The Paradigm Conflict and Convergence Between Artificial Intelligence and Evidence Law: A Study Through EU Regulation and Global Case Law

Xinyi Peng[1], Yueshan Hua[2]

[1]New South Wales University

[2]Sichuan University

xinyi.peng2@student.unsw.edu.au, 18724835930@163.com

Abstract

The integration of artificial intelligence (AI) into judicial fact-finding has triggered a global paradigm crisis in the law of evidence. This article, centred on the regulatory framework of the *European Union Artificial Intelligence Act* (‘*EU AI Act*’), critically analyses the dual character of AI in the evidentiary process, as both an enabler and a disruptor, through the lens of landmark cases from the United States, the United Kingdom and Canada.

The article argues that the *EU AI Act*, through its risk-based classification, transparency obligations, and fundamental rights impact assessments, provides a systemic blueprint for reforming evidentiary doctrine worldwide. Yet, from *State v Loomis* to *Regina v A*, leading cases have revealed a profound tension between legal frameworks and technological realities: algorithmic opacity undermines the right of cross-examination, data bias corrodes the fairness of trial, and generative AI destabilises the foundation of authenticity.

Ultimately, the article contends that the future of evidence law lies in developing a model of responsive regulation, which, by drawing on the EU’s regulatory logic, integrates mandatory algorithmic auditing, dynamic evidentiary disclosure, and algorithmic due process into a coherent governance framework. This, it is argued, is how evidence law can regain vitality in the digital age.

Opportunity: AI as an Evidentiary Enhancer and the Guidance of EU Regulation

AI’s role in legal fact-finding is not merely disruptive. When properly regulated, its analytical and inferential capacities can enhance the precision, efficiency and consistency of proof. The enduring significance of the *EU AI Act* lies in its attempt to balance innovation with safety, offering a trusted and structured framework rather than a prohibitionist response.

From Chain of Custody to Chain of Provenance

Traditional authentication of electronic evidence depends on demonstrating the integrity of a chain of custody — that the item has not been tampered with between collection and presentation in court. This logic collapses for AI-generated, “digitally native” content: a deepfake has no original reality to preserve.

Technological standards such as Content Provenance and Authenticity (C2PA) employ cryptographic metadata to record a file’s origin, editing history and software environment, effectively building a chain of provenance. Article 52(3) of the *EU AI Act* explicitly mandates that any AI system generating or manipulating image, audio or video content must clearly disclose that the content is AI-generated or AI-altered. This “transparency obligation” gives legal recognition to verifiable provenance, and shifts authentication from a static artefact toward an auditable narrative of origin. It equips courts to privilege content with traceable digital identity and to discount, or exclude, content of uncertain origin.

From Human Review to Computational Insight

In complex commercial fraud or organised-crime cases, digital and documentary evidence, including emails, chats, ledgers, can be so voluminous that it exceeds human capacity for meaningful review. Natural-language processing (NLP) and machine-learning models can classify text, detect sentiment, map relational structures and surface anomalies at a speed and scale no human team can match.

EU AI Act classifies forensic and justice-related AI tools as “high-risk” systems under Annex III(8). Because of that designation, providers and users must comply with the strict obligations in Chapter III: implementing risk-management systems; training on high-quality, representative datasets to minimise bias; maintaining detailed technical documentation; ensuring human oversight; and preserving auditable logs. This regime sets a reliability threshold for judicial AI, transforming it from “black-box technology” into an accountable forensic instrument (Gless 2020).

From Individual Discretion to Data-Informed Calibration

Judicial discretion is central to justice, but it is never fully immune from individual cognitive limits. AI can extract statistical regularities from large sentencing and liability datasets, offering courts a reference frame for consistency and proportionality (Law Council of Australia 2025).

EU AI Act is cautiously optimistic on this point. Even when deployed merely as a "decision-support" tool, such systems remain within the "high-risk" category, and must be designed around fairness to vulnerable groups and representativeness of data. The message is blunt: the promise of AI in judicial decision-making is conditional on non-discrimination. Without fairness controls, any "assistive" function risks entrenching historical injustice.

Risk: AI as an Erosive Force in the Foundations of Evidence Law

The same characteristics that make AI attractive to courts also endanger the structural commitments of evidentiary justice. The EU's regulatory response is built on an anticipation of those dangers, while global case law has already made them visible in practice.

