Check for updates

CASE STUDY

Facing the faces of AI in criminal evidence. Report on the seminar *EvidencIA*

[version 1; peer review: awaiting peer review]

Giulia Cascone

Derecho Procesal y Derecho Penal, Universidad Complutense de Madrid Facultad de Derecho, Madrid, Community of Madrid, Spain

v1 **First published:** 09 Oct 2026, **6**:409
https://doi.org/10.12688/openreseurope.25085.1
Latest published: 09 Oct 2026, **6**:409
https://doi.org/10.12688/openreseurope.25085.1

Abstract

On 6 July 2026, a seminar entitled '*EvidencIA – Facing the faces of AI in criminal evidence*' took place at the Complutense University of Madrid. The seminar marked the conclusion of the first year of a Marie Sklodowska Curie project named *Future* and had the objective to scrutinize the main problematic issues related to the use of AI in criminal evidence. These pages contain a report on the event, with the aim to present the reflections that took place during the seminar to the academic community and to encourage further debate on the fundamental issues that emerged. The key points drawn from the panellists' presentations are summarised and conceptualised within three broad frameworks: 'AI and criminal evidence at the crossroads between systemic balances and effective defence rights'; 'The pitfalls of AI applied to fact-finding and fact-checking: scenarios of *inquisitio generalis* and 'fake truths' between regulatory gaps and inconsistencies'; 'AI, cognitive processes and decision-making: between epistemic reliability and humanity of the decision'. Finally, some critical reflections on the main turning points of the seminar are advanced by the author.

Plain Language summary

On 6 July 2026, a seminar entitled 'EvidencIA – Facing the faces of AI in criminal evidence' took place at the Complutense University of Madrid. The seminar marked the conclusion of the first year of a Marie Sklodowska Curie research project named Future. The aim of the seminar was to explore the key issues surrounding the use of artificial intelligence in the context of criminal proceedings and, in particular, in relation to evidence. Professors of criminal procedure from various parts of Europe were invited to speak at the event. The purpose of this report is to summarise the main points of the presentations given by the speakers and the ensuing debate. The report concludes with a section by the author, setting out some critical reflections on the key issues that emerged during the day.

Open Peer Review

Approval Status *AWAITING PEER REVIEW*

Any reports and responses or comments on the article can be found at the end of the article.

Keywords
Artificial Intelligence; Criminal Justice; Criminal Evidence; EU Member States; Seminar EvidencIA

This article is included in the Horizon Europe gateway.

This article is included in the Marie-Sklodowska-Curie Actions (MSCA) gateway.

Corresponding author: Giulia Cascone (giuliacascone1@gmail.com)

Author roles: Cascone G: Conceptualization, Data Curation, Formal Analysis, Funding Acquisition, Investigation, Methodology, Project Administration, Writing – Original Draft Preparation, Writing – Review & Editing

Competing interests: No competing interests were disclosed.

Grant information: This project has received funding from the Horizon Europe Framework Programme under the Marie Skłodowska-Curie grant agreement No [101152771].
The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Copyright: © 2026 Cascone G. This is an open access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

How to cite this article: Cascone G. **Facing the faces of AI in criminal evidence. Report on the seminar *EvidencIA* [version 1; peer review: awaiting peer review]** Open Research Europe 2026, **6**:409 https://doi.org/10.12688/openreseurope.25085.1

First published: 09 Oct 2026, **6**:409 https://doi.org/10.12688/openreseurope.25085.1

On 6 July 2026, a seminar entitled 'EvidencIA – Facing the faces of AI in criminal evidence'[1] took place at the Complutense University of Madrid. The seminar, organised by Giulia Cascone, marked the conclusion of the first year of a Marie Sklodowska Curie project named 'Future – Fostering a common approach for the regulation of Artificial Intelligence in criminal procedural systems'[2].

Fernando Gascon Inchausti, the project supervisor, opened the seminar by emphasising the importance of advancing the debate on artificial intelligence in criminal justice. Building on the established foundations and principles developed to date by the scientific community, effective solutions to the issues concerning the impact of AI on the rights of the defence must be sought and found.

