Governing Algorithmic Evidence: A Comparative Legal Framework for AI-Based Forensic Systems Under the EU AI Act and U.S. Admissibility Standards*

Romana Afroze[1,*]

[1] SJD Candidate, Delaware Law School, 4601 Concord Pike, Wilmington, DE 19803, USA

Abstract

Artificial intelligence systems are increasingly deployed in forensic and law enforcement contexts, from facial recognition and probabilistic genotyping to acoustic gunshot detection, digital evidence reconstruction, and large-language-model-assisted case analysis. Yet the legal frameworks governing the admissibility, transparency, and accountability of such systems have not kept pace with their technical capabilities or societal consequences. This paper conducts a comparative legal analysis of the two most significant regulatory regimes now shaping forensic AI governance: the EU Artificial Intelligence Act (Regulation (EU) 2024/1689), which entered into force on 1 August 2024 and whose high-risk provisions are in their final compliance window as of May 2026, and the United States standards for expert evidence admissibility, principally the Daubert standard under Federal Rule of Evidence 702. The paper argues that the EU AI Act's classification of biometric and criminal justice AI as "high-risk" and its imposition of mandatory conformity assessments, technical documentation, bias-management requirements, and human oversight duties constitute a structurally more coherent framework for evidential integrity than the U.S.'s ad hoc, judicially managed Daubert regime. The paper identifies critical convergence points and proposes an original Forensic AI Evidence Standard (FAES): a four-pillar framework comprising Technical Transparency Disclosure, Computational Audit Trails, Human Oversight Attestation, and a Cross-Jurisdictional Mutual Recognition Protocol.

Keywords

forensic AI, EU AI Act, Daubert standard, algorithmic evidence, comparative AI law, evidential admissibility, law enforcement AI, explainability, high-risk AI systems

1. Introduction

The integration of artificial intelligence into forensic science and law enforcement has produced a governance paradox. AI-based tools promise unprecedented capabilities in pattern recognition, evidence analysis, risk stratification, and investigative triage; yet they are characterised by epistemic features (opacity, data-dependency, demographic variability, and resistance to traditional cross-examination) that challenge the foundational principles of due process, evidential reliability, and equality before the law on which legitimate criminal justice systems rest.

Courts across the United States, the European Union, and beyond increasingly confront AI-generated evidence yet lack clear standards for evaluating its validity, reproducibility, or susceptibility to bias. Facial recognition matches from proprietary algorithms trained on skewed data, probabilistic genotyping likelihood ratios shielded by trade secrecy, and contested recidivism risk scores are no longer hypothetical; they are active features of criminal proceedings across multiple jurisdictions, yet they routinely escape the principled epistemic scrutiny that the law demands of other forms of expert scientific evidence.

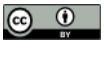
This paper addresses this governance deficit through a structured comparative legal analysis of two regulatory regimes. First, the EU Artificial Intelligence Act (EU AI Act) [1], which entered into force on 1 August 2024, has its high-risk AI provisions currently in their final compliance window

* *RELAF'26: Reasoning with Evidence in Law Enforcement and Forensics, Workshop at ICAIL 2026, Singapore.*

1* Corresponding author.

rafroze@widener.edu (R. Afroze)

0009-0000-1585-5464 (R. Afroze)



ahead of the August 2026 application. Second, the U.S. Federal Rules of Evidence, particularly FRE 702 and the Daubert trilogy [2–4]. The paper proceeds as follows: Section 2 surveys the AI forensics landscape and structural legal challenges; Section 3 examines the EU AI Act's architecture; Section 4 analyses Daubert doctrine; Section 5 conducts a comparative analysis; Section 6 develops the FAES; Section 7 concludes.

2. AI in Forensics and Law Enforcement: Landscape and Legal Challenges

2.1. Principal Deployment Modalities

Forensic AI systems are best understood through two analytical axes: their operational function (identification, prediction, reconstruction, or classification) and their evidentiary role (primary evidence, corroborating tool, or investigative triage mechanism). This taxonomy matters because different deployment modalities raise distinct legal concerns, and different aspects of the regulatory frameworks examined here are relevant to each modality. As Section 6 develops, it also provides the basis for calibrating the proposed Forensic AI Evidence Standard to the decisional weight of a given AI output.

Automated facial recognition technologies (FRT) are deployed by police forces in the United States and the EU for retrospective identification from CCTV footage. Research has consistently demonstrated significantly elevated false-positive rates for individuals with darker skin tones, for women, and for older persons [5]. The 2020 wrongful arrest of Robert Williams in Detroit, attributable to a facial recognition misidentification, illustrates the real-world consequences of these disparities [6].

