UDC 343.13
DOI https://doi.org/10.32782/forensic.science.2025.1.17

Hanna Kuczyńska
Professor at the Polish Academy of Sciences, Institute of Law Studies
https://orcid.org/0000-0002-1446-2244

THE ANALYSIS AND EVALUATION OF EVIDENCE BY ARTIFICIAL INTELLIGENCE IN A CRIMINAL TRIAL

In the increasingly digital environment of gathering and storing information about core crimes, enforcement authorities decided to use algorithms in order to more effectively analyze and manage evidence. These algorithms were designed as tools to be used in order to automatize evidentiary proceedings at all stages of collecting, storing, securing, and analyzing evidence. Enforcement authorities are functioning in an environment where hundreds of thousands of pieces of digital evidence and footages of core crimes are downloaded by potential witnesses. Alongside the social media there are other digital sources of evidence: such as digital audio and video-recordings, CCTV footage, aerial and satellite imagery, drone footage etc. Considering the drastic increases in the volume and velocity of data in the context of criminal investigations, AI has become indispensable in supporting the work of investigators. Not only algorithms became a part of digital forensics but also the use of AI have become central in modern digital investigations. However, investigators and judges must be aware of the problems that stem from the risks typical for the use of algorithms in data-analysis. Also, in consequence of such a rapid development of AI-derived and managed evidence, there is a need to assess admissibility of such evidence in a criminal trial.

***Keywords**: International Criminal Court, Artificial Intelligence, criminal trial, verification of evidence, admissibility of evidence.*

The importance of AI-derived and managed evidence

In the increasingly digital environment of gathering and storing information about core crimes, enforcement authorities decided to use algorithms in order to more effectively analyze and manage evidence. These algorithms were designed as tools to be used in order to automatize evidentiary proceedings at all stages of collecting, storing, securing, and analyzing evidence. Prosecutors are functioning in an environment where today hundreds of thousands of pieces of digital evidence and footages of core crimes are downloaded by potential witnesses [4, p. 283–336; 7, p. 102, 108–109). Alongside the social media there are other digital sources of evidence: such as digital audio and video-recordings, CCTV footage, aerial and satellite imagery, drone footage etc. Considering the drastic increases in the volume and velocity of data in the context of criminal investigations, AI has become indispensable in supporting the work of investigators [19]. Not only algorithms became a part of digital forensics but also the use of AI have become central in modern digital investigations. Whereas an algorithm defines the process through which a decision is made, AI uses training data to make such a decision - as AI models used in forensics have also ability to learn and adapt based on data, including historical and other criminal data that are available to law enforcement.

Algorithms are used in many purposes. First, they can extract specific data from an increasingly large number of data sources. They can even shift through social media and open sources. A rising number of law enforcement agencies are adopting automated OSINT tools for investigative purposes in order to investigate and reconstruct online criminal footprints. Secondly, they can classify information and manage vast amounts of data. AI can manage both structured and unstructured data. AI tools are trained to categorize

images based on the content or objects they detect. AI provides an advanced capability to sift through vast data bases, automating processes that would traditionally take human experts large amounts of time. While a human investigator must manually sort through thousands of files, AI can rapidly categorize, filter, and highlight relevant information based on predefined criteria or patterns (e.g. using image classification). An important part of such analysis relates to biometrics and face recognition: both in the area of identification of an unknown person and targeted searches of known persons. As a consequence of a surge of digital imagery from sources like CCTV cameras to personal devices, it is essential to use this vast visual data effectively. AI can also identify correlations between events and data, images and objects, identify correlations between different data types.

International experiences in AI-driven investigations

Presently, when it comes to international experiences in conducting investigations with the use of AI, two institutions are at the forefront. The first is the Office of the Prosecutor of the ICC (OTP ICC). The OTP 2023 Annual Report announced that: "This digital transformation is huge for us – it's like stepping into the future where our tools are smarter and our skills are always up to date" [18, p. 52]. This 'digital transformation' applied by the OTP consists of a digital tool that allows to use algorithms to analyze and manage data (Project Harmony). The system – as it is planned and currently described – is supposed to harness the advanced technology and artificial intelligence in the pursuit of justice. The algorithms used by the OTP allow to "handle larger information volumes utilizing Artificial Intelligence (AI) and Machine Learning (ML), significantly reducing the time required to review and act on it" [18, p. 52].