In this report, I set out the key points drawn from the panellists' presentations during the seminar, conceptualising them within three broad frameworks (paragraphs 1–3). Finally, I present some critical reflections on the main points that emerged from the presentations and the debate (paragraph 4).

The aim of this report is to present the reflections developed during the seminar to the academic community and to encourage further debate on the fundamental issues that emerged.

1. AI and criminal evidence at the crossroads between systemic balances and effective defence rights

Maria Lucia Di Bitonto started her presentation by observing that in democratic countries there is a political need to align the criminal justice systems with the adversarial model. Nevertheless, whilst in common law systems the idea of a trial where the prosecution and defence face each other on equal terms before the public and a judge dates back to the *Magna Charta Libertatum*, in many civil law systems – even where the choice has been made to develop adversarial models – the path towards that paradigm has not yet been completed. Consequently, the inquisitorial system continues to predominate, alongside the idea that a judicial ruling is the outcome of a solitary process of knowledge, the basis of which is the power of authority and its intrinsic reliability, rather than the adversarial relationship between the prosecution and the defence before the judge and the audience present in the courtroom.

Professor Di Bitonto fears the risk that AI may become the means by which the inquisitorial system will finally prevail over its long-standing adversarial counterparts. This is due to three characteristics of AI that constitute its essential features and are difficult to reconcile with the fundamental values of the adversarial paradigm: (1) AI usually operates as a black box; (2) AI systems are very expensive; (3) AI systems require a huge amount of energy to work. Characteristics (2) and (3) imply that the costs of AI make it a tool available only to the stronger party in criminal proceedings – which is typically the prosecution – thereby contradicting the spirit and purpose of criminal procedural rules, which is to mitigate the existing disparities between the prosecution and the defence. Characteristic (1) contradicts the three essentials of criminal evidence: evidence must be gathered in such a way as to ensure the quality of knowledge; it must safeguard the defendant's rights; and its acquisition must be equally accessible to the public present in the courtroom and to the judge who will deliver the judgement.

This does not mean that AI should be excluded from criminal proceedings. On the contrary, it can be usefully employed both for organisational purposes and during the investigative stage, where it may enhance the investigative capabilities of law enforcement authorities by directing them towards the most promising sources of evidence or by assisting them in analysing the material collected.

This would not contradict the fundamental principles of criminal proceedings, because the pressing need to ensure a balance between the prosecution and the defence arises in the trial.

Lorenzo Belvini used the example of facial recognition as a case study to address some key issues relating to the use of AI in criminal proceedings.

Starting from the premise that AI-generated results inevitably involve a degree of uncertainty and should not automatically be regarded as neutral, he emphasised the importance of meaningful human oversight, highlighting the shortcomings of the AI Act in this regard. The regulation assumes that humans are capable of supervising AI systems and correcting

[1]Recordings of the panellists' presentations are available on https://www.ucm.es/eufutureai/events.

[2]HORIZON-MSCA-2023-PF-01-01 - MSCA Postdoctoral Fellowships 2023 Marie Skłodowska-Curie Actions (MSCA), Grant agreement ID: 101152771.

potential errors, but meaningful oversight requires more than a review of the final output and a proper understanding of the process that generated the output is, in many cases, hindered by the fact that the internal working of a machine-learning system remains inaccessible.

The other key issue addressed in his presentation concerned the use of facial recognition in real time and on a large scale, which can create forms of surveillance unprecedented in both scale and intensity, as well as a potential 'chilling effect' on citizens. He highlighted the need to protect anonymity in public places, that requires us to move beyond the idea that being in public implies a waiver of privacy. This shift is now also mandated by the case law of the European Court of Human Rights.

In the debate that followed the presentation, Professor Belvini, together with Professor Karsai, highlighted that the right to a fair trial would require, in addition to access to the algorithm and even prior to that, that at the very least the use of the algorithm by law enforcement authorities be disclosed. They pointed out critically that, in many cases within their respective jurisdictions, there is no such obligation during the investigation phase.

