INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]

VOLUME 6 AND ISSUE 4 OF 2026

APIS – 3920 - 0001 *(and)* ISSN - 2583-2344

Published by

Institute of Legal Education

https://iledu.in

ALGORITHMIC EVIDENCE ON TRIAL: EVALUATING THE ADMISSIBILITY OF AI-GENERATED FORENSIC OUTPUTS UNDER DIVERGENT GLOBAL JUDICIAL STANDARDS

AUTHOR – DINESH KUMAR B* & MS. HEMAVATHY**

* STUDENT AT SCHOOL OF EXCELLENCE IN LAW, THE TAMILNADU DR.AMBEDKAR LAW UNIVERSITY

** PROFESSOR AT SCHOOL OF EXCELLENCE IN LAW, TNDALU

BEST CITATION – DINESH KUMAR B & MS. HEMAVATHY, ALGORITHMIC EVIDENCE ON TRIAL: EVALUATING THE ADMISSIBILITY OF AI-GENERATED FORENSIC OUTPUTS UNDER DIVERGENT GLOBAL JUDICIAL STANDARDS, *INDIAN JOURNAL OF LEGAL REVIEW (IJLR)*, 6 (4) OF 2026, PG. 87-98, APIS – 3920 – 0001 & ISSN - 2583-2344.

Abstract

The advent of artificial intelligence (AI) in forensic science has revolutionized evidence generation, from facial recognition to digital trace analysis and predictive modelling. Yet, AI-generated forensic outputs face unprecedented scrutiny in courtrooms worldwide due to divergent judicial standards governing admissibility. This research evaluates the challenges of introducing algorithmic evidence under frameworks such as the U.S. Daubert standard, which demands testability, peer review, and known error rates, contrasted with the United Kingdom's more flexible common law approach and the inquisitorial models of civil law jurisdictions across the European Union. Central tensions arise from AI's 'black box' opacity, where proprietary algorithms obscure reasoning, raising concerns over reproducibility and bias. Empirical analysis of landmark rulings reveals rejection rates exceeding 40% for unvalidated AI tools in adversarial proceedings, underscoring risks to judicial integrity. Key barriers include insufficient validation benchmarks, the absence of forensic-specific error metrics, cross-jurisdictional data privacy conflicts, and judicial unfamiliarity with AI limitations such as dataset skews that amplify racial biases. Proposed reforms advocate hybrid standards: mandatory AI explainability audits, international certification extending ISO 17025 frameworks, Rule 707-style disclosures for machine-generated evidence, and federated learning for privacy-preserving cross-border validation. By dissecting admissibility criteria through comparative legal lenses, this study charts pathways for harmonised protocols, ensuring AI enhances rather than erodes forensic trustworthiness. Balancing innovation with due process demands urgent, evidence-based judicial evolution.

Keywords: algorithmic evidence, AI admissibility, Daubert standard, forensic AI, judicial standards, black box opacity, evidentiary reliability, algorithmic bias, explainable AI, cross-jurisdictional forensics.

1. Introduction

The integration of artificial intelligence into forensic science represents one of the most consequential and consequentially under-regulated developments in the history of criminal justice. AI-driven tools now assist investigators in fingerprint identification, facial recognition, DNA analysis, gait analysis, digital trace reconstruction, voice pattern analysis, and predictive risk modelling. These technologies promise unprecedented speed, scalability, and pattern-recognition capabilities that substantially exceed the limits of unaided human analysis. The global forensic AI market was estimated at approximately USD 1.3 billion in 2022 and is projected to exceed USD 5 billion

INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]

VOLUME 6 AND ISSUE 4 OF 2026

APIS – 3920 - 0001 *(and)* ISSN - 2583-2344

Published by

Institute of Legal Education

https://iledu.in

by 2030, driven by law enforcement investment across North America, Europe, and the Asia-Pacific region. Yet the deployment of such tools within the adversarial framework of judicial proceedings raises profound questions about evidentiary reliability, algorithmic transparency, and the due process rights of defendants.

At the heart of this tension lies a fundamental incompatibility: AI systems, particularly those built on deep learning architectures, operate as 'black boxes' whose internal reasoning processes are opaque even to their own developers. This opacity conflicts directly with the core evidentiary requirements of adversarial legal systems, which demand that evidence be testable, reproducible, and subject to meaningful cross-examination.[123] The U.S. Supreme Court's landmark ruling in Daubert v. Merrell Dow Pharmaceuticals established that scientific evidence must be grounded in methodologies that are testable, peer-reviewed, possess known error rates, and enjoy general acceptance within the relevant scientific community. Applied to AI, each of these criteria presents acute interpretive and practical difficulties that existing evidentiary doctrine was not designed to resolve.

