23)*, Association for Computing Machinery, 2023.

Fattah E. A., *Criminology: Past, Present and Future*, Palgrave Macmillan UK, London, 1997.

Feuerbach L. *Kriminološka analiza recidivizma: etiologija, fenomenologija i potencijalna prevencija*, Master thesis, University of Zagreb, Faculty of Law, 2022, *https://repozitorij.pravo.unizg.hr/en/islandora/object/pravo:4484*.

Feuerbach L., "Criminological Insights into Recidivism Trends in Croatia", *Proceedings of the International Scientific Conference "LIFE IN PRISON: Criminological, Penological, Psychological, Sociological, Legal, Security, and Medical Issues"* (eds. Milićević M., Stevanović I., Ilijić Lj.), Institute of Criminological and Sociological Research, Belgrade, 2024.

Friendly M., "The life and works of André-Michel Guerry, revisited", *Sociological Spectrum*, 4–6/2022.

Getoš Kalac A. M., Bezić R., Šprem P., "'Ružno pače" hrvatskoga kaznenog pravosuđa – zatvorski sustav u svjetlu domaćih i europskih trendova', *Godišnjak Akademije pravnih znanosti Hrvatske*, 1/2021.

Getoš Kalac A. M., Feuerbach, L., "On (Measuring) Recidivism, Penal Populism and the Future of Recidivism Research", *Godišnjak Akademije pravnih znanosti Hrvatske*, 1/2023.

Gund P., "Investigating Crime a Role of Artificial Intelligence in Criminal Justice Investigating Crime", *The Online Journal of Distance Education and e-Learning*, 2/2023.

Hoffman P. B., Adelberg S., "The Salient Factor Score: A Nontechnical Overview", *Federal Probation*, 1/1980.

Hunter D., Bagaric M., Stobbs N., "A Framework for the Efficient and Ethical Use of Artificial Intelligence in the Criminal Justice System", *Florida State University Law Review*, 47/2020.

LaGrandeur K., "How safe is our reliance on AI, and should we regulate it?", *AI Ethics* 1/2021.

Lee D. D., "AI Detective: Solving Crimes with Artificial Intelligence", *SkyCuration*, 2024.

Lin Z., Jung J., Goel S., Skeem J. "The limits of human predictions of recidivism", *Science Advances*, 2020.

Mallari K., Adebayo J., Inkpen K., Wells M. T., Gordo A., Tan S., "Generative Models, Humans, Predictive Models: Who Is Worse at High-Stakes Decision Making?", *arXiv*, 2024 (last revised February 14, 2025).

Matić Bošković M., Implications of EU AI regulation for criminal justice, *Institute of Criminological and Sociological Research*, Belgrade, 2024.

Muggleton S., "Alan Turing and the development of Artificial Intelligence", *AI Communications*, 1/2014.

Neal T. M. S., Lienert P., Denne, E., Singh J. P., "A general model of cognitive bias in human judgment and systematic review specific to forensic mental health", *Law and Human Behavior*, 2/2022.

Noel L. H., "The Use of Artificial Intelligence in Gauging the Risk of Recidivism", *Judges' Journal*, 1/2019.

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*, 1/2022.

Passi, S., Vorvoreanu, M., "Overreliance on AI: Literature review", *AETHER AI Ethics and Effects in Engineering and Research*, 2022, *https://www.microsoft.com/en-us/research/wp-content/uploads/2022/06/Aether-Overreliance-on-AI-Review-Final-6.21.22.pdf?msockid=2040a001794f606c38a6b50478e66134,* 28 April 2025.

Patterson A., „Can AI really help predict recidivism and help with rehabilitation efforts?", *The Criminal Law Practitioner*, 2024.

Piquero A. R., Farrington D. P., Blumstein A., "The Criminal Career Paradigm", *Crime and Justice*, 30/2003.

Rigano C., "Using Artificial Intelligence to Address Criminal Justice Needs", *NIJ Journal*, 280/2019.

Ryberg J., "Artificial intelligence at sentencing: when do algorithms perform well enough to replace humans?", *AI Ethics*, 2024.