2. The pitfalls of AI applied to fact-finding and fact-checking: scenarios of *inquisitio generalis* and 'fake truths' between regulatory gaps and inconsistencies

Radina Stoykova emphasized that today law enforcement authorities regard the use of AI in the performance of their duties as an operational necessity. Automated digital forensics would be indispensable as investigations often require the analysis of enormous volumes of data.

Building on this observation, Dr. Stoykova highlighted several inconsistencies in the AI Act, particularly regarding the boundaries of the high-risk AI category. She pointed out that many routinely- used AI tools deployed by law enforcement units do not fall within the high-risk category and are therefore exempt from the more stringent safeguards applicable to such systems.

Among the examples she discussed, particular attention deserves the case of AI systems used in the assessment of the reliability of evidence. The Draft Commission Guidelines on the classification of high-risk AI systems[1][3] list the activities that fall outside the scope of the evaluation of the reliability of evidence in such broad terms that they risk substantially emptying the category. As a result, tasks that are not merely preparatory or ancillary, but instead involve the substantive assessment of the reliability of evidence, may fall outside the high-risk regime.

Another illustrative example concerns deepfake detectors used by law enforcement authorities. These systems should fall within the category of AI used to assess the reliability of evidence and, therefore, qualify as high-risk systems. However, a regulatory asymmetry can be detected that may undermine the effectiveness of AI systems intended to detect deepfakes. In practice, the development of an effective deepfake detector requires access to relevant information about the deepfake generation system against which it is designed to operate. Yet deepfake generators are not classified as high-risk AI systems under the AI Act and are therefore not subject to the information and transparency obligations applicable to that category.

Dr. Stoykova questioned whether the AI Act's classification framework ultimately makes sense in a sensitive domain like law enforcement. She argued that these inconsistencies may lead to regulatory incoherence and practical difficulties in the implementation of the regulation, thereby undermining the effectiveness of its risk-based approach.

In this scenario, the intended purpose of an AI system should be regarded only as a starting point and the success of the AI Act will ultimately depend on how it will be implemented in practice and on how it will interact with the legal framework governing data protection.

Stoykova therefore called for an extension of the procedural safeguards currently provided in some specific fields – for example, in relation to real-time biometric identification – to other categories of AI systems used in criminal justice.

Risk mitigation, human oversight, and explainability should not be understood merely as requirements relating to AI system design, but also as matters of criminal procedural guarantees. While the AI Act cannot and should not be regarded as a criminal procedural law, stronger procedural safeguards are nonetheless needed to ensure the protection of fundamental rights in criminal proceedings, also for AI systems falling outside the high-risk category.

---

[3]See point (365).

Alfonso Peralta warned us of the risks posed by deepfakes in criminal investigations and court proceedings, where the potential threat to the proper administration of justice is enormous, particularly considering that remote hearings have become a routine feature of judicial systems in many jurisdictions. The widespread availability of platforms that enable users to generate deepfakes has significantly increased the risk of procedural fraud, not only by facilitating a false representation of participants in judicial hearings but also by enabling the submission of fake evidence generated or manipulated through AI.

Today, detecting sophisticated deepfakes is becoming increasingly difficult, if not virtually impossible, due to the exceptionally high quality of AI-generated content. Moreover, an additional challenge arises as to how courts should treat AI-generated outputs that may, in principle, be admissible as evidence but are inaccurate not because of intentional human manipulation, but because they come from an AI hallucinations.

These concerns prompt calls for a common approach to AI regulation that would provide national legislators with a coherent regulatory framework on which to base the adoption of appropriate legal safeguards. Such a framework should be capable of addressing the multiple risks that the malicious or fraudulent use of AI poses to the integrity of judicial proceedings and the fairness of their outcomes.

3. AI, cognitive processes and decision-making: between epistemic reliability and humanity of the decision

Luca Pressacco structured his presentation around three key issues. He began by addressing the legal classification of algorithmic evidence, arguing that this has become an urgent matter for both legal systems based on the principle of freedom of proof and those in which the admissible forms of evidence are strictly defined by law. In both types of systems it is possible to identify existing evidentiary categories under which algorithmic evidence might be classified, the most intuitive ones being scientific evidence, electronic evidence, and – in jurisdictions where it is admitted – atypical evidence. Nonetheless, while useful to identify the key issues to be addressed, the traditional categories, considered in isolation, are not sufficient to fully grasp the specific characteristics of evidence generated or processed using AI.