The problem deepens in multi-jurisdictional investigations, such as international cybercrimes, cross-border organised crime cases, or terrorism proceedings, where conflicting evidentiary standards across legal systems create substantial and often unresolvable admissibility disputes. The European Union's General Data Protection Regulation (GDPR)[124] imposes stringent data minimisation and purpose limitation requirements that may directly conflict with the evidentiary collection and retention practices mandated by U.S. courts. This jurisdictional friction risks producing a fragmented global forensic landscape where AI tools are deployed inconsistently, with outcomes that simultaneously undermine investigative effectiveness and judicial integrity.

This paper undertakes a systematic comparative analysis of how three major legal traditions the United States, the United Kingdom, and the European Union's civil law jurisdictions approach the admissibility of AI-generated forensic evidence. It identifies the structural barriers impeding consistent and fair deployment, examines empirical data on rejection rates and landmark rulings, and proposes a coherent framework of reforms. The central argument is this: without harmonised international standards for the validation, disclosure, and judicial evaluation of forensic AI tools, the transformative potential of AI in criminal investigation will be irreparably undermined by legal unpredictability, perpetuating inequities and threatening the foundational credibility of criminal justice systems worldwide.

The paper proceeds as follows. Section 2 provides the theoretical framework, examining the principal judicial standards for scientific evidence across the three jurisdictions under study. Section 3 analyses the key barriers to admissibility, encompassing algorithmic opacity, bias, validation deficits, and cross-jurisdictional privacy conflicts. Section 4 presents empirical evidence from landmark rulings and rejection rate analysis. Section 5 develops the proposed reform framework. Section 6 discusses broader implications, and Section 7 offers conclusions.

2. Theoretical Framework: Judicial Standards for Scientific Evidence

2.1 The Daubert Standard (United States)

The Daubert standard, established by the U.S. Supreme Court in 1993 and subsequently refined through General Electric Co. v. Joiner (1997) and Kumho Tire Co. v. Carmichael (1999), fundamentally restructured the gatekeeping function of federal judges with respect to expert

[123]Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993). The Supreme Court held that federal judges must act as gatekeepers to ensure that expert scientific testimony rests on a reliable foundation and is relevant to the task at hand.

[124]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 (GDPR) [2016] OJ L 119/1, Articles 5(1)(b) and 5(1)(c) (purpose limitation and data minimisation principles).

scientific testimony. Under Federal Rule of Evidence 702,[125] as informed by Daubert, courts evaluate whether proposed expert testimony rests upon: (1) a testable and falsifiable theory or technique; (2) peer review and publication in recognised scientific outlets; (3) a known or knowable error rate; and (4) general acceptance within the relevant scientific community. This four-pronged framework was designed to prevent 'junk science' from unduly influencing jury deliberations, but it presents acute and arguably irresolvable challenges when applied to AI systems whose underlying architectures resist straightforward empirical validation.

Prior to Daubert, the prevailing admissibility standard derived from Frye v. United States,[126] which required only that a technique be 'generally accepted' within its field. The Frye standard remains applicable in several U.S. states, including California and Illinois, and its less demanding threshold has sometimes enabled AI tools to pass admissibility scrutiny that would fail under Daubert's more rigorous regime. The doctrinal divergence between Daubert and Frye jurisdictions within the United States itself creates an internal incoherence that mirrors the wider international fragmentation problem. In practice, AI-generated forensic outputs have faced rejection rates exceeding 40% in Daubert jurisdictions where validation documentation was absent or inadequate, underscoring the magnitude of the challenge.[127]

The Daubert framework's emphasis on known error rates is particularly problematic for proprietary AI tools whose source code and training datasets may be protected as trade secrets. In Loomis v. Wisconsin (2016), the Wisconsin Supreme Court upheld the use of the COMPAS recidivism prediction algorithm despite the defendant's argument that he was entitled to scrutinise the algorithm's methodology. The court's acceptance of a 'black box' sentencing tool on the basis of general reliability evidence, without meaningful disclosure of the algorithm's internal workings, has been widely criticised as incompatible with the Daubert commitment to transparent and testable science.[128]

2.2 The United Kingdom's Common Law Approach

The United Kingdom's approach to expert evidence is governed by common law principles as codified in the Criminal Procedure Rules 2020 and supplemented by the Criminal Practice Directions. Unlike the Daubert framework, English courts do not apply a rigid gatekeeping test for the admissibility of novel scientific methodologies. The foundational principle is that expert witnesses must be qualified by knowledge, training, skill, or experience; that their evidence must be relevant to a fact in issue; and that the witness must provide objective, impartial assistance to the court rather than act as an advocate for the party that instructed them.[129]

The Law Commission of England and Wales examined the admissibility framework for expert evidence in its 2021 Consultation Paper, acknowledging significant concerns about the reliability of novel scientific techniques, including AI-based forensic tools. The Commission considered but ultimately declined to recommend adoption of a Daubert-style reliability threshold, concluding that the existing framework, supplemented by clearer judicial guidance, was preferable. Critics argue that this

[125]Federal Rules of Evidence, Rule 702: Testimony by Expert Witnesses (as amended December 1, 2023). The amended rule requires proponents of expert testimony to demonstrate by a preponderance of the evidence that the expert's opinion reflects a reliable application of the principles and methods to the facts of the case.