Probabilistic genotyping (PG) platforms such as STRmix, TrueAllele, and ArmedXpert apply Bayesian models to compute likelihood ratios for complex, low-template, or mixed-contributor DNA profiles [7]. PG outputs are probabilistic, dependent on assumed population-genetics parameters, and often shielded from defence examination by trade secrecy claims [8]. The stakes are acute: PG outputs have been introduced in capital cases, yet the underlying methodology has, in some instances, been ruled non-discoverable.

Algorithmic risk assessment instruments (ARAIs), most prominently COMPAS, generate recidivism risk scores used in pretrial detention, sentencing, and parole determinations. Research has documented that COMPAS assigns systematically higher risk scores to Black defendants than to white defendants with equivalent recidivism histories [9]. Acoustic gunshot detection systems such as ShotSpotter triangulate gunshot events, producing outputs introduced in criminal proceedings despite documented false-positive rates and concerns about recalibration at the request of law-enforcement clients [10].

Most recently, large language model (LLM) systems have begun to appear in law enforcement contexts: for automated case summarisation, digital forensic report generation, suspect interview analysis, and intelligent document review in criminal defence. The governance deficit, absence of transparency requirements, audit obligations, and error accountability, is structurally identical across all these categories, even where the evidentiary mechanisms differ [11]. As Section 6.7 explains, however, the non-deterministic character of generative systems raises a distinct reproducibility problem that the audit-trail machinery developed below must address directly.

2.2. Structural Legal Challenges

The opacity problem. Most high-performing forensic AI systems rely on architectures, deep neural networks, gradient-boosted ensembles, and transformer-based LLMs, whose internal decision logic cannot be reduced to the kind of transparent, step-by-step methodology that adversarial cross-examination is designed to scrutinise [12]. This is constitutionally significant: the defendant's right to confront and challenge evidence is a foundational due process guarantee.

The auditability deficit. The absence of standardised audit trail requirements means there is no reliable mechanism for defence counsel or independent experts to reconstruct the computational chain of custody between raw data input and evidentiary output. Physical forensic evidence is subject to strict chain-of-custody protocols; AI-generated evidence typically is not, even though intermediary processing steps like data cleaning, feature extraction, model version selection, and threshold calibration can be outcome-determinative.

The bias amplification risk. AI systems trained on historically skewed datasets systematically disadvantage protected groups, as documented across facial recognition [5], risk assessment [9], and predictive policing [13]. Unlike human expert bias, which is subject to cross-examination and credibility assessment, algorithmic bias is embedded in model weights that are neither visible nor cross-examinable.

The versioning and reproducibility problem. The version of a system validated in peer-reviewed research may differ from the version deployed in the operational context that produced specific evidence. Without version-pinning disclosure requirements, courts cannot verify that validated performance characteristics apply to the actual evidence before them.

The cross-jurisdictional fragmentation problem. In the absence of harmonised standards, defendants may be subject to AI-generated evidence admitted in one jurisdiction and excluded in another, based on inconsistent applications of general-purpose admissibility doctrine rather than principled AI-specific criteria.

3. The EU AI Act Framework Applied to Forensic AI Systems

3.1. Legislative Status and Compliance Timeline

The EU AI Act entered into force on 1 August 2024 [1], following a staggered implementation schedule. Article 5 prohibitions on unacceptable-risk AI became applicable on 2 February 2025. The high-risk obligations under Chapter III analysed here (conformity assessment, technical documentation, bias management, and human oversight) apply from 2 August 2026, so as of this paper (May 2026) providers of high-risk forensic AI systems are in the final compliance window. Ongoing "Omnibus" simplification negotiations have raised the possibility of deferring certain high-risk obligations to 2027, but no such amendment has yet been adopted, and the August 2026 date remains the operative deadline.

3.2. Risk-Based Architecture and High-Risk Designation

The EU AI Act establishes a four-tier risk architecture: unacceptable-risk AI (prohibited), high-risk AI (subject to pre-market conformity assessment and ongoing obligations), limited-risk AI (transparency obligations only), and minimal-risk AI (no specific requirements).

Annex III lists categories of high-risk AI systems directly encompassing forensic and law enforcement AI: Category 1 (real-time and post-remote biometric identification used by law enforcement), Category 6 (crime analysis, prediction, or profiling of natural persons by competent authorities, squarely including ARAIs and predictive policing), and Category 8 (AI used in the administration of justice and legal proceedings) [1].