The second key issue concerned the admissibility of algorithmic evidence. After quoting the essential distinction between automatically generated evidence and evidence gathered or processed through AI [[2], p. 8], Professor Pressacco emphasized the need to determine what information – and how much information – about the AI system employed should be required before either category of evidence can be admitted or excluded in court. This would be ultimately a legal and policy choice, as it determines the balance between the investigative powers of public authorities and the rights of the defence. Once the relevant admissibility thresholds have been established, however, it remains essential to examine the technical features of the AI system in order to understand its limitations and assess its reliability.

The third issue addressed the evaluation of algorithmic evidence once admitted. With regard to AI-generated evidence, Professor Pressacco observed that Article 11 of the LED suggests treating such evidence as a form of circumstantial evidence. He also advocated the development of additional corroboration requirements beyond those already provided for in Article 11 of the LED. Finally, he stressed the need to establish a minimum set of information that should accompany algorithmic evidence to enable courts to assess its reliability effectively, much as has already been done with scientific evidence and digital evidence.

Krisztina Karsai focused on the possible use of AI algorithms in judicial decision-making, starting from the premise that, according to the current understanding of criminal justice, certain judicial decisions cannot – and should not – be made by, or with the support of, algorithmic systems. The key distinction for assessing whether algorithmic decision-making is both useful and compatible with the fundamental principles of criminal justice lies in the nature of the decision itself. In particular, AI support is conceivable only where the decision to be made is essentially statistical rather than inferential.

AI could provide meaningful assistance during the sentencing phase, where both the input and output datasets are finite and the judge's primary task is the assessment of sentencing circumstances. Most importantly, however, such support is appropriate only where the relevant evidence is already available, no extensive evidentiary investigation is required, and, above all, the evaluation of the evidence does not involve complex factual or legal assessments. Sentencing is, of course, a decision distinct from the determination of criminal liability, although it remains necessarily influenced by the findings reached at the adjudicative stage.

4. Everything must (really) change for everything to remain the same

When discussing AI in the context of criminal justice, the debate generally follows two opposing trajectories, reflecting two fundamentally different conceptions. On the one hand are those who argue that the protection of certain fundamental values requires excluding algorithmic solutions from criminal proceedings, since they alter the very nature of the judicial process[4]. On the other hand are those who argue that we must keep our feet on the ground, seeking compromise solutions capable of harnessing the efficiency promised by technological progress while preserving the essential features and underlying principles of criminal justice[5].

The latter perspective underpins the argument that the use of artificial intelligence has become an operational necessity, particularly in the field of law enforcement, where investigative authorities are routinely required to analyse vast quantities of data. While this claim is expected when viewed from the perspective of investigators and prosecutorial authorities, it raises a fundamental question concerning the direction in which criminal investigations are evolving.

If we accept that, without AI, law enforcement agents are incapable of processing all the data collected during a criminal investigation, we are implicitly acknowledging that the investigation has exceeded the principles of relevance and necessity that should govern the exercise of public power in the prosecution of crimes. The ever-increasing availability of enormous amounts of data – made possible by a constant and unsettling process of datafication[3] – expands investigative capabilities and raises the risk that criminal investigations may gradually evolve into a modern *inquisitio generalis*[6], incompatible with the principles that should regulate the relationship between the State and the individual in a liberal democratic society.

The exercise of public authority and its powers of intrusion into individuals' private lives should remain subject to the standard of proportionality. Within criminal investigations, one of the principal manifestations of that standard is that the activities carried out in the search for evidence be relevant to the facts under consideration. It is therefore legitimate to question whether these principles are respected when law enforcement authorities are unable to accomplish their tasks without resorting to software applications – such as the Hansken tool used in the Netherlands [[4], p. 45] – that search for patterns and correlations in immense volumes of information, much of which is irrelevant but nevertheless swept into the investigation simply because current technology makes its collection and analysis possible.