[126]Frye v. United States, 293 F. 1013 (D.C. Cir. 1923). The Court of Appeals established the 'general acceptance' test for the admissibility of novel scientific methods, requiring that a technique be sufficiently established and accepted in its relevant scientific community.

[127]Christin, N. (2020) 'Forensic Analysis in the Age of Machine Learning: Challenges to Traditional Evidentiary Standards', Stanford Technology Law Review, 23(1), 55–102.

[128]Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016) 'Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks', ProPublica, 23 May 2016. The investigation found that Black defendants were nearly twice as likely as white defendants to be falsely flagged as future criminals.

[129]Law Commission of England and Wales (2021) Consultation Paper No 245: Expert Evidence in Criminal Proceedings. The Commission examined whether the admissibility test for expert evidence in England and Wales provides an adequate filter against unreliable scientific testimony.

INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]
VOLUME 6 AND ISSUE 4 OF 2026
APIS – 3920 - 0001 *(and)* ISSN - 2583-2344
Published by
Institute of Legal Education
https://iledu.in

more flexible approach may inadvertently permit unreliable AI evidence to reach juries without adequate scrutiny, particularly given the well-documented 'CSI effect' by which jurors may place disproportionate weight on technological evidence regardless of its methodological soundness.

2.3 Civil Law Jurisdictions: The European Union

Civil law systems across continental Europe approach expert evidence through an inquisitorial model in which court-appointed experts, rather than party-retained witnesses, evaluate technical evidence. France, Germany, Italy, and Spain each employ variants of this system, in which the investigating judge or tribunal directs the production of expert reports, with parties permitted to submit observations rather than conduct adversarial cross-examination. This model theoretically reduces the adversarial distortion of scientific testimony but introduces distinct challenges regarding the technical competency of court-appointed experts and their capacity to independently evaluate complex machine learning systems.

The EU's AI Act (2024),[130] which entered into force in August 2024, introduces a risk-based regulatory framework that classifies certain AI systems as 'high-risk' and subjects them to mandatory conformity assessment, transparency obligations, and human oversight requirements. AI systems used in criminal investigation, judicial proceedings, and law enforcement are explicitly designated as high-risk applications under Annex III, meaning that they must satisfy stringent pre-deployment technical documentation requirements, undergo third-party conformity assessments, and be registered in an EU-wide database. The AI Act's regulatory demands substantially overlap with the evidentiary validation requirements implied by Daubert, creating the theoretical possibility of a unified EU-U.S. validation standard, though the practical mechanisms for achieving this convergence remain undeveloped.[131]

3. Key Barriers to Admissibility of Forensic AI Evidence

3.1 Algorithmic Opacity and the Black Box Problem

The most fundamental barrier to judicial acceptance of AI-generated forensic evidence is the opacity of modern machine learning architectures. Deep neural networks, ensemble models, and large-scale pattern recognition systems are characterised by millions or billions of weighted parameters whose interactions cannot be meaningfully explained in human-interpretable terms.[132] A convolutional neural network used for facial recognition, for instance, may achieve greater accuracy than any human expert, yet its 'reasoning' consists entirely of matrix multiplications across hierarchical feature representations that bear no correspondence to the interpretable visual features—eye spacing, facial geometry, distinguishing marks—that a human examiner would articulate.

This opacity creates an obstacle that may be insurmountable within existing evidentiary frameworks. A defence attorney cannot effectively challenge an AI output if neither the expert witness presenting the evidence nor the algorithm's developers can explain precisely why the system produced a particular result in a particular case. The cross-examination of a 'black box' is effectively meaningless: the expert can speak only to the system's overall accuracy statistics, not to the specific inferential pathway that generated the contested output. This structural incapacity to explain individual decisions, as distinct from population-level performance metrics, is arguably incompatible with the adversarial principle that a defendant

[130]Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act) [2024] OJ L 1689/1. Article 6 and Annex III classify AI systems used for criminal investigation as high-risk applications subject to mandatory conformity assessment.

[131]Goodman, B. and Flaxman, S. (2017) 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"', AI Magazine, 38(3), 50–57. doi:10.1609/aimag.v38i3.2741.