Erosion of the Right to Challenge: The Algorithmic Black Box

Cross-examination is the common law's canonical technique for testing truth. But when the "witness" is a neural network with millions of parameters, counsel cannot interrogate its perception, memory or reasoning. The opacity of the algorithmic "black box" can effectively neutralise the right to challenge adverse evidence and threaten equality of arms.

In *Loomis*, the defendant was sentenced in part on the basis of the proprietary COMPAS risk assessment tool. Neither he nor his lawyers could access or meaningfully scrutinise the model's internal logic. The Wisconsin Supreme Court acknowledged transparency concerns but upheld the sentence, in part on the ground that COMPAS was "only one factor" among many.

Loomis illustrates the collision between evidentiary doctrine and technological opacity. The judgment in effect lowered the reliability threshold for scientific or quasi-scientific evidence by tolerating reliance on an inscrutable model. It set a troubling precedent: where technology becomes too complex to explain, courts may defer to it rather than exclude it. This underscores the urgency of recognising explainability as a precondition for admissibility of algorithmic evidence.

Undermining Fair Trial: Bias and Structural Discrimination

AI models learn from historical data. If that data encodes racial, gender or socio-economic bias, the model can reproduce and amplify those patterns under a veneer of "objectivity".

In *Bridges*, the Court of Appeal held that the police's live facial-recognition deployment in public spaces was unlawful. The force had neither adequately assessed the technology's gender and racial bias, nor adopted sufficiently specific policies governing its use.

Bridges elevates algorithmic fairness from an ethics slogan to a procedural legal duty. The Court of Appeal held that public authorities using predictive or identification AI must undertake proactive, pre-deployment bias impact assessment.[20] This jurisprudence implies that AI-derived evidence which has not undergone bias assessment ought, at minimum, to face serious admissibility scrutiny.

The Collapse of Authenticity: Generative AI and the End of "Objective Reality"

The law of evidence ultimately aims at reconstructing factual reality. Generative AI, however, can fabricate photorealistic but wholly synthetic images, voices and videos. Such deepfakes sever the traditional ontological link between representation and the world. When "seeing is believing" no longer holds, authenticity — historically the evidentiary bedrock — becomes radically unstable.

The EU AI Act subjects generative AI systems, including general-purpose models, to heightened obligations. Articles 52 and following require clear labelling of AI-generated content, safeguards against generating illegal content and public disclosure of summaries of training data usage.

This package aims to rebuild a legal presumption structure around provenance. Mandatory transparency means synthetic content enters court flagged as such, shifting a heavier justificatory and probative burden to the party relying on it.

The Way Forward: Responsive Regulation for Law–Technology Co-Evolution

Confronted with both the promise and the danger of AI, evidence law needs structural renovation, not cosmetic patching. This article proposes that courts and legislators move toward a model of responsive regulation — one that is sensitive to risk level, procedurally dynamic, and explicitly rights-protective.

Risk-Based Rules of Admissibility

Evidentiary admissibility standards should be redesigned to reflect the European Union's tiered approach to AI risk. When courts are presented with high-risk AI evidence— for instance, sentencing algorithms, recidivism forecasts or offence-attribution models that speak directly to guilt — that material should not simply enter the record by default. Instead, it ought to attract a rebuttable presumption that it is inadmissible. The party seeking to rely on it then bears the burden of dislodging that presumption. Herke Csongor and David Toth (2024) point that doing so requires more

than assurances of accuracy: the proponent must supply full technical documentation, submit the system to genuinely independent audits of accuracy and fairness, and offer case-specific evidence that the tool operates reliably in the circumstances of the dispute.

Material that falls into a lower-risk category should be treated differently. AI-generated media used only illustratively or for contextual background does not demand the same exclusionary stance.It should, however, trigger a strict provenance disclosure obligation. The party tendering such material must be able to demonstrate where it came from and how it was produced — for example, by providing C2PA-style metadata or equivalent proof of origin. As Sabine Gless (2020) emphasised, if that provenance is missing, incomplete, or deliberately obscured, the court should either treat the material as having minimal probative value or refuse to admit it at all.

Seen as a whole, this model shifts the justificatory burden toward the party seeking to put algorithmic material before the court, and it does so in a way that scales with the potential of that material to distort fact-finding.