In such a scenario, significant concerns for the rights of the defence and the principle of equality of arms also emerge. A strong imbalance between the prosecution and the defence arises at the investigative stage itself, potentially influencing the entire course of the proceedings and prejudicing their ultimate outcome.

Therefore, the use of AI to analyse massive datasets (AI-filtered evidence) or to generate investigative leads cannot therefore be regarded as neutral [[2], pp. 7, 25 and 42] merely because it occurs during the pre-trial phase.

While it is obvious that strong procedural safeguards should operate during the trial, the European continental experience – shaped by an enduring inquisitorial legacy[7] – demonstrates that, even in jurisdictions commonly regarded as adversarial, criminal investigation and trial are not separate spheres, insulated from one another by more or less stringent standards of procedural impermeability[8]. And, apart from this observation, also adversarial systems characterised by a strong and rigid separation between the investigative file and the trial file[9], information obtained through AI during the investigative phase may still exert a significant indirect influence on the trial. Investigators may use AI-generated outputs to obtain information that is subsequently transformed into admissible evidence through conventional means, such as witness testimony. Likewise, AI-generated leads may enable investigators to identify new sources of evidence that can later be introduced at trial. Moreover, information produced through AI-assisted investigative tools may underpin

[4]For a strong defence of the human dimension of adjudication,[13,14]

[5]This is the approach endorsed both by leading international institutions,[19] and by prominent scholars who have long examined the interaction between law and technology.[19]

[6]In the sense that the fact that investigative activities extend to facts or persons unrelated to the *notitia criminis.*, and are therefore excessive in scope, is a revealing indicator of the existence of an *inquisitio generalis.*, in which criminal proceedings are instrumentalised in the service of an unlimited and objectively indeterminate investigation.[10], p. 85

[7]See, also for the necessary clarifications and nuances on the topic, the fundamental work of.[18]

[8]The Italian criminal justice system provides an example of a model strongly inspired by the adversarial tradition. It is based on the principle that evidence should be produced through adversarial proceedings at trial, while investigative acts are, in principle, not intended to have evidentiary value. The model introduced by the 1988 Code of Criminal Procedure, however, has gradually been eroded over time. Successive legislative reforms and judicial decisions have altered its original design, progressively blurring the boundary between the investigation and the trial – a boundary that was conceived as clear and, almost, impermeable. See[12]

[9]Like the Italian system. See.[11], p. 583

decisions taken during the pre-trial phase that directly affect fundamental rights, including orders for pre-trial detention and the imposition of other coercive measures.

Accordingly, the need to safeguard the rights of the defence and the equality of arms from the earliest stages of the proceeding should be recognised across all legal systems, including those in which evidence is formally and generally produced during the trial stage. Equality of arms cannot remain confined to the courtroom; it must be strengthened also in the investigative stage, where it is not justifiable that law enforcement authorities enjoy access to enormous quantities of data together with highly sophisticated analytical tools, while the defence has no comparable means at its disposal.

Procedural mechanisms capable of restoring procedural balance are therefore necessary and should be developped in two directions: on the one hand, they should prevent law enforcement authorities from transforming criminal proceedings into a modern *inquisitio generalis.* On the other hand, they should ensure that the defence has access to equivalent technological capabilities, where necessary through digital forensic assistance [[2], p. 56] [[5], p. 14][6] available already during the investigations.

Ultimately, we must abandon the assumption that the timing of procedural safeguards is neutral, and that almost any investigative practice is acceptable provided that some form of balancing occurs later during the subsequent phases of the proceeding. This idea – that can probably be traced back to the judgments of the European Court of Human Rights on the examination of sole prosecution witnesses[10] – must be reconsidered in light of the current technological revolution. While a later opportunity to examine a witness may effectively restore procedural balance, there are good reasons to doubt that the profound asymmetry created at the very outset of criminal proceedings by the prosecution's exclusive access to powerful AI-driven investigative technologies can ever be fully remedied at subsequent stages of the proceeding.

