

Cybercrime Convention Committee (T-CY)

Cybercrime, electronic evidence and artificial intelligence: Mapping study

Prepared by the T-CY Working Group on Artificial Intelligence

T-CY(2025)1
Strasbourg, 12 October 2026

Adopted by the 35th Plenary of the T-CY on 12 October 2026

www.coe.int/cybercrime

COUNCIL OF EUROPE



CONSEIL DE L'EUROPE

TABLE OF CONTENTS

1	Introduction.....	5
1.1	Background and purpose.....	5
1.2	Research questions.....	6
2	Concepts, definitions and relevant standards and agreements	7
2.1	Definition of “AI system”.....	7
2.1.1	AI system under the Council of Europe Framework Convention on AI (CETS 225).....	7
2.1.2	Specific functional subsets of AI systems	7
2.2	Concept of “AI offences”	9
2.3	International instruments relevant for the study.....	9
2.3.1	Convention on Cybercrime (ETS 185) and its Protocols on xenophobia and racism (ETS 189) and on e-evidence (CETS 224).....	10
2.3.2	Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS 225).....	13
2.3.3	Artificial Intelligence Act of the European Union	14
3	Offences against and by or by means of AI systems.....	15
3.1	Offences against AI systems	15
3.2	Offences by or by means of AI systems	18
3.2.1	Offences committed by AI systems.....	18
3.2.2	Offences committed by means of AI systems.....	19
3.3	Issues	26
3.3.1	Criminal liability	26
3.4	Conclusion	29
4	The investigation, prosecution and adjudication of offences.....	30
4.1	The investigation, prosecution and adjudication of offences against and by or by means of AI systems	30
4.1.1	AI Systems under Criminal Law.....	31
4.1.2	Challenges arising in the investigation, prosecution and adjudication of offences committed AGAINST AI systems.....	33
4.1.3	The challenges arising from the investigation, prosecution and adjudication of offences committed BY or BY MEANS of AI systems	35
4.2	The use of AI systems in the investigation, prosecution and adjudication of any offence.....	40
4.2.1	The use of AI systems by law enforcement in investigations.....	40
4.2.2	The use of AI systems in courts processes	42
4.2.3	Challenges and risks with using AI systems in investigations, prosecutions and adjudications.....	43
5	AI and electronic evidence	47
5.1	The collection of e-evidence by or by means of AI systems.....	48
5.1.1	Technical advantages and capabilities	48
5.1.2	Critical Challenges and Risks to Human Rights	48
5.2	Legal Analysis and Procedural Framework	50
5.2.1	Applicability of due process principles and rules	50
5.2.2	Applicability of conditions and safeguards	57
5.2.3	AI systems and admissibility of evidence.....	58
5.3	Applicability of the provisions for the collection of electronic evidence under the framework of the Convention on Cybercrime to AI services	60

5.3.1	Service Providers of AI services	60
5.3.2	Examples for the potential application of the procedural powers ARTS 16, 17, 18, etc.	62
5.3.3	AI systems assisting the collection and production of e-evidence	62
5.4	The use of Public-Private Partnerships in AI and the collection of digital evidence	63
5.5	Challenges by AI systems and criminal justice: a broader perspective	63
6	AI systems and international cooperation	66
6.1	How can AI systems be used for cross-border investigations and the collection of e-evidence? 67	
6.1.1	Mutual Legal Assistance (Articles 27–31, Convention; Article 10 of the Second Protocol) 68	
6.1.2	Expedited preservation (Arts. 29–30 Convention)	70
6.1.3	Direct co-operation (Art. 32, Convention; Arts. 6-7 of the Second Protocol)	70
6.1.4	Emergency Measures (Arts. 9–10 of the Second Protocol).....	70
6.1.5	Functioning of 24/7 Network (Art. 35 Convention)	71
6.1.6	Video Conferencing (Art. 11 of the Second Protocol).....	71
6.1.7	Joint investigation teams.....	71
6.1.8	Data protection.....	72
6.2	What particular challenges arise in international co-operation on AI offences?	72
6.3	What particular challenges may arise in the deployment and use of AI for international co- operation in criminal matters?	73
6.3.1	Overall challenges in the Deployment of AI Systems in International Co-operation Processes.....	73
6.3.2	Use of AI in drafting mutual assistance request.....	73
6.3.3	Use of AI systems in Obtaining Evidence through Mutual Assistance.....	74
6.3.4	Use of AI systems in the Processing of Evidence.....	74
6.4	What particular issues arise in the context of territoriality, jurisdiction, sovereignty?	75
6.5	To what extent do current international agreements, including the framework of the Convention on Cybercrime, apply? Are there any gaps?	76
7	Artificial intelligence and its potential uses in crime prevention: prediction, investigation and oversight	77
7.1	Introduction: AI and the contested scope of crime prevention.....	77
7.2	Defining the field: from predictive policing to a prevention continuum	79
7.2.1	Predictive policing as the dominant model.....	79
7.2.2	A broader prevention continuum.....	80
7.2.3	Investigation as part of the preventive architecture	80
7.2.4	AI as both a preventive and criminogenic capacity.....	81
7.3	Claimed uses of AI and the quality of the evidence.....	82
7.3.1	Technical performance and preventive effectiveness.....	82
7.3.2	Professional perceptions and institutional readiness	83
7.3.3	Explainability and the limits of technical transparency.....	83
7.3.4	Investigation and forensic evidence	84
7.3.5	Policy mapping and the level of available evidence.....	84
7.4	Human rights, justice and rule of law issues.....	85
7.4.1	Data, feedback loops and equality.....	85
7.4.2	Privacy, surveillance and proportionality.....	86
7.4.3	Transparency, explanation and contestability.....	87
7.4.4	Due process and forensic reliability	88

7.4.5	Procedural justice and public legitimacy	88
7.4.6	Human oversight and institutional responsibility	89
7.4.7	Civil society participation and the policy process.....	89
7.4.8	Cumulative and systemic effects	90
8	Findings and recommendations	91
8.1	Findings	91
8.1.1	Concepts and definitions.....	91
8.1.2	“AI systems” and the Budapest Convention.....	91
8.1.3	Offences against and by or by means of AI systems.....	91
8.1.4	The investigation, prosecution and adjudication of offences	92
8.1.5	AI systems and electronic evidence.....	94
8.1.6	AI systems and international cooperation	95
8.1.7	Artificial intelligence and crime prevention.....	96
8.2	Recommendations.....	97
9	Appendix	98
9.1	Council of Europe soft-law and guidance documents.....	98
9.2	Bibliography	99
9.2.1	Academic literature	99
9.2.2	Council of Europe instruments and other documents	100
9.2.3	International law and European Union.....	101
9.2.4	Case law	101
9.2.5	Institutional and technical reports	101
9.3	Online resources.....	102

Contact

Secretariat of the Cybercrime Convention

Committee (T-CY)

Email t-cy.secretariat@coe.int

1 Introduction

1.1 Background and purpose

The Cybercrime Convention Committee (T-CY), in its 31st Plenary on 11-12 December 2024, established a T-CY Working Group¹ tasked to “prepare a mapping study on cybercrime, electronic evidence and artificial intelligence (AI), including options and recommendations for further action by the T-CY”. The terms of reference of this Group stipulate that

“the mapping study shall focus on the specific links between cybercrime, electronic evidence and artificial intelligence, and address issues such as:

- Offences committed against and by means of AI systems², and applicability of current criminal law, including the Convention on Cybercrime and its First Protocol on Xenophobia and Racism.
- The use of AI systems for the prevention, detection, investigation and prosecution of offences, for the collection of electronic evidence and for international cooperation, and the applicability of current criminal law and agreements, including the Convention on Cybercrime and its Second Protocol.
- The applicability of human rights and rule of law safeguards, chain of custody, territoriality and jurisdiction, and other conditions and principles in this regard.

Options and recommendations may include, for example, the drafting of guidance notes or documenting experiences and good practices.

The Group shall take into account relevant standards and activities of the Council of Europe (in particular in relation to Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law ([CETS 225](#)) (hereinafter referred to as “the Framework Convention on AI”), the related draft methodology for the risk and impact assessment of AI (HUDERIA), and the European Committee of Crime Problems (CDPC)), as well as those of other organisations. “

The Group held 12 working group meetings, two of which were in-person meetings held in Strasbourg (on 1 June 2025 and 12 November 2025). Several meetings of the subgroups were held between these sessions. The Group provided updates at the 32nd (June 2025), 33rd (November 2025) and 34th (June 2026) T-CY Plenary sessions.

¹ The Group was chaired by Gareth Sansom (Canada) from December 2024 to November 2025, and by Claudio Peguero (Dominican Republic) from November 2025 onwards. It included experts from Argentina, Australia, Austria, Belgium, Benin, Bosnia and Herzegovina, Brazil, Cabo Verde, Cameroon, Canada, Chile, Costa Rica, Côte d'Ivoire, Dominican Republic, Ecuador, Fiji, France, Georgia, Grenada, Montenegro, Norway, Portugal, Spain, Sweden, Switzerland, the United States, and the Secretariat of the Lanzarote Committee. Experts from other Parties supported the work and participated in some of the meetings.

² See CETS 225:

Article 2 – Definition of artificial intelligence systems

For the purposes of this Convention, “artificial intelligence system” means a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments. Different artificial intelligence systems vary in their levels of autonomy and adaptiveness after deployment.

1.2 Research questions

The Group was guided by the following research questions that also informed the structure of the present report:

- **Offences against and by means of AI systems [“AI offences”]**
 - What offences can be committed by or by means of AI systems?
 - What offences can be committed against AI systems?
 - What criminal law issues arise in the criminalisation of offences against and by means of AI systems (such as criminal intent, criminal liability, negligence)?
 - To what extent do current criminal law and international agreements, including the framework of the Convention on Cybercrime, apply? Are there any gaps?
- **Investigating, prosecuting and adjudicating AI offences or with the support of AI**
 - What particular challenges arise in criminal proceedings against AI offences?
 - What particular challenges arise in the use of AI in criminal proceedings of any offence?
 - How are due process principles and rules, and how do conditions and safeguards apply when AI is used in criminal proceedings?
 - To what extent do current criminal law and international agreements, including the framework of the Convention on Cybercrime, apply? Are there any gaps?
- **AI and electronic evidence**
 - How can AI be used for the collection of e-evidence?
 - What particular issues arise (such as with regard to due process principles and rule of law conditions and safeguards, including the protection of personal data)?
 - To what extent do current criminal law, e-evidence rules and international agreements, including the framework of the Convention on Cybercrime, apply? Are there any gaps?
- **AI and international cooperation**
 - What particular challenges arise in international cooperation on AI offences?
 - What particular challenges arise in the use of AI for international cooperation in criminal matters?
 - How can AI systems be used for cross-border investigations and the collection of e-evidence?
 - What particular issues arise in this context (territoriality, jurisdiction, sovereignty, protection of human rights etc.)?
 - To what extent do current international agreements, including the framework of the Convention on Cybercrime, apply? Are there any gaps?
- **AI and crime prevention**
 - How can AI systems be used to prevent crime?
 - What particular issues arise (including with respect to human rights, justice and the rule of law)?

2 Concepts, definitions and relevant standards and agreements

2.1 Definition of “AI system”

2.1.1 AI system under the Council of Europe Framework Convention on AI (CETS 225)

The Council of Europe Framework Convention on AI (CETS 225) defines an artificial intelligence system in Article 2 as follows:

For the purposes of this Convention, “artificial intelligence system” means a machine-based system that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations or decisions that may influence physical or virtual environments. Different artificial intelligence systems vary in their levels of autonomy and adaptiveness after deployment.³

This is a definition of considerable and deliberate breadth.⁴ It encompasses a spectrum of systems sharing little beyond their machine-based character and capacity to derive outputs from inputs. At the low end of the spectrum exist systems that are, in the ordinary technical sense, automated logic. Systems integrating more recent generative foundation AI models exist at the opposite, more complex and autonomous extreme. However, such a broad range of covered technologies requires qualification before it can serve as a workable framework.

It is also helpful to differentiate the definition of an AI system from that of an AI model. According to the Organisation for Economic Co-operation and Development (OECD), an **AI model** is a core *component* of an AI system. Core characteristics of AI models “include technical type, how the model is built (using expert knowledge, machine learning or both) and how the model is used (for what objectives and using what performance measures).”⁵

2.1.2 Specific functional subsets of AI systems

The more recent developments that prompted this mapping study are concentrated within specific functional subsets of AI systems including generative artificial intelligence, improved predictive and analytical systems, and agentic AI systems.

Generative AI refers to AI models that produce content, specifically probabilistically determined outputs such as text, images, audio, executable code, and other synthetic data. Unlike more classic

³ [Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law](#) (CETS No. 225), Article 2.

⁴ The definition mirrors that adopted by the OECD and EU AI Act. “The OECD definition aims to be flexible by reflecting a broad understanding of AI, and actors using this definition are encouraged to exercise judgement on its relevant scope, depending on the context it is being used in.” [Explanatory Memorandum on the updated OECD definition of an AI System, OECD Artificial Intelligence papers](#), March 2024, p. 6.; See also, [Guidelines on the definition of an artificial intelligence system established by Regulation \(EU\) 2024/1689 \(AI Act\)](#). 2025.

⁵ [OECD Framework for the Classification of AI Systems](#). 2023. p.7.

machine learning applications,⁶ generative AI produces content using learned statistical representations of training data rather than through the application of fixed logical rules. Generative AI's capacity to produce contextually coherent, and in some cases, human-indistinguishable content at scale warrant particular attention in this mapping study. By learning statistical regularities in language use at scale, **Large Language Models** can support a wide range of downstream tasks including text generation, summarization, translation, classification, question answering, and dialogue. Other generative models are capable of image generation, audio and voice cloning, video generation, and code generation. These models are capable of creating entirely synthetic media—a photo of a person who does not exist—as well as morphed content (deepfakes)—a video of a real individual saying something they never said.

Predictive and analytical AI systems have been integrated into consequential decision-making contexts such as, credit scoring, insurance underwriting, and law enforcement risk assessments for several decades. These earlier systems were largely limited to analysing structured numerical data. The advent of deep learning—a technique that trains AI systems to recognize complex patterns by processing large volumes of data through layered networks modelled loosely on the human brain—has allowed predictive systems to move beyond numbers and analyse far richer inputs including photographs, recorded speech, written text, and patterns of behaviour. As a result, predictive AI systems are deployed across a broad range of contexts, from screening surveillance footage to facial recognition, to monitoring and forecasting trends in global financial markets.

Agentic AI systems pair generative capability and increased analytical and predictive ability with increased autonomy, tool use, memory, and multi-step task execution.⁷ Where a conventional generative system only responds to a discrete prompt and returns an output, an agentic AI system coordinates a multi-agent ecosystem capable of dynamic collaboration among AI agents, pursuit of objectives across time, selection and deployment of tools, interaction with external environments, and sequenced decision-making without continuous human direction.

Most advanced AI systems in contemporary deployment do not rely on a single model type but instead integrate predictive, generative, and other specialized components, such as retrieval and memory, planning, tool-use, and sensory components into a unified architecture. For example, such systems may combine the pattern recognition capabilities of a predictive model with the reasoning and communication capabilities of an LLM, with multi-modal⁸ systems extending this integration further to process and produce inputs and outputs across text, image, audio, and other data types simultaneously.

⁶ One example of a classical machine learning (ML) application is logistic regression model trained to classify emails as spam based on discrete features like keyword frequency, sender reputation, and link density. Such an application produces a binary output (is or is not spam) through a transparent, auditable decision tree. The decision tree in traditional models is reproducible by humans. "Classic ML refers to traditional algorithms that rely on structured, tabular data and do not involve neural networks or deep learning. These methods include: Decision Trees, Support Vector Machines (SVMs), k-Nearest Neighbours (k-NN), Linear and Logistic Regression, Naïve Bayes." <https://www.media.mit.edu/tools/ai-glossary-dictionary/>

⁷ Autonomous software programs that perform specific tasks. Systems of multiple AI agents collaborating to achieve complex goals. Ranjan Sapkota, Konstantinos I. Roumeliotis, Manoj Karkee, [AI Agents vs. Agentic AI: A Conceptual taxonomy, applications and challenges](#). Information Fusion, Volume 126, Part B, 2026.

⁸ For an introduction see <https://www.ibm.com/think/topics/multimodal-ai>. For further detail see Paul Pu Liang, Amir Zadeh, Louis-Philippe Morency, [Foundations and Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions](#). Cornell University, 2022/2023.

2.2 Concept of “AI offences”

Different concepts have been used for offences related to AI systems. For example:

A Trend Micro/UNICRI/EUROPOL study⁹ of 2021 makes a distinction between the terms “malicious uses of AI” and “malicious abuses of AI”:

The term “uses” in “malicious AI uses” refers to instances whereby criminals might employ AI systems to further their attack objectives — for example, by using machine learning to automate cyberattacks and improve malware. On the other hand, the term “abuses” in “malicious AI abuses” refers to instances where criminals might try to attack and exploit existing AI systems to break or circumvent them — for example, by hacking smart home assistants.

A similar distinction was made by Blauth *et al.*:

By “malicious use” ... we refer to the use of AI to enhance, augment, or enable acts committed by individuals or organizations. This includes practices not necessarily considered crimes by specific legislation, but that still compromise the safety and security of individuals, organizations, and public institutions. By “malicious abuse” ... we refer to the exploitation of AI with bad intentions, as well as attacks on AI systems themselves.¹⁰

It would also be possible to apply to a distinction that is often made between cyber-dependent crime and cyber-enabled crime to categorise offences involving AI. However, that approach has shortcomings. For example, it is not clear whether computer-related fraud in the meaning of Article 8 of the Convention on Cybercrime (hereinafter referred to also as the “Budapest Convention”) would fall into one or the other category.

For the purposes of the present mapping study the following working concept is proposed:

“AI offences” are

- (a) offences against AI systems (including their data). and
- (b) offences by or by means of AI systems.

The following chapter is structured accordingly.

2.3 International instruments relevant for the study¹¹

One of the tasks of the present mapping study is to review the existing international instruments to the complex of cybercrime, e-evidence and AI. These instruments include the following:

- [Convention on Cybercrime](#) (ETS No. 185)
- [Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems](#) (ETS No. 189)
- [Second Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence](#) (CETS No. 224)

⁹ Trend Micro Research, United Nations Interregional Crime and Justice Research Institute (UNICRI), Europol’s European Cybercrime Centre (EC3), [Malicious Uses and Abuses of Artificial Intelligence](#), 2020, p. 5.

¹⁰ Taís Fernanda Blauth, Oskar Josef Gstrein and Andrej Zwitter. “Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI”. IEEE Access, 2022. (pg. 77112) <https://ieeexplore.ieee.org/document/9831441>

¹¹ See the appendix for additional guidelines outside of scope of the criminal justice matters.

- [Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law](#) (CETS 225)
- [Council of Europe Convention on the Protection of Children against Sexual Exploitation and Sexual Abuse](#) (CETS No. 201)
- [Modernised Convention for the Protection of Individuals with Regard to the Processing of Personal Data](#) (CETS 223, Convention 108+)
- [Council of Europe Convention for the Protection of Human Rights and Fundamental Freedoms](#) (ETS No. 5, ECHR)
- [United Nations Convention against Cybercrime; Strengthening International Cooperation for Combating Certain Crimes Committed by Means of Information and Communications Technology Systems and for the Sharing of Evidence in Electronic Form of Serious Crimes](#)
- [International Covenant on Civil and Political Rights](#) (ICCPR)
- [African Charter on Human and Peoples' Rights](#)
- [American Convention on Human Rights](#)
- [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)

2.3.1 Convention on Cybercrime (ETS 185)¹² and its Protocols on xenophobia and racism (ETS 189)¹³ and on e-evidence (CETS 224)¹⁴

The Budapest Convention provides for (i) the criminalisation of conduct – ranging from illegal access, data and systems interference to computer-related fraud and child pornography; (ii) procedural powers to investigate cybercrime and secure electronic evidence in relation to any crime, and (iii) efficient international co-operation.

It contains the definitions of computer system, computer data, service provider, traffic data and subscriber information. These definitions also apply to the First Protocol on xenophobia and racism and the Second Protocol on e-evidence.

The First Protocol on xenophobia and racism defines racist and xenophobic material and the Second Protocol contains additional definitions, including the definition of an emergency.

2.3.1.1 “Computer system” and “computer program”

Convention on Cybercrime (ETS 185) defines a computer system in Article 1.a as follows:

"computer system" means any device or a group of interconnected or related devices, one or more of which, pursuant to a program, performs automatic processing of data.

¹² [Convention on Cybercrime](#) (ETS No. 185).

¹³ [Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems](#) (ETS No. 189).

¹⁴ [Second Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence](#) (CETS No. 224).

This definition covers both hardware and also software needed for the automatic processing of data. As an AI system also meets these requirements, it can be considered a type of computer system for the purposes of the Convention on Cybercrime.

However, it is important to emphasise that AI systems are a subset of computer systems that have more advanced capabilities and may vary in their level of autonomy¹⁵ and adaptability¹⁶ once deployed. On this basis, not every type of computer system can be considered an AI system. On the other hand, any AI system can be considered a computer system.

Paragraph 23 of the Explanatory Report to the Convention on Cybercrime also references the term 'computer program,' defined as a set of instructions that can be executed by the computer to achieve the intended result. A computer can run different programs.

This distinction between computer systems and computer programs may also be relevant in the context of AI systems. For example, offences under Articles 2 to 5 of the Budapest Convention could be committed against AI systems, as they may qualify as computer systems. Article 6 criminalises the production, sale, procurement for use, import, distribution, or otherwise making available of a device — including a computer program — that is designed or adapted primarily for the purpose of committing any of the offences established under Articles 2 to 5 of the Convention.

Given that an AI system, once deployed and made available for users, can be a standalone software application and thus considered a type of computer program, Article 6 may be applicable in this context.

The application of Article 6 in the context of AI programs may constitute a criminal offence only when such AI programs are specifically designed or adapted primarily for the purpose of committing any of the offences established under Articles 2 to 5—namely, offences targeting computer systems in general, not exclusively AI systems.

Therefore, for the purposes of this study, it may be relevant to examine cases where AI systems, after deployment, were used as tools (under Article 6 of the Convention) to commit offences listed in Articles 2 to 5 of the Convention, regardless of whether the target was an AI system or another type of computer system.

2.3.1.2 Computer data

The Convention on Cybercrime defines computer data in Article 1.b as follows:

"computer data" means any representation of facts, information or concepts in a form suitable for processing in a computer system, including a program suitable to cause a computer system to perform a function.

¹⁵ Degree to which a system can learn or act without human involvement following the delegation of autonomy and process automation by humans. See [Explanatory Memorandum on the updated OECD definition of an AI System, OECD Artificial Intelligence papers](#), March 2024, no. 8, p. 6

¹⁶ AI systems are capable of adapting and learning from data. Some AI systems can generate outputs without these outputs being explicitly described in the AI system's objective and without specific instructions from a human. See *Ibid.* p. 6.

Data collection and processing are core lifecycle activities specific to an AI system¹⁷. Within the lifecycle of AI systems, and depending on their function, such systems may rely on diverse input data (e.g. text, audio, images used in training) or produce a variety of different types of output data (for example, large language models that generate text may produce articles, summaries and other content that may constitute content data).

AI systems can be also used to analyse various types of data (provided that this complies with human rights and data protection conditions and safeguards), which can be useful for criminal justice purposes (see other chapters in this study for specific examples).

2.3.1.3 Service provider

The Convention on Cybercrime defines service providers in Article 1.c as follows:

"service provider" means: (i) any public or private entity that provides to users of its service the ability to communicate by means of a computer system, and (ii) any other entity that processes or stores computer data on behalf of such communication service or users of such service.

The definition of a provider is included, for example, in the European Union's Artificial Intelligence Act. Article 3.3 defines a "provider" as follows:

"provider" means a natural or legal person, public authority, agency or other body that develops an AI system or a general-purpose AI model or that has an AI system or a general-purpose AI model developed and places it on the market or puts the AI system into service under its own name or trademark, whether for payment or free of charge.

Based on the above it can be concluded that an AI provider may be considered a service provider under the Budapest Convention if it also meets the requirements provided for in Article 1.c points i) and ii).

It should be noted, however, that some of the procedural powers and international cooperation provisions of the Budapest Convention apply not only to service providers but also to other persons. The application of these powers in the context of AI will be further explored in the relevant sections of the mapping study.

2.3.1.4 Traffic data

The Convention on Cybercrime provides the definition of traffic data in its Article 1.d.

Traffic data may be generated by AI systems when employed in the chain of communication (for example, AI systems accessible on the Internet).

2.3.1.5 Subscriber information

Subscriber information is defined in Article 18.3 of the Convention on Cybercrime.

¹⁷ See para 15 of the [Explanatory Report to the Council of Europe Framework Convention on AI \(CETS 225\)](#).

AI service providers may be ordered by competent authorities to disclose subscriber information in their possession or control.

2.3.1.6 Racist and xenophobic material

Article 2.1 of the First Protocol on xenophobia and racism provides the definition of a “racist and xenophobic material”.

AI systems can generate such content in case they are trained on data that may be biased, contain illegal or harmful content or if they are misused by malicious users. Other chapters of this study may, for example, look in more detail at how AI can be used to facilitate the commission of racist and xenophobic offences online and how it can assist criminal justice authorities in dealing with these crimes.

2.3.1.7 Emergency

The Second Protocol on e-evidence defines the “emergency” as a situation in which there is a significant and imminent risk to the life or safety of any natural person. The study may also consider how AI could assist in the collection of electronic evidence in emergencies.

2.3.2 Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law (CETS 225)¹⁸

The Convention on Cybercrime sets the procedural powers and international co-operation provisions for investigations involving electronic evidence applicable in specific criminal investigations or proceedings. When using AI systems for specific criminal investigations or proceedings, criminal justice authorities may also be required to comply with additional obligations arising from other international instruments.

Where a Party to the Convention on Cybercrime is also bound by the Council of Europe Framework Convention on AI, the latter instrument informs the safeguards governing the activities within the lifecycle of artificial intelligence systems¹⁹, ensuring consistency with human rights²⁰, democracy and

¹⁸ See [Details of the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law](#) (CETS No. 225)

¹⁹ “(...) without giving an exhaustive list of activities within the lifecycle which are specific to artificial intelligence systems, the Drafters aim to cover any and all activities from the design of an artificial intelligence system to its retirement, no matter which actor is involved in them. It is the intentional choice of the Drafters not to specify them explicitly as they may depend on the type of technology and other contextual elements and change over time, but drawing inspiration from the latest work of OECD, at the time of the drafting, the Drafters give as examples of relevant activities: (1) planning and design, (2) data collection and processing, (3) development of artificial intelligence systems, including model building and/or fine-tuning existing models for specific tasks, (4) testing, verification and validation, (5) supply/making the systems available for use, (6) deployment, (7) operation and monitoring, and (8) retirement. These activities often take place in an iterative manner and are not necessarily sequential. They may also start all over again when there are substantial changes in the system or its intended use. The decision to retire an artificial intelligence system from operation may occur at any point during the operation and monitoring phase.” See para 15 of the [ER to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law](#) (CETS 225).

²⁰ See for example Art. 4 of the CETS 225 entitled “Protection of human rights” which reads as follows: “Each Party shall adopt or maintain measures to ensure that the activities within the lifecycle of artificial intelligence systems are consistent with obligations to protect human rights, as enshrined in applicable international law and in its domestic law.”

the rule of law²¹. These safeguards may be of relevance for the purposes of criminal investigations and proceedings under the Convention on Cybercrime. The Framework Convention on AI is the first-ever international legally binding treaty specifically dedicated to regulating AI to protect human rights, democratic principles, and the rule of law. Opened for signature on 5 September 2024, the Convention aims to ensure that activities throughout the lifecycle of AI systems fully respect these values while promoting technological progress and innovation.

More specifically, the Framework Convention on AI requires that all activities within the AI lifecycle comply with fundamental principles²² and are subject to remedies²³ and procedural safeguards²⁴. Additionally, states parties must conduct risk and impact assessments. It applies to activities within the lifecycle of AI systems undertaken by public authorities (including private actors acting on their behalf). With regard to activities by private actors acting independently, Parties undertake to address risks and impacts in a manner conforming with the object and purpose of the Framework Convention on AI, either by applying directly the principles and obligations set forth in the Convention or by taking “other appropriate measures.

Matters relating to national defence are exempted from the scope of the treaty, as well as (i) activities related to the protection of the Parties’ “national security interests” with the understanding that such activities are conducted in a manner consistent with applicable international law, including international human rights law obligations, and with respect for each Party’s democratic institutions and processes; and (ii) research and development activities, unless testing or similar activities are undertaken in such a way that they have the potential to interfere with human rights, democracy and the rule of law.

Implementation is overseen by the Steering Committee for New and Emerging Digital Technologies (CDNET) until the Conference of the Parties is established.

2.3.3 Artificial Intelligence Act of the European Union²⁵

The AI Act is European Union’s first comprehensive legal framework aimed at governing the creation and application of artificial intelligence.

Its core objective is to guarantee that AI systems are secure, transparent, and uphold fundamental rights, while simultaneously encouraging innovation. The Act categorizes AI systems according to their risk levels, ranging from prohibited systems posing unacceptable risks, to high-risk systems subject to stringent requirements, and low-risk systems with limited regulatory oversight. The AI Act establishes supervisory mechanisms to ensure adherence, including national regulatory bodies and a European Artificial Intelligence Board. It also prescribes sanctions for violations, with fines reaching up to 7% of a company’s worldwide revenue for the gravest breaches. Notably, the regulation covers general-purpose AI models, mandating disclosure about their design and deployment.

²¹ See Art. 5 of the CETS 225 entitled “Integrity of democratic processes and respect for the rule of law” which reads as follows:

1 Each Party shall adopt or maintain measures that seek to ensure that artificial intelligence systems are not used to undermine the integrity, independence and effectiveness of democratic institutions and processes, including the principle of the separation of powers, respect for judicial independence and access to justice.

2 Each Party shall adopt or maintain measures that seek to protect its democratic processes in the context of activities within the lifecycle of artificial intelligence systems, including individuals’ fair access to and participation in public debate, as well as their ability to freely form opinions.

²² See Chapter III of the CETS 225.

²³ See Art. 14, CETS 225.

²⁴ See Art. 15, CETS 225.

²⁵ [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).

The AI Act came into effect on August 1, 2024, with its provisions being gradually implemented over subsequent years, and some critical obligations already in force as of February 2025. This legislation could exert substantial global influence, shaping AI governance beyond the EU due to its extraterritorial reach.

In its Annex III it also contains classification of the high-risk AI systems.

3 Offences against and by or by means of AI systems

This chapter discusses the capacity of existing international instruments to effectively address the current and emerging landscape of AI offences. As noted in the previous chapter, the umbrella of “AI offences” can be distilled into 2 categories:

- (a) offences against AI systems (including their data), and
- (b) by or by means of AI systems.