Dynamic Evidentiary Disclosure

Conventional disclosure is static. Algorithmic systems evolve, are retrained, and behave differently across contexts. Inspired by the EU's ongoing-compliance obligations for high-risk systems, this article argues for dynamic disclosure:

The defence should have a qualified right, under court supervision, to conduct expert "red-team" testing of an AI system relied upon by the prosecution (Herke Csongor and David Toth 2024). That testing would probe for bias, error modes and instability in situ. This converts substantive scrutiny of an opaque model, which was impossible in *Loomis,* into a procedural entitlement.

A Right to Algorithmic Due Process

To protect human defendants and civil litigants in a justice system increasingly mediated by AI, the law should expressly recognise a set of algorithmic due process rights. First, parties should have a right to be notified whenever an AI system is used in relation to their case, so that the involvement of algorithmic tools is never hidden from those affected by their outputs. Lisa Messeri and M J Crockett(2024) argue that parties should also have a right to an explanation, meaning an entitlement to receive an intelligible, case-specific account of how the AI produced the relevant output and why that output is said to matter evidentially. In addition, parties should have a right to human review: they should be able to seek independent judicial reconsideration of any decision that relies, in whole or in part, on AI-generated analysis or assessment. These guarantees operate as procedural safeguards for dignity and personal agency. By insisting on transparency, intelligibility and human oversight, they help ensure that people are not quietly displaced in their own proceedings by statistical artefacts.

Conclusion

The *EU Artificial Intelligence Act*, alongside significant jurisprudence such as *Loomis* and *Bridges*, delineates the evolving landscape of evidence law in the digital era. This article posits that AI represents a paradigm challenge demanding a responsive reconstruction of evidentiary doctrine. Moving beyond passive adaptation, the proposed model of responsive regulation analytically synthesizes the EU's regulatory logic with comparative judicial lessons.

By translating these principles into tiered admissibility standards, dynamic disclosure, and algorithmic due process rights, courts and legislators can construct a next-generation evidentiary framework. This initiative transcends technical adjustments, representing a vital reaffirmation of rule-of-law values, as it harnesses AI's benefits while actively safeguarding procedural fairness and the very essence of substantive truth in the digital age.

References

Australian Institute of Judicial Administration. 2022. AI Decision-Making and Courts: A Guide for Judges, Tribunal Members and Court Administrators. https://aija.org.au/wp-content/uploads/woocommerce_uploads/2022/06. Accessed: 2025-10-11.

Content Provenance and Authenticity Coalition. 2024. C2PA Technical Specification. Released 21 December 2024. https://spec.c2pa.org/specifications/specifications/1.3/specs/C2PA_Specification.html. Accessed: 2025-11-01.

Csongor, H.; and Toth, D. 2024. Artificial Intelligence in Cybersecurity: Examining Liability, Crime Dynamics, and Preventive Strategies. EU and Comparative Law Issues and Challenges Series (ECLIC) 8(1): 730–760. doi.org/10.25234/eclic/32299.

European Parliament and Council of the European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council on Artificial Intelligence ('EU AI Act'). http://data.europa.eu/eli/reg/2024/1689/oj. Accessed: 2015-10-11.

Gless, S. 2020. AI in the Courtroom: A Comparative Analysis of Machine Evidence in Criminal Trials. Georgetown Journal of International Law 51: 211–246.

Law Council of Australia. 2025. Artificial Intelligence Use in the Federal Court of Australia. https://lawcouncil.au/media/news/artificial-intelligence-ai-use-in-the-federal-court-of-australia. Accessed: 2025-10-11.

Messeri, L.; and Crockett, M. J. 2024. Artificial Intelligence and Illusions of Understanding in Scientific Research. Nature 627: 50–52. doi.org/10.1038/d41586-024-00939-7.

National Institute of Standards and Technology. 2023. Artificial Intelligence Risk Management Framework (AI RMF 1.0).

https://nvlpubs.nist.gov/nistpubs/ai/NIST.AI.100-1.pdf. Accessed: 2025-10-11.

Pennsylvania v. Ritchie. 1987. 480 U.S. 39.

R (Bridges) v. Chief Constable of South Wales Police. 2020. EWCA Civ 1058.

Regina v. A. 2001.UKHL 25.

State v. Loomis. 2016. WI 68, 371 Wis. 2d 235, 881 N.W.2d 749.