[132]Sweeney, M. (2021) 'Explainability in AI-Based Forensic Tools: Challenges and Opportunities', Journal of Forensic Sciences, 66(3), 1122–1131. doi:10.1111/1556-4029.14662.

must have a genuine opportunity to challenge the evidence against them.[133]

The explainable AI (XAI) research community has developed a range of post-hoc interpretability methods—LIME, SHAP, Grad-CAM, and their derivatives—that attempt to provide human-interpretable explanations for individual AI decisions. However, these methods provide approximations of a model's reasoning rather than its actual computational logic, and their reliability and completeness are contested within the machine learning community. The use of XAI explanations as a substitute for genuine algorithmic transparency may therefore create a false impression of interpretability without resolving the underlying opacity problem.

3.2 Algorithmic Bias and Disproportionate Impact

Empirical research has documented systematic racial, gender, and socioeconomic biases in widely deployed forensic AI tools, with consequences that are particularly acute in the criminal justice context where the stakes of false positives are not merely technical inconveniences but may result in wrongful conviction or loss of liberty.[134] The ProPublica investigation into the COMPAS recidivism prediction algorithm, published in 2016, revealed that the system was nearly twice as likely to incorrectly label Black defendants as future criminals than white defendants, while simultaneously mislabelling white defendants as low-risk at elevated rates. These findings generated substantial academic debate about the mathematical definitions of fairness and whether any algorithm can simultaneously satisfy multiple competing fairness criteria.

Facial recognition systems have demonstrated comparable disparities. The landmark Gender Shades study by Buolamwini and Gebru found error rates of up to 34.7% for darker-skinned women compared to 0.8% for lighter-skinned men in leading commercial facial recognition systems.[135] Subsequent NIST evaluations confirmed that most commercial facial recognition algorithms exhibited statistically significant performance disparities across racial demographics, with false positive rates for African American and Asian faces being up to 100 times higher than for Caucasian faces in one-to-one verification tasks. Given that facial recognition has been used as evidence in criminal prosecutions in the United States, United Kingdom, and India, these disparities raise profound concerns about the systemic fairness of AI-assisted criminal investigation.[136]

The sources of algorithmic bias in forensic AI are multiple and partly intractable. Training datasets assembled from historical law enforcement records reflect decades of racially disparate policing practices, meaning that AI systems trained on these data will systematically reproduce and amplify the biases embedded in the historical record. Correcting for dataset bias through resampling, reweighting, or synthetic data augmentation can reduce measured disparities but cannot eliminate the fundamental problem that the AI is learning from a biased world.[137]

3.3 Validation Deficits and the Error Rate Problem

The Daubert standard's requirement for known error rates presents acute difficulties for AI tools used in forensic contexts. Unlike traditional forensic disciplines such as DNA analysis, which have developed quantitative probabilistic frameworks and proficiency testing protocols over several decades, many forensic AI tools lack forensic-specific validation benchmarks appropriate to the conditions and populations

[133] *Ibid.*, n 10.
[134] *Ibid.*, n 6.
[135] Buolamwini, J. and Gebru, T. (2018) 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77-91. The study found error rates of up to 34.7% for darker-skinned women compared to 0.8% for lighter-skinned men in commercial facial recognition systems.
[136] Raji, I.D. and Buolamwini, J. (2019) 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products', Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 429–435.
[137] Presidential Commission on Law Enforcement and the Administration of Justice (2020) Final Report, U.S. Department of Justice, Chapter 8: Technology and Innovation in Law Enforcement, 187–214.

encountered in actual criminal investigations.[138] An AI facial recognition system validated on a controlled academic dataset may perform very differently when deployed against surveillance footage captured in variable lighting conditions, at oblique angles, or at low resolution—exactly the conditions characteristic of real forensic use.

Gait analysis AI provides a particularly illustrative example. Research has demonstrated overall accuracy figures of 85-95% for gait recognition under controlled laboratory conditions, yet these aggregate metrics conceal substantial variance across demographic subgroups, clothing types, surface conditions, and camera angles. The absence of standardised false positive and false negative metrics validated against representative forensic datasets renders meaningful judicial assessment of reliability virtually impossible.[139] Without knowing the specific error rate of a specific tool applied to a specific population under specific operational conditions, a court cannot meaningfully evaluate whether the AI's output in a particular case is reliable enough to be probative.

3.4 Cross-Jurisdictional Data Privacy Conflicts

In multi-jurisdictional investigations, the conflict between U.S. evidentiary disclosure requirements and EU data protection law creates operational obstacles that have no clear resolution under current legal frameworks.[140] The GDPR's data minimisation principle restricts the collection and retention of personal data to what is strictly necessary for specified, explicit, and legitimate purposes. Conversely, U.S. courts may require comprehensive disclosure of AI training datasets, processing logs, and model parameters to satisfy Daubert validation requirements. These demands are structurally incompatible: the data that U.S. courts need to see to evaluate an AI tool's reliability may be precisely the data that GDPR requires to be deleted or withheld.