It must also be acknowledged, however, that achieving the kind of procedural balance just described may prove difficult given the current features of criminal proceedings in many European jurisdictions. As Belvini and Karsai have pointed out, investigators are often under no obligation to disclose the technologies and investigative tools employed in the course of their inquiries. This problem is particularly acute in legal systems where the pre-trial phase is generally conducted in secret and access to the case file – thereby enabling the effective exercise of defence rights – is granted only once the investigative phase has been completed. The same concern would arise, however, also in jurisdictions where the pre-trial stage is characterised by greater transparency and broader opportunities for defence participation[11]. The reason is obvious: certain uses of AI by law enforcement authorities – most notably the filtering of evidence and the identification of correlations within large datasets – are unlikely to be subject to disclosure obligations or adversarial scrutiny by the defence, because they are perceived merely as tools supporting simple desk investigative researches, rather than as measures affecting fundamental rights.

In short, we are confronted with a second *black box* problem. In this case, however, what is opaque is not the AI algorithm, but the criminal proceeding itself, at least during certain stages. This calls for serious reflection on whether broader disclosure obligations should be introduced with respect to the use of AI by law enforcement authorities [[2], p. 39]. Otherwise, guarantees of equality of arms and access to digital forensic assistance risk becoming useless.

The real challenge, however, lies in ensuring that such obligations are genuinely effective. It is one thing to require disclosure where an AI-generated output is intended to be relied upon as evidence, since the party seeking to introduce it must necessarily disclose it to the court. It is quite another where AI is used merely to identify correlations within large datasets, generate investigative leads, or perform other analytical tasks that never themselves become evidence. In such cases, compliance with the disclosure obligation ultimately depends on the good faith of law enforcement authorities, and the primary safeguard against breaches of that duty is likely to lie in the availability of effective sanctions.

Equally significant challenges arise when attempting to provide appropriate legal classifications of AI-evidence within criminal proceedings.

The use of AI in criminal justice can affect two distinct dimensions of fundamental rights: one inside and another outside the criminal proceeding. Biometric recognition technologies provide a good example. Their use implicates rights that extend beyond the proceedings themselves, since the processing of biometric data engages the right to data protection irrespective of whether criminal proceedings have been initiated. If the resulting output is relied upon as evidence, however, core procedural guarantees can also be jeopardised, including the equality of arms, the access to evidence, and

[10]*Ex multis.*, EctHR, Delta v. France, app. n. 11444/85, 19 December 1990; EctHR, Mamikonyan v. Armenia, app. n. 25083/05, 16 March 2010; EctHR, Al-Khawaja and Tahery v. the United Kingdom, app. n. 26766/05 and 22228/06, Grand Chamber, 15 December 2011; EctHR, Schatschaschwili v. Germany, app. n. 9154/10, 15 december 2015; EctHR, Alo v. Slovakia, app. n. 32084/19, 10 February 2022.
[11]For example, the Spanish one. See[17], p. 155

the right to challenge adverse evidence. For this reason, as Professor Pressacco rightly observed, we need legal basis and appropriate legal classifications.

Achieving these objectives, however, is not straightforward and AI compels us to rethink some of the traditional categories of the law of evidence.

Documentary evidence provides a first clear example. AI challenges this category at least in two ways: on the one hand, it is questionable whether a product generated by AI – a video, an audio recording, a photograph, or another form of digital content – can be classified as documentary evidence [7, p. 200], unless it is established that it constitutes a mere reproduction of reality and has not been subject to any autonomous processing or alteration by the AI system. A document represents; it does not evaluate.

On the other hand, the phenomenon of deepfakes compels us to reconsider the criteria governing the admissibility of documentary evidence in court. Strict admissibility rules have traditionally been a hallmark of Anglo-Saxon jurisdictions [8, p. 143]. Yet AI may also compel Continental systems to reconsider the relationship between rules governing the admissibility and the evaluation of evidence, with a view to strengthening the former, introducing more rigorous screening mechanisms before (at least, documentary) evidence is admitted at trial.

Another example can be found in the acquisition of expert evidence, both in adversarial systems – where it is usually carried out through expert testimony[12] – and in inquisitorial ones.

Can the category of expert testimony be relied upon to introduce into criminal proceedings knowledge generated or processed by AI?