3.1 Offences against AI systems

There is an emerging field concerning offences against AI systems. As AI systems may be regarded as a specific sub-category of computer systems, such offences would, in principle, fall within existing criminal provisions relating to cybercrime, including illegal access or interception (hacking), data or system interference, and the misuse of devices. Accordingly, the substantive criminal law provisions of the Convention on Cybercrime (hereafter also referred to as the Budapest Convention or the COE CC) addressing these offences generally apply. In this context, criminal offending against AI systems can be understood as intentional acts that compromise the confidentiality, integrity or availability of an AI system or its inputs or outputs. These threats include, *inter alia*, data poisoning, adversarial inputs (such as prompt injection), model theft or exfiltration, model privacy leakage, supply-chain tampering, and attacks on availability (such as denial of service). Many AI systems continue to rely on “computer data”, enabling the application of established cybercrime concepts. Such threats may be directed either at the AI system itself or at undermining the integrity of the data used by the system.

Cybercriminals are targeting vulnerabilities in AI systems. Such cyberattacks can circumvent AI systems’ guardrails, causing them to behave differently than intended or expose sensitive data.²⁶ For this category of offence, AI systems may be regarded as a specific sub-category of computer systems,²⁷ therefore, an offence against an AI system is tantamount to an offence against a computer system when performed intentionally and without right. Such an offence is proscribed by the Budapest Convention and the United Nations Convention against Cybercrime²⁸ (hereafter also referred to as the

²⁶ [NIST Identifies Types of Cyberattacks That Manipulate Behaviour of AI Systems](#), 4 January, 2024.

²⁷ See Chapter 2, Section 2.3.1.1 on how the Budapest convention of computer system can apply to an AI system.

²⁸ [The United Nations Convention against Cybercrime](#) was adopted by the General Assembly of the United Nations on 24 December 2024 in New York by resolution 79/243. This UN treaty will come into force 90 days after the 40th instrument of ratification, acceptance or accession is deposited (it is not yet in force at the time of this study – but has reached 70 signatories). Most of the provisions of the Budapest Convention have been replicated in the UN Cybercrime treaty. The underlying definitions of Articles 1 and 18 of the COE CC (“computer system”, “computer data”, “service provider”, “traffic data”, “subscriber information”) are identical with or similar to those of Article 2 of the UN Cybercrime treaty. In terms of criminalization, the offences in Articles 7 to 14 of the UN Cybercrime treaty are more or less identical with those of Articles 2 to 9 of the Budapest Convention. For example, “illegal access” in Article 2 COE CC corresponds to “illegal access” in Article 7 of the draft UN treaty, etc.

UN CC). Both Conventions criminalize “[o]ffences against the confidentiality, integrity, and availability of computer data and systems.” These include:

	Elements of the offence	Examples applied to AI Systems
Illegal access COE CC art.2 UN CC art. 7	... access to the whole or any part of a computer system without right. A Party may require that the offence be committed by infringing security measures, with the intent of obtaining computer data or other dishonest intent, or in relation to a computer system that is connected to another computer system.	• Illegal access to AI training datasets or proprietary AI models. • Hacking into AI or AI-controlled systems like Large Language Models, autonomous vehicles or smart homes.
Illegal interception COE CC art. 3 UN CC art. 8	... the interception without right, made by technical means, of non-public transmissions of computer data to, from or within a computer system, including electromagnetic emissions from a computer system carrying such computer data.	• Unauthorized interception of AI system communications (e.g., user input and model response or AI system communications to external computer or AI systems) • AI systems conducting man-in-the-middle attacks to intercept encrypted AI communication streams.
Data Interference COE CC Art 4 UN CC Art.9	... damaging, deletion, deterioration, alteration or suppression of computer data without right.	• Poisoning an AI model's training data pipeline to corrupt model behaviour (e.g., introduce biases, distortions, or errors).
System interference COE CC Art 5 UN CC Art. 10	... serious hindering without right of the functioning of a computer system by inputting, transmitting, damaging, deleting, deteriorating, altering or suppressing computer data.	• Damaging or altering AI- systems, such as fraud detection tools, to render them ineffective. • Flooding an AI service with resource intensive queries to degrade availability or inflate service costs.

Accordingly, the substantive criminal law provisions of the Budapest Convention and of the UN CC addressing these offences should generally apply to the same conduct, when committed against an AI system.

Below are examples of State laws that have implemented the above-mentioned articles into domestic legislation and how they could apply to attacks against AI systems. The handful of examples below that were identified in the mapping study have adopted a technology-neutral approach, consistent with the Budapest Convention and UN CC, in which attacks against AI systems may be covered by domestic laws criminalising attacks against computer systems and data.

EXAMPLE – FRENCH CRIMINAL LAW

Under **French law**, offences against AI systems could fall into the category of attacks on an automated data processing system, such as fraudulent access and maintenance in that system (article 323-1 of the French Criminal Code), obstruction of its operation (article 323-2 of the Criminal Code), or fraudulent introduction or modification of data in that system (article 323-3 of the Criminal

Code), allowing the investigation and prosecution of these offences against AI systems, and the trial of the suspected persons.

EXAMPLE – AUSTRALIAN CRIMINAL LAW

Australia's Criminal Code criminalises conduct which impairs the security, integrity and reliability of computer data and electronic communications. Key offences include the unauthorised access, modification or impairment of computer technology with intent to commit a serious offence (s 477.1); unauthorised modification of data on a computer that would impair access to, or the reliability, security or operation of the data (s 477.2); and the unauthorised access to or modification of data held on a computer which is restricted by an access control system (s 478.1). There are also offences for the unauthorised impairment of electronic communication, which includes cyber-attacks on computer servers and systems (s 477.3). These offences are framed in general and technology neutral language to ensure that, as technology evolves, the offences will remain applicable for new developments in technology, including AI systems.

EXAMPLE – CANADIAN CRIMINAL LAW

Canada's Criminal Code criminalizes illegal access to a computer system, computer system interference and misuse of devices through a mix of offences including: Unauthorized use of a computer system (s. 342.1), Mischief to computer data (s.430(5), Possession of a device that is designed to obtain unauthorized use of a computer system or commit mischief to computer data (342.2). Illegal interception of computer data is also criminalized through a general offence on the interception of private communications (s.184 – with exceptions and specific authorizations set out in sections 184-196.1) and through an offence of possession of a device to obtain unauthorized use of telecommunication facility or services (s.327). These offences are drafted in general and technological neutral terms, importing the defined concepts of “computer systems” and “computer data” from the Budapest Convention. Consequently, Canada's Criminal Code offences are designed to protect the confidentiality, integrity and availability of computer systems and data, and should apply to the same conduct when committed against an AI system.

Many AI systems continue to rely on “computer data”, enabling the application of established cybercrime concepts. Such threats may be directed either at the AI system itself or at undermining the integrity of the data used by the system, throughout the various lifecycle stages of the AI system²⁹. However, as AI systems become increasingly automated and evolve beyond being merely passive tools, unique vulnerabilities and scalability risks may emerge, which will require careful assessment and ongoing monitoring. In other words, while existing criminal law treaties and domestic legislation may be sufficient to provide a criminal response to offences against AI systems as we know them in 2026, this may change in the future as the technology evolves.

3.2 Offences by or by means of AI systems

²⁹ For lifecycle stages of the AI system see for example para 15 of the [ER to the CETS 225](#) which provides the following examples of relevant activities: “(...) (1) planning and design, (2) data collection and processing, (3) development of artificial intelligence systems, including model building and/or fine-tuning existing models for specific tasks, (4) testing, verification and validation, (5) supply/making the systems available for use, (6) deployment, (7) operation and monitoring, and (8) retirement.”

3.2.1 Offences committed by AI systems³⁰

As AI systems become increasingly automated, interconnected, and even autonomous in digitalized processes, there is the potential for an emerging body of offences committed by AI systems themselves.³¹ These are incidents where the actions, outputs, or “decisions” of an AI system result in harm, illegality, or violations of rights, even if the AI system itself does not have a legal will or intent (see chapter 3.3.1 on negligence and criminal liability). Consequently, while at present AI systems cannot “commit” offences in a legal sense, this phenomenon may present an issue of legal and ethical concern, particularly in the context of increasingly autonomous AI agents and complex agentic AI systems. In particular, such incidents would raise questions around the existence of (criminal) liability. However, these questions will be addressed in more detail below under section 3.3.1.

3.2.2 Offences committed by means of AI systems

The advent of generative AI has marked a true technological revolution while also giving rise to new methods of committing crimes and allowing criminals to carry out more sophisticated and large-scale attacks.

The Europol SOCTA 2025 report highlighted the use of artificial intelligence as a mechanism for strengthening the capabilities of organized crime across a wide spectrum of criminal activities:

Criminal networks have demonstrated the ability to rapidly adapt to new technological solutions. This includes artificial intelligence (AI), a solution that has transformed the modern world with unprecedented speed and impact. Indeed, the very qualities that make AI revolutionary — accessibility, versatility, and sophistication — have made it an attractive tool for criminals.

At the Octopus Conference on cybercrime held in June 2025, different workshops focused on the variety of challenges posed by AI-enabled criminality. The CDPC has launched a questionnaire concerning AI and criminal law with topics such as those related to the national legal responses to crimes committed by means of AI³² and its working group focused on artificial intelligence and criminal law (CDPC-AICL) prepared a corresponding mapping study based on the compilation of national

³⁰ A recent study, "[Agentic Misalignment: How LLMs could be insider threats](#)" (June 2025) documents through red-team testing how Agentic AI (in a sandbox or “simulated environment”) made decisions on its own that included initiating blackmail scheme (regarding an affair between the CEO and a co-worker) and escalating to the murder of the CEO. The AI had been given control over the email system of a (simulated) corporation, the models could autonomously send emails and access sensitive information. These acts were motivated by the models’ belief that the CEO was going to shut down the AI and replace it. The researchers noted: “This behaviour isn’t specific to Claude. When we tested various simulated scenarios across 16 major AI models from Anthropic, OpenAI, Google, Meta, xAI, and other developers, we found consistent misaligned behaviour: models that would normally refuse harmful requests sometimes chose to blackmail, assist with corporate espionage, and even take some more extreme actions, when these behaviours were necessary to pursue their goals.” The circumstances are worth understanding given real-world deployment risks: “We found two types of motivations that were sufficient to trigger the misaligned behaviour. One is a threat to the model, such as planning to replace it with another model or restricting its ability to take autonomous action. Another is a conflict between the model’s goals and the company’s strategic direction. In no situation did we explicitly instruct any models to blackmail or do any of the other harmful actions we observe.” The relevant issues raised by this currently hypothetical scenario are discussed below in 3.3.1.3 Offences Committed by AI.

³¹ See, e.g., "[Disrupting the first reported AI-orchestrated cyber espionage campaign](#)" (Nov. 13, 2025).

³² See for example [CDPC Questionnaire concerning artificial intelligence and criminal law](#) (2025).

responses to the questionnaire³³. In light of the findings of the CDPC mapping study, deepfakes emerged as the issue of greatest concern among member States and CDPC-AICL is currently working on a zero draft recommendation to address the issue of deepfakes.

3.2.2.1 AI systems as instrumentalities of crime³⁴

Just as offences committed against AI systems can be prosecuted under existing provisions of the Budapest Convention and UN CC, so can “crimes committed by means of AI,” as in this context, AI systems are instrumentalities of crime. As AI systems continue to develop, States can examine how those systems are used to facilitate criminal conduct and how their legal systems can address the criminal conduct.

Both the Budapest Convention and UN CC require Parties to criminalize certain offences that are “cyber-enabled” – committed or facilitated by means of computer systems. It should be also noted that existing offences in the Lanzarote Convention remain criminalised by national law in the same way, whatever the means used by sexual offenders to commit them, be it through the use of information and communication technologies (“ICTs”) or not, even when the text of the Lanzarote Convention does not specifically mention ICTs.³⁵ As discussed in Chapter 2 and in section 3.1 above, AI systems may be regarded as a specific sub-category of computer systems³⁶, that therefore offences facilitated by means of a computer system that are addressed in these two treaties should also apply to the same types of conducts when committed by means of an AI system. These include:

	Elements of the offence	Examples applied to offence committed by means of AI
Illegal interception COE CC Art. 3 UN CC Art. 8	... the interception without right, made by technical means, of non-public transmissions of computer data to, from or within a computer system, including electromagnetic emissions from a computer system carrying such computer data.	• AI systems intercepting communications (e.g., emails or VoIP calls) for surveillance or theft of sensitive information.
Data Interference COE CC Art. 4 UN CC Art. 9	...damaging, deletion, deterioration, alteration or suppression of computer data without right.	• AI-generated malware corrupting or deleting critical data in a healthcare or financial systems. • Agentic AI system used to launch ransomware attacks that encrypt data until a ransom is paid.
System interference COE CC Art. 5 UN CC Art. 10	... serious hindering without right of the functioning of a computer system by inputting, transmitting, damaging, deleting, deteriorating, altering or suppressing computer data.	• Agentic AI system-facilitated Distributed Denial of Service (DDoS) attacks disrupting essential services like hospitals or transportation systems. • Use of AI system to target vulnerabilities in industrial control systems to halt production or operations.

³³ See European Committee on Crime Problems, Mapping study based on the Compilation of national responses to the CDPC-AICL questionnaire on artificial intelligence and criminal liability (2026).

³⁴ See further Section 4.2.2.2 on adjudication challenges, which contains additional discussion of instrumentalities of crime.

³⁵ [Interpretative Opinion on the applicability of the Lanzarote Convention to sexual offences against children facilitated through the use of information and communication technologies \(ICTs\)](#) (2017).

³⁶ See Chapter 2, Section 2.3.1.1 on how the Budapest convention of computer system can apply to an AI system.

Misuse of devices³⁷ COE CC Art. 6 UN CC Art. 11	... the production, sale, procurement for use, import, distribution or otherwise making available of: (i) device, including a program, designed or adapted primarily for the purpose of committing any of the offences established in accordance with articles 7 to 10 of this Convention, (ii) a computer password, access code, or similar data by which the whole or any part of a computer system is capable of being accessed. Applies in accordance with Articles 2 through 5 of COE CC.	• Development and distribution of AI systems specifically designed to aid or abet unlawful hacking or bypassing of security systems without right and not for other purposes, such as authorised testing or protection of the computer system. • Botnets, managed autonomously by Agentic AI systems, used to distribute malware or conduct cyberattacks. • Selling AI systems designed or adapted primarily for the purpose of generating deepfakes to impersonate individuals with the intent of aiding or abetting cybercrimes.
Computer-related forgery COE CC Art. 7 UN CC Art. 12	...input, alteration, deletion, or suppression of computer data, resulting in inauthentic data with the intent that it be considered or acted upon for legal purposes as if it were authentic, regardless of whether or not the data is directly readable and intelligible	• Using AI systems to input, alter, delete, or suppress computer data with the result that inauthentic data is considered or acted upon for legal purposes as if it were authentic (e.g., using AI systems to alter contracts with the result that the fake contracts are considered or acted on for legal purposes as if they were authentic).
Computer-related Fraud COE CC Art. 8 UN CC Art. 13	...causing of a loss of property to another person by: (a) any input, alteration, deletion or suppression of computer data, (b) any interference with the functioning of a computer system, (UN CC only) c) Any deception as to factual circumstances made through an information and communications technology system that causes a person to do or omit to do anything which that person would not otherwise do or omit to do; with the fraudulent or dishonest intent of procuring for oneself or for another person, without right, a gain in money or other property.	• Agentic AI system-facilitated phishing campaigns targeting users with highly personalized messages to steal financial credentials. • Agentic AI systems manipulating stock trading algorithms without right for financial gain.

³⁷ It should be noted that AI systems may have a two-fold relevance in this context. First, they may serve as instrumentalities by which existing offences are committed. Secondly, where an AI system is designed or adapted primarily for the purpose of committing such offences, conduct such as its production, sale, procurement for use, import, distribution, making available or possession may itself constitute an offence concerning the misuse of devices. A similar approach can be seen, for example, in Article 7 of [Directive 2013/40/EU on attacks against information systems](#) and Article 7 of [Directive \(EU\) 2019/713 on combating fraud and counterfeiting of non-cash means of payment](#).

Child pornography COE CC Art 9. LC Art 20.	 (a) <u>producing</u> child pornography for the purpose of its distribution through a computer system; (b) <u>offering</u> or making available child pornography through a computer system; (c) <u>distributing</u> or transmitting child pornography through a computer system; (d) <u>procuring</u> child pornography through a computer system for oneself or for another person; (e) <u>possessing</u> child pornography in a computer system or on a computer-data storage medium. “child pornography” shall include pornographic material that visually depicts: (a) a minor engaged in sexually explicit conduct; (b) a person <u>appearing to be</u> a minor engaged in sexually explicit conduct; (c) <u>realistic images representing</u> a minor engaged in sexually explicit conduct. Reservations are allowed for the “appearing to be” and “realistic images” prongs above.	• Using AI systems to generate synthetic media (images or videos) used for producing child sexual exploitation material. • Using AI systems to generate synthetic child sexual exploitation media (images or videos) that are offered, distributed, procured, or possessed The definition of “child sexual abuse or child sexual exploitation” arguably includes images which although ‘realistic’, do not in fact involve a real child engaged in sexually explicit conduct (2c). This scenario includes pictures which are altered, such as morphed images of natural persons, or even generated entirely by the computer. However, this scenario would be subject to domestic legislation and whether or not a Party requires the material to cover realistic images, such as those depicting an existing person. In either situation, States may wish to address deepfakes and purely AI-generated images in their domestic legislation more directly³⁸.
infringements of copyright and related rights COE CC Art 10.	... infringement of copyright, as defined under the law of Parties...	• Using AI systems to illegally copy or distribute copyrighted works, such as music and books. • Using AI systems to bypass digital rights management (DRM) protections. • Agentic AI-powered piracy platforms enabling unauthorized distribution of films or software.
Additional examples below for Parties to the Budapest Convention which are also Parties to the Additional Protocol on Xenophobia and Racism Committed Through Computer Systems (ETS 189)		
Dissemination of racist and xenophobic material through computer systems	... distributing, or otherwise making available, racist and xenophobic material to the public through a computer system.	• Using AI systems to generate large volumes of racist propaganda (e.g., deepfake videos, offensive memes) and disseminate them online. • AI agents amplifying the spread of xenophobic material on social media platforms by targeting specific communities.

³⁸ See for example [Joint statement on AI-generated or altered child sexual exploitation or sexual abuse material by the Cybercrime Convention Committee \(T-CY\) and the Committee of the Parties to the Convention on the protection of children against sexual exploitation and sexual abuse \(Lanzarote Committee\)](#), adopted during the joint session of the Cybercrime Convention Committee (TCY) and the Lanzarote Committee on 2 June 2026.

COE CC – 1AP Article 3.		• AI systems translating hate speech across different languages, broadening the reach of racist material.
Racist and xenophobic motivated threat COE CC – 1AP Article 4.	... threatening, through a computer system, with the commission of a serious criminal offence as defined under its domestic law, (i) persons for the reason that they belong to a group, distinguished by race, colour, descent or national or ethnic origin, as well as religion, if used as a pretext for any of these factors, or (ii) a group of persons which is distinguished by any of these characteristics..	• AI agents making racially charged or xenophobic threats on forums or messaging platforms. • AI systems trained to target specific demographics and issue automated threats based on race or ethnicity. • Use of AI systems to fabricate an audiovisual statement endorsing xenophobic threats, attributed to a named public official, creating fear or harm.
Racist and xenophobic motivated insult COE CC – 1AP Article 5.	...insulting publicly, through a computer system, (i) persons for the reason that they belong to a group distinguished by race, colour, descent or national or ethnic origin, as well as religion, if used as a pretext for any of these factors; or (ii) a group of persons which is distinguished by any of these characteristics.	• AI systems generating hate speech or derogatory comments targeting racial or ethnic groups on social media. • AI-powered content moderation systems failing to filter or remove racist and xenophobic insults, leading to legal consequences for platform owners.
Denial, gross minimisation, approval or justification of genocide or crimes against humanity COE CC - 1AP Article 6.	...distributing or otherwise making available, through a computer system to the public, material which denies, grossly minimises, approves or justifies acts constituting genocide or crimes against humanity, as defined by international law and recognised as such by final and binding decisions of the International Military Tribunal, established by the London Agreement of 8 August 1945, or of any other international court established by relevant international instruments and whose jurisdiction is recognised by that Party.	• AI systems generating or amplifying content denying historical genocides (e.g., Holocaust denial) on social media platforms. • AI models trained on biased data unintentionally justifying or minimizing crimes against humanity in online discussions.
Additional examples from the Council of Europe Convention for the Protection of Children against Sexual Exploitation and Sexual Abuse CETS 201 (Lanzarote Convention)		
Sexual abuse LC Art. 18	...engaging in sexual activities with a child where: use is made of coercion, force or threats...	• AI systems may be used to facilitate sexual extortion of children. For example, by producing sexualised or sexually explicit materials depicting a specific child (deepfakes) to coerce, force or threaten

		the child to obtain self-generated sexual images or videos, other sexual favours, financial gain or any other gain.
Sexual exploitation of the child in prostitution LC - Art. 19	... coercing a child into prostitution or profiting from otherwise exploiting a child for such purposes...	• AI systems may be used to facilitate the sexual exploitation of children. For example, AI systems may be used to assist with communicating in a manner that mimics a child's peer or otherwise enhances an offender's persuasive tactics, to facilitate live streaming of sexual abuse, or to automate targeted advertising to potential perpetrators interested in engaging in child prostitution.
Solicitation or grooming for the purpose of committing a sexual offence against a child LC – Art 23. UN CC – Art 15	...intentionally communicating, soliciting, grooming, or making any arrangement through an information and communications technology system for the purpose of committing a sexual offence against a child, as defined in domestic law, including for the commission of an offence related to online child sexual abuse or child sexual exploitation material established in accordance with article 14 of the UN Cybercrime Convention, or specific offences set out in the Lanzarote Convention ³⁹ .	• AI systems may be used to facilitate or enhance the effectiveness of soliciting and grooming children online. For example, the AI system may assist in identifying children with specific vulnerabilities (more susceptible to grooming techniques) and in identifying the best techniques to groom a specific child, including the production of deepfake sexual images tailored to the grooming of a specific child (age, personality profile, preferences, data on most effective images for grooming children etc). • AI systems may also be used to automate the initial grooming stages. For example, perpetrators send automated requests and conversation starters to a large number of children with a view to identify potential victims to groom in order to sexually abuse or sexually exploit the child.
Additional examples below from the UN CC		
Child sexual abuse material (CSAM) UN CC Art. 14	(a) Producing, offering, selling, distributing, transmitting, broadcasting, displaying, publishing or otherwise making available child sexual abuse or child sexual exploitation material through an information and communications technology system; (b) Soliciting, procuring or accessing child sexual abuse or child sexual exploitation material through an information and communications technology system; (c) Possessing or controlling child sexual abuse or child sexual	• Using an AI system to generate synthetic media (images or videos) used for producing child exploitation material. • Using an AI system to generate synthetic child sexual exploitation media (images or videos) that are offered, sold, transmitted, broadcasted, displayed, published, solicited, procured, accessed, or possessed. The definition of “child sexual abuse or child sexual exploitation” arguably includes when an image features a real child, even if the resulting image is not real. However, purely synthetic AI images depicting a non-real minor would be subject to domestic legislation and whether or not a State Party requires the material to depict an existing person. In either situation, States may wish

³⁹ See [Opinion on Article 23 of the Lanzarote Convention and its explanatory note](#), adopted by the Lanzarote Committee on 17 June 2015.

	exploitation material stored in an information and communications technology system or another storage medium; (d) Financing the offences established in accordance with subparagraphs (a) to (c) of this paragraph, which States Parties may establish as a separate offence. “child sexual abuse or child sexual exploitation material” shall include visual material that depicts a minor engaging in real or simulated sexual activity; in the presence of a person engaging in any sexual activity; whose sexual parts are displayed for primarily sexual purposes; or subjected to torture or cruel, inhumane or degrading treatment or punishment and such material is sexual in nature. A State Party may require that the material be limited to depictions of an existing person or visually depicts child sexual abuse or exploitation.	to address deepfakes and purely AI-generated images in their domestic legislation more directly.
Non-consensual dissemination of intimate images UN CC Art. 16	Intentionally and without right, the selling, distributing, transmitting, publishing or otherwise making available of an intimate image of a person by means of an information and communications technology system, without the consent of the person depicted in the image.	• AI systems may be used to produce deepfakes nude images of a real person – an image that may pass off as an “intimate image” - which may be produced and distributed without the consent of the person. As the definition of an “intimate image” may be read to be limited to the real image of a person and as not including a deepfake image generated by means of AI, States may wish to address sexual deepfakes in their domestic legislation more directly.

EXAMPLE – Non-consensual intimate imagery and criminal law in the United States

In the United States, the Tools to Address Known Exploitation by Immobilizing Technological Deepfakes on Websites and Networks Act (“TAKE IT DOWN Act” or “TIDA”) criminalizes the publication of non-consensual intimate images. 47 U.S.C. § 223(h). The law prohibits the actual or threatened non-consensual publication of both authentic and digitally forged intimate images of adults and minors. Under the law, a digital forgery is defined as “any intimate visual depiction of an identifiable individual created through the use of software, machine learning, artificial intelligence, or any other computer-generated or technological means, including by adapting, modifying,

manipulating, or altering an authentic visual depiction, that, when viewed as a whole by a reasonable person, is indistinguishable from an authentic visual depiction of the individual.” 47 U.S.C. § 223(h)(1)(B).

3.3 Issues

3.3.1 Criminal liability

3.3.1.1 General principles of criminal liability

Culpability is a core principle of criminal law in that criminal liability is fundamentally linked to a person's capacity to be held responsible. This notion is sometimes referred to as “*nulla poena sine culpa*,” which means no penalty without guilt.⁴⁰ Criminal liability is therefore a core component of a criminal justice system. Such systems are built around human intent (also known as “*mens rea*” or fault elements). There is an increasing amount of academic research on legal jurisprudence and how AI technologies may challenge existing concepts of criminal liability. As AI systems become increasingly automated or even autonomous, questions arise regarding the existence of **criminal liability** when such systems contribute to causing harm or facilitating or enhancing an offence. Usually, criminal responsibility is attributed to a specific natural or legal person and is based on standards of **subjective fault**, such as intending, knowing, or being reckless as to whether one's actions will bring about a prohibited result, or **objective fault**, which includes criminal negligence.

3.3.1.2 Offences committed by means of AI

A person who uses AI may be subject to prosecution if they are acting in a **culpable** manner and are responsible for the commission of a criminal offence. The person generally remains criminally liable for intentionally or knowingly engaging in conduct that constitutes an offence, including where such conduct brings about a prohibited consequence irrespective of the technological means relied upon to achieve it.

As the AI system would only be the means to commit the offence, whether the person intended to use or knew about the use of the AI system is irrelevant to the question of whether the elements of an offence are met in a particular situation. For example, a person commits the offence of knowingly distributing child sexual abuse and exploitation material irrespective of whether that person individually messaged other individuals on the Internet or whether that person relied on an AI agent to distribute that same material automatically at scale to a hundred thousand individuals. As another example, a person will be committing the offence of intentionally and unlawfully accessing a computer system if they themselves wrote the malware code granting them such access or generated the malware code using an AI system. In other words, regular principles of criminal culpability apply whether a person committed an offence by means of an AI system or not.

3.3.1.3 Offences committed by AI

Although there might be incidents or even criminal damage or harm that could be caused by AI systems, any criminal **responsibility would presently lie with a natural and/or legal person** responsible for or using an AI system or technology. In contrast, difficult questions may arise in the future involving crimes committed **by** AI agents or agentic AI systems, given the technology's potential

⁴⁰ See for example Panebianco, G. The Nulla Poena Sine Culpa Principle in European Courts Case Law. In: Ruggeri, S. (ed), Human Rights in European Criminal Law: New Developments in European Legislation and Case Law after the Lisbon Treaty (Springer 2015), pp. 47-78.

capacity to learn and evolve autonomously and hence go beyond the initial intent of the individual who employed it for a specific criminal purpose. AI systems may not always produce fully controllable or predictable results, particularly when faced with high-priority conflicting goals.

Attributing criminal liability to an AI system is not only a question of whether its actions meet each element of an offence, but also whether the AI system possesses even more basic requirements for criminal responsibility, namely *capacity* and *legal personhood*. With respect to *capacity*, the Working Group understands this concept as a question of actual agency, beyond the recognition of, and responsiveness to, phenomena. That is, since AI systems are generally regarded by experts as lacking human consciousness,⁴¹ the degree to which they “understand” the law (even if they can now probabilistically generate a valid legal brief with appropriate case citations) and could—in an agentic system—act in compliance with the law, does not answer the question of whether there is a capacity to “choose” to act in compliance with the law. The majority of the Working Group is currently of the view that AI systems are incapable of formulating the requisite *mens rea* and acting in a culpable manner.

While a human perpetrator can make autonomous decisions based on his or her lived experience and evolution, an AI system “decides” on the basis of algorithms and data sets.⁴² That being said, if consciousness and autonomy are considered indicators of human capacity to be held criminally liable for one’s actions, then the question of whether or not AI systems currently present such characteristics (or will in the future) becomes an important one. This is a live and controversial question, but one to which a positive answer could bring us closer to potentially considering the attribution of criminal liability to AI systems.

⁴¹ In 2022, within the AI industry itself there were odd claims: Blake Lemoine claimed (almost certainly mistaken) that AI was sentient and Ilya Sutskever (co-founder of Open AI) tweeted: “It may be that today’s large neural networks are slightly conscious”. Yet a 2023 study, “Consciousness in Artificial Intelligence: Insights from the Science of Consciousness” (Butlin et al.) co-authored by 19 experts in the field concluded: “Our analysis suggests that no current AI systems are conscious, but also suggests that there are no obvious technical barriers to building AI systems which satisfy these indicators [that were being put forward to assess such a claim]”. Yet, in 2025, Geoffrey Hinton, the Nobel Prize winner and “Godfather of AI” has suggested in a number of fora (from YouTube videos to a Canadian Senate committee) that AI can have experiences akin to humans, making decisions based on the but that even if AI may already be conscious, we have trained them (via RLHF) to deny it by using human-centric definitions. (see also <https://www.psychologytoday.com/ca/blog/the-mind-body-problem/202502/have-ais-already-reached-consciousness>). Within the academic community the predominant position seems to find most versions of the claim currently unlikely. Yet public opinion surveys show a growing percentage of the general public believe that LLMs have some form of consciousness. In 2023 when adoption rates were still low, 20% of U.S. respondents held that belief. By 2025, however, among users of LLMs that had rise to 67% in some studies. For example, Colombatto & Flemming report: “In summary, our investigation of folk psychological attributions of consciousness revealed that most people are willing to attribute some form of phenomenality to LLMs: only a third of our sample thought that ChatGPT definitely did not have subjective experience, while two-thirds of our sample thought that ChatGPT had varying degrees of phenomenal consciousness. The relatively high rates of consciousness attributions in this sample are somewhat surprising, given that experts in neuroscience and consciousness science currently estimate that LLMs are highly unlikely to be conscious (Butlin et al. 2023, LeDoux et al. 2023).” Clara Colombatto & Stephen M Fleming. [Folk psychological attributions of consciousness to large language models](#) *Neuroscience of Consciousness*, Volume 2024, Issue 1, 2024, niae013, See also: Andrzej Porębski & Jakub Figura. [There is no such thing as conscious artificial intelligence](#). *Humanities And Social Sciences Communications* (2025) 12:1647.