Schwerzmann K., "Abolish! Against the Use of Risk Assessment Algorithms at Sentencing in the US Criminal Justice System", *Philosophy & Technology*, 4/2021.

Simmons A. B., Chappell S. G. "Artificial Intelligence – Definition and Practice", *IEEE Journal of Oceanic Engineering*, 2/1988.

Solanke A.A., "Explainable digital forensics AI: Towards mitigating distrust in AI-based digital forensics analysis using interpretable models", *ScienceDirect, Forensic Science International: Digital Investigation* 42/2022.

Sushina T., Sobenin A., "Artificial Intelligence in the Criminal Justice System: Leading Trends and Possibilities", *Proceedings of the 6th International Conference on Social, economic, and academic leadership (ICSEAL-6-2019)*, Atlantis Press, Paris, 2020.

Thomson Reuters Institute/National Center for State Courts AI Policy Consortium for Law and Courts, *Principles and Practices for Using AI Responsibly and Effectively in Courts: A Guide for Court Administrators, Judges, and Legal Professionals*, 2025.

Tiarks E., "Report on Artificial Intelligence and the Administration of Justice in the United Kingdom Predictive Justice", *e-Revue Internationale de Droit Pénal*, 2023.

van der Knaap L. M. *et al.*, "Reevaluating Interrater Reliability in Offender Risk Assessment", *Crime & Delinquency*, 1/2012.

van Dijck G., "Predicting Recidivism Risk Meets AI Act", *European Journal on Criminal Policy and Research*, 3/2022.

von Eschenbach, W.J., "Transparency and the Black Box Problem: Why We Do Not Trust AI.", Springer, Philosophy and Technology 34/2021.

Wang C. *et al.*, "In Pursuit of Interpretable, Fair and Accurate Machine Learning for Criminal Recidivism Prediction", *Journal of Quantitative Criminology*, 2/2023.

Yoon D., Spehr A., Briken P., "Structured assessment of protective factors: a German pilot study in sex offenders", *Journal of Forensic Psychiatry & Psychology*, 6/2011.

OTHER SOURCES

European Commission for the Efficiency of Justice, *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment*, Strasbourg, 2018. *https://www.europarl.europa.eu/cmsdata/196205/COUNCIL%20OF%20EUROPE%20-%20European%20Ethical%20Charter%20on%20the%20use%20of%20AI%20in%20judicial%20systems.pdf.*

European Commission, Proposal for a regulation of the European Parliament and of the Council laying down harmonised rules on artificial intelligence and amending certain Union legislative acts (COM/2021/206 final), 2021, *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=COM%3A2021%3A206%3AFIN,* 15. April 2025.

European Parliament, Resolution 2020/2016 (INI) on artificial intelligence in criminal law and its use by the police and judicial authorities in criminal matters, Official Journal of the European Union, 2020, *https://www.europarl.europa.eu/doceo/document/TA-9-2021-0405_EN.html*, 15. April 2025.

European Union, Charter of Fundamental Rights of the European Union, Official Journal of the European Union, C 326/391, 2012, https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A12012P, 15. April 2025.

European Union, Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), 2016, *https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=CELEX%3A32016R0679*, 15. April 2025.

European Union, Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act)

High-Level Expert Group on Artificial Intelligence, Ethics guidelines for trustworthy artificial intelligence. European Commission, 2019, *https://ec.europa.eu/newsroom/dae/document.cfm?doc_id=60419*, 15 April 2025.

Loomis v. Wisconsin, 881 N.W.2d 749 (Wis. 2016) 137 S. Ct. 2290 (2017).

National Institute of Standards and Technology, Artificial Intelligence Risk Management Framework (AI RMF 1.0), AI 100-1, 2023, https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf, 15. April 2025.

U.S. Department of Justice, *Final report on Artificial Intelligence and Criminal Justice*, 2024. *https://www.justice.gov/olp/media/1381796/dl.*

United States Congress, CHIPS and Science Act of 2022, Public Law No: 117–167, *https://www.congress.gov/bill/117th-congress/house-bill/4346/text*, 15 April 2025.