This enhancement relates to three areas of evidence-analysis and management. The first development is the technical analysis of data, including comparing biometric traits, e.g. facial identification, vocal recognition; image enrichment, multimedia file translations, automatic transcription (and transliteration); and video and image analysis (rapid pattern identification). While none of these innovations are new in themselves, when combined in this way they may prove invaluable to the OTP's effectiveness in collecting, storing, securing, analyzing, and reviewing evidence. For example, facial identification tools can help investigators obtain potential forensic versions by allowing them to more quickly compare multiple images that may show the same person [12, 15]. The second improvement is the management of data stored in the database. The algorithm is said to be able to easily filter out irrelevant information, allowing to focus on the most credible and relevant information. The third improvement is a search-engine, allowing for targeted searches of source materials. Such a tool should also be able to cross-match the results of analyses. These are automatic innovative algorithms (therefore assumed to be AI), which can contribute in the search for evidence, their analysis, and management.

The second institution conducting core crime investigations (but not only such) is Europol, who announced in public reports how the algorithms are being used in the process of evaluation and analysis of evidence for the purposes of a criminal trial [14]. Within Europol, analysts are supported by the data science team use a set of AI models to classify images by automatically assigning tags to millions of pictures or to extract named entities from text, including the names of people, locations, phone numbers, or bank accounts. AI models also allow analysts to search for images of specific objects.

However, in both cases, the algorithms do not take the place of humans as the entity taking a decision. They do not fully control the procedure, but provide data and results of analyses; quickly analyzing big data and extracting information that can be useful to investigators and establishing correlations between pieces of information that are invisible to the human eye. Later, when algorithms present the result of the automated search, the analysts can validate the AI-generated information and start looking for leads, such as pictures of certain objects. Analysts can then narrow down their search, cross-check the information with other databases, and begin to build a chart connecting different suspects and their activities. Specifically, Europol reports rightly observe, that these are decisions that require specific expert knowledge and, for this reason, will always be performed by human analysts.

Classifying millions of pictures and extracting container numbers, however, is tedious, and previously time-consuming, work that AI can support to free up scarce human resources that can be put to better use elsewhere.

What are the dangers that the AI-derived evidence brings?

Although the use of AI has become crucial for core crimes digital forensics, investigators and judges must be aware of several problems that stem from the risks typical for the use of algorithms in data-analysis. The first such risk is the bias built into the algorithm's operation mechanisms (15; 9, p. 10). Serious concerns are expressed in the literature relating to the fact that the analysis of evidence performed by an algorithm must assume that the algorithm is properly focused on the specific data sets in question and in accordance with the appropriate specific patterns. As a result, the acquired analysis may be susceptible to "algorithm bias" (e.g. in relation to racial, ethnic or gender issues). For example, in the case of facial recognition systems, the model may be trained on less diverse data sets and lead to inaccurate and biased recognition of people of a nationality or race whose representatives are more likely to commit crimes. Gender bias may also occur due to gaps in the documentation regarding harm to men and women, or due to social norms [16]. Algorithmic systems can only be as good as the data they are trained on. That is why the scope and nature of the data fed to the algorithm is crucial. Therefore, a diverse set of training data, error mitigation techniques, and regular evaluation of the machine learning model should be employed in order to mitigate these risks.

Given the digital context in which the algorithms operate, the second problem is feeding the algorithm with intentionally falsified data. For instance, a machine learning model may be intentionally misled to incorrectly classify or identify an object or person. In the era of deepfakes, small, intentional data disruptions cannot be ruled out. Every photo, every video and digitally stored information can be predisposed to be false. Also, large amounts of data may be created intentionally, which may lead to the creation of a false narrative – these may be campaigns sponsored by states or private entities pursuing specific goals [16]. Deepfakes can be created not only manually, but they can be also created by Generative AI (a term that refers to "any tool based on a deep-learning software model that can generate text or visual content based on the data it is trained on"). There are even special tools available online for this purpose; tools that have emerged in recent years and are capable of generating images realistic enough to create disinformation (but also tools being able to discover this technology: e.g. AI or Not app). Such intentional disinformation could involve, for example, digitally replacing a specific uniform with another, or changing a face to look like another person's face. When digital sources of information are fed disinformation, the mere threat or suspicion of information modification can lead to undermining the very possibility of obtaining evidence in this way, thus negating its procedural value for fact-finding [15].