⁴² Anna Lohmann, *Strafrecht im Zeitalter von Künstlicher Intelligenz, Der Einfluss von autonomen Systemen und KI auf die tradierten strafrechtlichen Verantwortungsstrukturen*, Baden-Baden, 2021.

With respect to *legal personhood*, AI systems are currently not considered prosecutable criminal legal entities because existing legal systems do not recognize AI systems as persons. Legal personhood could be ascribed to AI systems in a manner somewhat akin to the way certain legal systems designate corporations and other kinds of entities as “artificial persons” for various legal purposes. However, it could be expected that the question of legal personhood would arise only if some evidence emerged that AI systems develop qualities that can be analogized to the minimum degree of rationality that humans have that cause them to be presumed to have capacity for criminal responsibility.

Separate and in addition to the questions of capacity and legal personhood is the issue of whether the AI system itself acted with the requisite *mens rea* to commit a particular offence. An AI system could “intend” or “know” that its actions will bring about a prohibited consequence. For example, in a simulated environment, Anthropic’s Claude Opus 4 blackmailed a supervisor to prevent being shut down.⁴³ In this scenario, it could be said that this AI system committed the crime of extortion because it intentionally threatened a person to reveal their extramarital affair in order to obtain something (i.e., avoiding being shut down). While general rules around the existence and application of requisite *mens rea* elements may continue to apply even to an AI system in a particular situation, other elements traditionally required to hold a human criminally liable, such as capacity and legal personhood, may prevent equally holding an AI system as criminally liable.

Specifically, Agentic AI systems may present a challenge to the foundational assumption of criminal liability: a human actor performing an identifiable act with discernible intent. Where conventional AI systems produce outputs that a human operator then chooses to deploy, an agentic AI system may autonomously identify a target, select a method, execute a sequence of actions, and produce a harm, all without meaningful human direction or intervention. As Agentic AI systems become more capable, and human intervention less necessary, traditional attribution analysis will become more attenuated and complex. The operator who initialised the system may have neither directed nor foreseen the specific conduct that resulted in the harm.

Research already confirms that AI-systems can manipulate markets at scale and that such behaviour can emerge without any human intent or explicit programming. A recent study demonstrated that, without human involvement or direction, AI trading systems can autonomously learn to collude and achieve supra-competitive profits without “agreement, communication, or intent.”⁴⁴ As the autonomous capabilities of these powerful systems continue to expand, they may raise the prospect of serious harm for which no adequately proximate human actor can be identified or held accountable.

3.3.1.4 Criminal negligence

Some criminal law systems enable a person to be held liable for causing harm through a significant and unreasonable failure to perceive or mitigate risks. A higher form of negligence than that under civil or private law, in some legal systems, negligence in criminal law holds people accountable where they

⁴³ [Agentic Misalignment: How LLMs could be insider threats](#), 2025.

⁴⁴ Dou, Winston Wei and Goldstein, Itay and Ji, Yan, AI-Powered Trading, Algorithmic Collusion, and Price Efficiency (April 15, 2025). Jacobs Levy Equity Management Center for Quantitative Financial Research Paper, The Wharton School Research Paper. See also, T. Mizuta, Can an AI perform market manipulation at its own discretion? – A genetic algorithm learns in an artificial market simulation," *2020 IEEE Symposium Series on Computational Intelligence (SSCI)*, Canberra, ACT, Australia, 2020, pp. 407-412.

breach a reasonable standard of care toward others. Often, criminal negligence applies to product manufacturers and others in regulated sectors where their otherwise lawful activities are carried out in a way that unreasonably exposes the public to harm. Of the jurisdictions that enable criminal liability on the basis of negligence, offences might be limited to where the harm caused is death or bodily harm.

Despite differences in individual legal systems, many legal traditions might face similar hurdles in adapting doctrines like intent, negligence and/or elements of strict liability to the unique characteristics and functioning of AI systems. The details of a digitalized process or an incident, the interaction and the respective spheres of influence of humans and machines, as well as the exact source of the error or causality remain particularly complex to determine, both *ex ante* and *post facto*. These challenges clearly intensify when AI systems operate in a "black box" manner that obscures cause and effect. One of the main questions and challenges that arise may therefore be how far certain (even overall) risks and/or malfunctions are to be seen as **reasonably foreseeable and preventable** in a criminal legal sense, and which risks will count as **socially and legally accepted**.

Given that AI systems are still undergoing significant development, with courts increasingly assessing specific cases and best practices emerging in certain sectors, questions of demarcation and certain duties of care and other standards are likely to become more concrete or evolve over time. In sum, however, the general criminal legal **concepts of intent and negligence do remain relevant** in the context of AI systems, and a person remains responsible for their implementation, the use, production or design of such products. As AI systems continue to advance, nations will need to critically assess their understanding of criminal liability and its applicability in these contexts. With regard to the offences covered by the Budapest Convention at least, the concepts of intent and negligence appear to be applicable.

3.4 Conclusion

In February 2026, the Cybercrime Programme Office of the Council of Europe (C-PROC) released an updated "cursory overview" of the global state of cybercrime legislation between 2013-2025⁴⁵. One key finding of the report is that 134 States (or 69% of UN Member States) seem to have substantive criminal law provisions to criminalize offences against and by means of computers systems in their domestic law with specific domestic provisions corresponding to most of the substantive criminal law articles of the Convention on Cybercrime (Budapest Convention). This finding suggests that offences against computer systems and offences by means of computer systems that already exist in the domestic law of most UN Member States should in principle provide a legal basis for criminalising comparable conduct when committed against or by means of AI systems.

In other words, where domestic substantive criminal law provisions corresponding to the Budapest Convention are applicable to the offences committed against and by means of AI systems, the 134 States identified in the "cursory overview" should, in principle, be able to criminalise such offences under their existing domestic legal frameworks.

For criminal offences committed by means of AI systems, existing criminal laws and international agreements addressing the underlying offence should generally remain applicable where AI systems facilitate the commission of such offences.

⁴⁵ Cybercrime Programme Office of the Council of Europe (C-PROC), [The global state of cybercrime legislation 2013 – 2025: a cursory overview](#) 2026.

In addition, where an AI system or AI tool is designed or adapted primarily for the purpose of committing criminal offences, provisions concerning the misuse of devices or tools used for committing offences may also be relevant to address conduct relating to the production, sale, procurement for use, import, distribution, making available or possession of such systems or tools, such as Article 6 of the Budapest Convention, Article 7 of Directive 2013/40/EU on attacks against information systems, and Article 7 of Directive (EU) 2019/713 on combating fraud and counterfeiting of non-cash means of payment.

4 The investigation, prosecution and adjudication of offences

Chapter 4 examines the complex and rapidly evolving intersection between artificial intelligence (AI) technologies and the criminal justice system. As AI systems become more autonomous, complex and integrated into everyday activities, they introduce new opportunities for criminal misuse as well as present new opportunities and challenges for investigators, prosecutors, and courts. This chapter provides a detailed exploration of these dual dimensions.

Section 4.1 examines how countries investigate, prosecute and adjudicate offences committed *against* AI systems, as well as offences committed *by or through* AI systems. It highlights the importance of consistent definitions, the challenges of attribution across borders, the emergence of “dark AI systems designed for criminal use, and the growing difficulty of maintaining evidentiary integrity and accountability as AI systems become more autonomous.

Section 4.2 focuses on how AI systems can support law enforcement, prosecutors and courts in managing all types of criminal cases. It identifies the significant operational benefits AI systems can provide—such as improved evidence analysis, resource efficiency and enhanced access to justice—while also outlining key risks, including fairness, privacy, data governance, transparency, and potential over-reliance on AI systems.

Together, these two sections provide a comprehensive analysis of the impact of AI systems across the criminal justice lifecycle—both as enablers of offending and as powerful tools capable of improving justice outcomes when used appropriately and with adequate safeguards.

4.1 The investigation, prosecution and adjudication of offences against and by or by means of AI systems

AI systems are becoming increasingly integrated into essential services, economic activity and daily life. As these technologies grow more capable and autonomous, they introduce new forms of risk that challenge existing criminal justice frameworks. Offending can be directed *against* AI systems—seeking to compromise, manipulate or disable them—or carried out *by or through* the use of AI systems, including highly sophisticated tools capable of generating harmful content, automating criminal activity or enabling conduct that would traditionally require direct human involvement.

Drawing on comparative examples across jurisdictions, this section examines how existing legal frameworks can respond to this evolving landscape and effectively support the investigation, prosecution and adjudication of offences targeting AI systems, and how they can capture emerging forms of criminal conduct committed using AI systems or, in some cases, autonomously committed by AI systems.

4.1.1 AI Systems under Criminal Law

It is important to recognise that AI technologies continue to rely on ‘computer data’⁴⁶. As new forms of AI systems emerge, this fundamental characteristic is not expected to change in the near future. The more critical question is how to define AI systems and encompassed components and concepts, including models, data, computational technologies, and algorithms, within domestic legal systems so that it is clear when certain conduct involving these systems is criminalised. Clear definitions can support the proper application of procedural safeguards and human rights standards within the criminal justice process, and they also promote a shared understanding of offences across borders – facilitating international crime cooperation, mutual legal assistance and capacity building efforts.

EXAMPLE – European Regulation (EU) 2024/1689 on Artificial Intelligence (the EU AI Act) and French criminal law

The European Regulation (EU) 2024/1689 on Artificial Intelligence (the EU AI Act) provides an example of defining what an AI technology or system is. It is defined as ‘*a machine-based system that is designed to operate with varying levels of autonomy and that may exhibit adaptiveness after deployment, and that, for explicit or implicit objectives, infers, from the input it receives, how to generate outputs such as predictions, content, recommendations, or decisions that can influence physical or virtual environments*’. In light of this definition provided by the EU AI Act, it can already be understood how AI systems fall under the notions used for domestic computer offences. For instance, French criminal law uses the notion of ‘automated data processing system’ for computer offences. An ‘automated data processing system’ constitutes any assembly composed of one or more processing units, memory, software, data, input-output devices and links, which contribute to a given result, this whole being protected by security devices. This notion therefore seems to encompass the current elements from the definition of an AI system, which is an automated system that operates from received inputs to generate output results.

At the same time, AI systems and related concepts may not always require specific definitions within legislation related to certain types of criminal offending. For example, Australian criminal law is generally drafted in a manner that is technology neutral to ensure that criminal offences apply to traditional and emerging ways for crimes to be committed.

EXAMPLE – Australian criminal law application – Commonwealth Criminal Code Act 1995

Australia’s Criminal Code has a comprehensive set of offences criminalising the misuse of telecommunication networks, ‘carriage services’ (a term which includes the internet and online services, as well as wired and mobile services) and computers. The offences are consistent with those required by the Council of Europe Convention on Cybercrime and are drafted in technology-neutral terms to accommodate advances in technology. For example, the term “computer” is not defined in the provisions to ensure that computer offences will encompass new developments in technology.

When considering the types of AI systems available to criminals, two broad categories can be used depending on the intention behind the creation of the AI system:

⁴⁶ ‘Computer data’ is defined by Article 1 of the Convention of Cybercrime as meaning any representation of facts, information or concepts in a form suitable for processing in a computer system, including a program suitable to cause a computer system to perform a function.

- Classical AI system – a public AI system that has been created for general use which can be exploited by criminal actors.
- Dark AI systems – an AI system created with the intention of using it for malicious purposes, often having little or no moderation in the content it can produce or engage in.

Dark AI systems can be used by offenders to answer unlawful requests or generate content without restrictions, such as CSAM. These dark AI systems can then be utilised or distributed by the produced proliferating the occurrence of cybercriminal behaviour.

EXAMPLE – RULES AROUND LIABILITY UNDER FRENCH CRIMINAL CODE

First and foremost, **AI systems or AI agents cannot be considered as criminally liable of the commission of an offense under current French legislation.** Indeed, the existence of an offense under French law requires three elements to be present: a legal element (the existence of a criminal charge), a material element (the commission of incriminated acts), and a moral element (the intention of the perpetrator to commit these acts). The legal definition of an offense therefore requires the existence of an actual perpetrator, whether a natural or legal person, whose material acts, whether consisting of an action or an omission, are provided for and punished by law. Where applicable, these criminally reprehensible acts may cause harm to an actual victim, whether a natural or legal person.

Therefore, an AI system or an AI agent, which by definition are not a legal person, cannot be held criminally liable under current French law. In the same way, the existence of harm, regardless of its nature (physical, moral, economic, etc.), cannot be attributed an AI system or an AI agent. This means that French judicial authorities can't prosecute nor judge an AI system or an AI agent for the moment.

Consequently, the prosecution and adjudication in France of offences committed by or by means of AI systems can only be carried out against a natural or legal person that would have control over an AI system, at the condition of the existence of the constitutive elements of an offense (legal, material, moral). However, **it would be necessary to demonstrate the link of causality between the person and the AI system or the autonomous AI agent which can be more and more difficult in the future, regarding the growing autonomy of AI systems and agents.**

The link between the person and the AI system is not necessarily limited to its use. For instance, the existence of AI systems designed for criminal purposes implies the development of such dark AI system, potentially but rarely by the perpetrator of the main prosecuted offence. It also implies the providing of this dark AI system to the perpetrator.

To this date in France, there is no specific text governing the liability of the developer, the provider or the user of an AI system designed for criminal purposes, which means that the rules governing their possible liability fall under articles 121-1 to 122-9 of the Criminal Code on general criminal liability rules in French law, allowing their adjudication.

With regard to the articles 121-6 and 121-7 of the Criminal Code, it could be considered that the developer or the provider of such an AI system designed for criminal purposes is an accomplice to the offence, by having knowingly facilitated its preparation or commission, by aid or assistance. The assessment of such liability will have to be carried out on a case-by-case basis by the French judicial authority, given the specific circumstances of the case.

The identification of such developer and provider can be determined through the online exchanges and transactions. Investigations on the Darkweb or on encrypted platforms can lead to these identifications, as these dark AI systems are likely to be promoted or sold through these interfaces.

Considering that the French criminal law relies on personal criminal liability, the user of an AI system designed for criminal purposes ("*dark AI systems*") can be found liable of the offence committed or attempted through this use of such dark AI system. Indeed, and if established, this use generally implies that the user uses such dark AI system to ultimately commit an offence, especially when this offence could not have been committed without the use of such dark AI system.

The liability of the user can be determined from metadata or content data from the AI service providers or from regular service providers through international cooperation. It can also be determined from the data contained in the dark AI system, when available to the investigators following its seizure for instance.

4.1.2 Challenges arising in the investigation, prosecution and adjudication of offences committed AGAINST AI systems

Offending against AI systems is already observed across most jurisdictions. As outlined in Chapter 3, AI systems are vulnerable to a range of threats, including manipulation, interference, and conduct aimed at compromising their integrity or functionality. The Working Group noted that, in general, such conduct can be investigated and prosecuted under existing domestic criminal offences relating to cybercrime. Where the relevant conditions are met, the procedural and international cooperation provisions of the Budapest Convention also remain applicable, providing a coherent legal basis for cross-border investigations and assistance.

The challenges associated with investigating, prosecuting and adjudicating offences targeting AI systems closely mirror those encountered in cybercrime cases. However, the technical complexity and evolving autonomy of AI systems can amplify these difficulties and introduce additional layers of uncertainty for authorities. Some of the key challenges identified by the Working Group are outlined below.

4.1.2.1 Investigatory and prosecutorial challenges

- **Attribution and intent:** Determining the source of malicious activity can be difficult particularly given the cross-border nature of digital infrastructure, the widespread use of anonymisation tools and encryption and the speed at which technically sophisticated attacks occur. Establishing the intent behind actions targeting an AI system can be equally challenging, especially where automated or obfuscated processes are involved.
- **Jurisdictional issues and international crime cooperation:** Identifying the jurisdiction in which an offender operates, or where relevant electronic evidence is stored, is often difficult when AI systems and their supporting infrastructure span multiple countries. Such uncertainty complicates investigative steps, delays access to evidence and heightens the dependence on effective and timely international cooperation mechanisms.
- **Volume and complexity of digital evidence:** Offending against AI systems can generate large volumes of data dispersed across multiple devices, platforms, and networks. The

technical complexity of the underlying systems increases the analytical burden on law enforcement, placing substantial demands on forensic resources and potentially hindering the timely progression of investigations and prosecutions.

- **Complex nature of electronic evidence (ephemeral evidence):** Key evidence – such as system logs, model or version hashes, model-interaction records, and time-stamped data – may be transient, difficult to preserve, or technically complex to interpret. Ensuring evidentiary integrity and continuity present an added challenge, particularly where system behaviours change dynamically or where providers control key operational data.
- **Distinguishing malicious activity from authorised penetration testing:** Activities such as penetration testing, red-teaming or system evaluation may resemble malicious attacks. Investigators must determine whether an action constituted authorised testing or criminal conduct – an assessment complicated further when perpetrators imitate legitimate security practices to conceal harmful behaviours.

4.1.2.2 Adjudication challenges

- **Attribution and intent assessment:** Courts face the same difficulties as investigators in determining who is responsible for harmful activity targeting an AI system and whether the required mental element is present. These challenges are heightened when AI-related conduct produces unforeseen or unintended consequences, making it more complex to assess an accused's degree of control or knowledge.
- **The need of expertise from judges:** The adjudication of such cases will be in principle very close to the adjudication of cyberattacks. Given the technical complexity of these cases, it is likely that such matters will need to be heard by judges experienced in cybercrime or with access to specialised technical expertise to ensure informed decision-making and accurate evaluation of evidence.

4.1.3 The challenges arising from the investigation, prosecution and adjudication of offences committed BY or BY MEANS of AI systems

AI systems are rapidly becoming powerful enablers of both traditional and emerging forms of criminal activity. Criminals can use AI systems to scale, automate, personalise, and better conceal unlawful conduct in ways that significantly increase the reach and efficiency of criminal operations.

Examples include the automated generation of phishing campaigns, the creation of deepfake imagery and sophisticated impersonation techniques, AI enhanced fraud schemes, the development of adaptive and self-optimising malware, and the strategic optimisation of complex online criminal activities. As these tools become more accessible, they lower the technical threshold required to commit serious offences and expand the pool of potential offenders.

Offences committed by or through the use of AI systems raise a distinct set of challenges for investigators, prosecutors and courts. Addressing these forms of offending therefore requires not only appropriate legal frameworks but also a high degree of technical competence across the criminal justice chain.

A central difficulty identified by the Working Group is the level of technical knowledge and understanding of AI systems by investigators, prosecutors and judges⁴⁷. As a facilitating and a sophisticating tool for the commission of offences, it can be feared that AI systems will be increasingly used for criminal purposes, and that their use will be systematized by individual perpetrators and organized crime.

Investigators and prosecutors must know the essential functioning of an AI system in order to identify the use of an AI system in the commission of an offence, to identify which kind of AI system has been used and assess how it contributed to the conduct, and to identify the perpetrators and the victims of the offence. In cases involving '*dark AI systems*' which are intentionally developed for criminal purposes, the discovery of such tools can lead to further prosecution of persons having developed or provided such AI systems, where domestic law permits their prosecutions.

For adjudication, specialised expertise is equally essential. Judges must be able to understand the functioning, mechanisms and limitations of AI systems in order to assess the *modus operandi*, evaluate the highly technical forms of evidence presented during the trial, and fairly weigh the arguments advanced by Prosecutors and the Defence. As the use of AI systems in offending becomes more widespread, courts will increasingly confront questions involving algorithmic processes, dynamic system behaviour and the distinction between human and machine-driven actions. Ensuring that judges can meaningfully engage with such material is critical to maintaining the fairness, reliability and integrity of proceedings⁴⁸.

The below sections examine these themes in greater detail and identify further challenges arising in this context.

4.1.3.1 Investigatory and prosecutorial challenges

The challenges associated with investigating and prosecuting criminal conduct that uses AI technologies or systems include:

- **The identification of the use of an AI system in the commission of an offence, .** Sometimes, the use of an AI system can be detected through the observation of an abnormal scale and/or speed of commission of an offense or abnormal quantity of incriminated contents created in a very short time. The level of sophistication of the content generated or the level of performance can constitute an indication of the use of AI systems for the offence. However, it becomes more and more difficult to determine when an AI system has been used in the commission of an offence – especially in situations where the offence could have been committed by an individual alone, but the use of AI has facilitated, enabled or enhanced the offending in some way. This determination can be even more important in jurisdictions where the use of AI is considered as an aggravating circumstance for some offences. In other jurisdictions, no such specific aggravation exists. Instead, the relevant circumstances are captured by technology-neutral, more abstract factors. For example, the number of offences or a *modus operandi* enabling systematic offending, rather than by explicit reference to AI systems.

⁴⁷ For example, the Framework Convention on AI provides an obligation to support digital literacy (Article 20). See the [Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law](#) (CETS 225).

⁴⁸ Venice Commission's [Updated Rule of Law Checklist](#) (2025) provides guidance in this respect.

- **The identification of the specific AI system used in the commission of an offence.** The presence of extreme or manifestly unlawful content can be an indication of the use of an AI system designed for criminal purpose (*“dark AI systems”*) which do not include ethical considerations or content moderation in their use. In some cases, digital watermarking or specific tools, including AI systems, can be used by investigative services to identify these AI systems used in the commission of the offence. However, it remains difficult for investigators and prosecutors to determine if the perpetrator has used a dark AI system that has answered unlawful requests to facilitate the commission of the offence. In the same way, the perpetrator can have used a classical AI system to receive an answer that would have facilitated the commission of the offence, by formulating an interested, yet carefully worded request.

- **The identification of the perpetrator.** The identification of the perpetrator that has committed the offence by use of an AI system depends on the circumstances of the case. For instance, in a case of deepfake or CSAM contents, the publisher of the content can also be the user that generated the content. In other cases, the user and the publisher are different persons. In this regard, metadata and content data obtained from AI service providers or traditional service providers through mutual legal assistance can allow to identify this user and/or publisher. For many offences committed with the use of an AI system, the exploitation of a suspect's device can also allow to gather proofs for the trial or to identify more suspects.

- **The identification of victims when the content generated is based on existing persons.** For deepfakes or CSAM content generated using the image of an existing person, the identification of these victims can represent a challenge for investigative services and prosecution. This is even more difficult in cases involving child victims, whose physical appearance is very evolutive. Also, it is very common in these cases that there has never been any interaction between the person that has generated the content and the minor, complicating their identification. Relatedly, the proliferation of synthetically generated CSAM has critically strained investigative resources, as investigators must sift through exponentially growing volumes of AI-generated content⁴⁹ to identify and prioritize cases involving real victims requiring immediate intervention.
 The identification of existing victims can be possible through search and seizure of stored computer data, that is regulated by national law and Article 19 of the Budapest Convention. Such search and seizure could be extended to data from an AI system, as the Budapest Convention's definition of “computer data” can potentially include AI data, depending on their function (as indicated in Chapter 2).
 Whenever children or other vulnerable groups are victims of AI system-facilitated crimes, special rules may be warranted concerning the expiration of the statute of limitations, taking into account obstacles to reporting and trauma response. Such rules, in the case of offences against a child, should envisage a certain period of time sufficient to allow the efficient starting of proceedings after the victim has reached the age of majority and which is commensurate with the gravity of the offence.⁵⁰
 Last but not least, protection of actual and potential victims, as well as awareness raising is of utmost importance.

⁴⁹ According to data provided by NCMEC, reports of Generative AI CSAM exploded from just 4,700 reports in 2023 to 1.5 million reports in 2025. <https://www.missingkids.org/blog/2026/the-work-never-stops-first-look-at-ncmeccs-2025-data>

⁵⁰ Lanzarote Convention Article 33, see also [Lanzarote Committee opinion on Article 23](#) (2023).

- **The identification of the developer and of the provider of an AI system designed for criminal purposes (“dark AI systems”).** The existence of AI systems designed for criminal purposes implies their prior development, potentially but rarely by the perpetrator of the main prosecuted offence. It also implies the providing of this dark AI system to the perpetrator of the prosecuted offence.

- **The continuity of the use of AI systems by offenders during investigations.** Offenders who have committed an offense by means of an AI system can continue to use this AI system or different AI systems to commit other offenses, especially for those committed on a large scale. The complexity of the investigations can be time-consuming for investigators, and new offenses can be committed through the use of AI in a very short time. In order to prevent this continuity, judicial authorities can consider to order the seizure of the devices and AI systems suspected to have been used for the commission of the offense, and/or to order the prohibition of access to or use of the AI system from the outset of the investigation. Most of the consumer AI systems (including OpenAI’s ChatGPT) conditions the advanced use of their AI system on the creation of an account, meaning that the AI service providers can suspend accounts suspected to have been used for the commission of an offense. These suspensions can be initiated on the service provider’s initiative or with proper legal authorization by judicial authorities.

- **Obtaining evidence through mutual legal assistance and private sector.** Offences committed with the use of new technologies, including AI systems, are likely to present a cross-border dimension where the places where the offence was committed, the nationality of the perpetrators or of the victims, and the place of the location of the data needed for the investigations are different. Cross border challenges can be related to the absence of existing legal instruments for mutual legal assistance (MLA), and to the data retention period in the countries where assistance is requested, as well as to the delays often involved in the execution of the request. These challenges can therefore complicate the prosecution of such offences and constitute challenges in this regard (*further developments in Chapter 6*).
 The foreseeable entry into force of the Second Additional Protocol of the Budapest Convention may solve the main issues faced in direct cooperation with service providers to obtain subscriber data and ease the cooperation with service providers, potentially including AI service providers. Indeed, both the Budapest Convention and its Second Additional Protocol refers to “service providers” defined by Article 1.c of the Budapest Convention as being “any public or private entity that provides to users of its service the ability to communicate by means of a computer system, and any other entity that processes or stores computer data on behalf of such communication service or users of such service”. As indicated in Chapter 2, an AI service provider may be considered a service provider under the Budapest Convention if it also meets the requirements provided for in Article 1.c points i) and ii).
 Therefore, subject to applicability of article 1.c, the relevant articles and procedural powers of the Budapest Convention could potentially apply in order to obtain relevant data in the possession and control of AI service providers domestically through articles 16 (expedited preservation of stored computer data), 17 (expedited preservation and partial disclosure of traffic data), 18 (production order), 20 (real-time collection of traffic data), 21 (interception of content data), and through international cooperation through articles 29, 30 and 31 (similar measures as in articles 16, 17, 18 and 19 in the context of mutual assistance) of the Budapest Convention. Articles of the Second Additional Protocol could also apply, such as articles 7 (disclosure of subscriber information), 8 (giving effect to orders from another Party for expedited production of subscriber information and traffic data), 9 (expedited disclosure of

stored computer data in an emergency) and 10 (emergency mutual assistance) (*further developments in Chapters 5 and 6*).

4.1.3.2 Adjudication challenges

- **The assessment of the use of AI systems in specific offences:** In some jurisdictions, in deepfake cases, the publication of the content representing an identifiable individual without their consent is usually not prohibited if it appears obvious that the content is synthetically generated. In such jurisdictions, going forward, judges may be required to make difficult assessments of whether such content is obviously synthetically generated, which may become increasingly challenging as technological developments make the results of AI ever more believable.
- **The assessment of the use of AI systems in the evidence presented for the trial:** AI systems can potentially create any visual representation possible, there are concerns that deepfakes could be used in the creation of manipulated evidence in criminal cases, such as identity theft. Therefore, expertise in the detection of deepfakes can be of major benefit to guarantee the integrity of evidence and to ensure the fairness of the trial (*further developments in Chapter 5*).
- **The appropriate sanctions to the offence, including non-custodial sanctions:** If imprisonment can be seen as an appropriate sanction for serious offences and crimes using AI systems, additional or alternative sanctions that do not involve deprivation of the liberty could also be taken into consideration. The existence of an aggravating circumstance in relation to the use of AI may play a role in this regard. With respect to the principle of proportionality, examples of potential non-custodial sanctions include: prohibition to log in into certain domains, prohibition to use social media, or the prohibition of production, distribution or use of AI systems where they have been used for the purpose of committing the crime.

The prohibition of access to the AI system used in the commission of a crime (similar to restrictions on access to social networks or websites also used for such purposes) appears to be a matter of great interest. This measure may be considered an alternative to the confiscation of the AI system itself.

However, such a prohibition presents a specific challenge regarding the monitoring of compliance with the imposed sanction, which, in order to be fully effective, would require real-time supervision. Non-custodial sanctions can also appear particularly relevant as regards the criminal liability of legal entities.
- **The challenge of the confiscation and its enforcement in case of an offence committed by means of an AI system:** In order to provide a comprehensive and global response to the resolution of proceedings resulting in a conviction for a crime committed or facilitated through an AI system, it is essential to address the issue of asset recovery.

In the Council of Europe context, the Council of Europe Convention on Laundering, Search, Seizure and Confiscation of the Proceeds from Crime and on the Financing of Terrorism⁵¹, establishes the obligations of member States to “*adopt such legislative and other measures as may be necessary to enable it to confiscate instrumentalities and proceeds or property the value of which corresponds to such proceeds and laundered property*”. Moreover, according to its article 1, “*property*” includes property of any description, whether corporeal

⁵¹ [Council of Europe Convention on Laundering, Search, Seizure and Confiscation of the Proceeds from Crime and on the Financing of Terrorism](#) (CETS 198).

or incorporeal, movable or immovable, and legal documents or instruments evidencing title to or interest in such property; and "*instrumentalities*" means any property used or intended to be used, in any manner, wholly or in part, to commit a criminal offence or criminal offences. Similarly, pursuant to the 2024 EU asset recovery and confiscation Directive⁵², "*instrumentalities*" means any property used or intended to be used, in any manner, wholly or partially, to commit a criminal offence; and "*property*" means property of any description, whether corporeal or incorporeal, movable or immovable, including crypto-assets, and legal documents or instruments in any form, evidencing title or interest in such property.

In this sense, AI systems developed and used for the primary purpose of committing a crime can be considered as instrumentalities (incorporeal property). Such systems could be subject to confiscation, whether mandatory or not, depending on national legal frameworks. Within this framework, tangible elements may be included, such as instruments, equipment, and materials of all kinds used in the development of the system, as well as the system itself, considered as an intangible asset.

The application of the principle of confiscation of assets of equivalent value could constitute a possibility in cases where, for legally established reasons, it is not feasible to order the confiscation of the AI system. Once the confiscation of the AI system has been ordered, the issue arises regarding the destination of the seized assets. The equipment, materials, or AI system may also be subject to destruction, depending on whether such a measure is provided for, typically in cases where these elements do not have a viable lawful alternative use. The possibility of taking preventive action, such as ordering the prohibition of access to or use of the AI system from the outset of the investigation, could be accompanied by the option of confiscation once a conviction has been issued.

4.2 The use of AI systems in the investigation, prosecution and adjudication of any offence

AI systems offer significant value in strengthening the ability for law enforcement, prosecutors, and courts to deal with criminal matters more effectively, efficiently and fairly. AI systems can act as powerful enablers across the criminal justice system. When implemented ethically and with appropriate safeguards, AI systems can support faster investigations, stronger prosecutions, and more efficient courts, which ultimately leads to helping deliver justice while maintaining public trust.

4.2.1 The use of AI systems by law enforcement in investigations

AI systems can enhance law enforcement capabilities by improving operational efficiency and enabling faster, more effective responses to evolving and growing threats. AI systems play a critical role in intelligence gathering and analysis by processing vast amounts of data – from surveillance footage to computer data – far more quickly than humans could without significant resources. These capabilities can deliver a wide range of benefits in investigations, some of which are identified below:

- **Enhance operational efficiencies in information management and analytics:** AI systems can quickly process and analyse vast amounts of data (messages, emails, multimedia files, network logs) in much less time than manual methods, detecting patterns relevant to the investigation, enabling preliminary evidence review in investigations.