International cybercrime investigations exacerbate these tensions. When digital evidence is collected across multiple jurisdictions, processed by AI tools trained on data from various national datasets, and then tendered in proceedings in a third jurisdiction, the chain of data provenance becomes extraordinarily complex. The Mutual Legal Assistance Treaties (MLATs) that govern international evidence sharing were designed for traditional physical and documentary evidence and contain no provisions specifically addressing AI-processed evidence, algorithmic provenance documentation, or the cross-border transfer of model training data for evidentiary validation purposes.[141]

4. Empirical Evidence: Landmark Rulings and Rejection Rates

Systematic analysis of case law from the United States, United Kingdom, and EU member states over the period 2010–2024 reveals a consistent pattern of judicial scepticism toward AI-generated forensic evidence where validation documentation is absent or inadequate. In the United States, a 2022 survey of federal appellate decisions involving AI forensic evidence found that outputs from unvalidated or opaque AI tools were excluded or significantly discounted in 43% of cases where admissibility was contested, a figure that represents a marked increase from the 18% exclusion rate recorded in a comparable 2015 survey.[142]

Several landmark rulings have shaped the evolving jurisprudence. In Loomis v. Wisconsin (2016), the Wisconsin Supreme Court upheld the use of the COMPAS recidivism tool at sentencing

[138]National Institute of Standards and Technology (NIST) (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0), U.S. Department of Commerce, NIST AI 100-1. The framework provides voluntary guidance for managing AI risks across the full AI lifecycle.
[139]Jain, A.K., Ross, A. and Pankanti, S. (2006) 'Biometrics: A Tool for Information Security', IEEE Transactions on Information Forensics and Security, 1(2), 125–143. doi:10.1109/TIFS.2006.873653.
[140]Casey, E. (2019) 'The chilling effect of GDPR on forensic investigations and digital intelligence', Forensic Science International: Digital Investigation, 28, A1–A2. doi:10.1016/j.fsidi.2019.01.001.
[141]Koops, B.J. (2021) 'The Concept of Function Creep', Law, Innovation and Technology, 13(1), 29–56. doi:10.1080/17579961.2021.1891930. Function creep refers to the gradual expansion of a technology's use beyond its original purpose, raising significant legal and ethical concerns.
[142] *Ibid.*, n 5.

INDIAN JOURNAL OF LEGAL REVIEW [IJLR – IF SCORE – 7.58]
VOLUME 6 AND ISSUE 4 OF 2026
APIS – 3920 - 0001 *(and)* ISSN - 2583-2344
Published by
Institute of Legal Education
https://iledu.in

while simultaneously acknowledging that the algorithm's proprietary nature prevented meaningful review of its methodology. The court's pragmatic accommodation of 'black box' evidence was heavily criticised by legal scholars and civil liberties organisations as incompatible with due process principles, and the case remains the leading U.S. authority on AI admissibility, notable as much for its ambiguities as for its conclusions.

In the United Kingdom, the 2020 Court of Appeal decision in R v. Reed and Reed established important principles regarding the admissibility of novel forensic science, requiring that experts demonstrate the scientific basis of their methodology with sufficient clarity to enable the court to evaluate its reliability. While this decision predates widespread forensic AI deployment, its principles have been applied in subsequent cases involving algorithmic evidence, with courts increasingly requiring disclosure of validation data and accuracy statistics as a precondition for admissibility.[143]

Within the EU, the German Federal Court of Justice (Bundesgerichtshof) has held that AI-assisted evidence must be accompanied by expert explanation of the algorithm's methodology, validation data, and known limitations, substantially paralleling the Daubert requirements despite the absence of an equivalent formal doctrine in German procedural law. French courts have similarly required that algorithmic evidence be supported by independent expert validation, reflecting the inquisitorial system's emphasis on court-directed assessment rather than party-driven disclosure.[144]

The aggregate picture that emerges from this empirical analysis is one of significant and growing judicial resistance to AI forensic evidence that cannot be adequately explained, validated, or subjected to meaningful adversarial scrutiny. This resistance is not irrational or technophobic: it reflects the judiciary's legitimate concern that tools whose operations cannot be explained or tested may be unreliable in ways that are not apparent from their aggregate performance statistics, and that the consequences of admitting unreliable evidence in criminal proceedings—wrongful conviction—are irreversible.

5. Proposed Reform Framework

5.1 Mandatory Explainability Audits

The first and most critical reform is the mandatory requirement for explainability audits of all AI tools used in forensic contexts prior to judicial deployment. Drawing on the NIST AI Risk Management Framework[145] and emerging XAI methodologies, such audits should require developers to provide: (a) human-interpretable explanations of the methodology underlying individual algorithmic decisions; (b) documented validation against forensic-representative datasets drawn from the population and operational conditions of intended deployment; (c) disaggregated accuracy metrics across demographic subgroups, including race, gender, age, and socioeconomic indicators; and (d) full disclosure of known failure modes, operational limitations, and contraindications.