I think there is a fundamental obstacle to this approach. In adversarial systems, ever since *Daubert*[13], the principles governing the admissibility (or, in certain jurisdictions, the evaluation of reliability[14]) of scientific evidence have rested on the assumption that the expert's knowledge and methodology can be tested through adversarial examination. The credibility of expert evidence depends precisely on the possibility of examining and cross-examining the expert on the knowledge he is called to bring in the proceeding. AI, however, cannot be examined and cross-examined, so the only available option to apply the scheme of expert evidence would be to examine a human expert on the reliability of the AI system. Anyway, if an expert were called to explain the functioning of an AI system, the knowledge that would actually be introduced in court would not be related the AI-generated content itself, but rather limited to the the functioning and operability of its source [7, p. 214 and 222].

But does examining an expert on the reliability of AI really amount to testing the AI-generated or AI-processed evidence itself? Or are we, instead, merely introducing a form of 'evidence on the evidence' [9, p. 15 ff. and 29 ff.]? This seem to be the case: an expert opinion is not fully tested simply because another expert has been examined on the reliability of the methodology on which it is based. A crucial stage in the assessment of expert evidence would still be missing: the application of that methodology in the specific case under consideration. But – this is the problem – that segment would inevitably remain opaque because AI is not capable to explain how a certain result was produced.

And, moving the focus to inquisitorial systems, where expert knowledge is usually acquired through experts' written reports, we would have exactly the same problem, that is the impossibility of testing that crucial segment of the object of the evidence concerning the single operation realised by the machine instead of its overall functioning.

In other words, in both cases we would not exhaust the entire object of the expert evidence.

If we assume this conclusion – that is, that a certain degree of opacity will inevitably affect the evidentiary output introduced in the court – is there any basis for treating that output as a form of circumstantial evidence[15]?

---

[12]As Sabine Gless noted, 'Machine evidence and expert testimony are inextricably linked due to the fact that AI-generated data must be explained',.[7], p. 239

[13]Daubert v. Merrel Dow Pharmaceutical, Inc., 509, U.S. 579 (1993).

[14]The so called *Daubert.* text extended its influence beyond the US jurisdiction. In Italy, for example, the criteria underlying that fundamental judgment have been adopted by the Court of Cassation since the decision *Cozzini.*, Sez. IV, 17 September 2010, n. 43786. Differently from the US, in Italy those criteria are usually applied in the stage of the evaluation – rather than the admission – of evidence.

[15]In this sense, for example,[19], p. 17

My own view is that we cannot, for we corroborate inductions when we do not have deductions, but it is not an induction what we would corroborate in this case.

Corroboration is a cognitive process through which a proposition derived from strong – although not certain – premises acquires greater credibility because it is reinforced by other elements that are external to those premises and conclusions. For such process to be sound, however, the premises themselves must be true, even if they do not logically entail the conclusion. The difficulty with AI is that this assumption is often absent. In many cases, we cannot confidently regard the premises underlying an AI-generated output as true in the epistemological sense. Put differently, would we ever accept to introduce as evidence in a criminal proceeding the pronouncement of an oracle simply because it is corroborated by other items of evidence?

As Stoykova has pointed out during her presentation, not all AI systems are opaque black boxes. Some achieve levels of transparency that may be considered acceptable. The challenge, therefore, is twofold. First, we must determine what degree of transparency should be regarded as sufficient for use in criminal proceedings. Secondly, we must ask whether that degree of transparency is technologically attainable. It is precisely at this stage that a meaningful dialogue between lawyers and computer scientists becomes indispensable.

Both of these questions, however, presuppose that a more fundamental issue has first been addressed. Before asking *how* AI-generated outputs should be introduced into criminal proceedings, we should perhaps ask *whether* we want them to be introduced at all.

This is, in the first place, a political and ideological question rather than purely legal. It is probably for this reason that the European legislator has avoided confronting it. The AI Act provides a telling example. It classifies the use of AI to assess the reliability of evidence as a high-risk. Yet is that categorisation really adequate? Are we not dealing here with the very core of criminal adjudication? The assessment of the reliability of evidence, its internal coherence, and its consistency with other evidentiary records lies at the heart of the judicial function. Can judicial decision-making truly be separated from the evaluation of the evidence, as though the latter were not an integral component of the former?