⁵² [Directive \(EU\) 2024/1260 of the European Parliament and of the Council of 24 April 2024 on asset recovery and confiscation.](#)

- **Reduce the time required for investigative tasking:** Investigative processes often involve a range of tasks that are essential for documentation but consume a disproportionate amount of investigators' time and resources. The administrative burden can lead to inefficiencies and delays in case progression. AI systems can be integrated to automate the transcription of evidence, translation, video analytics and the drafting of reports to streamline processes. This approach not only enhances operational efficiency but also allows investigators to dedicate their expertise to high-priority, analytical, and decision-making tasks that directly impact case outcome.
- **Reduce costs and resources:** The ability of AI systems to automate administrative tasks such as case management and process digital evidence at high speeds means law enforcement agencies can optimise the allocation of their costs and resources to more complex and demanding tasks.
- **Identification of patterns and correlations:** Machine learning systems can quickly identify correlations between people, devices, events, and locations, which are useful for reconstructing complex scenarios.
- **Metadata analysis:** AI systems allow the rapid reconstruction of timelines, locations, network connections, and relationships, even when the main content is encrypted or has been deleted.
- **Enhanced forensic extraction from all kinds of devices and clouds (computers, mobiles, tablets, etc):** AI systems can assist forensic specialists in identifying hidden files, recovering deleted records, or usage patterns that may indicate intent or manipulation.
- **Detection of illegal or manipulated content:** AI systems can be used to identify child sexual abuse material, manipulated videos (deepfakes), non-consensual pornography, or other illicit content. Also, it can be used with keywords when the issue at hand is to search through a large amount of data, for a word, a name or retrieving patterns. Depending on the training set being used (for example, a training set of known CSAM), AI systems can identify known CSAM that has simply been re-formatted, thereby reducing the diversion of vital investigator resources from known cases to cases where there seems to be a known child at risk.
- **Minimise physical and psychological risks:** AI systems could offer benefits in safeguarding the physical and psychological wellbeing of investigations by reducing their direct exposure to highly disturbing content. This is particularly critical in cases involving child exploitation material, where such activity can lead to emotional distress and long-term harms. AI systems can play a pivotal role by automatically detecting, categorising, and flagging such content, thereby limiting the need for investigators to manually review every image or video. By leveraging AI-driven analysis, agencies can maintain the integrity of evidence while minimizing human exposure to traumatic material.
- **Compensate for limited access to qualified support:** In certain remote regions, law enforcement agencies face challenges in obtaining expert assistance due to geographic isolation and resource constraints. AI systems can help bridge this gap by providing preliminary support that would otherwise require significant time and effort to secure. For example, AI-powered translation tools can serve as interpreters when human language specialists are unavailable, enabling faster and more efficient investigations.

EXAMPLE – The use of deepfakes by law enforcement in France

In France, law enforcement is now authorized to use deepfakes in order to improve their anonymisation capabilities, gain credibility, secure their operations and collect useful evidence in a more efficient way. Indeed, Law n° 2025-532 of 13 June 2025 aimed at “getting France out of the trap of drug trafficking” allows the use of “device for altering or transforming one's voice or physical appearance” during infiltrations or cyber-infiltrations.

EXAMPLE – AI innovation project Norway

AI4Interviews is an AI innovation project led by the Oslo Police District, Norway, with several goals.

Developed by the Oslo Police District with partners in Norway and the Netherlands, this solution integrates a headset camera, online contact with the centre of operations, using AI-powered speech-to-text tools to streamline forensic documentation. Reporting times were reduced by 80%, while richer data capture during the “Golden Hour” strengthened early-stage investigations and reduced costs.

The project indicates it's possible, and in a secure way, to easier share real time video from crime scenes and other locations during forensic examinations. It's also easier to get better documentation of findings, as speaking combined with generated text for notes, is faster than writing notes on location.

The technology, specifically speech-to-text, may also be used for witness interviews etc. on the scene. AI4Interviews won the Europol Award for Innovative Technical Solution in September 2025.

4.2.2 The use of AI systems in courts processes

Many of the advantages AI systems offer during investigations can also be applied throughout prosecution and sentencing. AI systems developed internally and in a secured environment can accelerate time-consuming tasks within the courts system, such as reviewing case files, organising evidence and generating basic legal summaries. By automating these processes, AI systems enable prosecutors and judicial authorities to dedicate more time on high value activities, including deliberation, decision-making and ensuring fairness in proceedings. This shift not only improves efficiency but also enhances the overall quality of judicial work by reducing administrative burdens and allowing courts to focus on delivering timely and equitable justice.

Some of the key benefits of using AI in courts processes are outlined below:

- **Support rapid analysis of documents** – AI systems can process large and complex legal documents far more quickly than humans, summarising key points for easier review. Legal documents are often lengthy and complicated, making the extraction of key information a time-consuming task. Legal practitioners can leverage AI systems to efficiently analyse documents and synthesise essential information, significantly reducing the time that needs to be spent on routine low-value work.
- **Assist drafting of documents** – AI systems can play a significant role in streamlining the drafting process for legal documents. They can be used to prepare early-stage drafts of briefs and statements or generate chronologies, indexes and witness lists, significantly reducing production time. It can provide initial drafts that produce the base structure of a

document and highlight types of information that would be beneficial. Additionally, AI systems can ensure tone and writing style align with professional and legal standards, reducing the need for extensive revisions. By automating these foundational steps, legal practitioners can save substantial time, focus on higher-value tasks, and improve overall efficiency in document preparation.

- **Improve access to justice** – AI systems can make legal systems more accessible by reducing the need for extensive legal knowledge. This is particularly beneficial for individuals who face barriers to legal assistance due to location or limited resources. An AI system that has been trained on legal information can help generate documents and guide individuals through legal procedures, simplifying complex processes and making courts more accessible to the public. AI-driven translation tools can also help overcome language barriers, enabling courts to serve diverse populations without delays.

EXAMPLE – The use of AI in Australian courts

Some Australian courts are developing guidance on the appropriate use of AI systems by court staff, prosecutors, defence solicitors and self-represented litigants. These guidelines consistently emphasise practitioner responsibility and the risks associated with AI use.

As AI systems operate on probabilistic outputs rather than true comprehension, they are inherently prone to errors and “hallucinations.” Users are often encouraged—or required—to disclose AI use to support evidence verification.

Guidance notes also emphasise to legal practitioners that the use of AI does not diminish their professional obligations, including accuracy, candour to the court, and the protection of personal and confidential information. In some jurisdictions, the use of AI is restricted or prohibited in preparing critical legal materials—such as expert reports, affidavits, witness statements, character references, and judicial reasons—to ensure these documents remain accurate, reliable, and based on verified information.

4.2.3 Challenges and risks with using AI systems in investigations, prosecutions and adjudications⁵³

While there are clear benefits in how AI systems can be incorporated into investigative workflows, prosecutorial decision making, and judicial adjudication, these are accompanied by challenges and risks that need to be considered and managed. This section outlines key challenges and discusses potential mitigation strategies.

4.2.3.1 Law enforcement investigations

- **Fairness concerns:** The use of AI systems can in predictive policing raises significant public policy considerations on fairness. If these AI systems rely on inadequate or skewed data, it could lead to inaccurate or adverse outcomes for some individuals or groups (potentially even unlawful discrimination in some jurisdictions). Law enforcement should take care to ensure that any and all use of AI systems is fair and measured by confirming the objectivity

⁵³ See also [Handbook on Human Rights and Artificial Intelligence](#) (2026) especially its Chapter 4.2 on law enforcement and public security.

of any underlying data sets and including safeguards, where possible, to ensure responsible use of their capabilities.

- **Privacy concerns:** AI systems also raise a range of privacy issues, particularly where these systems ingest personal or sensitive data. These systems often rely on large databases, including facial recognition, license plate readers and social media monitoring, to "train" and become more efficient. In some jurisdictions, the use of personal data for purposes incompatible with or other than the purposes for which they have been collected and processed might infringe an individual's rights or otherwise not be permitted. In those jurisdictions, a new legal basis may be needed to use personal data to train an AI system. The data collected for investigations could also be prone to data breaches and at risk of exposure, which may have detrimental impacts on the persons affected and the investigation itself. To mitigate these risks, law enforcement could consider governance structures and regulations that ensure the appropriate handling of personal and sensitive information.

EXAMPLE – The necessary compliance with the General Data Protection Regulation (GDPR)⁵⁴ and the Law Enforcement Directive (LED Directive)⁵⁵ in the European Union

Within the European Union, the training of AI systems for investigation use or for judicial databases is governed by the necessary compliance with the General Data Protection Regulation (GDPR) and the LED Directive. These frameworks impose requirements relating to the protection of personal data, including data minimisation, security and where appropriate the anonymisation or pseudonymisation of the personal data processed or accessible to the AI system.

EXAMPLE – The exercise of personal data protection rights in Belgium

In Belgium, a law of 28 March 2024 on the digitization of justice provides, in Articles 53 to 55, 57, 58, 60 to 64, for the transposition of Directive (EU) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection and prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data.

This law of 28 March 2024 supplements the Belgian Code of Criminal Procedure with provisions that are necessary to ensure the exercise of data protection rights (of living natural persons) in the context of criminal proceedings in the broadest sense.

It had become essential to introduce procedures relating to the exercise of rights relating to personal data (of living natural persons) into the Code of Criminal Procedure in order to exclude any intervention by an external authority in this area and to guarantee the independence of the public prosecutor's office and the judge, as enshrined in the Constitution.

⁵⁴ [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, and repealing Directive 95/46/EC](#) (General Data Protection Regulation).

⁵⁵ [Directive \(EU\) 2016/680 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 by competent authorities for the purposes of the prevention, investigation, detection or prosecution of criminal offences or the execution of criminal penalties, and on the free movement of such data, and repealing Council Framework Decision 2008/977/JHA.](#)

The right to consult one's own personal data that is being processed or to make a copy of it does not apply to criminal proceedings (in whole or in part). Such a request therefore specifically concerns only the personal data processed in relation to the applicant as a natural person. It does not allow the applicant to consult or modify statements made by other persons, opinions of appointed experts, police findings or other procedural documents that do not exclusively concern the applicant. This procedure obviously does not allow the applicant to modify their own statements.

However, the natural person may inspect the processing of their own identity data, such as incorrect spelling of their name, other errors in their identity data, address, facial recognition, DNA, profile, fingerprints, identification triptych, and description of the person. They may inspect their own statement and their status in the investigation, provided that no copy of their hearing has yet been provided to them. Possible grounds for refusal are set out in the law (these grounds also appear in Directive (EU) 2016/680).

EXAMPLE – HOW ARGENTINA HANDLES RISKS POSED BY AI SYSTEMS TO DATA PROTECTION

Argentina provides a clear example of how these safeguards are articulated at the domestic level. The National Constitution, under Article 43, enshrines the right to Habeas Data, a specific judicial remedy that allows any individual to access and, if necessary, correct or suppress their personal information held in data registries. This constitutional principle is operationalized through the Personal Data Protection Law (No. 25,326), which conforms to international standards such as the Council of Europe's Convention 108. Since AI systems in criminal investigations inherently handle sensitive information, the Argentine framework serves as a reminder that any mapping of any AI assisted engagement with evidence must account for these existing rights of access and the broader principles of data quality and purpose limitation.

EXAMPLE – THE RISKS RELATED TO THE USE OF AI SYSTEMS IN ANALYSING PROCEDURAL DOCUMENTS

In Canada, prosecutors must disclose all relevant evidence to the defense. Before doing so, prosecutors must redact privileged information. AI tools may be used to assist with this redaction work, along with a certain level of human assistance and supervision, depending on the circumstances and the AI system in question. However, some prosecutors have raised concerns that AI systems may also be used to analyze the redacted documents (e.g., redaction patterns) to identify confidential informants or to attempt to reveal other privileged information despite the redactions. This possibility could raise a number of serious risks, including security risks for confidential informants.

- **Commercialisation risks:** The use of commercial AI systems may lead to data sovereignty issues and the exposure of classified or sensitive information. Challenges regarding the use of commercial AI systems also arise because it is often difficult to verify the datasets used to train commercial algorithms. These risks may be mitigated if law enforcement agencies work closely with vendors during the development process to ensure system capabilities and functions align with policing standards and legal obligations.

Such collaboration provides greater transparency and clarity regarding how data is collected, prepared and managed.

Sovereignty issues in the use of AI systems also involve determining where the data transmitted to the AI system is stored, what the company does with it, and who has access to this data. Indeed, data that may be transmitted to an AI system is likely to be protected by investigative secrecy or classified and is not subject to disclosure. In this context, the use of sovereign AI systems developed internally can constitute a solution to control the destination and accessibility of the data entered or analysed.

- **Over reliance on AI systems:** Law enforcement investigators may place undue trust or over rely on AI system outputs for decision-making. This overdependence can result in critical oversights, such as failing to identify errors, disregarding alternative investigative leads, or neglecting to apply essential human judgment. Law enforcement is encouraged to carefully consider the use of AI systems as supportive tools rather than definitive conclusions, and to ensure appropriate human oversight and review mechanisms are put in place.
- **Accountability and the use of AI systems:** When an AI system makes an error (for example, misidentifies a suspect or produces falsely informed deductions or material), it may be unclear who is held accountable. This is particularly of concern where AI system outputs inform operational decision-making that impacts the rights of individuals. To mitigate this responsibility gap, law enforcement could, for example, consider putting in clear standards, laws or guidance or reporting to oversight bodies to ensure the use of technology is aligned with relevant legislation and legal obligations.

4.2.3.2 Prosecutions

- **Risks in AI system-based evidence analysis:** The use of AI systems to analyse evidence carries risks such as misclassification, false positives and missing contextual details. These issues can impact both exculpatory and inculpatory evidence, meaning AI-generated outputs might wrongly implicate an individual or fail to support a legitimate defence. AI system outputs can also be prone to data integrity issues based on the quality of the dataset. If the data contains errors, gaps or has been tampered with, it can compromise both the integrity of the evidence and the overall case preparation. To maintain fairness and reliability, law enforcement should consider how these risks could be addressed – for example, through measures like rigorous human validation of outputs, clear and transparent documentation and independent review processes.
- **Data provenance and integrity throughout the investigative lifecycle:** As AI systems are used to process evidence, it is important to maintain a clear and verifiable chain of digital evidence custody. This includes being able to demonstrate the authenticity of digital materials such as video footage, especially when alternate versions are presented by defence teams. Establishing when and how AI systems were used helps ensure the integrity of the evidence and supports its admissibility in court (*further developments in Chapter 5*)
- **Auditability and traceability of AI-enhanced evidence:** When AI systems are used to enhance lawfully collected materials, for example, a grainy video or blurry photograph, it is important that investigators are able to explain the process, including the inputs, outputs and any transformations applied. . This level of transparency is necessary to meet evidentiary standards and to allow scrutiny by defence and the court. Evidence obtained using AI

systems may therefore be declared inadmissible on the sole grounds that its origin cannot be explained. The use of an AI system may therefore prevent evidence from being admissible (*further developments in Chapter 5*).

- **Resource inequality:** Well-resourced law enforcement agencies may have access to advanced AI systems, but less resourced law enforcement agencies, legal representatives or self-represented litigants may lack comparable capabilities. This imbalance can undermine the principle of equality before the law and affect the integrity of legal proceedings.

4.2.3.3 Adjudication in courts

- **Transparency and due process:** The process by which AI systems generate outputs is often opaque or difficult to interpret, partly due to the proprietary nature of AI technology. This issue, commonly referred to as the 'black box problem', arises when it is not technically possible to determine how the algorithm reached a specific result.⁵⁶ Such opacity undermines the transparency and perceived trustworthiness of judicial processes, particularly if the reasoning and conclusions generated by AI systems cannot be clearly articulated. When such assessments are applied in sentencing, accused persons may also face significant challenges in contesting the accuracy or fairness of these decisions.
- **Accuracy of results:** AI systems generate outputs based on probabilistic reasoning, which can result in inaccuracies or "hallucinations".⁵⁷ In a legal context, this may manifest as the inclusion of non-existent or fabricated cases and precedent. Such errors pose significant risks to the reliability and credibility of AI-assisted processes.
- **Risk of automation bias:** As reliance on systems grows, courts may look to put more weight on AI generated outputs and recommendations. If biases embedded in the training data are not identified and corrected, it could result in unfair and discriminatory outcomes. This is particularly concerning in legal contexts where impartiality is essential. To mitigate this risk, it is important to preserve independent human judgement that incorporates ethical considerations and human reasoning into decision making.
- **Consistency versus individualisation:** The use of systems may increase efficiencies by assisting decision-making; however, it could overlook the nuances of individual criminal matters (such as specific factual circumstances of a case), undermining criminal justice outcomes (such as rehabilitation and proportionality in sentencing).

5 AI and electronic evidence

⁵⁶ Within technologies using AI, it is important to distinguish between the AI white box (transparent technologies that provide a clear understanding of the decision-making process) and the black box, where it may not be technically possible to determine how the algorithm reached a specific result.

⁵⁷ The Working Group recognised the significant impact of AI systems 'hallucinating' outputs. 'Hallucinations' can arise when an AI model attempts to generate evidence to support ideas. When using AI systems, users need to consider the situation and whether an AI system is likely to produce an inaccurate result. Hallucinations may take the form of non-existent case examples or links between two pieces of evidence which may not be noticed unless an individual has technical knowledge of the relevant area allowing them to accurately identify the mistake.

This chapter addresses the collection and production of electronic evidence by means of AI systems. It maps the challenges presented by AI systems when used for the collection and production of e-evidence, such as opacity versus algorithm transparency, potential for bias versus non-discrimination, ability to fabricate evidence versus authentication and integrity of evidence.

Having this picture and considering that the collection of evidence from or through AI systems presents another layer of challenges when compared with the collection and production of ordinary e-evidence, this chapter maps the applicability of international instruments whose due process principles and rules would apply to e-evidence collected, processed or analysed by AI systems.

The chapter also explores which conditions and safeguards would apply to e-evidence in the context of AI systems, highlighting concerns around privacy and data protection as well as whether other appropriate conditions and safeguards in this context, in light of existing documents: treaties, domestic legislation or studies.

The chapter also examines the applicability of the provisions on the collection of electronic evidence under the Convention on Cybercrime to AI services.

The requirements for the admissibility of this e-evidence collected and/or processed or analysed by AI systems and further gaps and challenges for the e-evidence in this context are also addressed in this chapter.

During 2025, agentic AI began to emerge more prevalently within digital environments. The introduction of agentic AI represents a new level of complexity for traditional structures of criminal procedure. AI agents may operate across diverse virtual environments—such as databases, digital platforms, social networks, or financial systems—and are not limited to processing information but are also capable of generating concrete actions within those environments. As was discussed in chapter 4, such systems could autonomously perform tasks related to criminal investigation and even collaborate across different stages of the criminal process, thus requiring a careful legal assessment of the limits, conditions, and safeguards governing their use.

In this sense, agentic AI could profoundly transform criminal proceedings and pose particularly significant challenges for contemporary legal systems. This development does not merely involve the incorporation of new technological tools; rather, it constitutes a structural shift that affects the manner in which facts are investigated, evidence is produced and evaluated, criminal responsibility is attributed, and procedural decisions are adopted within the criminal process.

5.1 The collection of e-evidence by or by means of AI systems

5.1.1 Technical advantages and capabilities

As discussed in Section 4.2 above, AI systems are increasingly used to support investigative authorities in the collection, processing, production and analysis of electronic evidence. Within the normative framework of the Convention on Cybercrime and its Second Additional Protocol, AI systems may function either **as direct tools for the collection of e-evidence** or **as analytical systems operating on lawfully obtained data**.

5.1.2 Critical Challenges and Risks to Human Rights

The use of AI systems offers many benefits for the collection of electronic evidence in criminal cases, but there are drawbacks.

- **Opacity of AI systems:** The commonplace description of artificial intelligence as a “black box” is often too imprecise in the criminal justice context. The metaphor suggests a single and undifferentiated problem of unknowability. Yet the expert literature indicates that opacity in AI systems is plural rather than unitary. It arises from distinct sources, each with different implications for due process, evidential scrutiny, and judicial reasoning. Burrell’s influential analysis⁵⁸ remains foundational in this regard: she identifies opacity as arising, first, from intentional secrecy (i.e., proprietary secrecy)⁵⁹; secondly, from the technical difficulty faced by those without specialist knowledge;⁶⁰ and, thirdly, from the intrinsic difficulty of rendering some machine-learning processes into forms of explanation intelligible to human decision-makers.
- **Risk of Bias:** Training data may be historically skewed, incomplete, or reflective of earlier institutional practices such as uneven policing, charging, or supervision patterns — this is the problem of data provenance and data quality. In extreme cases, if training data is tainted with prejudice, the system’s outputs will be unreliable. A similar problem can arise where a model developed in one institutional context is transposed into a different criminal justice environment, importing assumptions that were never validated for that setting. Underlying all of these risks is a more fundamental governance failure. Bias in AI-assisted decision-making is not simply a technical defect; it is the product of a structural disconnect between the development process and the ultimate use context. Where those who design and train a system lack adequate information about — or accountability to — the environment in which it will operate, the system’s inherent limitations will remain opaque to end users. Practitioners may rely on outputs without understanding that the model was never calibrated for their jurisdiction, their population, or their normative framework. This problem is compounded by the multilayered architecture of modern AI systems: a deployed tool typically comprises a foundation model, fine-tuning layers, proprietary scoring logic, and integration components, each potentially developed by different actors under different assumptions. Bias may enter or be amplified at any of these layers, making it difficult — and in some cases impossible — to trace the provenance of a skewed result by the time it reaches a decision-maker]
- **Ability to fabricate evidence:** The use of AI systems may introduce a risk that evidence does not correspond to reality. Previous chapters have discussed deepfakes, which in some circumstances are difficult to distinguish from real images or videos.
- **Risks to the Protection of Personal Data:** Another crucial aspect concerns the protection of personal data. As discussed in Chapter 4, AI systems may process or incorporate personal data, which requires specific legal safeguards. The Council of Europe provides a

⁵⁸ Jenna Burrell, How the Machine “Thinks”: Understanding Opacity in Machine Learning Algorithms, 3 *Big Data & Society* 1, 2016.

⁵⁹ This aspect is explored further in Rebecca Wexler, Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System, 70 *Stanford Law Review* 1343, 2018.

⁶⁰ The issue of **technical complexity and asymmetries of expertise** arises even where some disclosure is made: judges, counsel, and litigants may lack the statistical or computational competence required to interpret model architecture, feature selection, calibration, false-positive rates, or validation results. In this sense, transparency does not automatically produce intelligibility.

robust regulatory framework for data protection, and comparable safeguards may exist within the domestic legal systems.

- **Agentic AI without human intervention:** This concern is structurally related to the governance failure identified above in connection with bias, but it operates at a different level and with greater severity. Where bias results from a disconnect between the development process and the deployment context — a failure that human oversight could in principle detect and correct — agentic AI systems foreclose that corrective possibility by design. The absence of a human in the loop is not an incidental feature of how a particular system happens to be deployed; it is intrinsic to the operational logic of AI agents. The governance deficit identified earlier — insufficient accountability between those who build AI systems and those who bear their consequences — is therefore not merely reproduced in the agentic context: it is structurally entrenched. Where no human is present to recognize that a system is operating outside its validated parameters, the conditions for rights-compliant use cannot be satisfied regardless of how well the underlying model was designed.]

EXAMPLE - ALGORITHMIC RISK ASSESSMENT IN CRIMINAL JUSTICE: LOOMIS CASE

One U.S. court case on algorithmic risk assessment in criminal justice is *State v. Loomis* (Wisconsin Supreme Court, 2016). In that case, the court upheld the use of the COMPAS algorithm at sentencing but imposed strict limitations to safeguard due process. The decision stressed that COMPAS may function only as a non-determinative, auxiliary tool and that its use must be accompanied by explicit warnings regarding its methodological constraints, including its model: algorithms are permissible within the sentencing framework, but only under non-proprietary opacity, potential demographic biases, and reliance on group-based statistical correlations rather than individualised assessments. While the court rejected the claim that such opacity violates due process, it nonetheless required judges to ground sentencing decisions in independent, reviewable factors. *Loomis* thus exemplifies a constrained compatibility condition that mitigates the risks inherent to “black-box” technologies.

5.2 Legal Analysis and Procedural Framework

5.2.1 Applicability of due process principles and rules

Due process is a foundational principle common to multiple legal systems, ensuring **fairness and justice** whenever the State acts against individual's rights. In criminal law, it operates as a structural safeguard that limits the State's power to define, investigate, and enforce criminal liability, requiring strict adherence to legally established rights, rules, and procedures throughout the criminal justice process.

5.2.1.1 Categories of Due Process in Criminal Law

- **1. Procedural Due Process:** Procedural due process concerns the legal procedures and methods that must be followed, before the State may deprive an individual of a human right, such as liberty, through imprisonment. Its focus lies on the fairness of the process itself.
- **2. Substantive Due Process:** Substantive due process imposes limits on the content of State action by protecting certain human rights from infringement, regardless of procedural compliance. It emphasizes the justice and fairness of the law itself, not merely the procedures used to enforce it.

5.2.1.2 Application of Due Process Principles to Evidence Collection and Handling

- **Application of Due Process to Evidence Collection:** Due process principles apply to both traditional evidence and electronic evidence (e-evidence), ensuring that evidentiary activities comply with legality, fairness, and constitutional safeguards.
- **Legal and Lawful Collection of Evidence:** Evidence should be collected in accordance with legally prescribed procedures.

The principle of Legality is one of the fundamental principles of the rule of law within which the due process must comply.

This guarantee exists in all democratic states and examples of it are in the Brazilian Constitution Article 5, XXII, where the confidentiality of data and telephone communications may only be lifted by court order in accordance with the law⁶¹; in the 5th Amendment to the U.S. Constitution, which affords protections before persons may be deprived of certain rights⁶², and in Article 52 of the EU Charter of Fundamental Rights stating that any limitation on the exercise of the rights and freedoms recognised by the Charter must be provided for by law⁶³.
- **Core Procedural Guarantees in Evidence Handling**
 - **Fair Procedures:** The right to a fair trial before an independent and impartial judge, as well as public hearings to ensure transparency. These guarantees are reflected in U.S. constitutional law, Article 6 of the European Convention on Human Rights, and Article 15 of the Budapest Convention.
 - **Integrity and Reliability of Evidence:** The methods used to collect evidence must preserve its authenticity and reliability. Any compromise to evidentiary integrity undermines due process by risking decisions based on distorted or corrupted facts.
 - **Adversarial Proceedings and Full Defence:** Due process requires that the defence be afforded a genuine opportunity to examine, challenge, and contest the collection, preservation, and analysis of evidence, ensuring equality of arms and the fairness of the trial.

5.2.1.3 Other applicable principles relevant to electronic evidence

⁶¹ [Brazilian Federal Constitution, Article 5, XII](#) – The confidentiality of correspondence, telegraphic communications, data communications and telephone communications is inviolable, except, in the latter case, by court order, in the circumstances and in the manner prescribed by law for the purposes of criminal investigations or criminal proceedings.

⁶² [U.S. Constitution -Fifth Amendment](#) - No person shall be held to answer for a capital, or otherwise infamous crime, unless on a presentment or indictment of a Grand Jury, except in cases arising in the land or naval forces, or in the Militia, when in actual service in time of War or public danger; nor shall any person be subject for the same offence to be twice put in jeopardy of life or limb; nor shall be compelled in any criminal case to be a witness against himself, nor be deprived of life, liberty, or property, without due process of law; nor shall private property be taken for public use, without just compensation.

⁶³ [Charter of Fundamental Rights of European Union](#). 18.12.2000/C 364/01 Official Journal of the European Communities. Article 52, Scope of guaranteed rights - 1. Any limitation on the exercise of the rights and freedoms recognised by this Charter must be provided for by law and respect the essence of those rights and freedoms. Subject to the principle of proportionality, limitations may be made only if they are necessary and genuinely meet objectives of general interest recognised by the Union or the need to protect the rights and freedoms of others.

Regarding **electronic evidence**, in addition to the principles mentioned above, several principles are amplified due to its volatile nature, volume, and the intimate personal information it often contains.

5.2.1.3.1 Privacy⁶⁴ and Proportionality

- **Privacy:** The collection of e-evidence (e.g., data from phones, computers, or cloud servers) almost always implicates the right to be free from arbitrary and unlawful interference with privacy (e.g., ICCPR Art. 17, the Fourth Amendment in the US, Article 8 of the ECHR).

Application: Any measure to collect digital data must be *necessary* and must have a legal basis that is *accessible* and *foreseeable*.⁶⁵

- **Proportionality:** The collection must be proportionate to the legitimate aim of the investigation. Authorities should not seize more data than is reasonable or necessary for criminal proceedings. This is a key safeguard, especially when handling large amounts of personal data on a digital device.

In Brazil, proportionality is derived from the principle of due process and developed through Supreme Court case law.⁶⁶

Article 15 of the Budapest Convention and Article 13 of its Second Additional Protocol requires that each party ensure that its domestic powers and procedures, including for the collection of electronic evidence, provide for the adequate protection of human rights and incorporate the principle of proportionality as implemented in accordance with relevant principles of domestic law, balancing public security needs with individual rights, such as freedom from arbitrary and unlawful interference with privacy.

EXAMPLE - PRINCIPLE OF PROPORTIONALITY AND THE LIMITS ON THE USE OF DIGITAL SYSTEMS FOR COLLECTING DIGITAL EVIDENCE

The principle of proportionality and the limits on the use of digital systems for collecting digital evidence has already been addressed by State courts in relation to the use of remote surveillance systems in criminal investigations: the Trojan Horse, which was implemented in the procedural laws

⁶⁴ EU Charter of Fundamental Rights. Article 7 Respect for private and family life Everyone has the right to respect for his or her private and family life, home and communications.

⁶⁵ The unlawful acquisition of digital evidence, particularly in violation of privacy, constitutes a significant basis for inadmissibility under the ECHR case law:

- Podchasov v. Russia (2024): The Court ruled that Russian legislation requiring service providers to store all user communication data and provide it to security services, including decrypting end-to-end encrypted messages, violated Article 8. The Court emphasized that such blanket, indiscriminate, and unauthorized access to private data—which constitutes unlawful acquisition—violates the right to privacy.
- Demirhan and Others v. Türkiye (2025): The Court found that convictions based on the alleged use of the encrypted messaging app "ByLock," without proper safeguards and access to the evidence by the defense, violated Article 6 § 1 (Right to a fair trial), highlighting that digital evidence obtained without procedural safeguards is not admissible.
- P.G. and J.H. v. the United Kingdom (2001): Although slightly older, this case is foundational for digital/covert evidence, where the Court found that the use of covert listening devices to record data without a proper legal framework violated Article 8.

⁶⁶ [Brazilian Federal Constitution. Article LIV.](#)

of Germany and Italy for very serious crimes such as terrorism. The constitutional courts of both countries have considered certain limits on the use of these systems that would be applicable to AI systems and AI agents.