Another crucial consideration in the accountable and effective use of AI models is the need to be aware of their uncertainty to explain their outputs. It is rightly claimed that AI models should be turned into "explainable AI" (XAI), which allows for more transparent foundations. Specifically, trustworthy human-centric AI should be used by enforcement authorities, where the "human in the loop" will allow for increased fairness and accountability. Notably, XAI explores methods that provide humans with the ability of intellectual oversight over AI algorithms. Machine learning (ML) algorithms used in AI can be categorized as white-box or black-box. White-box models provide results that are understandable to experts in the domain. Black-box models, on the other hand, are extremely hard to explain and may not be understood even by domain experts. Black-box takes place where even the AI's designers cannot explain why it arrived at a specific decision. On the other hand, XAI algorithms follow the three principles of transparency, interpretability, and explainability. The main focus is on the reasoning behind the decisions or predictions made by the AI algorithms to make them more understandable and transparent for the user of end-results. Therefore, not only algorithms must be written in a way that is understandable, and logically explicable to those whose task is to develop or maintain the system, but also the results of AI-analysis should be explainable to the judge, or defense counsel. The EU Commission rightly observed that "The lack of transparency (opaqueness of AI) makes it difficult to identify

and prove possible breaches of laws, including legal provisions that protect fundamental rights, attribute liability and meet the conditions to claim compensation" [21, p. 14]. Without the possibility to explain the rules which governed the data analysis and led to results presented in trial, it is hard to guarantee effectiveness of the right of defense. The fact, that it is not possible to question or cross-examine an algorithm should not lead to the consequence that the defense has no possibility to learn about the techniques of data-analysis.

Verification of AI-derived evidence during criminal trial

In consequence of such a rapid development of AI-derived and managed evidence, in a criminal trial there is a need to assess admissibility of algorithmically-derived evidence. Verification of digital evidence should become an obligatory stage in criminal investigations, allowing for the accuracy of the source and validity of a piece of evidence to be established. As rightly observed in the literature, it is necessary "to ensure the integrity of the evidentiary material and preserve the history of its transmission through continuous instrumental controls during data retrieval" [17; 10, p. 100–102]. Moreover, "any action taken on electronic evidence must be documented so that an independent third party can repeat the action and obtain a similar result" [11, p. 198]. During trial it is necessary to verify the authenticity of digital evidence. This can be done by internal investigators or algorithms – checking the data by following available sources and links. The methods for doing so are various, all rooted in the digital environment: they can include comprehensive metadata checks, reverse image searches, as well as more sophisticated tools and techniques, making it possible to reveal potential tampering, misattribution, and authorship and authenticity – executed by an expert. Another question is, however, who should be responsible for this verification phase: whether they should be experts appointed by the parties or independent experts called by the court; external or internal experts [6, p. 673; 8, p. 1235]. It is also possible that law enforcement agencies could apply technological solutions capable of discovering deepfakes without a need to call an expert [5, p. 122].