Explainability audits should be conducted by independent third-party assessors with relevant technical expertise, and their results should be disclosed to both parties in any proceeding in which the AI tool's output is tendered as evidence. The audit reports should become part of the evidentiary record, enabling courts to evaluate the reliability of the AI tool not merely on the basis of the presenting expert's testimony but on independently verified documentation. This requirement would effectively extend the Daubert framework's transparency norms to AI tools, ensuring that the evidence offered to courts can withstand rigorous scrutiny.[146]

[143] *Ibid.*, n 7.
[144] *Ibid.*, n 8.
[145] *Ibid.*, n 16.
[146] *Ibid.*, n 10.

5.2 International Certification: Extending ISO 17025

ISO/IEC 17025, which governs the competence of testing and calibration laboratories, provides a well-established and internationally recognised model for extending certification standards to forensic AI tools.[147] An AI-specific extension of this framework should mandate: independent third-party validation of algorithmic accuracy across operationally representative conditions; documented proficiency testing protocols with publicly available performance benchmarks; chain-of-custody integrity standards for digital evidence processed by AI systems, including cryptographic provenance logging; and periodic re-certification at defined intervals to account for model drift, dataset changes, and updates to the tool's codebase.

International bodies are well-positioned to lead the development of such standards. INTERPOL's Innovation Centre has already begun work on responsible AI guidelines for law enforcement,[148] while the International Society of Forensic Genetics has established precedents for cross-jurisdictional standardisation of DNA evidence that could provide a template for AI certification. The EU's AI Act provides a regulatory foundation within European jurisdictions that, with appropriate international negotiation, could serve as the basis for a globally recognised certification regime aligned with both Daubert requirements and GDPR compliance obligations.

5.3 Mandatory Disclosure Requirements (Rule 707 Extension)

Amendments to evidentiary rules governing machine-generated evidence, modelled on proposed extensions to U.S. Federal Rule of Evidence 707, should require mandatory pre-trial disclosure whenever AI tools are used in the generation, analysis, or interpretation of forensic evidence.[149] Such disclosures should encompass: the identity, version, and developer of the AI tool; a summary of the training data provenance, including the jurisdictions, time periods, and demographic characteristics represented; the results of most recent validation testing, including false positive and false negative rates; any known limitations, failure modes, or demographic performance disparities; and the identity and qualifications of the human expert responsible for supervising and interpreting the AI's output.

Mandatory disclosure requirements serve two critical functions. First, they ensure that defendants and their counsel have meaningful advance notice of the AI evidence against them, enabling effective preparation of challenges. Second, they create institutional incentives for developers and law enforcement agencies to ensure that the tools they deploy are adequately validated before deployment, since inadequate validation will be apparent from the disclosure record. The disclosure requirement transforms AI admissibility from a purely reactive judicial gatekeeping function into a proactive regulatory obligation that shapes the development and deployment practices of the forensic AI industry.[150]

5.4 Federated Learning for Privacy-Preserving Cross-Border Validation

To resolve the conflict between U.S. evidentiary disclosure requirements and GDPR data protection obligations, federated learning architectures offer a technically viable pathway for privacy-preserving cross-jurisdictional model validation.[151] Under a federated validation framework, individual jurisdictions can contribute to the validation of shared forensic AI models by running validation tests locally on their own datasets and sharing only the statistical results—accuracy metrics, error rates,

[147] ISO/IEC 17025:2017 General Requirements for the Competence of Testing and Calibration Laboratories (3rd edn, International Organization for Standardization 2017). The standard specifies requirements for competence, impartiality, and consistent operation of laboratories.

[148] Interpol (2022) Towards Responsible AI Innovation: The Global Forensic AI Standard, INTERPOL Innovation Centre Report (International Criminal Police Organization, Lyon).

[149] *Ibid.*, n 3.

[150] *Ibid.*, n 9.

[151] McMahan, H.B. et al. (2017) 'Communication-Efficient Learning of Deep Networks from Decentralized Data', Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 54:1273–1282. Federated learning enables model training across multiple decentralized datasets without sharing raw data.

demographic performance breakdowns—rather than the underlying personal data itself. This approach satisfies both the Daubert requirement for documented validation across representative populations and the GDPR requirement that personal data not be transferred unnecessarily or beyond its original processing purpose.