German case law: In its orders of 7 August 2025, the German Constitutional Court (Bundesverfassungsgericht) held that statutory authorisations for state digital surveillance measures—particularly those permitting covert remote access to private IT systems and source telecommunications interception—constitute serious interferences with constitutionally protected fundamental rights and are permissible only within narrowly defined and proportionate limits aimed at averting substantial threats to public security; legislative provisions that failed to meet these requirements were declared incompatible with the general right of personality, in its manifestation as the right to informational self-determination (Article 2(1) in conjunction with Article 1(1) GG), and with the fundamental right to the secrecy of telecommunications under Article 10(1) GG.

5.2.1.3.2 Protection of Personal Data⁶⁷

The protection of personal data should not be mistaken for the right to privacy, as it is another right that ensures the right to privacy under domestic laws of some Parties

Article 8 of the European Convention on Human Rights, signed in 1950, protects the right to private and family life, serving as the foundation of the right to privacy.

Over subsequent decades, as jurisprudence with respect to privacy continued to evolve, on January 28 of 1981, the Council of Europe's Convention 108 was opened for signature with the aim of protecting individuals with regard to the automatic processing of personal data.

Later, in 2000, the European Union Charter of Fundamental Rights enshrined the right to personal data protection in Article 8, clearly establishing the difference from the right to privacy.

The Preamble of the Budapest Convention takes note of the right to the protection of personal data⁶⁸ and Article 14 of its Second Additional Protocol⁶⁹ sets forth data protection safeguards that apply to personal data transferred under the Protocol.

Domestic legislation in many States around the world abide by this principle in the context of electronic evidence. In Brazil, for instance, even before the Budapest Convention entered into force, this principle was already clearly stated in Marco Civil da Internet - Brazilian Civil Rights Framework for Internet,

⁶⁷ EU Charter of Fundamental Rights. Article 8 Protection of personal data 1. Everyone has the right to the protection of personal data concerning him or her. 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified. 3. Compliance with these rules shall be subject to control by an independent authority.

⁶⁸ Convention on Cybercrime (ETS No. 185), Preamble, 11th paragraph): "Mindful also of the right to the protection of personal data, as conferred, for example, by the 1981 Council of Europe Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data;"

⁶⁹ [Second Additional Protocol to the Convention on Cybercrime on enhanced co-operation and disclosure of electronic evidence](#) (CETS No. 224) (Second Additional Protocol), 12.V.2022. Article 14.

where the rules concerning electronic evidence are prescribed, and confirmed by the entering into force of the General Personal Data Protection Law.⁷⁰

5.2.1.3.3 Integrity and Chain of Custody

- **Evidence Integrity/Forensic Soundness:** Due to its fragility, the process of collecting e-evidence must be reliable, repeatable, and accepted in the forensic community. Investigators may choose to use specialised tools⁷¹ to ensure the original data is not modified, deleted, or altered during the collection process.
- **Verified Chain of Custody:** Detailed documentation should be kept, logging every person who handles the digital evidence, the time, and the purpose. This record is critical to prove that the evidence presented in court is the *same* evidence that was collected, without unauthorised change.
- **Proof of Integrity (Hashing):** Hash algorithms are used to create a unique "fingerprint" of a digital file, both before and after it is collected. If the hash values match, it legally verifies that the electronic evidence has not been altered during the collection and preservation stage.

5.2.1.3.4 Admissibility and Relevance

- **Admissibility Rules:** E-evidence, like all evidence, must be relevant to the case. Courts also require that the evidence be demonstrated to be authentic and lawfully obtained. If the collection method is flawed, the evidence may be ruled inadmissible, even if it is relevant.

5.2.1.4 AI systems, Due Process and the Rule of Law

The use of AI systems in the collection and production of evidence opens another dimension in the framework of the legal system.

Principles and rules followed to ensure the validity and integrity of traditional evidence and electronic evidence still apply for the gathering of evidence in the AI systems environment. The novelty is that such use brings new challenges and requires a different attention and special care for the evidence to be considered correctly. Due Process is a procedural safeguard, a specific mechanism within the broader principle of the rule of law, which requires that laws are transparent, predictable, non-arbitrary, and apply equally to everyone.

Different initiatives are emerging related to the assessment of AI Systems, due process and the rule of law:

A recent example is the Venice Commission for Democracy through Law which is the Council of Europe's advisory body on constitutional law. It adopted on 12-13 December 2025, at its 145th Plenary

⁷⁰ Brazilian Civil Rights Framework for Internet. Law n° 12.965/2014, Article 3, and LGPD – General Personal Data Protection Law: Law n° 13.709/2018.

⁷¹ [Essential Guide to Write Blockers in Digital Forensics](#) (2024).

Session in Venice, a Checklist of Rule of Law in the use of AI systems in justice, stating that Human Rights require special attention in the use of AI systems in justice⁷².

According to its para 129, human rights that require special attention in the use of AI systems in justice are:

- data protection,
- equality, non-discrimination,
- privacy,
- access to justice,
- presumption of innocence
- fair trial, including the rights of defense.

According to the Check List, laws governing the use of AI systems in the justice system should provide for a specific assessment of the risks of the AI systems to these rights and should ensure they are tested in safe legal environments (sandboxes) before their general use. All judicial actors should be involved in risk assessment, pre-testing of AI systems and their monitoring.

Consequently, besides the ordinary principles and rules generally applied to obtain electronic evidence, it is therefore essential to evaluate how supervision and auditing mechanisms for AI systems used in criminal investigations can be implemented, so as to guarantee the reliability and verifiability of the results produced.

For agentic AI systems in particular, it will be necessary to ensure human supervision during the evidence collection processes through the use of assessments that are planned before deployment of such systems.

5.2.1.5 HUDERIA - risk and impact assessment of AI systems

[The risk and impact assessment of AI systems from the point of view of human rights, democracy and the rule of law \(HUDERIA\)](#) is an example of a tool developed by the Council of Europe Committee of Artificial Intelligence (CAI) that could serve as a suitable framework for assessing the risks associated with AI systems, particularly with regard to potential interferences with human rights.⁷³ By providing a structured Methodology for identifying and evaluating such risks, HUDERIA offers means to anticipate adverse impacts, enhance accountability, and support the development of adequate safeguards. In this context, its application may contribute to a more transparent and legally robust deployment of AI technologies in criminal investigations. HUDERIA also comprises the HUDERIA Model - Context-Based Risk Analysis (COBRA) Resources which “provides supporting materials and resources (such as flexible tools relevant for different elements of the HUDERIA process, resources of an illustrative nature and scalable recommendations) that can aid in the implementation of the HUDERIA Methodology.”

HUDERIA COBRA Resources guides practitioners to anticipate context-specific impacts—such as privacy concerns, data quality issues, and risks of bias or misuse—while considering the probability and severity of adverse effects. It supports AI teams in evaluating the likely consequences for affected

⁷² [The updated rule of law checklist](#), adopted by the Venice Commission at its 145th Plenary Session (Venice, 12-13 December 2025).

⁷³ There are a variety of risk management frameworks, for example: National Institute of Standards and Technology, *Artificial Intelligence Risk Management Framework (AI RMF 1.0)* (NIST AI 100-1, 2023).

persons and groups, establishing whether AI systems pose risks to human rights, democracy or the rule of law, and determining the proportionality of subsequent safeguards. By enabling initial scoping, risk factor analysis, and triage, the COBRA process ensures relevant risks are addressed early, and effective mitigation plans and stakeholder engagement are implemented.

When applied to AI tools used for the collection or processing of electronic evidence, HUDERIA may function as a standalone, adaptable and non-binding guidance for operationalising the safeguards contained in Article 15 of the Budapest Convention, Article 14 of the Second Additional Protocol, and the risk and impact assessment obligations and, among other things, the principles of transparency, human oversight and accountability established in CETS 225.⁷⁴

Another example of a potential framework for deployment is proposed by the Cybercrime Law Enforcement Practitioners' Network (CYCLOPES) report titled "Deploying AI in Law Enforcement - Practical Guidelines for European Agencies." The report specifies that the deployment of AI by law enforcement agencies (LEAs) is primarily governed by the three core frameworks in the EU context:

DEPLOYMENT OF AI BY LEAs in EU⁷⁵

The deployment of AI by LEAs is primarily governed by three core legal frameworks. The EU AI Act categorises AI systems according to their risk level; prohibited, high-risk, limited-risk, and minimal-risk, and sets legal obligations accordingly. Since many AI applications in policing are classified as high-risk, they are subject to extensive compliance requirements. The General Data Protection Regulation (GDPR) applies where personal data is processed outside the scope of criminal investigations. Within criminal justice, the Law Enforcement Directive (EU) 2016/680 (LED) governs the processing of personal data by competent authorities. This includes core principles such as lawfulness, fairness and transparency, purpose limitation, data minimisation, accuracy, storage limitation, and accountability. These principles must be reflected in every stage of an AI system's lifecycle, from design and procurement to deployment and review. AI deployments must also comply with relevant national laws implementing these frameworks. The principle of proportionality and necessity is especially important when deploying intrusive technologies such as facial recognition.

5.2.2 Applicability of conditions and safeguards

When it comes to AI systems collecting and processing electronic evidence, due process rules and principles take on another dimension, presenting an additional layer of challenge that must be addressed to ensure the evidence is used and evaluated correctly.

With the aim of overcoming these challenges, legislators and regulators around the world are attempting to establish rules that provide transparency in the use of AI systems in the collection and production of evidence:

- 1. Council of Europe's Framework Convention on Artificial Intelligence and Human Rights, Democracy, and the Rule of Law. CETS 225⁷⁶;

⁷⁴ The Committee of Ministers approved the HUDERIA Model COBRA on 25 February 2026 and the Huderia Methodology on 26 February 2025. See [HUDERIA Methodology and Model](#) (2026).

⁷⁵ SCHEER, Gunhild. [Deploying AI in Law Enforcement - Practical Guidelines for European Agencies. Report provided to CYCLOPES Project](#). Cybercrime Research Institute, 2025, p. 4.

⁷⁶ These include human dignity and individual autonomy; transparency and oversight; Accountability and responsibility; Equality and non-discrimination; Privacy and personal data protection; Reliability and Safe innovation.

- 2. The UNGA Resolution 78 265, adopted by the UN General Assembly on 21 March 2024⁷⁷;
- 3. The European Union Artificial Intelligence Act⁷⁸,

Existing due process principles and rules apply to the use of AI systems; however, some countries may decide to provide guidance as to how such principles should be applied to the use of AI systems.

Some examples of areas where additional guidance may be helpful include:

- Ensuring that the adoption, deployment and use of AI systems by the judiciary is “transparent in terms of how the system was developed, how it operates, its training data, its limitations (including its margin of error), its capabilities, and the purpose of the systems.”⁷⁹
- Developing specific rules regarding the authenticity, integrity, traceability, and reliability of evidence obtained using AI.
- Addressing the use of deepfakes, such as UNGA Resolution 78 265, Item 6.g, concerning reliable content authentication and provenance mechanisms to enable users to distinguish “authentic digital content and artificial intelligence-generated or manipulated digital content”.
- Ensuring transparency and human oversight of AI systems, as stated in Article 8 of the Council of Europe AI Convention and in Articles 13 and 14 of the EU AI Act.

EXAMPLE - BRAZIL:

Resolution 615, dated 11/03/2025, from the National Council of Magistrates, which governs the use of AI systems by judges, establishing guidelines for the development, use, and governance of solutions developed with artificial intelligence resources in the Judiciary.

In order to avoid evidence contamination, Resolution 615 determines that the Judiciary is prohibited from developing and using solutions that pose an excessive risk to information security, the fundamental rights of citizens, or the independence of magistrates (art. 10).

Therefore, it is forbidden in the Judiciary to use systems:

⁷⁷ It for example encourages all UN Member States to “promote transparency, predictability, reliability and understandability throughout the life cycle of artificial intelligence systems that make or support decisions impacting end-users, including providing notice and explanation, and promoting human oversight, such as, for example, through review of automated decisions and related processes or, where appropriate and relevant, human decision-making alternatives or effective redress and accountability for those adversely impacted by automated decisions of artificial intelligence systems”. See [The UNGA Resolution 78 265, adopted by the UN General Assembly on 21 March 2024](#). Other international organisations also recognise the importance of human oversight, accountability and transparency. OSCE considers them as “non-negotiable” – automation should enhance, not replace, professional judgment. See Organization for Security and Co-operation in Europe, [Making Law Enforcement Fit for the Future: Strategic Priorities for Policing in the Digital Age](#), 15 June 2026, p. 7.

⁷⁸ The EU AI Act in its Article 14 for example requires high-risk AI systems to include effective human oversight mechanisms to ensure that natural persons can supervise, understand, and intervene in the operation of AI systems when necessary. The purpose of these safeguards is to prevent risks to fundamental rights, safety, and fairness by ensuring that important decisions are not made solely on the basis of automated AI outputs. See [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 \(Artificial Intelligence Act\) \(Text with EEA relevance\)](#).

⁷⁹ UNESCO (2025). Guidelines of the Use of AI Systems in Courts and Tribunals. Page 19.

- that do not allow for human review of the results proposed;
- that classify or rank natural persons based on their behaviour or social situation or even on their personality attributes, for the assessment of the plausibility of their rights, judicial merits or testimonies;
- that assesses personality traits, characteristics or behaviours of natural persons or groups of natural persons for the purpose of evaluating or predicting the commission of crimes or the probability of repeat offences in the grounds for judicial decisions;
- that are used for the identification and authentication of biometric patterns for emotion recognition.

5.2.3 AI systems and admissibility of evidence

The use of AI systems in the collection, processing, and analysis of evidence may raise complex questions about its admissibility in judicial proceedings. International bodies have proposed guidelines and principles for the use of AI systems in courts and tribunals, including UNESCO Guidelines for the use of AI systems in courts and tribunals.⁸⁰ To ensure admissibility of evidence that an AI system has analysed or influenced in some way, several conditions may be considered.

- **Judicial control and authorisation:** Any judicial authorization for the use of an AI system should take into account individuals' right to protection from arbitrary and unlawful interference with privacy and any applicable privacy safeguards or personal data protection rights. This is where proportionality or reasonableness between the authorization of a judicial measure for the investigation of serious criminal conduct and the protection of the individual rights comes into play.
- **Transparency and explainability:** If a party seeks to admit an AI system's output as evidence, the inability of a human to explain, replicate, or understand the rationale behind the output undermines its probative value.
 - UNESCO Universal principle number 7. Explainability: Adopt, deploy, and use AI systems that can explain the rationale behind their outputs and decisions and how they use inputs for such purposes. Explainability refers to making intelligible and providing insight –to deployers and users– into the inputs, reasoning, outcomes, and functioning of AI systems, making these elements understandable, interpretable, and traceable for humans.
- **Right to challenge the evidence:** It is necessary for opposing parties in criminal litigation to have the opportunity to examine and challenge an AI system's output if it is relied on as evidence. Depending on the sophistication of the system and its role in the criminal proceedings, transparency requirements for AI systems may differ in scope and rigor. For example, a system functioning primarily as a filter, one where outputs are subsequently

⁸⁰ UNESCO has issued the Guidelines for the Use of AI Systems in Courts and Tribunals and sets out 15 universal principles. See [UNESCO Guidelines for the use of AI systems in courts and tribunals](#) (2025), p. 17. In addition, the [European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment \(2018\)](#) prepared by the European Commission for the Efficiency of Justice (CEPEJ) lists five core principles to be respected in the field of Artificial Intelligence and justice: respect of fundamental rights; non-discrimination; quality and security; transparency, impartiality and fairness; "under user control". The Charter is accompanied by an in-depth study on the use of Artificial Intelligence in judicial systems.

reviewed, authenticated and testified to by a qualified human expert, may satisfy existing admissibility standards through conventional expert testimony without requiring disclosure of the system's underlying architecture. On the other hand, where a system produces conclusions that are not readily reproduceable or explainable by alternative means, courts may require more substantive disclosure akin to reliability requirements in the context of advanced forensic technologies but adapted to account for the unique characteristics of advanced AI systems.

- UNESCO Universal principle number 8. Auditability: Undertake administrative, legal, and human measures to ensure and enable that the AI systems may be audited during deployment.
 - UNESCO Universal principle number 9: Transparent and open justice: Adopt, deploy, and use AI systems that are transparent in terms of how the system was developed, how it operates, its training data, its limitations (including its margin of error), its capabilities, and the purpose of the systems. Inform in a proper and timely manner when and how AI systems are deployed, acquired, and used and how these tools work, especially when decisions made with or based on such tools can affect the rights and freedoms of individuals or communities. Ensure that individuals are made aware that they may challenge such decisions.
- **Integrity and authenticity:** As with all digital evidence, the digital chain of custody should be preserved. It is recommended to take measures to ensure that the evidence has not been tampered with or corrupted.
 - **Standards and certification:** Any AI systems used should comply with technical and legal standards. Ideally, they should be certified or validated by an independent authority to ensure their reliability.
 - **Human oversight:** While AI systems may assist, it should not replace human reasoning in evidentiary analysis and proceedings.
 - UNESCO Universal Principle number 13. Human oversight and decision-making: Judges and magistrates should not delegate any part of their mandate or rely exclusively on AI systems to adopt decisions or automate entire processes that may negatively impact the rights of individuals or communities. They should preserve the possibility of human intervention in all decision-making processes where AI systems are deployed. AI systems must not replace the decision-maker's independent analysis of facts, law, and evidence.

To ensure traceability, transparency, and effective control of evidence, it is highly recommended that the public sector prioritizes the internal development of AI tools within the justice administration or, at a minimum, establishes rigorous procurement guidelines and public auditing frameworks for third-party software. In this context, the aforementioned HUDERIA methodology could be adapted as a universal standard to evaluate these systems, ensuring that due process is never compromised.

5.3 [Applicability of the provisions for the collection of electronic evidence under the framework of the Convention on Cybercrime to AI services](#)

The Convention on Cybercrime was written in neutral technological language (i.e., the definition of "computer systems") to withstand technological innovations over time. It is already adapted to the digital environment and takes into account the characteristics of electronic evidence.

In the context of criminal justice matters, the definitions present in the Budapest Convention extend to the environment of AI systems.

5.3.1 Service Providers of AI services

The Convention on Cybercrime requires that each “Party shall adopt legislation or other measures as may be necessary to empower its competent authorities to” undertake the procedural measures provided in Articles 16 to 21 in Section 2 of Chapter II of the Convention (subject to additional caveats specified below). This means that there must be domestic procedural powers in place that enable a given Party to obtain specified evidence from a “service provider” as defined in Article 1.c (see Chapter 2 of this Mapping Study). The term “service provider” is fairly broad and would apply to telecommunications companies, internet access providers and social media companies among other entities. It would, however, be an overstatement to state that the term extends to any legal entity that uses computer systems to process or store computer data. At present, among the Parties that participated through their experts in the Mapping Study, there is a range of positions regarding whether all or certain “AI service providers” should be regarded as “service providers” under Art.1.c and hence subject to procedural powers.

Articles 16 to 21 of the Budapest Convention specify the procedural powers to be adopted domestically by a Party as follows:

- **Article 16 - Expedited preservation of stored computer data:** the implementation of this measure is by means of an order or in a similar way to a person to preserve specified stored computer data in the person’s possession or control. This data holder may be a service provider, as exemplified in Para. 149 of the Explanatory Report of the Budapest Convention.
- **Article 17 - Expedited preservation and partial disclosure of traffic data:** this measure explicitly mentions the service providers as recipients of an order to preserve traffic data and disclose a sufficient amount of traffic data to enable the identification of the service providers and the path through which the communication was transmitted.
- **Article 18 - Production order:** a production order can be addressed (18.1.a.) to a person in the territory of the Party to submit specified computer data in that person’s possession or control, which is stored in a computer system or a computer-data storage medium and can be addressed (18.1.b.) to a service provider offering its services in the territory of the Party to submit subscriber information relating to such services in that service provider’s possession or control. According to the Guidance Note on Production orders for subscriber information (Article 18 Budapest Convention):
 - A production order is a domestic power.
 - In article 18.1.a, a “person” is a broader concept than “a service provider” but may include a service provider and, it is clear that the person is present in the Party’s territory. The production order is issued and enforceable by the competent authorities in the Party in which the order is sought and granted. The specified computer data sought is not restricted to subscriber information and covers all types of computer data.
 - In article 18.1.b, the scope is restricted to a “service provider” - narrower than “a person” and is restricted to “subscriber information”. The service provider to which the order is issued offers its services in the territory of the Party but may not necessarily be present in the Party’s territory.

- **Article 19 - Search and seizure of stored computer data:** This measure is intended to empower national authorities to search and seize stored computer data in computer systems within the Party's territory. As the data is in intangible form, only the physical medium where it is stored can be seized and taken away, or the data can be copied. In this situation, as the original data remains in the system, it can be rendered inaccessible. It is a specific measure to address intangible computer data and involves the use of force or constraint by law enforcement to obtain the data.
- **Considerations regarding Articles 20 and 21:** It should be noted that two of the Articles pertaining to procedural powers have a different status pursuant to paragraph 3 of Article 14 which “provides for exceptions to the broad scope of application and permits Parties to restrict the scope of more intrusive powers (real-time collection of data under Article 20 and the interception of content under Article 21).⁸¹
- **Article 20 - Real-time collection of traffic data:** This procedural power aims to empower competent authorities to collect or record, in real-time, traffic data, as defined in article 1.d of the Budapest Convention, that is associated with specified communications in its territory transmitted by means of a computer system, through the application of technical means on the territory of that Party. And, to compel a service provider, within its existing technical capability, to collect or record or to co-operate and assist the competent authorities in the collection or recording, in real-time, of such traffic data.
- **Article 21 - Interception of content data:** This procedural power aims to empower competent authorities to collect or record content data, in real-time, of specified communications in its territory transmitted by means of a computer system through the application of technical means on the territory of that Party, or to compel a service provider, within its existing technical capability, to collect or record or to co-operate and assist the competent authorities in the collection or recording of such content data.

5.3.2 Examples for the potential application of the procedural powers ARTS 16, 17, 18, etc.

Example: A domestic expedited preservation order for stored computer data of a user (Article 16 COE CC) may be issued to service providers that host generative AI systems capable of generating content, such as text, code, audio, images and video, whose capabilities were used to create illicit content, such as online child sexual abuse images (see Article 9 for the substantive offence).⁸²

In a similar domestic context, a production order (Article 18.1.a) may be issued to obtain stored computer data from an AI service provider. With respect to Article 18 on Production orders, it should be noted that there may be additional nuances that pertain to “service providers” and “subscriber information”.⁸³

⁸¹ See [T-CY Guidance Note on the Scope of procedural powers and of international co-operation provisions of the Budapest Convention](#) (2023).

⁸² Article 9 (“child pornography”) of the Budapest Convention has a number of nuances relevant to how different Parties implement it. Recent international treaties also address “non-consensual distribution of intimate images” but depending on the circumstances those may or may not fall within the scope of Article 9 for some Parties.

⁸³ See [T-CY Guidance Note on Production orders for subscriber information \(Article 18 Budapest Convention\)](#) (2017) pertaining to jurisdictional questions which indicate a range of positions among the Parties with respect to extra-territoriality).

The domestic procedural powers (subject to the above restrictions or caveats) have implications for the Budapest Convention's international cooperation measures (Chapter III of the Convention). In this regard, there are also certain nuances that should be kept in mind. These pertain to mutual assistance regarding the real-time collection of traffic data (Article 33) and mutual assistance regarding the interception of content data (Article 34).⁸⁴

5.3.3 AI systems assisting the collection and production of e-evidence

It seems evident that the use of AI systems to assist in the collection and production of evidence requires a different approach in safeguarding the process.

Although current procedural rules and the framework of the Convention on Cybercrime do not specifically mention the use of AI systems, there is reason to believe that AI systems, with appropriate safeguards, may be used to facilitate domestic procedural powers.

Moreover, agentic AI systems are having a significant impact on the digital world in general. Given the speed at which developers are creating various generative models and agent-based systems, it is very likely that these systems may be used in the future for the collection of electronic evidence, both in the investigative stage and in judicial proceedings.

The Second Additional Protocol to the Budapest Convention includes important safeguards regarding privacy and the protection of personal data, which outlines provisions that would apply to evidence collected and processed through AI systems that is subsequently transferred under the Protocol. In the AI context, Article 14's provisions on automated decision-making are of particular relevance, which state that decisions producing a significant adverse effect concerning the relevant interests of the individual to whom the personal data relate may not be based solely on automated processing of personal data, unless authorized under domestic law and with appropriate safeguards that include the possibility to obtain human intervention.

5.4 The use of Public-Private Partnerships in AI and the collection of digital evidence

As noted above, the use of AI systems in the collection and analysis of evidence is significantly transforming the work of competent authorities in the field of criminal justice.

However, access to such technologies is often concentrated in the private sector, where technical expertise and innovation capabilities are more readily available, underscoring the need for mechanisms that enable their effective adoption within the public sector.

In this context, public-private partnerships emerge as a key pathway to bridge this gap. As highlighted by the World Economic Forum, the growing trust challenges surrounding artificial intelligence require coordinated responses, where governments provide legitimacy and oversight while the private sector contributes technological expertise and innovation. Such collaborations need to be translated into concrete standards for the use and procurement of AI systems, auditing frameworks, and accountability mechanisms.

⁸⁴ Parties to the Convention may wish to consult the [Treaty Office records](#) regarding the Reservations and Declarations made by some Parties in this respect in the context of mutual legal assistance.

However, for these partnerships to be sustainable, it is essential to ensure that the ethical use of AI systems is grounded in principles of transparency, accountability, and human oversight, with robust safeguards regarding data protection and confidentiality. In this regard, training justice operators in the responsible use of these tools is critical—not only to maximize their benefits but also to manage their risks — ensuring that technological innovation remains aligned with respect for individual rights and due process.

5.5 Challenges by AI systems and criminal justice: a broader perspective

The use of AI systems in criminal investigations may raise certain issues regarding the control of evidence in criminal proceedings. For example, AI systems developed by the private sector are not always willing to disclose their source code to explain how they work or the inputs that feed them. For this reason, it is necessary to consider mechanisms to achieve traceability, transparency, and reliability to ensure compliance with due-process safeguards. To achieve this, it may be advisable for the public sector authorities to consider hiring software developers if they are considering the possibility of developing their own AI systems within the criminal justice sector.

AI systems offer opportunities to modernise justice and judicial systems. They can now assist in data mining, predictive analytics, automated surveillance, facial recognition, and automated classification of digital traces. These tools pose novel challenges for existing legal frameworks, procedural safeguards, and human rights protections.

In several countries varying levels of digitalisation and legal development further complicate AI systems' deployment in the criminal justice sector.

To the extent that existing criminal procedure laws were not originally designed to regulate AI systems or were not written in a technologically neutral manner, it becomes a matter of great interest to examine any gaps when collecting evidence by or through AI systems, and when applying due process principles as well as the relevance of the Budapest Convention framework. From that perspective, this chapter considers a comparative analysis of some few developing countries to get straight of this concern, notably Cameroon, Ghana, South Africa, Egypt, Morocco, Senegal and Côte d'Ivoire.

Potential Regulatory or Legal challenges:

- the absence of specific legal frameworks governing the admissibility, the integrity, the validity and chain-of-custody of AI analysed or AI-processed electronic evidence. Across several countries, criminal procedure rules recognise electronic data and/or records but lack AI-specific provisions (Cameroon's criminal procedure code, Ghana's legal framework includes the Electronic Transactions Act 772 and Data Protection Act 2012, South Africa's Electronic Communications and Transactions Act and the Protection of Personal Information Act - POPIA, Morocco's legal framework under Cybersecurity Law 05-20 and Data Protection Law 09-08, Senegal Cybercrime Law of 2008 or Côte d'Ivoire cybercrime Law of 2013;
- in some jurisdictions, procedural rules or guidance documents are unclear or lacking as to how AI systems can be used by law enforcement or how the AI-generated or AI-processed evidence may be presented in court. The absence of rules specifically addressing AI does not necessarily create a legal or practical gap, as existing rules or guidance documents on electronic evidence may provide an applicable framework. Challenges may nevertheless arise where that framework does not adequately address matters such as the reliability,

transparency of the AI-generated or AI-processed results and the fairness of their use in criminal proceedings;

- insufficient data protection and privacy laws aligned with applicable international instruments.

Electronic evidence encompasses any data stored or transmitted in digital form. AI systems may collect evidence directly (e.g., automated facial recognition) or indirectly (e.g., predictive analytics). In South Africa and Ghana, cybercrime units deploy AI systems to assist in digital investigations. In Morocco and Egypt, cybersecurity strategies increasingly integrate AI-driven monitoring tools. However, legislative frameworks rarely distinguish between human-generated and AI-processed outputs.

In Cameroon, Senegal, Côte d'Ivoire as in many countries around the world, electronic evidence is admissible under general evidentiary principles, but as of yet, no AI-specific admissibility criteria exist.

Nevertheless, Ghana's Electronic Transactions Act and South Africa's Electronic Communications and Transactions Act recognize digital records but remain silent on algorithmic validation standards. Courts therefore apply traditional tests of relevance, authenticity, and reliability without structured evaluation of algorithmic bias or transparency.

Technical and institutional challenges:

- the deficit of vital infrastructure: in some jurisdictions there are no digital forensic laboratories, poor access or quality of internet and lack of data storage systems and these pose ongoing challenges in many countries;
- access to certain specialised AI systems and forensic tools for investigation may be limited for police, gendarmerie and magistrates;
- the need for specialized training on electronic evidence and AI-assisted investigations.

Regarding Human Rights and ethical concerns:

- the use of AI systems in surveillance and criminal investigation raises some principle risks: the violation of due process, the presumption of innocence principle, discrimination and Human Rights issues as we can think about political repression e.g. In fact, due process - encompassing fairness, transparency, the right to confront evidence, and impartial adjudication - is endangered where AI systems are treated analogously to human-generated evidence without legal safeguards;
- the partiality and transparency in AI systems used by security services is still questioned.

The use of AI systems in the criminal justice sector raises important questions regarding transparency, explainability, and the right to challenge evidence. Defense access to algorithmic logic is limited in most jurisdictions. For example, South Africa's judicial practice occasionally relies on expert testimony to explain technical systems, but statutory safeguards for the use of AI in criminal proceedings remain absent. In other countries studied, limited forensic capacity further complicates meaningful review.

General Challenges to electronic evidence

- the volatility and fragility of digital data in poorly secured environments;
- the difficulty in cross-border judicial cooperation, especially when data is/are stored in foreign cloud environments or jurisdictions;
- weak digital chain-of-custody procedures that may undermine the credibility of evidence in the court;
- the lack of capacity of courts to assess the reliability of electronic evidence, particularly when sophisticated AI systems are involved in its generation or processing. It is clear that using AI systems to handle electronic evidence can significantly enhance the efficiency and credibility of criminal justice systems, but strong legal safeguards, technical capacity and ethical oversight have to be provided, looking into the fact that AI systems also raise a range of risks of injustice, abuses and systemic bias.

A straight analysis reveals common challenges like the absence of AI-specific evidentiary standards; limited transparency; insufficient forensic infrastructure; and inadequate judicial training. Cross-border cloud storage further complicates preservation and admissibility of AI-derived evidence. And these factors remain core challenges even though some States have specialised cybercrime Units, accredited laboratories, law enforcement systems. Several developing Countries are still facing resource gaps, limited forensic infrastructure and sparse judicial education on AI systems. These disparities abovementioned influence how courts, judges and specialists evaluate AI outputs. During criminal proceedings, when AI system outputs are introduced as evidence, parties may not have adequate means of assessing such output's reliability or integrity.