Moreover, it must be remembered that before the AI deployment, state authorities must complete a fundamental rights impact assessment and register the system in the EU database. A special group – AP4AI (Accountability Principles for Artificial Intelligence) project, which is a joint endeavor of Europol and the Centre of Excellence in Terrorism, Resilience, Intelligence and Organized Crime Research (CENTRIC) and members of the EU Innovation Hub for Internal Security – presented a report with answers to concerns about data bias, fairness, and potential encroachments on privacy, accountability, human rights protection and discrimination [14]. This group also took into consideration the requirements of the EU's Artificial Intelligence Act. It should be noted that in Chapter II, Art. 5 EU AI Act, the following types of AI system are prohibited: e.g. biometric categorization systems inferring sensitive attributes (race, political opinions, trade union membership, religious or philosophical beliefs, sex life, or sexual orientation), except labelling or filtering of lawfully acquired biometric datasets or when law enforcement categorizes biometric data; compiling facial recognition databases by untargeted scraping of facial images from the internet or CCTV footage; 'real-time' remote biometric identification (RBI) in publicly accessible spaces for law enforcement, except when: searching for missing persons, abduction victims, and people who have been human trafficked or sexually exploited, preventing substantial and imminent threat to life, or foreseeable terrorist attack; or identifying suspects in serious crimes (e.g., murder, rape, armed robbery, narcotic and illegal weapons trafficking, organized crime, and environmental crime, etc.). Before deployment of biometric identification, police also must obtain authorization from a judicial authority or independent administrative authority, though, in duly justified cases of urgency, deployment can commence without authorization, provided that authorization is requested within 24 hours. If authorization is rejected, deployment must cease immediately, deleting all data, results, and outputs.

In consequence, it is visible, that the use of AI in criminal investigations leads to creation of a totally different landscape from the one we already know and use. It does not mean however,

that such technology should be avoided. As long as certain guarantees are provided, especially taking into consideration the UE AI Act, the right to a fair trial and the effectiveness of the right to defense, AI models can simplify the work of investigators and make it more efficient. Possibly, it would also require a proper national legislative framework, providing rules for the use of AI-derived evidence. Also, it is inevitable, that investigators and judges dealing with core crimes should be trained in the new technology and aware (and prepared for) all the risks that it carries.

Funding for the research: The research project was financed from the funds of the National Centre of Science (Narodowe Centrum Nauki) granted on the basis of a contract No. UMO-2023/49/B/HS5/02623 for a project entitled "In search of justice for core crimes in the digital age."

Bibliography:

1. AKSAMITOWSKA, Karolina. Digital Evidence in Domestic Core International Crimes Prosecutions: Lessons Learned from Germany, Sweden, Finland and The Netherlands, Journal of International Criminal Justice, Volume 19, Issue 1, 2021.
2. D'ALESSANDRA, Federica; SUTHERLAND, Kirsty. The Promise and Challenges of New Actors and New Technologies in International Justice, Journal of International Criminal Justice, Vol. 19, Issue 1, 2021.
3. FREEMAN Lindsay. Digital Evidence and War Crimes Prosecutions: The Impact of Digital Technologies on International Criminal Investigations and Trials. Fordham International Law Journal, Vol. 41, 2018.
4. FREEMAN Lindsay. Weapons of War, Tools of Justice Using Artificial Intelligence to Investigate International Crimes, Journal of International Criminal Justice, 2021.
5. GARRIE, Daniel B.; MORRISSY, David J. Digital Forensic Evidence in the Courtroom: Understanding Content and Quality, Journal of Technology and Intellectual Property, Vol. 12, Issue 2, 2014.
6. GILLETT, Mathew; FAN, Wallace. Expert Evidence and Digital Open Source Information Bringing Online Evidence to the Courtroom, *Journal of International Criminal Justice*, Vol. 21, Issue 4, 2023.
7. KHAN, Karim. Innovation and Technology in Building Modern Investigations and Prosecutions at the ICC. In: The International Criminal Court in Its Third Decade. Reflecting on Law and Practices. STAHN, Carsten; BRAGA DA SILVA, Rafael (eds.). Brill, 2023.
8. KOENIG, Alexa; FREEMAN, Lindsay. Cutting-Edge Evidence: Strengths and Weaknesses of New Digital Investigation Methods in Litigation, Hastings Law Journal, Vol. 73, 2022.
9. RICHMOND, Karen. AI, Machine Learning, and International Criminal Investigations: The Lessons From Forensic Science (November 10, 2020), iCourts Working Paper Series, No. 222.
10. RUGGIERI, Franco. Security in digital data preservation, Digital Evidence and Electronic Signature Law Review, Vol. 11, 2014.
11. BLAHUTA, Roman; MOVCHAN, Anatolii; MOVCHAN, Maksym. Use of Electronic Evidence in Criminal Proceedings in Ukraine, *Advances in Social Science, Education and Humanities Research. Proceedings of the International Conference on Social Science, Psychology and Legal Regulation*, 18.11.21, Use of Electronic Evidence in Criminal Proceedings in Ukraine | Atlantis Press (atlantis-press.com) (23.12.2024).
12. CRAWFORD, Julia; PETIT, Franck. Insights on the digital revolution for war crimes probes in Ukraine, JusticeInfo.Net, <https://www.justiceinfo.net/en/93111-insights-digital-revolution-war-crimes-probes-ukraine.html (23.12.2024).
13. DIGITAL LOCKERS: Archiving Social Media Evidence of Atrocity Crimes 2021, Human Rights Center, UC Berkeley School of Law (https://humanrights.berkeley.edu/sites/default/files/digital_lockers_report5.pdf).
14. Europol, AI and policing The benefits and challenges of artificial intelligence for law enforcement, AI and policing | Europol.
15. EVANS, Hayley; HAZIM, Mahir. Digital Evidence Collection at the Int'l Criminal Court: Promises and Pitfalls OTPLink, Project Harmony, and Digitalization Efforts, *JustSecurity*, July 5, 2023, https://www.justsecurity.org/87149/digital-evidence-collection-at-the-intl-criminal-court-promises-and-pitfalls/ (23.12.2024).
16. MIMRAN, Tal; WEINSTEIN, Lior. Digitalize It: Digital Evidence At the ICC, *Lieber Institute West Point*, 14.08.23, Digitalize It: Digital Evidence at the ICC - Lieber Institute West Point, (23.12.2024).