United States Congress, National Artificial Intelligence Initiative Act of 2020, H.R. 6216, 116th Congress*, https://www.congress.gov/bill/116th-congress/house-bill/6216/text*, 15. April 2025.

United States Office of Science and Technology Policy, Blueprint for an AI Bill of Rights. The White House, 2022, *https://marketingstorageragrs.blob.core.windows.net/webfiles/Blueprint-for-an-AI-Bill-of-Rights.pdf*, 15. April 2025.

United States President, Executive Order 13960: Promoting the Use of Trustworthy Artificial Intelligence in the Federal Government, 85 FR 78939, 2020, *https://www.federalregister.gov/documents/2020/12/08/2020-27065/promoting-the-use-of-trustworthy-artificial-intelligence-in-the-federal-government*, 15. April 2025.

United States President, Executive Order 14110: Safe, Secure, and Trustworthy Development and Use of Artificial Intelligence, 88 FR 75191, 2023, https://www.federalregister.gov/documents/2023/11/01/2023-24283/safe-secure-and-trustworthy-development-and-use-of-artificial-intelligence, 15. April 2025.

Lea Feuerbach*
Doris Skaramuca**

ULOGA UMJETNE INTELIGENCIJE U PREDVIĐANJU RECIDIVIZMA

Rezime

Rad istražuje primjenu umjetne inteligencije (UI) u predviđanju recidivizma među počiniteljima kaznenih djela, uzimajući u obzir potencijalne prednosti, ali i etičke izazove. Analiziraju se različiti modeli strojnog učenja koji se koriste za predviđanje recidivizma. Svaki pristup donosi specifične prednosti i nedostatke u pogledu točnosti, transparentnosti i primjene u praksi. Primjerice, dok neki modeli postižu visoku razinu prediktivne točnosti, često su nedovoljno interpretabilni, što otežava sucima i članovima uvjetnih odbora da u potpunosti vjeruju tim predikcijama. S druge strane, modeli koji nude veću interpretabilnost ponekad kompromitiraju razinu točnosti, ali omogućuju jasniji uvid u način na koji su predviđanja generirana. Posebna pažnja posvećena je zakonodavnim okvirima koji uređuju upotrebu umjetne inteligencije u kaznenopravnom sustavu. Uspoređuju se pristupi Sjedinjenih Američkih Država i europskih zemalja, pri čemu se naglašava kako različiti pravni i etički standardi oblikuju razvoj sustava. U SAD-u su alati umjetne inteligencije izazvali

* Doktorandica Pravnog fakulteta Sveučilišta Miškolcu, istraživačica na Central European Academy, Városmajor utca 12-14, 1122 Budimpešta, Mađarska; leafeuerbach07@gmail.com.

** Doktorandica Pravnog fakulteta Sveučilišta Miškolcu, istraživačica na Central European Academy, Városmajor utca 12-14, 1122 Budimpešta, Mađarska; doris.skaramuca@centraleuropeanacademy.hu.

značajne rasprave o odgovornosti i diskriminaciji, osobito s obzirom na povijest pristranosti unutar sustava. Nasuprot tome, europske zemlje često daju prednost zaštiti osobnih podataka i privatnosti, što utječe na njihove metode implementacije prediktivnih modela. Rad se također bavi ključnim pitanjem pristranosti unutar sustava umjetne inteligencije. Povijesni podaci koji se koriste za treniranje ovih modela mogu zadržati postojeće obrasce pristranosti i potencijalno rezultirati neproporcionalnim predikcijama za određene demografske skupine. Zaključno, ističe se potreba za interdisciplinarnom suradnjom između tehnologa, pravnih stručnjaka i etičara u razvoju pravednih aplikacija umjetne inteligencije. Ovaj rad zagovara odgovornu upotrebu alata umjetne inteligencije u svrhu predviđanja recidivizma, kako bi se unaprijedio, a ne ugrozio, kaznenopravni sustav.

Ključne riječi: prevencija recidivizma, umjetna inteligencija, objašnjiva umjetna inteligencija, COMPAS, pristranost, transparentnost.