Professor Karsai's reflections deserve to be taken seriously. As long as our understanding of criminal adjudication remains unchanged, certain uses of AI will probably remain incompatible with the very conception of adjudication on which contemporary criminal justice is founded. The real question, however, is what, in an era of rapid technological transformation, prevents us from rethinking that conception.

As already suggested, the answer is probably a political and ideological one. It depends on the model of criminal procedure we wish to preserve and on what we expect the administration of criminal justice to be. If we are committed to preserving the current conception of criminal justice, however, the application of existing rules, safeguards, and procedural guarantees will not suffice. Rather, a number of fundamental and far-reaching regulatory choices will inevitably have to be made.

AI disclosure

AI (Deepl) was used to review the accuracy of the English language in drafting this report. The drafting of the text, the reconstruction carried out, and the opinions expressed are solely attributable to the author.

Ethics and consent

All the people mentioned in this report have reviewed the text, confirmed that the ideas and views attributed to them are consistent with those expressed during the seminar, and have given their written consent to its publication.

Data and Software Availability Statement

This report does not contain underlying or extended data.

References

1. E. Commission: *ANNEX III to the COMMUNICATION FROM THE COMMISSION, Draft Commission guidelines on the classification of high-risk AI systems under Article 6 of Regulation (EU) 2024/1689 (AI Act) for stakeholder consultation.* Brussels: 2026.

2. Ligeti K: *Unpacking AI Evidence and (Re-)Defining Procedural Safeguards in Digital Investigations.* HART; 2026.
Publisher Full Text

3. Mayer-Schönberger V, Cukier K: *Big Data: A Revolution that Will Transform how We Live, Work, and Think.* Houghton Mifflin Harcourt; 2013.
4. Galič M, Das A, Schuilenburg M:*AI and Administration of Criminal Justice. Report on the Netherlands, e-Revue Internationale de Droit Pénal.* 2023.
5. Gless S, Lederer F, Weigend T: **AI-based Evidence in Criminal Trial? .** *Tulsa Law Review.* 2024.
6. Gless S: *Rules on Expert Testimony Based on a Comparative Perspective on Device Evidence.* 2024.
7. Gless S: **AI in the Courtroom: a Comparative Analysis of Machine Evidence in Criminal Trial.** *Georgetown Journal of International Law.* 2020.
8. De Hert P, Bouchagiar G: *AI-Generated Evidence in European Courtrooms.* Cambridge University Press; 2026.
Publisher Full Text
9. Gascón Inchausti F: **El control de la fiabilidad probatoria: "prueba sobre la prueba en el proceso penal".** *Ediciones Revista General de Derecho.* 1999.
10. Aguilera Morales M: *Proceso penal y causa general. La inquisitio generalis en el Derecho español.* Thomson Reuters-Civitas; 2008.
11. Negri D: *I fascicoli.* Cedam; 2023.
12. Aprile E, Catalano EM, Chinnici D, *et al.*: *Le erosioni silenziose del contraddittorio.* Giappichelli; 2017.
13. Garapon A: *La despazializzazione della giustizia.* Mimesis; 2021.
14. Garapon A, Lassègue J: *Justice digitale. Révolution graphique et rupture anthropologique.* PUF; 2018.
15. CEPEJ: *European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and their environment.* Council of Europe; 2018.
16. Floridi L, Cowls J: **A Unified Framework of Five Principles for AI in Society.** *Harvard Data Science Review.* 2019.
Publisher Full Text
17. Asencio Mellado JM, Fuentes Soriano O: *Derecho procesal penal.* Tirant lo Blanch; 2024.
18. Damaška M: *The Faces of Justice and State Authority: A Comparative Approach to the Legal Process.* Yale University Press; 1986.
19. Gialuz M: **Quando la giustizia penale incontra l'intelligenza artificiale: luci e ombre dei risk assessment tools tra Stati Uniti ed Europa.** *Diritto Penale Contemporaneo.* 2019.