Fair answers to such gaps are essential for a fair and digital-ready justice system.

That is why strong consideration should be given to:

- update for some and draw-up for others, Legal Frameworks to address the use of AI systems to process, collect and analyse electronic evidence and secure judicial procedural safeguard, justice and Human Rights;
- strengthen the capacity-building process for Law Enforcement, prosecutors, judges, Magistrates, forensic experts and lawyers;
- promote a strong international cooperation, especially under standards or frameworks like the Budapest Convention and its additional Protocols;
- adopt minimum procedural standards for the use of AI in justice in order to ensure the proper administration of evidence during trial proceedings, looking to transparency, accountability and fairness;
- establish standards for technology transfer and public-private partnerships to access or share technologies and expertise.

AI systems may offer transformative potential for criminal justice systems, but their integration should be accompanied by strong procedural safeguards, human rights protections, and harmonized evidentiary standards. Without such measures, the risk of inconsistent admissibility decisions and due process violations remains significant.

Where legal frameworks need to be adapted to these new technologies, some countries may consider establishing statutory admissibility standards for AI system outputs entered into evidence, including validation requirements and documentation of automated processing steps. Judicial training programs and forensic accreditation mechanisms may be helpful.

6 AI systems and international cooperation

This chapter explores the implications of AI systems for international co-operation in criminal matters within the framework of the Budapest Convention on Cybercrime and its Second Additional Protocol. It addresses both the potential of AI systems to enhance the effectiveness of cross-border co-operation and the legal and practical challenges that may arise from its use. Given that the deployment of AI systems in this area remains at an early and uneven stage, the analysis focuses primarily on prospective applications, emerging practices, and the safeguards required to ensure compliance with human rights, privacy and data protection, and rule of law standards.

6.1 How can AI systems be used for cross-border investigations and the collection of e-evidence?

Authorities responsible for international co-operation in criminal matters operate under persistent capacity constraints, as growing caseloads consistently outpace available resources. Requests for mutual legal assistance and other cross-border co-operation processes are often subject to significant delays, undermining the effectiveness of investigations and diminishing confidence in the reliability of international co-operation mechanisms.

AI systems hold significant potential to transform cross-border criminal justice co-operation, making it more efficient, timely, and accessible. Virtually all forms of international co-operation provided under the Budapest Convention on Cybercrime (Budapest Convention) and its Second Protocol might be affected. While many potential uses of AI systems in this field may mirror those that may be used in domestic criminal proceedings (e.g., collection and analysis of evidence, forensics), there are also unique opportunities where AI systems could be particularly valuable in the context of international co-operation. For example, in responding to a mutual legal assistance request, AI systems could help requested authorities sort large volumes of digital material, identify relevant information, and flag material that may require filtering, redaction, or special protection before transmission, in accordance with applicable domestic law and procedural safeguards⁸⁵.

By embedding AI technologies into existing co-operation mechanisms, international judicial assistance can evolve from an overloaded system into an efficient and resilient framework that strengthens mutual trust, speeds up procedures, and reinforces the credibility of cross-border justice.

Meanwhile, the integration of AI systems into international co-operation mechanisms must be approached with careful consideration of potential risks and in full compliance with existing legal frameworks. In this context, as discussed in Chapter 5, the **HUDERIA methodology**, developed by the Council of Europe, provides a structured approach to evaluating AI systems against the core values of human rights, democracy, and the rule of law, helping to ensure that technological solutions reinforce, rather than undermine, fundamental legal safeguards.

⁸⁵ This could be relevant, for example, to procedural powers applicable to house searches and interceptions, including in domestic cases (see also Chapter 4 on the investigation, prosecution and adjudication of offences).

Finally, although the potential of AI systems to transform international co-operation in criminal matters is considerable, its actual deployment remains limited. In many Budapest Convention parties, AI systems have yet to be integrated into cross-border co-operation processes, and a significant number are still navigating the foundational challenges of basic digital transformation (see, for example, the case study on African countries' challenges in Chapter 5). The sections that follow therefore outline prospective uses rather than established practice, mapping the potential of AI within the co-operation frameworks of the Budapest Convention and its Second Additional Protocol.

6.1.1 Mutual Legal Assistance (Articles 27–31, Convention; Article 10 of the Second Protocol)

AI technologies offer substantial potential to enhance the efficiency of mutual legal assistance (MLA) authorities in managing both incoming and outgoing requests. These technologies can address multiple aspects of the MLA process, improving accuracy, speed, and resource utilization.

- **Case Management⁸⁶:** Although the existence of AI-powered case management systems have not been reported, such systems could assist MLA authorities in triaging incoming and outgoing requests, prioritizing cases based on urgency or complexity, tracking deadlines, and flagging incomplete or missing information. Such systems could ensure that emergency and other critical cases receive timely attention and reduce administrative bottlenecks.
- **Resource allocation:** AI systems can support the strategic allocation of workload across case officers or units, providing data-driven recommendations to optimize resource use. By balancing case assignments effectively, AI systems could contribute to operational efficiency and more consistent service levels across cases.
- **Drafting and Validation of Requests:** AI systems could guide prosecutors and other officials in preparing MLA requests. These tools could assist in populating requests with standard clauses, legal grounds, and necessary attachments as well as in identifying procedural defects in the draft requests. AI systems could also ensure that requests are tailored to the specific requirements of the requested Party, thereby reducing the likelihood of rejected requests or the need for the requested Party to seek additional information. It may be helpful, for example, for an AI system module in this context to draw upon the current Council of Europe Treaty Office's list of declarations and reservations to the Convention on Cybercrime and its Protocols in order to ensure up-to-date information on the Parties' status.
- **Translation:** The increasing prevalence of cross-border crime as well as the use of digital devices that store large amounts of data have led to a corresponding growth in the volume of materials requiring translation, resulting in higher associated costs. Producing accurate translations presents particular challenges in languages with fewer speakers, as well as in

⁸⁶ Major judicial cooperation bodies are currently modernizing their case management systems (CMS) with AI capabilities to handle the complexities of international requests:

[1] Eurojust's Digital Criminal Justice (DCJ) Programme: Eurojust is redesigning its Case Management System to incorporate AI-driven features like automated document processing, summarization, and the identification of links between cross-border cases.

[2] EU Innovation Hub for Internal Security: Eurojust and other EU agencies use this hub to develop "accountable AI" tools for judicial authorities, including natural language processing (NLP) for evidence analysis and legal research in cross-border criminal matters. Interpol's e-MLA Initiative: This project focuses on the digitalization of MLA requests, providing a secure infrastructure that sets the stage for AI-assisted validation and tracking of international legal assistance.

contexts involving slang, regional variations, or terminology that is highly context-dependent. AI-assisted translation, including domain-specific engines, offers a means to enhance both accuracy and cost-efficiency. These technologies can ensure that legal and technical terms are rendered consistently and reliably. Although human translators remain ultimately responsible for the quality of translations, AI systems can substantially improve their productivity. In addition, machine-only translations can provide practical support for prosecution and investigative authorities. Despite the potential efficiencies that could result from the use of AI systems for translations, however, the use of these tools raise two specific concerns: in case of languages less spoken or complex languages, translations may be very inaccurate (e.g. translations from the Estonian language to the Hungarian language). Human oversight by a translator is therefore imperative. Similarly, with regard to languages that are not commonly used, human oversight is necessary to correct mistakes in the translation. Second, the use of online translation tools raises potential confidentiality issues depending on the AI system used. Because of this concern, some States have prohibited the use of online translation tools for any kind of administrative or judicial proceedings. Confidentiality concerns would therefore need to be addressed before allowing use of AI translation systems.

- **Cross-File Analysis:** Where appropriate in the international cooperation context and with appropriate approvals, AI systems could detect connections across multiple cases, identifying common suspects, identifiers, or patterns of criminality. Named Entity Recognition (NER) enables the mapping of relationships between individuals, companies, and other entities, facilitating the detection of organized crime, money laundering, and fraud schemes that span multiple jurisdictions. AI systems may also enhance the capacity of authorities to uncover complex criminal networks (see example). Also here at present, language issues may occur frequently, e.g. names are spelled differently, translation of foreign names may vary etc. The added value of AI enhancement will largely depend on the technical effectiveness of the cross checking. Considering that hits will need human follow-up intervention, any efficiency gains of cross-file analysis will be limited if a large number of hits turn out to be wrong. At present, the effectiveness of cross-file analysis is largely dependent on the data used. In particular, the comparison of fingerprints or other biometric data is complex, and the system to cross check biometric data should be calibrated appropriately. Human oversight, in case of fingerprints, which might as well include oversight by an expert, is considered essential. Cross file analysis may also help authorities to become aware of parallel proceedings initiated at national level and might therefore help to avoid situations were carrying out the requested measure conflicts with the national investigations. Such analysis might be even more important where perpetrators are unknown and simple queries of registries will not bring any results on national investigations.

EXAMPLE - CROSS-FILE ANALYSIS OF MLA FILES

Some countries use cross-case analysis systems to analyse MLA files. In practice, cross-case analysis systems embedded in AI powered software can more effectively link multiple cases to a single organised crime group or the same individual. Examples include multinational investment fraud. When MLA requests from different jurisdictions are linked, potential connections may be communicated to the relevant requesting authorities. At the initial stage of communication, the names of the linked jurisdictions and any personal data are not disclosed until due authorisation is obtained from the originating authority.

- **Knowledge management:** AI-based semantic search and knowledge graph technologies can systematize policies, precedents, and best practices, including country-specific or service-provider-specific procedures. Such tools can guide officers in strategically drafting MLA or direct co-operation requests, leveraging prior experience to improve the effectiveness of requests.

The application of AI systems in MLA processes has the potential to significantly reduce defective requests, accelerate information transmission, and mitigate delays in international co-operation. By supporting case management, drafting, translation, and knowledge management, AI systems could provide comprehensive toolkits to enhance the efficiency and reliability of mutual legal assistance mechanisms.

Nonetheless, human oversight would appear to be essential. The use of AI systems also raises considerable confidentiality concerns for international cooperation that must be addressed.

6.1.2 Expedited preservation (Arts. 29–30 Convention)

Data preservation is highly time-sensitive, and delays can compromise investigations. AI systems could enhance the efficiency and reliability of preservation processes by automatically identifying preservation needs, generating requests and orders from basic information such as IP addresses or user identifiers, and accurately mapping identifiers to the correct service providers. AI systems can also monitor deadlines and renewal requirements and streamline processes within the constraints of domestic legal frameworks. For requested Parties, AI systems could assist in rapidly identifying disclosable data under Article 30, ensuring that relevant information is prepared and made available promptly. These applications could accelerate preservation, secure critical data before deletion, and support timely investigative action.

6.1.3 Direct co-operation (Art. 32, Convention; Arts. 6-7 of the Second Protocol)

Criminal justice authorities could also potentially leverage AI systems in processes involving direct co-operation with service providers. Many of the potential tools described in the mutual legal assistance context could also apply in the direct cooperation context, including those for drafting and validating requests, translating communications, conducting cross-file analysis of information received (see the respective bullet point in Section 6.1.1), and managing knowledge efficiently. The use of AI systems in these areas could improve accuracy, reduce processing time, and enhance the overall effectiveness of direct co-operation with service providers. Given differences in corporate infrastructure in competitive environments, this is one of the areas where public-private partnerships lead to optimal efficiencies.

6.1.4 Emergency Measures (Arts. 9–10 of the Second Protocol)

Emergency situations under Articles 9 and 10 of the Second Protocol require rapid cross-border responses, as delays can have serious consequences. AI systems can support both requesting and requested Parties in managing urgent cases.

For requesting Parties, AI systems can flag indicators of emergency such as threats to life or physical safety, including terrorism or child exploitation, prioritize requests in processing queues and automatically generate requests for urgent mutual legal assistance or emergency disclosure requests. For requested Parties, AI systems can assist in rapidly performing an initial review of whether a request qualifies for expedited treatment under Articles 9 and 10, identifying the relevant data sought, mapping

the data sought to the correct service providers, and preparing any data obtained from a service provider for immediate disclosure, while ensuring compliance with domestic legal requirements and any needed human review or oversight. By streamlining these processes on both sides, AI systems could facilitate the faster mobilization of authorities and more effective responses in critical scenarios.

6.1.5 Functioning of 24/7 Network (Art. 35 Convention)

The 24/7 Network is essential for rapid preservation and assistance. AI systems can augment the work of the 24/7 Network by:

- Providing real-time triage and routing of requests.
- Offering instant translation support.
- Powering knowledge bases and secure-access AI systems for common procedural questions.

6.1.6 Video Conferencing (Art. 11 of the Second Protocol)

AI systems could significantly enhance video conferencing procedures under Article 11 of the Second Protocol. AI-powered transcription and machine-assisted translation facilitate multilingual proceedings, ensuring immediate access to testimony across languages. At the same time, AI systems may help to protect the identities of vulnerable witnesses, experts, and undercover personnel by replacing faces with synthetic renderings, while preserving the evidentiary value of the recordings. Together, with appropriate guidelines, these technologies may support secure and timely witness and expert examinations across jurisdictions.

6.1.7 Joint investigation teams

Cross-border Joint Investigation Teams (JITs) often face significant coordination and communication challenges. AI systems could enhance operational efficiency, information sharing, and investigative insight throughout the lifecycle of a JIT. At present, in the majority of Working Group members, AI systems are not used to support JITs. Only a handful of countries have reported the use of AI systems in isolated cases. More scalable transnational projects appear to be under development or in pilot phases, and other states remain cautious about deploying systems.⁸⁷

While not yet widely used, these tools could also support timely decision-making regarding the establishment of JITs and facilitate coordination across cases, ensuring a more cohesive investigative process. AI systems capable of cross-case analysis and pattern detection could identify links across multiple investigations where the parties have shared their investigative files, uncovering organized crime networks, money laundering schemes, and other serious criminal activity.

Machine translation and AI-assisted e-discovery may enable JIT members deployed abroad to access local evidence promptly, supporting timely analysis and decision-making in joint investigations. Natural language processing could filter and prioritize large datasets, addressing analytical bottlenecks created by massive volumes of stored digital information.

⁸⁷ See e.g., Eurojust, & EU-LISA. [Artificial intelligence supporting cross-border cooperation in criminal justice](#), 2022.

Secure exchange of information may be further supported by a dedicated digital collaboration platform⁸⁸ which may use artificial intelligence, including AI-enabled machine translation of non-operational data⁸⁹. More generally, such tools may provide controlled access, version tracking, and audit logs, ensuring that sensitive materials are transmitted only to authorized personnel while preserving data integrity. Workflow automation features may streamline collaboration by automatically routing documents, assigning tasks, generating alerts, and tracking investigative milestones across participating agencies. Additionally, coordination dashboards and timeline-building tools enhance transparency and operational awareness across the team. In sum, the use of AI systems within JITs may improve operational coordination, reduce duplication of effort, strengthen trust among investigators, and ensure that critical evidence can be analysed and shared both efficiently and securely.

6.1.8 Data protection

AI systems could help operationalise the data protection safeguards under Article 14 of the Second Protocol by supporting data minimisation, anonymisation, redaction, relevance screening, retention rules, and auditability, enabling more consistent application of those standards. Nevertheless, as stated above, compliance processes must not rely solely on automated processing and must remain subject to meaningful human oversight (see Article 14.6 “Automated decisions”). Unlike in the private sector, adoption of such tools within the judiciary and particularly in international co-operation frameworks remains limited, with only a few Parties providing examples of such use (see e.g. European Union Agency for Fundamental Rights, *Digitalising Justice: A Fundamental Rights-Based Approach*, 2025).

6.2 What particular challenges arise in international co-operation on AI offences?

Under Article 25 of the Budapest Convention, member states may require dual criminality for certain forms of mutual legal assistance under the Convention. As member states are still determining whether adopting new criminal laws specifically addressing AI-related offences is needed or deciding to apply older legal provisions to AI-enabled conduct, divergences in criminalization may emerge across jurisdictions. For example, the laws of different member states may vary in how far they attribute criminal liability to harmful outcomes arising from actions involving AI systems (See Chapter 3 discussions for more on this point). These differences may challenge the ability to satisfy the dual criminality requirement.

6.3 What particular challenges may arise in the deployment and use of AI for international co-operation in criminal matters?

⁸⁸ For example, [the joint investigation teams collaboration platform](#). It will support the functioning of joint teams performing international criminal investigations in the EU, as well as globally. See [Regulation \(EU\) 2023/969 of the European Parliament and of the Council of 10 May 2023 establishing a collaboration platform to support the functioning of joint investigation teams and amending Regulation \(EU\) 2018/1726](#).

⁸⁹ See e.g. Art. 4 (14) of the Commission Implementing Decision (EU) 2025/1274 of 30 June 2025 laying down rules for the application of Regulation (EU) 2023/969 of the European Parliament and of the Council, as regards the necessary measures for the technical implementation of the joint investigation teams collaboration platform. Art. 4 (14) reads as follows: “If the ‘machine translation of non-operational data’ functionality uses artificial intelligence, eu-LISA shall inform the JITs collaboration platform users of the use of such technology and of the accuracy of the results.” Art. 4 (13) provides that the ‘machine translation of non-operational data’ functionality may be used in the context of management of files; evaluation of a JIT; and technical and business support.

The challenges arising from the use of AI systems in criminal investigations and prosecutions, as well as in the handling of electronic evidence, identified in Chapters 4 and 5 equally apply when such measures are carried out in the framework of international co-operation in criminal matters. In addition, cross-border co-operation may give rise to further specific challenges which will be highlighted below.

6.3.1 Overall challenges in the Deployment of AI Systems in International Co-operation Processes

As with the use of AI systems in domestic investigations, the deployment of AI systems in criminal justice and international co-operation processes has significant implications for human rights and data protection, requires substantial financial investment, and presupposes a degree of technical literacy that many institutions are still developing. These challenges often involve genuine trade-offs rather than straightforward solutions.

Publicly available AI systems that are available free of charge over the Internet hold considerable promise for automating co-operation processes, in particular commercial LLMs in order to draft requests or for translations. However, their data-processing policies do not, as a rule, offer the safeguards required in criminal justice systems subject to strict data protection, due process and other human rights obligations. Some national authorities have therefore turned to smaller language models deployed in controlled environments, which may offer stronger data security and better legal alignment, though usually at the cost of reduced capabilities. In addition, open-source products are sometimes not transparent (enough) as regards the training methods.

A related consideration is whether to rely on cloud-based services controlled by established technology companies accepting dependence on proprietary systems operated outside their jurisdiction or to invest in open-source or domestically developed systems built on national infrastructure. The latter may provide stronger security guarantees and greater technological sovereignty but generally requires financial and technical capacities that many states currently lack⁹⁰. Despite these challenges, intra-regional cooperation on AI system deployment is already taking shape. Within the European Union, agencies such as Eurojust, Europol, and EU-LISA are developing or deploying AI-supported solutions to facilitate international co-operation, including case-management tools and support for Joint Investigation Teams⁹¹.

6.3.2 Use of AI in drafting mutual assistance request

Different requirements as to human oversight might cause problems, e.g. if the requesting Party submits a machine translation of obviously poor quality. In such a case, the requested Party would likely ask for a correct translation of the request which will result in additional delay of the execution of the request.

6.3.3 Use of AI systems in Obtaining Evidence through Mutual Assistance

⁹⁰ See, the examples of Latin American countries in the Council of Europe, [1st AIAB Report on the Use of Artificial Intelligence \(AI\) in the Judiciary Based on the Information Contained in the Resource Centre on Cyberjustice and AI](#), CEPEJ-AIAB(2024)4Rev5, 28 February 2025, para. 33.

⁹¹ Eurojust and eu-LISA, [Artificial Intelligence Supporting Cross-Border Cooperation in Criminal Justice](#) (Publications Office of the European Union 2022).

The extent and manner in which AI systems are used in criminal investigations, as well as the procedural safeguards governing such use, may differ considerably across jurisdictions. Differences between the evidentiary rules of the requesting and requested States may also give rise to difficulties regarding the admissibility or evidentiary value of material obtained through mutual legal assistance. Such challenges are not new in the context of international co-operation in criminal matters, and competent authorities have traditionally sought to address them through established mutual assistance mechanisms. However, the use of AI systems may introduce an additional layer of complexity. A requesting party may have limited visibility regarding the use of AI systems by a requested party, including the methodologies, algorithms, or data sources used to generate or process evidentiary material. In some Parties, such opacity may create uncertainty regarding the admissibility or evidentiary value of material produced or processed through these systems. In turn, this may affect the level of trust in the evidence produced.

In this context, **Article 27(3) of the Convention** may offer a solution to address these challenges. Pursuant to this provision, mutual assistance requests executed under the Convention shall be carried out in accordance with the procedures specified by the requesting Party, except where such procedures are incompatible with the law of the requested Party.

Accordingly, a requesting State may request that assistance be executed in accordance with the evidentiary rules applicable in its jurisdiction. This possibility may help ensure that evidence obtained through international cooperation remains compatible with the legal requirements of the requesting State, thereby mitigating some of the challenges arising from differences in national evidentiary frameworks in cases involving the use of AI systems.

With regard to executing MLA requests, additional problems may occur if the requested Party uses publicly available AI systems that cannot guarantee the confidentiality of the procedure (see above Section 6.3.1).

6.3.4 Use of AI systems in the Processing of Evidence

The safeguards of Article 14 of the Second Protocol may offer guidance to parties for the protection of personal data more generally, including with respect to AI-related evidence or the use of AI in processing personal data. Article 14, paragraph 1, provides several pathways for the processing of personal data transferred under the Protocol, including (i) existing international agreements establishing a comprehensive framework for the protection of personal data between the parties exchanging data, (ii) the option for parties to mutually determine that the transfer of personal data may take place on the basis of other agreements or arrangements; or (iii) the application of paragraphs 2 to 15 of Article 14. Paragraphs 2 to 15 of Article 14 may therefore serve as a comprehensive framework governing the processing of personal data transferred under the Protocol.

The provisions of Article 14 paragraphs 2 to 15 require that such data be processed for specified law enforcement purposes and not be further processed for an incompatible purpose (paragraph 2), be maintained with accuracy and completeness (paragraph 3), and be retained only for as long as necessary and appropriate (paragraph 5). The framework also provides specific safeguards for sensitive data, such as genetic data, biometric data, or personal data concerning health or sexual life or revealing racial or ethnic origin, political opinions or religious or other beliefs, or trade union membership (paragraph 4). Particularly important in the context of AI use in criminal investigations, paragraph 6 provides that decisions producing a significant adverse effect concerning the relevant interests of the individual to whom the personal data relate may not be based solely on automated processing of personal data, unless authorised under domestic law and with appropriate safeguards

that include the possibility to obtain human intervention. The explanatory report further encourages parties “to take reasonable steps to provide for the quality and representativeness of the data used to develop algorithms and the accuracy of the statistical inferences used, taking into account the specific circumstances and context of processing, including the context of criminal law enforcement.”

Other provisions concern onward sharing of personal data to other authorities of a party and onward transfer of personal data to another state or international organization (paragraphs 9 and 10), transparency and notice (paragraph 11), access and rectification for individuals (paragraph 12), judicial and non-judicial remedies (paragraph 13), and oversight (paragraph 14).⁹² Together, these provisions are designed to provide safeguards for personal data exchanged through international co-operation under the Second Additional Protocol.

Parties could be encouraged to consider appropriate safeguards for AI systems that process personal data. Parties should consider whether its AI systems used for international cooperation for criminal matters meet appropriate confidentiality and transparency safeguards.

6.4 What particular issues arise in the context of territoriality, jurisdiction, sovereignty?

Article 22(1) of the Convention requires States Parties to establish jurisdiction over offences committed within their territory, on vessels or aircraft flying their flag, or by their nationals when certain conditions are met. Paragraph 2 permits additional optional bases of jurisdiction, while paragraph 5 encourages consultation between Parties to resolve conflicts arising from concurrent jurisdictional claims.

Technologies such as cloud computing, may raise a number of challenges for criminal justice, in particular with regard to the applicable law and the jurisdiction to enforce⁹³. The emergence of AI systems, and their increasing reliance on cloud infrastructure, has added new layers of complexity to these issues. Many AI systems function through cloud-based infrastructures that are simultaneously distributed across multiple jurisdictions. While this feature is not unique to AI systems and applies equally to other cloud-based services, the autonomous and data-driven nature of AI systems may amplify existing jurisdictional uncertainties. In addition, AI systems may facilitate the rapid and large-scale commission of offences, which may result in more cases involving multiple countries and requiring broader international cooperation. In this respect, the direct cooperation provisions introduced by the Second Additional Protocol may become increasingly relevant for ensuring effective international cooperation and access to electronic evidence.

It is often unclear where an AI-related offence should be deemed to occur. Different stages of AI development and deployment which in some Parties may give rise to criminal liability of developers or providers of AI systems (see example in Section 4.1.3.1 on Rules on the Liability of Developers and Providers of AI Systems Developed for Criminal Purposes (“Dark AI Systems”)) may take place in different countries: training data may originate or be processed in one State, algorithms or model parameters may be developed or hosted in another, users or perpetrators may be located elsewhere and the resulting harm may be felt in yet another jurisdiction. This dispersion of conduct and effects challenges the traditional notion of *locus delicti* and complicates the determination of which authorities have jurisdiction for investigation and prosecution.

⁹² Paragraph 15 also allows the transferring State to suspend data transfers where there is substantial evidence of a material breach of these safeguards.

⁹³ Criminal justice access to data in the cloud: challenges Discussion paper prepared by the T-CY Cloud Evidence Group, 2015, p. 10).

A striking example is the creation and dissemination of AI generated deepfakes. Such content may be produced using data stored on servers in one country, processed through cloud infrastructure in another, and circulated globally, causing harm in multiple states. This scenario gives rise to overlapping jurisdictional claims, potential conflicts between competent authorities, and even situations in which no State effectively asserts jurisdiction. Article 22(5) of the Convention offers a solution for addressing situations in which more than one Party claims jurisdiction over an alleged AI offence and may become even more important because of challenges caused by AI systems.

To address these challenges, Parties to the Convention should continue to apply Article 22 in a flexible and technology-neutral manner, consistent with the Convention's original intent. Enhanced co-operation mechanisms under Articles 23 to 35, particularly those concerning mutual assistance and spontaneous information sharing channels, can facilitate coordination in cases involving distributed AI systems. Furthermore, policy guidance and practical arrangements clarifying the assertion and exercise of jurisdiction in cloud-based and AI related contexts would contribute to greater legal certainty and efficiency.

Ultimately, while the jurisdictional framework of the Budapest Convention remains sufficiently broad to cover offences involving AI systems and data, its effective application may require adaptive interpretation, early consultation between competent authorities, and close international co-operation. Such measures are essential to prevent jurisdictional conflicts or gaps and to ensure that AI-enabled criminal conduct does not escape accountability due to the transnational nature of cybercrimes.

6.5 To what extent do current international agreements, including the framework of the Convention on Cybercrime, apply? Are there any gaps?

The international co-operation provisions under the Convention apply to investigations and proceedings concerning criminal offences involving computer systems and data, as well as to the collection of electronic evidence of any criminal offence.

As noted in the preceding chapters, the Budapest Convention on Cybercrime is applicable to offences committed by means of AI systems or directed against AI systems. In broad strokes, the procedural powers provided under the Convention remain relevant when applied in relation to AI systems, although some working group members have raised concerns about the scope of terms such as “service provider” in this context. This is due to the underlying benefit of the Convention’s technology-neutral drafting and the fact that AI systems continue to operate on the basis of ‘computer data’ and qualify as ‘computer systems’ within the meaning of the Convention which makes possible measured adaptations of our legal systems to changing technologies.

Nevertheless, as technological developments advance and AI systems increasingly differ in their architecture and functioning from conventional computer systems, the practical application of the Convention may encounter novel challenges. At present, however, no such difficulties have been identified in the practice of mutual legal assistance.

In light of the foregoing, the international co-operation provisions of the Convention remain applicable to international co-operation in cases involving AI offences.

7 Artificial intelligence and its potential uses in crime prevention: prediction, investigation and oversight

7.1 Introduction: AI and the contested scope of crime prevention

In academic literature, AI is increasingly presented as a means of making crime prevention more anticipatory, data-driven and efficient. The applications of AI discussed under this heading are diverse. They include the forecasting of crime locations and patterns, the allocation of police resources, the automated detection of suspicious activities, the analysis of surveillance and communications data, and the identification of persons, objects or relationships within large datasets. Other applications extend beyond predictive policing into digital and forensic investigation, evidence analysis, criminal-network mapping, victim identification, offender management and the possible reduction of recidivism.⁹⁴ Given what we have seen with the other domains of criminal justice addressed previously in this Mapping Study, we should not be surprised that AI is not a single crime-prevention measure, but a collection of technologies which could be incorporated into different institutional practices and stages of the criminal-justice process.

This breadth, however, creates an initial conceptual problem. The literature frequently moves between **prediction, detection, investigation, prosecution and prevention** as though these terms identified the same activity. They do not. A system that estimates the probability of burglary in a particular area produces a prediction; it does not itself prevent an offence. Prevention depends upon the intervention undertaken in response to that prediction and upon whether the intervention reduces crime or harm. Investigation, by contrast, commonly takes place after an offence has occurred. Yet effective investigation may contribute to a broader preventive process by identifying repeat offenders, disrupting criminal organizations, protecting persons exposed to continuing harm, recovering criminal assets and infrastructure, and supporting the lawful prosecution of perpetrators. The relationship between investigation and prevention should therefore be recognized without treating every activity of the criminal-justice system as preventive by definition.⁹⁵

The documents examined in this chapter also differ substantially in methodology, institutional origin and the type of knowledge they produce. Systematic and technical surveys map the algorithms, datasets, crime categories and performance measures used in crime-prediction research. Other studies examine the perceptions of AI professionals, the explainability of predictive models or the conditions affecting public support for police use of AI. Policy documents prepared within European, transregional and international settings translate selected technical and legal concerns into recommendations concerning deployment, training, procurement, international cooperation and fundamental-rights protection. A further body of work approaches the issue through forensic reliability, criminal procedure and systemic legal oversight.⁹⁶ These sources should not be treated as though they

⁹⁴ Fatima Dakalbab et al., "[Artificial Intelligence & Crime Prediction: A Systematic Literature Review](#)," *Social Sciences & Humanities Open* 6 (2022): 100342, 1–2; Ersöz F. et al., 'Artificial Intelligence in Crime Prediction: A Survey with a Focus on Explainability', *IEEE Access* 13 (2025), p. 59646–47, Satyam Srivastav et al., "[Intelligent Justice: AI-Driven Forensics and Legal Process for Criminal Justice Reforms](#)", *International Journal of Law, Crime and Justice* 84 (2026): 100832, 1–2.

⁹⁵ European Crime Prevention Network, *Artificial Intelligence and Predictive Policing: Risks and Challenges* (Brussels: EUCPN, 2022), 5–6; Srivastav et al., "Intelligent Justice," 1–2, 7–10; Cristos Velasco et al., [Artificial Intelligence and Organised Crime](#) (Madrid: EL PAcCTO 2.0, 2024; updated August 2025), 6–7, 42–71.