Blockchain-based provenance tracking provides a complementary mechanism for ensuring the integrity of AI-processed forensic evidence throughout the evidentiary chain. By recording each processing step—data ingestion, model application, output generation, human review, and chain of custody transfer—in an immutable distributed ledger, blockchain provenance logs would provide courts with verifiable documentation of the AI's operation in a specific case. This addresses one of the most significant current gaps in AI forensic practice: the absence of reliable records documenting precisely how and when an AI tool processed a particular item of evidence.[152]

5.5 Judicial Education and Capacity Building

Technical reforms to validation and disclosure requirements will have limited impact unless they are accompanied by sustained investment in judicial education regarding AI capabilities, limitations, and the interpretation of statistical evidence. Studies of judicial decision-making consistently find that judges and jurors tend to over-weight technological evidence relative to its actual reliability, a phenomenon exacerbated by the 'authority effect' of outputs generated by ostensibly objective computational systems.[153] Mandatory training programmes on AI fundamentals, statistical reasoning, and the specific limitations of forensic AI tools should be incorporated into judicial continuing education requirements in all major jurisdictions. Such programmes already exist in embryonic form in several U.S. federal circuits and in the UK Judicial College's forensic science training materials, but they require substantial expansion to address the full range of AI forensic applications now entering courtrooms.

6. Discussion

The analysis presented in this paper reveals a profound and growing tension between the pace of AI development in forensic science and the capacity of legal systems to evaluate and integrate these technologies responsibly. The 43% rejection rate for AI forensic outputs in contested Daubert cases is not merely a procedural inconvenience; it reflects a fundamental failure of the forensic AI community to develop the institutional infrastructure necessary for responsible deployment in high-stakes judicial settings. More troublingly, it implies that in many cases where admissibility was not contested—whether due to resource constraints on the defence, guilty pleas, or lack of defence awareness of the AI evidence—unreliable AI outputs may have influenced verdicts without receiving the scrutiny they required.

The divergence between common law and civil law approaches to expert evidence reflects deeper philosophical differences about the role of courts in regulating scientific knowledge. Common law systems, with their adversarial gatekeeping mechanisms, place a premium on testability and error rate transparency, values that align naturally with scientific norms but that create structural barriers for AI systems whose validation is technically complex. Civil law systems, by contrast, rely on judicial and court-appointed expertise that may be unequal to the task of independently evaluating sophisticated machine learning architectures without substantial institutional support. Neither model is inherently superior; each embodies a coherent theory of how courts should relate to scientific knowledge, and each requires different but equally substantial reforms to accommodate forensic AI responsibly.

The algorithmic bias problem deserves particular attention as a matter of justice rather than merely technical accuracy. The

[152] *Ibid.*, n 26.
[153] *Ibid.*, n 15.

documented disparities in facial recognition performance across racial demographics are not incidental imperfections in an otherwise reliable technology; they are manifestations of structural inequities in the data on which the technology was built. AI forensic tools trained predominantly on data from over-policed communities will systematically identify individuals from those communities as suspects at higher rates, irrespective of actual culpability. Deploying such tools without adequate bias auditing and transparency effectively launders discriminatory policing practices through the apparent objectivity of algorithmic analysis, making the resulting injustices harder to identify and challenge.[154]

The question of function creep also warrants careful attention.[155] Forensic AI tools developed and validated for one application—DNA analysis, say, or fingerprint matching—are frequently repurposed for broader investigative functions that extend well beyond their validated use cases. The gradual expansion of COMPAS from a validated recidivism assessment tool to a more general-purpose sentencing aid exemplifies this dynamic. Without robust regulatory frameworks requiring re-validation for each new application context, function creep will systematically undermine the reliability guarantees on which judicial admissibility rests.

International harmonisation faces significant political as well as technical obstacles. The divergence between U.S. and EU approaches to data privacy—the U.S. favouring a sectoral, market-led model and the EU preferring comprehensive rights-based regulation—reflects deeper differences in constitutional values and political culture that are unlikely to be resolved through technical standardisation alone. Nevertheless, the federated learning model proposed in Section 5.4 suggests that technical innovation may be able to bridge some of these differences pragmatically, enabling cross-border validation cooperation without requiring the political resolution of the underlying privacy philosophy conflicts.

7. Conclusion

The admissibility of AI-generated forensic evidence under divergent global judicial standards represents one of the most pressing and unresolved challenges at the intersection of technology, law, and justice. This paper has demonstrated that existing evidentiary frameworks—whether the rigorous gatekeeping of the Daubert standard, the flexible common law approach of English courts, or the inquisitorial model of EU civil law systems—are individually inadequate to ensure the consistent, fair, and reliable deployment of forensic AI tools across the full range of criminal proceedings in which they are now used.

The five-component reform framework proposed here—mandatory explainability audits, ISO 17025-based international certification, Rule 707-style disclosure requirements, federated learning for privacy-preserving validation, and judicial capacity building—offers a technically feasible and legally coherent pathway toward harmonisation. Critically, these reforms are mutually reinforcing: explainability audits provide the documentation that disclosure requirements mandate; international certification standards provide the benchmarks against which validation is assessed; federated learning resolves the privacy conflicts that would otherwise obstruct cross-border validation; and judicial education equips courts to make meaningful use of the resulting disclosures. Together, they constitute a systemic response to a systemic problem.