17. MOLINA GRANJA, Fernando; RODRIGUEZ Glen Dario. The Preservation of Digital Evidence and Its Admissibility in the Court, International Journal of Electronic Security and Digital Forensics, Vol. 9, 2017, <https://www.researchgate.net/publication/312934498_The_preservation_of_digital_evidence_and_its_admissibility_in_the_court> (23.12.2024).
18. OTP 2023 Annual Report Delivering Better – Office of the Prosecutor Annual Report 2023 (icc-cpi.int) (23.12.2024).

19. Policing in an AI-Driven World, Police Chief Online, April 24, 2024, Policing in an AI-Driven World - Police Chief Magazine (23.12.2024).
20. 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), OJ L, 2024/1689, 12.7.2024.
21. European Commission White Paper on Artificial Intelligence; A European approach to excellence and trust, (EU Commission, 19th February 2020.

Ганна Кучинська. Аналіз та оцінка доказів штучним інтелектом у кримінальному процесі

У все більш цифровому середовищі збору та зберігання інформації про основні злочини правоохоронні органи вирішили використовувати алгоритми для більш ефективного аналізу й управління доказами. Ці алгоритми були розроблені як інструменти для автоматизації процесу доказування на всіх етапах збирання, зберігання, захисту й аналізу доказів. Правоохоронні органи функціонують в умовах, коли потенційні свідки завантажують сотні тисяч цифрових доказів і відеозаписів основних злочинів. Окрім соціальних мереж існують інші цифрові джерела доказів: цифрові аудіо- та відеозаписи, записи з камер відеоспостереження, аеро- та супутникові знімки, зйомка з дронів тощо. Враховуючи різке збільшення обсягу та швидкості передачі даних у контексті кримінальних розслідувань, ШІ став незамінним помічником у роботі слідчих. Не лише алгоритми стали частиною цифрової криміналістики, а й використання штучного інтелекту стало займати центральне місце в сучасних цифрових розслідуваннях. Однак слідчі та судді повинні усвідомлювати проблеми, пов'язані з ризиками, характерними для використання алгоритмів в аналізі даних. Крім того, унаслідок такого стрімкого розвитку доказів, отриманих і керованих штучним інтелектом, виникає потреба в оцінці допустимості таких доказів у кримінальному процесі.

Ключові слова: *Міжнародний кримінальний суд, штучний інтелект, кримінальний процес, перевірка доказів, допустимість доказів.*