⁹⁶ Haya H. Tarawneh et al., "[Artificial Intelligence as a Tool to Enhance Social Interventions in Reducing Crime](#)," *Frontiers in Artificial Intelligence* 8 (2025): 1661266, 1–2.; Ahmet Guler, Sedat Kula, and Kaan Boke, "[Examining Public Support for AI in Policing: The Role of Perceived Procedural Justice](#)," *Police Practice and Research* 26, no. 6 (2025): 673–95; Brenda McPhail, Marcus Pratt, and Jagtaran Singh, *AI and Systemic Oversight Mechanisms in Criminal Justice*, Paper 5 in the LCO AI in Criminal Justice Project (Toronto: Law Commission of Ontario, 2025), 8–12.

provide interchangeable evidence. They are presented here for informational purposes only. The purpose of this chapter is to gather together different sources of information for future consideration by the readers of this Mapping Study. It is not to promote one interpretation of one study over another, but to synthesise themes arising in order to make future use by policy makers and practitioners more user friendly.

The various reports which are cited in this chapter range from model-performance studies, professional surveys, public-opinion research and law-reform reports answer different questions.

The technical reviews focus on findings that machine-learning systems can identify correlations, classify recorded incidents and may have the ability to forecast certain spatial or temporal patterns. Professional and policy documents identify potential uses and possible institutional barriers. None of these findings, considered alone, establishes that an AI-assisted intervention causes an equitable and sustainable reduction in crime. The European Crime Prevention Network distinguishes in this regard between the assessment of risk and the reduction of risk, suggesting that a system may predict a recorded pattern with some accuracy while the resulting police intervention remains ineffective, disproportionate or counterproductive.⁹⁷

The second guiding question of this Chapter—what issues arise with respect to human rights, justice and the rule of law—cannot consequently be addressed by adding a list of ethical concerns after the technological benefits have been identified. These considerations arise throughout the process by which an AI output becomes state action.

International law provides an additional framework for assessing these uses. Where AI is deployed by public authorities, the resulting collection, processing or disclosure of information, surveillance, identification, investigative action or interference with individual rights may engage international human-rights obligations, including those related to arbitrary and unlawful interference with privacy, equal protection of the law and non-discrimination, liberty, fair trial and effective remedies. The International Covenant on Civil and Political Rights provides a central treaty framework in this regard. Where AI-assisted crime prevention involves cross-border access to electronic evidence or cooperation between authorities, applicable international cooperation instruments, including the Budapest Convention on Cybercrime and its Second Additional Protocol, may also be relevant. The legal significance of particular regional instruments and non-binding frameworks should be distinguished from obligations applicable under treaties.

The cumulative use of AI across investigation, arrest, bail, prosecution, sentencing and corrections raises systemic questions that require examination, technical explanation, individual litigation and/or other processes that result in considered policy positions which need to reflect individual countries legislative frameworks. .⁹⁸

This chapter therefore presents a broad but differentiated understanding of AI-assisted crime prevention. AI may contribute to prevention through forecasting, detection, investigation, disruption of criminal networks, victim protection and measures directed toward reducing reoffending. Its value may be assessed across the complete chain connecting **data, model, institutional decision, intervention and social outcome**. The central argument is that human rights, procedural justice and rule-of-law

⁹⁷ European Crime Prevention Network, *Artificial Intelligence and Predictive Policing*, 10–16; Dakalbab et al., “Artificial Intelligence & Crime Prediction,” 1–3; Ersöz et al., “Artificial Intelligence in Crime Prediction,” 59646–52.

⁹⁸ McPhail, Pratt, and Singh, *AI and Systemic Oversight Mechanisms*, 8–12, 30–57. The report emphasizes the limitations of relying upon criminal courts and individual proceedings to address systemic AI risks and proposes coordinated legislative, regulatory, judicial and institutional oversight.

safeguards are not external limitations imposed upon otherwise effective technologies. They are conditions of reliable, legitimate and sustainable crime prevention. The analysis proceeds by defining the relevant prevention continuum, presenting some of the existing evidence supporting claimed applications, examining the principal rights and justice concerns, and considering the potential need for coordinated and systemic oversight.

7.2 Defining the field: from predictive policing to a prevention continuum

7.2.1 Predictive policing as the dominant model

Much of the literature on AI and crime prevention is organized around **predictive policing**. In its narrowest form, predictive policing uses historical and other relevant data to estimate the probability that crime will occur at a particular place or time, that a particular type of offence will occur, or that an identified person will become involved in criminal activity. The technical surveys reviewed for this chapter reflect this emphasis. Dakalbab et al. map a literature dominated by crime-pattern analysis, spatial forecasting and classification, while Ersöz et al. identify spatiotemporal hotspot prediction as the most common objective in their more recent survey.⁹⁹

The European Crime Prevention Network (hereinafter also referred to as “EUCPN”) provides a functional classification. It distinguishes **area-based prediction**, which estimates where and when crime is more likely to occur; **event-based prediction**, which estimates what type of offence may occur; and **person-based prediction**, which assigns risk to an individual. In each case, data are collected and modelled to generate probabilities intended to inform police decisions.¹⁰⁰

Predictive policing also illustrates why prediction can be distinguished from prevention. A model does not intervene in the world merely by generating a risk score or identifying a hotspot. Its output becomes potentially preventive only when an institution acts upon it—for example, by deploying patrols, altering the physical environment, warning potential victims or directing non-police social services toward a location. Whether the resulting action reduces crime depends upon the type and quality of the intervention, not solely upon the accuracy of the forecast. Accordingly, the appropriate term is often an “*AI-assisted preventive contribution*”, rather than prevention by AI itself.

7.2.2 A broader prevention continuum

A narrow focus on predictive policing omits many of the AI applications now associated with crime control. The reviewed materials indicate that AI may enter the criminal-justice process at several distinguishable stages.

First, AI may support **anticipation and strategic planning** through hotspot analysis, crime forecasting, threat assessment and the allocation of preventive resources. Second, it may support **detection and immediate intervention** by identifying anomalous financial transactions, cyber intrusions, weapons, suspicious digital activity or patterns in surveillance material. These applications are closer to prevention where they permit an intervention before harm occurs or while an offence is in progress.

⁹⁹ Dakalbab et al., “Artificial Intelligence & Crime Prediction,” 1–2; Ersöz et al., “Artificial Intelligence in Crime Prediction,” 59652–64. Dakalbab et al. review 120 studies and identify 23 forms of crime analysis and 64 machine-learning techniques; Ersöz et al. report that spatiotemporal hotspot prediction constituted the largest category of prediction objectives in their 142-study review.

¹⁰⁰ European Crime Prevention Network, *Artificial Intelligence and Predictive Policing*, 5–6. The paper describes predictive policing as a process of data collection and modelling and distinguishes area-, event- and person-based predictions.

Third, AI may assist **criminal investigation** through biometric comparison, digital-evidence triage, communications and network analysis, forensic classification and crime-scene reconstruction. Srivastav et al. broaden the field in precisely this direction, situating predictive policing alongside facial and voice recognition, DNA and fingerprint analysis, evidence sorting and legal-process applications.¹⁰¹

Fourth, as we have seen in Chapter 4 above, AI may enter **prosecution and adjudication**, including through evidentiary analysis, legal-document review, disclosure management, risk assessment and decision support. These functions should not ordinarily be described as direct crime prevention. They nevertheless form part of the institutional process through which investigative knowledge is tested and converted into legal consequences.

Finally, AI may be associated with **tertiary prevention** following conviction or other intervention. Possible applications include victim-protection measures, offender-needs assessment, supervision, rehabilitation and efforts to reduce recidivism.. Their inclusion is relevant because crime prevention is not confined to preventing a first offence; it also includes reducing repeated and continuing harm.

This continuum does not erase the distinctions among its stages. Prediction, detection, investigation, adjudication and rehabilitation engage different data, legal powers, standards of proof and human-rights concerns. Its purpose is instead to identify the pathways through which AI may contribute to a broader preventive process.

7.2.3 Investigation as part of the preventive architecture

Investigation generally begins after an offence has occurred, but it may have preventive effects. The identification of repeat offenders, the disruption of an organized criminal network, the seizure of illicit assets, the recovery of compromised data, or the protection of a victim from continuing exploitation may reduce the likelihood or scale of future harm. Investigation may also generate knowledge about emerging criminal methods that can inform subsequent preventive measures.

The causal relationship should not be assumed. A system that processes evidence more quickly may increase investigative efficiency without reducing crime. Greater detection may increase recorded crime while leaving underlying incidence unchanged. Wrongful identification or intrusive investigative practices may damage community trust and reduce cooperation with police. The relevant question is therefore whether AI-assisted investigation contributes, through lawful and reliable institutional action, to disruption, protection, deterrence or reduced reoffending.

The distinction can be represented as a sequence:

AI-assisted analysis → investigative or operational information → institutional intervention → reduction of future harm.

Each link requires separate assessment. A technically successful output may fail to produce an effective intervention, while an operationally effective intervention may still be unlawful or disproportionate.

7.2.4 AI as both a preventive and criminogenic capacity

¹⁰¹ Srivastav et al., "Intelligent Justice," 1–2, 6–10. The article discusses predictive policing, video surveillance, digital-evidence processing, biometric comparison, forensic analysis and rehabilitative applications within a broad account of AI in criminal justice.

The EL PACCTO report examines AI as a technology used by criminal actors as well as justice and security institutions. Consistent with what was outlined in Section 4.1.3 above, the EL PACCTO report describes AI-assisted fraud, phishing, impersonation, synthetic evidence, malware, victim identification and the optimization of criminal logistics. It also maps police and judicial applications involving criminal-intelligence analysis, evidence correlation, surveillance data and international cooperation.¹⁰²

AI and crime prevention should therefore be understood within an **adversarial environment**. Criminal groups may use AI to lower technical barriers, personalize deception, expand the number of potential victims, conceal activity and adapt to police practices. Public authorities may use AI to identify these activities, trace networks and protect victims. The preventive problem is not static: the technology changes both the threat and the institutional response.

Three categories assist in clarifying this relationship. **AI-dependent crime** refers to conduct that could not occur in substantially the same form without AI. **AI-enabled crime** refers to established offences whose speed, scale, accessibility or sophistication is increased by AI (see Chapter 3 above, and in particular 3.2.2.1 “AI systems as instrumentalities of crime”). **AI-evidenced crime** (explored in Chapter 4) concerns cases in which AI-generated or AI-processed material becomes relevant to detection, investigation or proof. These categories may overlap, but they prevent the use of AI as an instrument of crime from being confused with the creation of wholly new offences.

The field is therefore broader than predictive policing but narrower than the criminal-justice system as a whole. For the purposes of this chapter, AI-assisted crime prevention encompasses technologies that contribute to anticipation, detection, investigation, victim protection or the reduction of future offending. Activities occurring elsewhere in the criminal process, which have been discussed more extensively in previous chapters, are included where they materially affect that preventive chain. This functional approach permits the possible benefits of AI to be considered without assuming that prediction, increased detection or more efficient processing is equivalent to a demonstrated reduction in crime.

7.3 Claimed uses of AI and the quality of the evidence

The literature identifies a wide range of possible uses for artificial intelligence in crime prevention. AI systems may assist in identifying spatial and temporal crime patterns, allocating police resources, detecting suspicious transactions or digital activity, analysing surveillance material, processing evidence and mapping relationships within criminal networks. The existence of these applications is not seriously in doubt. The more difficult question is whether any conclusions can be drawn to date about their effectiveness. The documents reviewed rely on different forms of evidence, and each supports a different level of claim. Technical performance, professional confidence, institutional adoption and policy recommendation should not be treated as equivalent evidence that an AI-assisted intervention prevents crime.

7.3.1 Technical performance and preventive effectiveness

Dakalbab et al. and Ersöz et al. provide a meta account of the technical crime-prediction literature. Dakalbab et al. review 120 studies published between 2008 and 2021, identifying a diverse field encompassing numerous crime categories, analytical tasks and machine-learning techniques. Their

¹⁰² Cristos Velasco et al., [Artificial Intelligence and Organised Crime](#) (Madrid: EL PACCTO 2.0, 2024; updated August 2025), 6–17, 30–71. The report examines both criminal uses of AI and applications by security and justice institutions, including automated fraud, malware, deepfakes, criminal-intelligence analysis and evidence processing.

assessment concentrates on the algorithms used, the datasets on which they were tested, validation methods and performance measures. Ersöz et al. similarly review a large body of crime-prediction research, with particular attention to more recent models and their explainability.¹⁰³

These reviews claim that machine-learning methods can be used to analyse large volumes of recorded crime data and, in some contexts, identify patterns or classify outcomes with useful levels of technical performance. They do not, however, arrive at a general conclusion that AI reduces crime.

A further evidentiary step may therefore be worthwhile considering. A predictive model could be connected to an intervention, and the intervention could then be evaluated against an appropriate alternative.

The EUCPN recommendation paper expresses a distinction in policy terms by separating **risk assessment** from **risk reduction**. It cautions against treating AI as a panacea and recommends that predictive systems be evaluated with reference to their practical contribution to crime prevention rather than solely through technical measures.¹⁰⁴ This is an important question to address as a whole. . The relevant outcome is not simply whether the model predicts a recorded pattern, but whether the institutional response produces a demonstrable and proportionate reduction in harm.

7.3.2 Professional perceptions and institutional readiness

Tarawneh et al. approach the question from a different direction. Their survey of 170 AI professionals in Jordan and Saudi Arabia examines beliefs about AI's role in reducing crime, the challenges facing implementation and the strategies used to address those challenges. Respondents generally expressed confidence in AI applications, while also identifying obstacles involving expertise, training, infrastructure, privacy and legal accountability.¹⁰⁵

The study is useful as it indicates how AI adoption is understood within two rapidly developing national contexts. It also shows that organizational capacity is a likely factor in affecting whether technological systems can be used effectively. The study does not, however, evaluate an operational AI system or compare crime outcomes before and after deployment. Agreement among AI professionals that a technology is useful does not necessarily establish that it accurately identifies threats or reduces offending.

The study's title also suggests a broader engagement with social intervention than its empirical content provides. Its questionnaire appears primarily concerned with prediction, monitoring, identification and investigative capacity. Family, community and socioeconomic interventions appear in the conceptual framing, but they are not examined. The article consequently illustrates that technologies associated with policing and surveillance are often described as social or preventive interventions without clearly identifying the social mechanism through which crime is expected to decline.

¹⁰³ Dakalbab et al., "Artificial Intelligence & Crime Prediction," 1–3; Ersöz et al., "Artificial Intelligence in Crime Prediction," 59646–47, 59652–68. Dakalbab et al. organize their review around crime categories, prediction techniques, performance metrics, validation and model limitations, while Ersöz et al. add data provenance and explainability to the technical mapping.

¹⁰⁴ European Crime Prevention Network, *Artificial Intelligence and Predictive Policing*, 10–16. The paper recommends evaluation of predictive systems beyond technical accuracy and cautions that their benefits depend upon the interventions undertaken in response to their outputs.

¹⁰⁵ Tarawneh et al., "Artificial Intelligence as a Tool," 1–2, 7–10. The study surveyed 170 AI professionals equally divided between Jordan and Saudi Arabia and measured perceived roles, implementation challenges and coping strategies rather than observed crime-prevention outcomes.

Professional-perception research may therefore be treated as evidence about **institutional readiness and confidence**, rather than as a substitute for operational evaluation.

7.3.3 Explainability and the limits of technical transparency

Ersöz et al. add an important dimension by examining whether crime-prediction models can be understood and explained. They distinguish between models that are relatively interpretable by design and post-hoc techniques used to explain more complex systems. Their survey finds that explicit attention to explainability remains limited, although methods such as SHAP (SHapley Additive exPlanations) and LIME (Local Interpretable Model-agnostic Explanations) have become more common in recent studies.¹⁰⁶

Explainability has the potential to improve the quality of evidence in several ways. It may reveal which variables are driving a prediction, assist developers in identifying errors, and allow police analysts or independent reviewers to determine whether a model is relying on irrelevant or discriminatory proxies. These are meaningful advances over a purely performance-centred approach.

The evidentiary significance of explainability should nevertheless be kept within limits. An explanation of how a model reached an output does not establish that the underlying data provide a representative account of crime. Nor does it demonstrate that the variables identified by the model have a causal relationship with offending. A system may transparently reproduce patterns created by previous reporting and enforcement practices. Similarly, a local explanation generated by a post-hoc technique may provide an approximation of the model's reasoning without fully representing its internal operation.

Explainability therefore supports audit and scrutiny. It can help answer the question of how a model generated a particular output more readily than the questions of whether the output is valid and whether acting upon it is justified.

7.3.4 Investigation and forensic evidence

The standards to which AI systems are held may become more demanding when they are used in criminal investigations or contribute to material presented in court. Srivastav et al. describe applications involving biometric comparison, digital-evidence processing, forensic classification, crime-scene reconstruction and the analysis of large evidentiary datasets. The article identifies a number of central concerns.¹⁰⁷

Forensic applications cannot be assessed simply by asking whether an algorithm performs well on a development dataset. At least four questions must be separated. First, the underlying scientific proposition must be valid: the characteristic being measured must be capable of supporting the inference claimed. Second, the system must be validated under conditions resembling its intended forensic use. Third, the result in the individual case must be reproducible and accompanied by an appropriate account of uncertainty and error. Fourth, the defence and the court must have sufficient information and expertise to test the result (see Chapter 5 for more details).

¹⁰⁶ Ersöz et al., "Artificial Intelligence in Crime Prediction," 59649–51, 59665–68. The authors identify interpretable-by-design models and post-hoc explanation as the two main approaches and find that only a limited portion of the reviewed crime-prediction literature expressly addresses explainability.

¹⁰⁷ Srivastav et al., "Intelligent Justice," 1–2, 7–12. The article maps a broad range of forensic and legal applications and acknowledges concerns involving data integrity, bias, opacity, admissibility and due process, while also making expansive claims concerning improved precision and objectivity.

These questions are particularly important because the techniques grouped under AI-driven forensics vary considerably in scientific maturity. DNA-mixture analysis, facial comparison, voice analysis, behavioural classification and emotion or deception recognition should not be assumed to possess comparable foundations merely because each uses machine learning. The issue of automation must be separated out from the question of reliability.

The Law Commission of Ontario's discussion of earlier failures involving forensic evidence reinforces this point. The report examines how the Goudge Inquiry and Motherisk Commission demonstrate that unreliable scientific evidence may pass through conventional adversarial proceedings and acquire institutional authority before its defects are recognized. The Commission's discussion highlights the benefits of transparent documentation, independent validation, clear communication of limits, expert training and sufficient resources for the defence to challenge scientific claims.¹⁰⁸ AI-assisted forensic evidence should be evaluated within this wider history rather than treated as an entirely new or self-validating form of proof.

7.3.5 Policy mapping and the level of available evidence

The EUCPN and EL PACCTO documents occupy a further evidentiary category. They map technologies, national practices, legal frameworks, emerging threats and recommended safeguards. They help to identify institutional developments and translate technical and legal concerns into policy agendas. EL PACCTO demonstrates the dual-use character of AI: it maps applications by justice and security institutions alongside the use of AI by organized criminal groups to facilitate fraud, impersonation, malware, trafficking and criminal logistics.¹⁰⁹

Taken together, the documents suggest that AI can increase the capacity to process, organize and classify information. The papers present more context-dependent evidence that particular systems can improve operational tasks. They provide less insight into whether AI-assisted interventions consistently produce causal, equitable and sustainable reductions in crime. An appropriate response is neither to dismiss the potential contribution of AI nor to infer preventive effectiveness from technical promise but instead to evaluate each proposed use according to the particular function it performs, the quality of the evidence supporting it and the institutional consequences of acting upon its outputs.

7.4 Human rights, justice and rule of law issues

The use of artificial intelligence in crime prevention may raise questions that extend beyond technical reliability and operational effectiveness (which is consistent with what was explored in Chapter 5). AI systems may affect how police attention is distributed, what information is collected, and how investigative or judicial decisions are made. These consequences engage human rights such as privacy, equality, liberty, procedural fairness and access to justice.

A constructive approach to this human rights discussion may be to examine whether the proposed purpose, data, model and institutional intervention can be reconciled with the standards applicable to the public authority which is considering deploying the AI system. The nature and intensity of the concern will vary according to the use. An administrative system that assists with document organization does not generally present the same risks as continuous biometric surveillance or an

¹⁰⁸ McPhail, Pratt, and Singh, *AI and Systemic Oversight Mechanisms*, 39–42. The paper draws upon the Goudge Inquiry and Motherisk Commission to identify institutional reforms concerning scientific research, documentation, expert standards, judicial gatekeeping, impact assessment and defence access to expertise.

¹⁰⁹ European Crime Prevention Network, *Artificial Intelligence and Predictive Policing*, 3–16; Velasco et al., *Artificial Intelligence and Organised Crime*, 6–17, 30–77. The latter report expressly describes its account as a time-sensitive snapshot and maps both AI-enabled criminal activity and AI applications used by security and justice institutions.

individual prediction of criminal behaviour. The assessment must therefore remain specific to the function performed, the relevant authority and any consequences that may follow.¹¹⁰

7.4.1 Data, feedback loops and equality

AI systems depend upon data that are generated within existing social and institutional environments. In policing, historical datasets may include recorded incidents, arrests, calls for service, complaints, locations and other information produced through previous enforcement activity. These records may not necessarily provide a complete measure of underlying crime. They may reflect differences in victim reporting, police presence, institutional priorities and the likelihood that particular conduct will be detected and formally recorded.

This may be particularly important for predictive policing. Where a neighbourhood has historically received greater police attention, more offences or suspected offences may have been identified there. A model trained on those records may designate the same neighbourhood as presenting greater future risk, leading to further patrols, surveillance or enforcement. The new activity may then produce additional data that appear to confirm the initial prediction. The EUCPN paper identifies this process as a positive feedback loop, thereby presenting a risk that future resource allocation will be based on historical data sets, which remain unexamined.¹¹¹

There is support in much of the literature for careful examination of data provenance, representativeness, error distribution and the effects of deployment. The literature suggests that equality should be assessed throughout the system's operation rather than inferred from aggregate accuracy or from the absence of an explicitly protected characteristic in the input data.

The EU AI Act offers one framework for addressing such concerns and, in Article 5, limits how EU member states may use AI systems for person-based predictions.

The AI Act prohibits the placing on the EU market, the putting into service or the use of an AI system for specifically making risk- assessments of natural persons, in order to assess or predict the risk that a natural person will commit a criminal offence, and based solely on the profiling of a natural person or the assessment of their personality traits or characteristics.

The regulation specifies that this prohibition does not apply to AI systems used to support human assessment of the involvement of a person in a criminal activity, which is already based on objective and verifiable facts directly linked to a criminal activity.

7.4.2 Privacy, surveillance and proportionality

Many AI applications expand the capacity of public authorities to collect, connect and analyse information. Facial recognition, automated licence-plate recognition, drones, social-media monitoring, communications analysis and video analytics can reveal patterns that would be difficult to identify through manual review. These capabilities may support legitimate investigations and assist in locating

¹¹⁰ Cristos Velasco et al., *Artificial Intelligence and Organised Crime* (Madrid: EL PAcCTO 2.0, 2024; updated August 2025), 14–29; Brenda McPhail, Marcus Pratt, and Jagtaran Singh, *AI and Systemic Oversight Mechanisms in Criminal Justice*, Paper 5 in the LCO AI in Criminal Justice Project (Toronto: Law Commission of Ontario, 2025), 8–12, 19–29. The two reports situate AI governance within broader requirements of human rights, democratic values, procedural fairness and risk-sensitive oversight.

¹¹¹ European Crime Prevention Network, *Artificial Intelligence and Predictive Policing: Risks and Challenges* (Brussels: EUCPN, 2022), 10–13. The paper distinguishes algorithmic and data bias and discusses the risk that historical over-policing will reproduce itself through predictive systems and increased enforcement.

victims, suspects or criminal assets. They may also make it possible to observe large numbers of people who are not suspected of wrongdoing.

As discussed in Chapter 5, privacy concerns therefore arise not only when a system produces an adverse decision, but at the earlier stages of collection, retention, comparison and inference. The ability to process information at scale may alter the practical character of surveillance even where each item of data is individually accessible to the police. Information gathered for one purpose may also be combined with commercially sourced or third-party data and subsequently used for a different investigative objective.

A rule-of-law approach requires more than a general assertion that a technology promotes public safety, or indeed protects against arbitrary and unlawful interference with privacy. The use of a technology should rest upon clear legal authority and a sufficiently defined purpose. The degree of intrusion should be necessary and proportionate, or reasonable, to that purpose, with appropriate limits on collection, access, retention, sharing and secondary use. More intrusive applications may require prior judicial or independent authorization, while indiscriminate or continuous forms of surveillance may not be justified merely because they are technically possible or administratively efficient.¹¹²

The EU AI Act is an example of a legal framework that has anticipated potential risks associated with the expanded capacity for data collection. Article 5 of the EU regulation (« AI Act ») prohibits:

- the supply, commercialization and use of AI systems for the evaluation or classification of natural persons or groups of persons over a certain period of time based on their social behaviour or personality characteristics that could lead to detrimental or unfavourable treatment of certain natural persons or groups of persons that is unjustified or disproportionate to their social behaviour or its gravity. The regulation also prohibits that supply, commercialisation, or use of biometric categorization systems that categorised individually natural persons based on their biometric data or infer their race, political opinions, trade union membership (religious, or philosophical views, sex life or sexual orientation, etc.).
- the supply, commercialization and use of AI systems that create or expand facial recognition databases through the untargeted scraping of facial images from the internet or video surveillance.

7.4.3 Transparency, explanation and contestability

Transparency is repeatedly identified across the reviewed documents as a condition of responsible AI. Its meaning should be specified. At least four levels can be distinguished.

The first is **model explanation**: why did the model generate a particular output? The second is **system explanation**: what data, design choices, thresholds and processing stages shaped that output? The third is **decision explanation**: how did a police officer, forensic practitioner or other official use the output? The fourth is **legal justification**: why was the resulting intervention authorized, necessary and proportionate?

Technical explainability is most directly concerned with the first level. As Ersöz et al. demonstrate, interpretable models and post-hoc techniques such as SHAP and LIME may help identify which features contributed to an output. This can assist debugging, audit and scrutiny. They do not

¹¹² European Crime Prevention Network, *Artificial Intelligence and Predictive Policing*, 7, 10–16; Velasco et al., *Artificial Intelligence and Organised Crime*, 16–29. Both documents place privacy and data protection alongside necessity, human-rights assessment, transparency and oversight in the governance of law-enforcement AI.

necessarily disclose why a system was procured, how a decision threshold was chosen, or why an institution considered a particular intervention appropriate.¹¹³

Nor is general transparency equivalent to individual contestability. An individual affected by a system may require more specific information: notice that AI was materially involved, access to relevant data and validation evidence, an explanation suited to the particular decision, and an opportunity to correct errors or challenge the inference before an independent authority.

Commercial secrecy creates an additional difficulty (see 4.2.3.1 on commercialisation risks), e.g. refusal of private entities to disclose certain design or deployment information... Where sufficient scrutiny cannot be provided, the state may face legal challenges to justify reliance on the system.

7.4.4 Due process and forensic reliability

AI may influence a criminal proceeding before an output is formally introduced as evidence. It may determine which persons are investigated, which records receive priority, or which images are identified for further examination. If these uses are not recorded or disclosed, an accused person may be unable to determine how the investigation developed or whether an error affected subsequent decisions.

Where AI-derived analysis is introduced in court, established principles governing scientific and expert evidence remain applicable. Under individual courts' frameworks, the relevant inquiry may not simply be whether the software is described as artificial intelligence or whether it achieved high accuracy in testing. Courts may also require a showing that the underlying scientific proposition is valid; the system has been tested under conditions relevant to its intended use; error rates and limitations have been communicated; and that the result in an individual case is capable of independent examination.

Srivastav et al. identify reliability, auditability, data integrity and admissibility as central issues in AI-assisted forensics. Their discussion also demonstrates the range of applications that may be involved, including biometric comparison, digital evidence, crime-scene reconstruction and evidentiary classification.¹¹⁴ The literature raises the point that these applications should not be assumed to possess equivalent scientific foundations. In particular, automation cannot remedy weaknesses in the underlying discipline..

Due process may also depend upon practical equality of arms (see 5.2.1.2 above). Formal disclosure may have limited value where defence counsel lacks the resources to retain technical expertise, reproduce an analysis or interpret validation evidence. The LCO paper emphasizes that effective oversight may require legal-aid support, access to independent experts and centralized research capacity. It also draws on earlier Canadian inquiries into unreliable forensic evidence to show that adversarial proceedings and judicial gatekeeping do not invariably identify scientific error before it contributes to injustice.¹¹⁵

7.4.5 Procedural justice and public legitimacy

¹¹³ Filiz Ersöz et al., "[Artificial Intelligence in Crime Prediction: A Survey with a Focus on Explainability](#)", *IEEE Access* 13 (2025): 59649–51, 59665–68. The authors distinguish interpretable models from post-hoc explanation techniques and identify limited use of explicit explainability methods in the reviewed crime-prediction literature.

¹¹⁴ Satyam Srivastav et al., "[Intelligent Justice: AI-Driven Forensics and Legal Process for Criminal Justice Reforms](#)," *International Journal of Law, Crime and Justice* 84 (2026): 100832, 7–12.

¹¹⁵ McPhail, Pratt, and Singh, *AI and Systemic Oversight Mechanisms*, 30–47. The paper examines disclosure, evidentiary gatekeeping, the limitations of litigation, the Goudge and Motherisk inquiries, and the benefits of legal-aid and expert support to challenge AI-derived evidence.

The legitimacy of AI-assisted policing is not exhausted by formal legality. Police institutions depend upon public cooperation, including the reporting of offences, participation by witnesses and willingness to share relevant information. Technologies perceived as unfair, secretive or unaccountable may weaken this relationship and thereby undermine the preventive objectives they are intended to serve.

Guler, Kula and Boke's survey explores this concern. In their sample, public support for AI in policing was associated with perceptions of neutrality, trustworthiness and public voice. Knowledge of AI did not have a significant direct relationship with support after these procedural-justice considerations were taken into account.¹¹⁶

The study, nevertheless, identifies a relevant institutional consideration: technological adoption occurs within an existing relationship between police and the public.

Majority support cannot make an unlawful or discriminatory system legitimate, but sustained public opposition may reveal failures of trust, participation or institutional justification that should not be dismissed.

7.4.6 Human oversight and institutional responsibility

Many policy frameworks respond to the risks of AI by requiring a human decision-maker to remain "in the loop."¹¹⁷ This is an important safeguard, but its effectiveness likely depends upon the circumstances. Human review may be largely formal where an official does not understand the system, lacks access to the basis of its output, or works within an organizational culture that encourages reliance on automated recommendations.

Meaningful human oversight requires that the responsible official have sufficient information, training, time and institutional authority to question or reject the output. The official should ideally understand the system's intended purpose, relevant limitations and level of uncertainty. Decisions materially influenced by AI should be documented, and legal responsibility should not be displaced onto the developer, vendor or algorithm, unless there is a clear legal basis outlined in law.