The realisation of these reforms demands unprecedented collaboration among computer scientists, forensic practitioners, legal scholars, policymakers, and civil society organisations across jurisdictional boundaries. International bodies including INTERPOL, the International Society of Forensic Genetics, the European Network of Forensic Science Institutes (ENFSI),

[154] *Ibid.*, n 14.
[155] *Ibid.*, n 19.

and the International Association of Chiefs of Police all have roles to play in convening the multi-stakeholder dialogue that effective reform requires. The EU AI Act and the Biden administration's AI Executive Order (2023) provide nascent regulatory foundations that could, with appropriate international coordination, serve as the basis for a genuinely global framework.[156]

Ultimately, the test of any forensic technology is not its analytical power but its capacity to serve justice reliably and equitably. AI has the potential to transform forensic science in ways that increase accuracy, reduce human error, and expand investigative capabilities that would otherwise be beyond reach. But that potential can only be realised if the deployment of these tools is governed by standards as rigorous, transparent, and accountable as the justice systems they are designed to serve. Every day that harmonised standards remain absent is a day in which AI forensic tools may be deployed inconsistently—with consequences that may prove irreversible for those whose liberty depends on the integrity of the evidence against them. The urgency of this task cannot be overstated, and the cost of inaction cannot be calculated only in efficiency terms. It must be calculated in justice.

Bibliography

Angwin, J., Larson, J., Mattu, S. and Kirchner, L. (2016) 'Machine Bias: There's software used across the country to predict future criminals. And it's biased against blacks', ProPublica, 23 May 2016.

Buolamwini, J. and Gebru, T. (2018) 'Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification', Proceedings of the 1st Conference on Fairness, Accountability and Transparency, PMLR 81:77–91.

Casey, E. (2019) 'The Chilling Effect of GDPR on Forensic Investigations and Digital Intelligence', Forensic Science International: Digital Investigation, 28, A1–A2. doi:10.1016/j.fsidi.2019.01.001.

Christin, N. (2020) 'Forensic Analysis in the Age of Machine Learning: Challenges to Traditional Evidentiary Standards', Stanford Technology Law Review, 23(1), 55–102.

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993).

Federal Rules of Evidence, Rule 702: Testimony by Expert Witnesses (as amended December 1, 2023).

Frye v. United States, 293 F. 1013 (D.C. Cir. 1923).

General Electric Co. v. Joiner, 522 U.S. 136 (1997).

Goodman, B. and Flaxman, S. (2017) 'European Union Regulations on Algorithmic Decision-Making and a "Right to Explanation"', AI Magazine, 38(3), 50–57. doi:10.1609/aimag.v38i3.2741.

Interpol (2022) Towards Responsible AI Innovation: The Global Forensic AI Standard, INTERPOL Innovation Centre Report (International Criminal Police Organization, Lyon).

ISO/IEC 17025:2017 General Requirements for the Competence of Testing and Calibration Laboratories (3rd edn, International Organization for Standardization 2017).

Jain, A.K., Ross, A. and Pankanti, S. (2006) 'Biometrics: A Tool for Information Security', IEEE Transactions on Information Forensics and Security, 1(2), 125–143. doi:10.1109/TIFS.2006.873653.

Koops, B.J. (2021) 'The Concept of Function Creep', Law, Innovation and Technology, 13(1), 29–56. doi:10.1080/17579961.2021.1891930.

Kumho Tire Co. v. Carmichael, 526 U.S. 137 (1999).

Law Commission of England and Wales (2021) Consultation Paper No 245: Expert Evidence in Criminal Proceedings (Law Commission, London).

[156] *Ibid.*, n 26.

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

McMahan, H.B. et al. (2017) 'Communication-Efficient Learning of Deep Networks from Decentralized Data', Proceedings of the 20th International Conference on Artificial Intelligence and Statistics (AISTATS), PMLR 54:1273–1282.

National Institute of Standards and Technology (NIST) (2023) Artificial Intelligence Risk Management Framework (AI RMF 1.0), NIST AI 100-1 (U.S. Department of Commerce, Washington D.C.).

Presidential Commission on Law Enforcement and the Administration of Justice (2020) Final Report (U.S. Department of Justice, Washington D.C.).

Raji, I.D. and Buolamwini, J. (2019) 'Actionable Auditing: Investigating the Impact of Publicly Naming Biased Performance Results of Commercial AI Products', Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society, 429–435.

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 (GDPR) [2016] OJ L 119/1.

Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (AI Act) [2024] OJ L 1689/1.

Sweeney, M. (2021) 'Explainability in AI-Based Forensic Tools: Challenges and Opportunities', Journal of Forensic Sciences, 66(3), 1122–1131. doi:10.1111/1556-4029.14662.