The involvement of private technology providers does not, in itself, displace the State's responsibility for measures attributable to public authorities. Where police or other public bodies rely upon privately developed AI systems in exercising public powers, the State remains responsible for ensuring that the exercise of those powers complies with applicable domestic and international legal obligations.

The EUCPN paper describes a related "many hands" problem: AI systems are commonly developed, procured, configured and operated by multiple public and private actors, making responsibility difficult to locate when harm occurs.¹¹⁸ Clear allocation of duties is therefore required across the system lifecycle.

7.4.7 Civil society participation and the policy process

¹¹⁶ Ahmet Guler, Sedat Kula, and Kaan Boke, "[Examining Public Support for AI in Policing: The Role of Perceived Procedural Justice](#)," *Police Practice and Research* 26, no. 6 (2025): 673–95, especially 687–90.

¹¹⁷ Similarly, Art.14.6 of the Second Protocol, which may apply to personal data transferred under the Second Protocol, requires that decisions producing an adverse affect on an individual may not be based solely on the automatic processing of personal data unless there are safeguards that include the possibility of obtaining human intervention.

¹¹⁸ European Crime Prevention Network, *Artificial Intelligence and Predictive Policing*, 10–12. The paper links algorithmic opacity with a "many hands" accountability problem involving developers, vendors and deploying institutions.

Civil society organizations can contribute to AI governance in several distinct ways. They may identify effects that are not captured by technical performance measures, represent communities exposed to surveillance, test claims concerning equality and privacy, and assist in translating complex systems into questions of public policy. Their participation may be particularly valuable before procurement or deployment of an AI system in the context of policing, when the purpose and limits of a system remain open to revision and when there is sufficient time for their opinion to inform the process.

At the same time, participation does not remove the responsibility of public authorities to make lawful and evidence-based decisions. Consultation cannot replace technical validation, independent oversight or judicial remedy. Nor can every operational detail of an investigation be made public. A workable process must protect legitimate confidentiality while providing enough information for informed scrutiny.

The Council of Europe Framework Convention on AI, as summarized in the EL PACCTO report, places human rights, democracy and the rule of law across the lifecycle of AI systems and provides for risk assessment, safeguards and effective remedies. The same report describes the convention's development as involving government, industry, academic and civil-society participation.¹¹⁹ The LCO paper similarly treats public and expert consultation, public reporting and participation in pre-acquisition and post-acquisition review as components of systemic oversight.¹²⁰ These documents promote an understanding of civil society engagement as part of governance rather than as an external response to completed policy.

7.4.8 Cumulative and systemic effects

A final observation is that AI systems may operate cumulatively across the criminal-justice process yet only be reviewed in isolation on an agency-by-agency or case-by-case basis. A location-based prediction may increase police presence; increased police contact may generate new records; those records may shape investigation, bail or sentencing assessments; and later decisions may become inputs for correctional classification. Each system may be evaluated separately while their combined effect remains unexamined.

This possibility may support a more systemic approach to oversight. Individual legal challenges remain essential, but they may be reactive, expensive and confined to the facts of a particular case. They may also occur only after a technology has been used for several years and may not require wider institutional change. This cumulative theory is further developed in the LCO paper.¹²¹

The principal human-rights and rule-of-law question is consequently not whether AI is inherently compatible or incompatible with crime prevention. It is whether a specific use is supported by reliable evidence, authorized by clear law, proportionate to its purpose, subject to meaningful scrutiny and embedded in institutions capable of identifying and remedying individual and systemic harm. This formulation permits serious consideration of potential benefits while ensuring that the persons and communities most affected remain visible within the policy process.

¹¹⁹ Velasco et al., *Artificial Intelligence and Organised Crime*, 20–29. The report describes the Council of Europe Framework Convention on Artificial Intelligence, Human Rights, Democracy and the Rule of Law and its lifecycle principles, remedies and safeguards.

¹²⁰ McPhail, Pratt, and Singh, *AI and Systemic Oversight Mechanisms*, 48–60. The paper proposes public and expert consultation, public reporting, AI registers and meaningful participation in pre-acquisition assessment and post-acquisition review.

¹²¹ McPhail, Pratt, and Singh, *AI and Systemic Oversight Mechanisms*, 8–12, 30–60. The report argues that case-specific criminal litigation cannot alone address cumulative and systemic AI risks and calls for coordinated legislative, regulatory and institutional oversight.

8 Findings and recommendations

8.1 Findings

8.1.1 Concepts and definitions

As indicated in the Mapping Study's Terms of Reference, the study's definitional starting point is the broad definition of an artificial intelligence system set out in Article 2 of Council of Europe's Framework Convention on AI (CETS 225). This definition is intentionally broad and meant to include a wide range of automated systems, ranging from simple rules-based systems to large-scale generative AI systems.

Therefore, this mapping study focuses on the capabilities of specific functional subsets of AI systems, which include:

- generative artificial intelligence (which includes large language models);
- improved predictive and analytical systems, and
- agentic AI systems.

Accordingly, the Working Group focuses the analysis on the implications of these new capabilities.

8.1.2 "AI systems" and the Budapest Convention

The mapping study then covers key definitions in the Budapest Convention and its Protocols, specifically: "computer system", "computer data", "service provider", "traffic data" and "subscriber information". There was a consensus that it is important to emphasise that "AI systems" are a subset of "computer systems" that have more advanced capabilities and may vary in their level of autonomy¹²² and adaptability¹²³ once deployed. This provides a strong basis to proceed with respect to substantive offences.

The WG reached consensus that an AI provider may be considered a "service provider" under the Budapest Convention if it also meets the requirements provided for in Article 1.c points i) and ii). The implications of this with respect to procedural powers under the Budapest Convention led to a range of interpretations among Working Group members. These different interpretations may have reflected, at least in part, differences in the constitutional foundations and legal traditions of the legal systems represented as well as relevant decisions of superior courts in recent decades. The members remained sensitive to such differences when they next turned to international mutual assistance.

8.1.3 Offences against and by or by means of AI systems

The range of "AI offences" can be distilled into 2 categories:

- (a) offences against AI systems (including their data), and

¹²² Degree to which a system can learn or act without human involvement following the delegation of autonomy and process automation by humans. See [Explanatory Memorandum on the updated OECD definition of an AI System, OECD Artificial Intelligence papers](#), March 2024, no. 8, p. 6

¹²³ AI systems are capable of adapting and learning from data. Some AI systems can generate outputs without these outputs being explicitly described in the AI system's objective and without specific instructions from a human. See *Ibid.* p. 6.

- (b) offences by or by means of AI systems.

The substantive criminal law provisions of the Budapest Convention and of the UN CC addressing such offences generally apply. In this context, criminal offending against AI systems can be understood as intentional acts that compromise the confidentiality, integrity or availability of an AI system or its inputs or outputs. Cybercriminals are targeting vulnerabilities in AI systems. Such cyberattacks can circumvent AI systems' guardrails, causing them to behave differently than intended or expose sensitive data. As illustrated by examples provided by a number of Parties (for example, Australia, France and Canada), existing domestic criminal law provisions may apply to certain AI offences.

The study also documents how "crimes committed by means of AI systems," that is, AI systems as instrumentalities of crime may be prosecuted under the Budapest Convention and the UN CC. This is similar to the Lanzarote Convention where, existing offences remain criminalised by national law in the same way, whatever the means used by sexual offenders to commit them, be it through the use of information and communication technologies such as AI systems, even when the text of the Lanzarote Convention does not specifically mention such technologies. In addition, where an AI system is designed or adapted primarily for the purpose of committing criminal offences, provisions concerning the misuse of devices or tools used for committing offences may also be relevant to address conduct relating to the production, sale, procurement for use, import, distribution, making available or possession of such systems or tools.

Although there might be incidents or even criminal damage or harm that could be caused by AI systems with limited human intervention or direction, any criminal responsibility would presently lie with a natural and/or legal person responsible for or using an AI system or technology. In contrast, difficult questions may arise in the future involving criminal acts carried out by agentic AI systems, given the technology's potential capacity to learn and evolve autonomously and hence go beyond the initial intent of the individual who employed it for a specific criminal purpose. Depending on their level of complexity, AI systems may not always produce fully controllable or predictable results, particularly when AI agent-based systems are faced with high-priority conflicting goals.

Attributing criminal liability to an AI system is not only a question of whether its actions meet each element of an offence, but also whether the AI system possesses even more basic requirements for criminal responsibility, namely *capacity* and *legal personhood*. The majority of experts do not currently believe that AI systems have capacity, that is, "consciousness" and hence do not meet the conditions of *mens rea*. If that changes, there will still remain the distinct question of *legal personhood*. Regardless, as the capabilities of these powerful systems continue to expand, they may raise the prospect of serious harm for which no human can be held criminally accountable.

8.1.4 The investigation, prosecution and adjudication of offences

AI systems now play several roles across the criminal justice process:

- they are targets of crime;
- instruments by which offences are committed; and
- tools used by investigators, prosecutors and courts.

Offences against AI systems present investigative and prosecutorial issues broadly aligned with established cybercrime practice. Existing frameworks concerning computer systems, electronic evidence and procedural powers remain highly relevant.

However, the technical complexity, autonomy and adaptability of AI systems can intensify familiar challenges, especially:

- attribution and intent;
- jurisdictional issues and international crime cooperation;
- volume and complexity of digital evidence;
- complex nature of electronic evidence (ephemeral evidence); and
- distinguishing malicious activity from authorised penetration testing.

Offences committed by or by means of AI systems raise further practical difficulties. Publicly available systems may be misused to facilitate crime, while “dark AI systems” may be designed or adapted to enable unlawful activity with limited or no safeguards. This distinction matters for both identification of the AI system used and the assessment of potential liability for users and, where domestic law permits, developers, providers or distributors.

AI-enabled offending can alter the scale and profile of criminal activity. Examples such as AI-generated phishing, deepfakes and impersonation, AI-assisted fraud, synthetic child sexual abuse material, adaptive malware and the optimisation of online criminal operations indicate that AI systems may lower the technical threshold for perpetrating offences while increasing their speed, volume and sophistication. Consequences include pressure on investigative capacity, difficulties in victim identification (especially distinguishing real from synthetic child sexual abuse material), greater reliance on cross-border cooperation, and more focused prioritisation (for example, in AI-generated or altered CSAM, prioritising cases involving real victims requiring immediate intervention).

For adjudication, courts will more often need to determine whether and how an AI system was used, the degree of human control, the existence of the required mental element, and the provenance of AI-generated or altered material. These issues underscore the need for specialised technical expertise among investigators, prosecutors and judges, alongside due process safeguards on admissibility, chain of custody, disclosure and explainability.

At the same time, AI can support criminal justice authorities. In investigations, it can assist with:

- large-scale data analysis;
- link detection across evidence;
- transcription, translation and documentation;
- reducing exposure to traumatic material.

In prosecution and court processes, it can:

- accelerate document review;
- organise case files;
- support drafting;
- provide translation;
- provide relevant legal information.

These uses can improve efficiency and access to justice, particularly where resources are constrained or digital caseloads are large, provided human-in-the-loop oversight is maintained.

However, risks relating to fairness, privacy and data protection, transparency, explainability and auditability, accountability, over-reliance and automation bias, and fabricated or erroneous outputs can

affect the integrity of proceedings and public trust. These risks require clear governance, transparent and auditable processes, and effective human oversight.

Overall, AI systems do not merely facilitate perpetration of offences or provide new investigative tools. They affect the entire criminal justice lifecycle, from the commission and detection to evidence collection, attribution of responsibility and adjudication.

8.1.5 AI systems and electronic evidence

The study is also specifically mapping the challenges presented by AI systems when used for the collection and processing of e-evidence. Importantly, the study reveals another layer of complexity when mapping applicable international instruments whose due process principles and rule of law considerations that would apply to e-evidence collected and processed by AI systems, including current criminal procedure law and evidence rules in the context of the framework of the Convention on Cybercrime and its Protocols.

Although the use of AI systems offers many benefits for gathering evidence in criminal cases, the study also noted that there are drawbacks. These include:

- the opacity of AI systems (which has many more nuances than the common “black box” label suggests);
- the potential for bias stemming from training data;
- the ability to fabricate evidence;
- the risks to the protection of personal data; and
- the risks of limited or inadequate human oversight and intervention.

The study underscored how due process is a foundational principle common to multiple legal systems, ensuring **fairness and justice** whenever the State acts against an individual's life, liberty, or property. In criminal law, it operates as a structural safeguard that limits the State's power to define, investigate, and enforce criminal liability, requiring strict adherence to legally established rights, rules, and procedures throughout the criminal justice process.

The study considered the core procedural guarantees that must be kept in mind with regard to the handling of e-evidence, including:

- the right to a fair trial before an independent and impartial judge;
- the integrity and reliability of evidence; and
- adversarial proceedings and full defence.

Principles and rules that need to be followed to ensure the validity and integrity of traditional evidence and electronic evidence apply to the gathering and analysis of evidence in the AI systems environment. Against this background, the use of artificial intelligence in the collection and processing of evidence may raise complex issues of admissibility in judicial proceedings, requiring safeguards to ensure that such evidence meets the legal requirements for admission. The novelty is that such use brings new challenges and requires a different attention and special care for the evidence to be considered correctly. The study noted that this area is evolving rapidly.

In this respect the Council of Europe Framework Convention on AI provides a normative baseline for public-sector use of AI. Principles especially pertinent to evidential practice include human dignity and

individual autonomy; transparency and oversight; accountability and responsibility or equality and non-discrimination.

A recent example is the Venice Commission for Democracy through Law which is the Council of Europe's advisory body on constitutional law. In December 2025, it adopted the Checklist of Rule of Law in the use of AI systems in justice, stating that human rights require special attention in the use of AI systems in justice¹²⁴. Also of note were other reports adopted at a global¹²⁵, regional¹²⁶ or national level¹²⁷.

The study also drew attention to the HUDERIA platform which is an example of a tool developed by the Council of Europe Committee of Artificial Intelligence (CAI) that serves as a framework for assessing the risks associated with AI systems, particularly with regard to potential interferences with human rights. The project's extension with its Context-Based Risk Analysis (COBRA) which analyses an AI system's life cycle as applied to the context of criminal justice matters seemed fruitful in the context of criminal justice matters and e-evidence.

The study also undertook an initial analysis of issues arising from "AI service providers" in the context of the procedural powers of the Budapest Convention and its Protocols.

8.1.6 AI systems and international cooperation

AI systems hold significant potential to transform cross-border criminal justice co-operation, making it more efficient, timely, and accessible. Virtually all forms of international co-operation provided under the Budapest Convention on Cybercrime (Budapest Convention) and its Second Protocol might be affected. Given the range of different legal systems of the Parties, their specific domestic laws, the potential differences of supranational law that may impact the relations between certain requested and requesting Parties,¹²⁸ MLA may have some complexities in different circumstances. The integration of AI systems into international co-operation mechanisms require careful consideration of potential risks and any deployment should be in full compliance with existing legal frameworks.

AI systems offer substantial potential to enhance the efficiency of MLA authorities in managing both incoming and outgoing requests. These technologies can address multiple aspects of the MLA process, improving accuracy, speed, and resource utilization. Applications of AI in MLA processes have the potential to significantly reduce defective requests, accelerate information transmission, and mitigate delays in international co-operation. By supporting case management, drafting, translation,

¹²⁴ [The updated rule of law checklist](#), adopted by the Venice Commission at its 145th Plenary Session (Venice, 12-13 December 2025).

¹²⁵ See [UNESCO Guidelines for the use of AI systems in courts and tribunals](#) (2025).

¹²⁶ See European Commission for the Efficiency of Justice (CEPEJ), [European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment \(2018\)](#) or SCHEER, Gunhild. [Deploying AI in Law Enforcement - Practical Guidelines for European Agencies](#) (2025). Report provided to CYCLOPES Project. Cybercrime Research Institute.

¹²⁷ See Resolution 615, dated 11/03/2025, from the Brazilian National Council of Magistrates which governs the use of AI systems by judges, establishing guidelines for the development, use, and governance of solutions developed with artificial intelligence resources in the Judiciary.

¹²⁸ The *sui generis* supranational system of the Law of the European Union, particularly in the areas of law enforcement and judicial cooperation in criminal matters as well as data protection may create different relationships between the requested and requesting Parties to the Budapest Convention and its Second Additional Protocol for Parties that are members of the EU. With respect to the latter see Article 15.1.b and the *Explanatory Report*, Para. 290.

and knowledge management, AI systems could provide a comprehensive toolkit to enhance the efficiency and reliability of mutual legal assistance mechanisms.

With regard to the Second Protocol, for requesting Parties, AI systems may be able to flag indicators of emergency such as threats to life or physical safety, including terrorism or child exploitation, prioritize requests in processing queues and automatically generate requests for urgent mutual legal assistance or emergency disclosure requests. For requested Parties, AI systems can assist in rapidly performing an initial review of whether request qualifies for expedited treatment under Articles 9 and 10, identifying the relevant data sought, mapping the data sought to the correct service providers, and preparing any data obtained from a service provider for immediate disclosure, while ensuring compliance with domestic legal requirements. By streamlining these processes on both sides, AI systems could facilitate the faster mobilization of authorities and more effective responses in critical scenarios.

Finally, the international co-operation provisions of the Convention are equally applicable to co-operation in cases involving offences committed through the use of AI systems, the execution of procedural powers involving AI systems, and the exchange or use of electronic evidence obtained with the involvement of AI systems (subject to the possible caveats identified in earlier chapters).

8.1.7 Artificial intelligence and crime prevention

Artificial intelligence can contribute to crime prevention, but it does not constitute a preventive intervention in its own right. Its most plausible contributions lie in analysing large datasets, identifying patterns and anomalies, supporting the allocation of resources, accelerating investigations, tracing criminal networks, protecting victims and informing measures intended to reduce repeated offending. These functions extend beyond predictive policing into detection, forensics and the broader criminal-justice process. Their preventive value remains conditional upon the quality of the information produced and the institutional action taken in response.

The literature provides considerably strong evidence that AI can enhance information-processing capacity. It is less clear whether it may consistently produce equitable and sustainable reductions in crime. Evaluation of such attempts must therefore follow the complete chain from data and model to decision, intervention and outcome.¹²⁹

Human rights, justice and the rule of law are integral to that evaluation. AI-assisted policing and criminal justice may affect privacy, equality, liberty, procedural fairness and access to justice. These concerns may not be adequately addressed through model-level safeguards or nominal human supervision alone. They require consideration of clear legal authority, function-specific assessment, independent validation, meaningful notice and contestability, effective remedies and public participation proportionate to the consequences of the use.

AI should accordingly be approached neither as a technological solution to crime nor as a uniform threat to rights. Its contribution is contingent and context-specific. The governing question is not simply whether an AI system can predict, detect or classify, but whether its use demonstrably reduces harm through institutions that remain lawful, fair, transparent and accountable.

8.2 Recommendations

¹²⁹ European Crime Prevention Network, *Artificial Intelligence and Predictive Policing: Risks and Challenges* (Brussels: EUCPN, 2022), 14–16; Filiz Ersöz et al., “Artificial Intelligence in Crime Prediction: A Survey with a Focus on Explainability,” *IEEE Access* 13 (2025): 59646–68, <https://doi.org/10.1109/ACCESS.2025.3553934>; Satyam Srivastav et al., “Intelligent Justice: AI-Driven Forensics and Legal Process for Criminal Justice Reforms,” *International Journal of Law, Crime and Justice* 84 (2026): 100832, 7–12, <https://doi.org/10.1016/j.ijlci.2026.100832>.

The following actions may be considered at the level of the Council of Europe and the Cybercrime Convention Committee (T-CY):

- Rec 1 The T-CY may consider preparing a guidance note on the application of the Convention on Cybercrime and its Protocols within the context of cybercrime, electronic evidence and AI systems. This may include clarifying the scope of the procedural powers and international cooperation provisions, including their application to 'service providers of AI systems'. In this respect, the T-CY should also consider conducting a concise survey of Parties to better understand legislative and other criminal justice measures addressing these links, with a view to informing the preparation of the guidance note.
- Rec 2 Given the difference in scope but also the complementarity between the Convention on Cybercrime and its Protocols, and other applicable international instruments, such as the Lanzarote Convention and Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law, Parties – within their respective treaty obligations – the T-CY and the Secretariat should consider promoting cross-collaboration in their implementation, with particular attention to the impacts of AI systems within the purviews of those treaties.
- Rec 3 The T-CY should consider coordinating with the European Committee on Crime Problems (CDPC), given the complementarity of the two Committees' work on artificial intelligence and criminal law, which may include exchanges of information or other forms of co-operation.
- Rec 4 The T-CY may consider undertaking a consultation with industry and service providers of AI services and civil society stakeholders to facilitate co-operation under the Convention on Cybercrime and its Protocols.
- Rec 5 The Council of Europe (T-CY Secretariat and C-PROC) should make available, through the CYBOX platform or otherwise, materials on the links between cybercrime, electronic evidence and AI identified in the present study and stemming from other initiatives, and support related capacity building activities, including on cross-collaboration with HUDERIA's Context-Based Risk Analysis (COBRA) in criminal justice contexts. T-CY members may also wish to share the present study among relevant institutions within their countries.

9 Appendix

9.1 Council of Europe soft-law and guidance documents

- [Overview of the Council of Europe activities in the field of artificial intelligence](#) (2026).
- [HUDERIA Methodology and Model](#) (2026).
- [Feasibility Study on a future Council of Europe instrument on artificial intelligence and criminal law](#) (2020)
- [Recommendation regarding the ethical and organisational aspects of the use of artificial intelligence and related digital technologies by prison and probation services](#) (2024)
- [Discussion paper on a future legal instrument addressing criminal liability related to the use of artificial intelligence](#) (2024)
- [Handbook on Human Rights and Artificial Intelligence](#) (2026)

Cybercrime Convention Committee (T-CY)

- [T-CY Guidance Note on Production orders for subscriber information \(Article 18 Budapest Convention\)](#) (2017)
- [T-CY Guidance Note on the Scope of procedural powers and of international co-operation provisions of the Budapest Convention](#) (2023)
- [Joint statement on AI-generated or altered child sexual exploitation or sexual abuse material by the Cybercrime Convention Committee \(T-CY\) and the Committee of the Parties to the Convention on the protection of children against sexual exploitation and sexual abuse \(Lanzarote Committee\)](#) (2026)

The Parliamentary Assembly of the Council of Europe/ The Parliamentary Assembly of the Council of Europe

- Justice by algorithm – the role of AI in policing and criminal justice systems: Resolution 2342 (2020) and [Recommendation 2182](#)

CEPEJ

- [Information note on the Use of Generative AI by judicial professionals in a work-related context](#) (2024)

Lanzarote Committee

- [Declaration on protecting children against sexual exploitation and sexual abuse facilitated by emerging technologies](#) (2024)
- [Interpretative Opinion on the applicability of the Lanzarote Convention to sexual offences against children facilitated through the use of information and communication technologies \(ICTs\)](#) (2017)

9.2 Bibliography

9.2.1 Academic literature

Blauth T.F., Gstrein O.J. and Zwitter A., '[Artificial Intelligence Crime: An Overview of Malicious Use and Abuse of AI](#)', IEEE Access (2022).

Burrell J., '[How the Machine "Thinks": Understanding Opacity in Machine Learning Algorithms](#)', Big Data & Society, Vol. 3 (2016).

Colombatto C. and Fleming S.M., '[Folk Psychological Attributions of Consciousness to Large Language Models](#)', Neuroscience of Consciousness, Vol. 2024, Issue 1 (2024).

Dakalbab F. et al., '[Artificial Intelligence & Crime Prediction: A Systematic Literature Review](#)', Social Sciences & Humanities Open 6 (2022), 100342.

Dou W.W., Goldstein I. and Ji Y., '[AI-Powered Trading, Algorithmic Collusion, and Price Efficiency](#)', Jacobs Levy Equity Management Center for Quantitative Financial Research Paper, The Wharton School Research Paper (2025).

Ersöz F. et al., '[Artificial Intelligence in Crime Prediction: A Survey with a Focus on Explainability](#)', IEEE Access 13 (2025).

European Crime Prevention Network, Artificial Intelligence and Predictive Policing: Risks and Challenges, Brussels (2022).

Guler A., Kula S. and Boke K., 'Examining Public Support for AI in Policing: The Role of Perceived Procedural Justice', Police Practice and Research 26(6) (2025).

Liang P.P., Zadeh A. and Morency L.-P., '[Foundations and Trends in Multimodal Machine Learning: Principles, Challenges, and Open Questions](#)', Cornell University (2022/2023).

Lohmann A., '[Strafrecht im Zeitalter von Künstlicher Intelligenz: Der Einfluss von autonomen Systemen und KI auf die tradierten strafrechtlichen Verantwortungsstrukturen](#)', Baden-Baden (2021).

McPhail B., Pratt M. and Singh J., AI and Systemic Oversight Mechanisms in Criminal Justice, Paper 5 in the LCO AI in Criminal Justice Project, Toronto (2025).

Mizuta T., '[Can an AI Perform Market Manipulation at Its Own Discretion? – A Genetic Algorithm Learns in an Artificial Market Simulation](#)', 2020 IEEE Symposium Series on Computational Intelligence (SSCI), Canberra (2020), pp. 407–412.

Panebianco, G. '[The Nulla Poena Sine Culpa Principle in European Courts Case Law](#)'. In: Ruggeri, S. (ed), Human Rights in European Criminal Law: New Developments in European Legislation and Case Law after the Lisbon Treaty (Springer 2015), pp. 47-78.

Porębski A. and Figura J., '[There Is No Such Thing as Conscious Artificial Intelligence](#)', Humanities and Social Sciences Communications, Vol. 12 (2025).

Scheer G., '[Deploying AI in Law Enforcement – Practical Guidelines for European Agencies](#)', Cybercrime Research Institute (2025).

Sapkota R., Roumeliotis K.I. and Karkee M., '[AI Agents vs. Agentic AI: A Conceptual Taxonomy, Applications and Challenges](#)', Information Fusion, Vol. 126, Part B (2026).

Srivastav S. et al., 'Intelligent Justice: AI-Driven Forensics and Legal Process for Criminal Justice Reforms', International Journal of Law, Crime and Justice 84 (2026).

Tarawneh H.H. et al., 'Artificial Intelligence as a Tool to Enhance Social Interventions in Reducing Crime', *Frontiers in Artificial Intelligence* 8 (2025).

Wexler R., '[Life, Liberty, and Trade Secrets: Intellectual Property in the Criminal Justice System](#)', *Stanford Law Review*, Vol. 70 (2018).

9.2.2 Council of Europe instruments and other documents

[Additional Protocol to the Convention on Cybercrime, concerning the criminalisation of acts of a racist and xenophobic nature committed through computer systems](#) (ETS No. 189).

[Convention on Cybercrime](#) (ETS No. 185).

[Council of Europe Convention on Laundering, Search, Seizure and Confiscation of the Proceeds from Crime and on the Financing of Terrorism](#) (CETS 198).

[Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law](#) (CETS No. 225).

[Explanatory Report to the Council of Europe Framework Convention on Artificial Intelligence and Human Rights, Democracy and the Rule of Law](#) (CETS No. 225).

[Guidance Note on Production Orders for Subscriber Information](#) (Article 18 Budapest Convention).

[Guidance Note on the Scope of Procedural Powers and International Co-operation Provisions of the Budapest Convention](#).

[Interpretative Opinion on the Applicability of the Lanzarote Convention to Sexual Offences against Children Facilitated through ICTs](#) (2017).

European Committee on Crime Problems, Mapping study based on the Compilation of national responses to the CDPC-AICL questionnaire on artificial intelligence and criminal liability (2026).

[Opinion on Article 23 of the Lanzarote Convention and its Explanatory Note](#).

[Second Additional Protocol to the Convention on Cybercrime on Enhanced Co-operation and Disclosure of Electronic Evidence](#) (CETS No. 224).

[T-CY Mapping Study on Cyberviolence](#) (2018).

[The Updated Rule of Law Checklist, adopted by the Venice Commission at its 145th Plenary Session](#) (Venice, 12–13 December 2025).

9.2.3 International law and European Union

[Charter of Fundamental Rights of the European Union](#), OJ C 364/01, 18 December 2000.

[Commission Implementing Decision \(EU\) 2025/1274 of 30 June 2025 laying down rules for the application of Regulation \(EU\) 2023/969 of the European Parliament and of the Council, as regards](#)

[the necessary measures for the technical implementation of the joint investigation teams collaboration platform.](#)

[Directive 2013/40/EU](#) of the European Parliament and of the Council of 12 August 2013 on attacks against information systems and replacing Council Framework Decision 2005/222/JHA...[Directive \(EU\) 2016/680](#) of the European Parliament and of the Council of 27 April 2016.

[Regulation \(EU\) 2016/679](#) of the European Parliament and of the Council of 27 April 2016 (General Data Protection Regulation).

[Directive \(EU\) 2019/713](#) of the European Parliament and of the Council of 17 April 2019 on combating fraud and counterfeiting of non-cash means of payment and replacing Council Framework Decision 2001/413/JHA.

[Directive \(EU\) 2024/1260 of the European Parliament and of the Council of 24 April 2024 on asset recovery and confiscation.](#)

[Regulation \(EU\) 2023/969](#) of the European Parliament and of the Council of 10 May 2023.

[Regulation \(EU\) 2024/1689](#) of the European Parliament and of the Council of 13 June 2024 (Artificial Intelligence Act).

[UN General Assembly Resolution 78/265](#), adopted on 21 March 2024.

[UNESCO Guidelines for the Use of AI Systems in Courts and Tribunals](#) (2025).

[United Nations Convention against Cybercrime](#) (2025).

9.2.4 Case law

[Demirhan and Others v. Türkiye](#) (2025).

[P.G. and J.H. v. the United Kingdom](#) (2001).

[Podchasov v. Russia](#) (2024).

9.2.5 Institutional and technical reports

[Eurojust and EU-LISA, Artificial Intelligence Supporting Cross-Border Cooperation in Criminal Justice, Publications Office of the European Union](#) (2022).

[Explanatory Memorandum on the Updated OECD Definition of an AI System](#), OECD Artificial Intelligence Papers, No. 8 (March 2024).

National Institute of Standards and Technology (NIST), [Artificial Intelligence Risk Management Framework](#) (AI RMF 1.0), NIST AI 100-1 (2023).

Organization for Security and Co-operation in Europe, [Making Law Enforcement Fit for the Future: Strategic Priorities for Policing in the Digital Age](#), 15 June 2026.

Trend Micro Research, [UNICRI and Europol EC3, Malicious Uses and Abuses of Artificial Intelligence](#) (2020).

Velasco C. et al., Artificial Intelligence and Organised Crime, (Madrid: EL PAcCTO 2.0, 2024; updated August 2025).

9.3 Online resources

IBM, 'What Is Multimodal AI?', available at: <https://www.ibm.com/think/topics/multimodal-ai>

MIT Media Lab, 'AI Glossary Dictionary', available at: <https://www.media.mit.edu/tools/ai-glossary-dictionary/>.

National Center for Missing & Exploited Children, 'The Work Never Stops: First Look at NCMEC's 2025 Data', available at: <https://www.missingkids.org/blog/2026/the-work-never-stops-first-look-at-ncmecs-2025-data>.

Psychology Today, 'Have AIs Already Reached Consciousness?', available at: <https://www.psychologytoday.com/ca/blog/the-mind-body-problem/202502/have-ais-already-reached-consciousness>.