3.3. High-Risk Obligations as a Forensic Evidence Standard

Article 9 mandates a risk management system operating throughout the AI system's lifecycle, encompassing identification of all known and reasonably foreseeable risks to fundamental rights, including those arising from system misuse or interaction with other systems.

Article 10 requires that training, validation, and testing datasets be subject to appropriate data-governance practices, including examination for biases that could lead to prohibited discrimination or otherwise affect fundamental rights. This imposes an affirmative duty to identify, measure, and

mitigate the demographic-accuracy disparities documented across FRT, PG, and risk-assessment systems.

Article 11 mandates comprehensive technical documentation, kept up to date and sufficient to enable competent authorities to assess compliance, a mandatory disclosure package providing exactly the technical specification that courts applying Daubert currently cannot access in the face of trade-secrecy defences.

Article 12 requires that high-risk AI systems automatically generate logs of their operation to the extent technically feasible, enabling identification of the period of operation, relevant input data, and, for biometric identification systems, the persons involved. These automatic logs form the statutory basis for the Computational Audit Trail requirement proposed in Section 6.

Article 14 requires that high-risk AI systems be designed to enable effective human oversight, including the ability to decline to use the system, override its outputs, and interrupt its operation. For forensic AI, no AI output should be determinative of an evidentiary conclusion without a human expert's substantive engagement and endorsement.

Read together, these obligations are not merely administrative; each carries an evidentiary consequence that this paper argues domestic courts should recognise. The absence of the Article 11 technical documentation should support a presumption against admissibility, because without it neither the court nor the opposing party can assess the system's reliability; an unaddressed Article 10 bias finding should go to the weight of the evidence and trigger the demographic disaggregation requirement developed in Section 6; the absence of Article 12 logs should be treated as a failure of the computational chain of custody; and a deficient Article 14 oversight arrangement should preclude treating the AI output as the substantive basis of an evidentiary conclusion. This interpretive move from regulatory obligation to evidentiary effect mirrors the reading of Article 5 as an upstream exclusionary rule developed in Section 3.5, and it is what converts the EU AI Act from a compliance catalog into a source of admissibility criteria. Because the Act has no direct effect in U.S. proceedings, these consequences are advanced as normative criteria for domestic adoption, most concretely through the Federal Forensic AI Evidence Act proposed in Section 6.6, rather than as operative law binding on U.S. courts.

3.4. Post-Market Monitoring and Incident Reporting

Article 72 requires providers of high-risk AI systems to implement a post-market monitoring plan analysing system performance in deployed conditions, and Article 73 requires them to report serious incidents to national market surveillance authorities within prescribed timelines. A parallel obligation for deployers, such as police departments operating a forensic AI system, to monitor performance and report serious incidents to both the provider and the competent national authority is established under Article 26(5). Together, these provisions create a continuous, multi-party governance cycle with no counterpart in the Daubert framework, which is triggered only by adversarial challenge in individual proceedings.

3.5. Article 5 Prohibitions and the Upstream Exclusionary Rule

Article 5 establishes absolute prohibitions on certain AI practices. Most significantly for law enforcement, it prohibits real-time remote biometric identification in publicly accessible spaces, subject to three strictly circumscribed exceptions, each requiring prior judicial or independent administrative authorisation and subject to temporal and geographic limits: (i) the targeted search for a specific missing person or for victims of abduction, trafficking, or sexual exploitation; (ii) the prevention of a specific and imminent terrorist threat or a substantial and imminent threat to the life or physical safety of persons; and (iii) the localisation or identification of a person suspected of one of the serious offences enumerated in Annex II (including murder, kidnapping, armed robbery, rape, trafficking, and terrorism) [1]. This prohibition constitutes an upstream exclusionary rule grounded in the EU Charter's fundamental rights framework: evidence obtained through a prohibited AI application cannot lawfully be collected.

4. U.S. Standards for AI Evidence Admissibility

4.1. The Daubert Trilogy and the Gatekeeping Function

Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579 (1993), restructured the federal standard for the admissibility of expert scientific testimony [2]. Rejecting the Frye general-acceptance standard as insufficiently rigorous, the Supreme Court assigned to federal district courts an active gatekeeping role, requiring assessment of whether proffered evidence is grounded in a reliable methodology and relevant to the dispute. The non-exhaustive reliability factors (testability, peer review and publication, known or potential error rate, and general acceptance in the relevant scientific community) were refined in *General Electric Co. v. Joiner*, 522 U.S. 136 (1997) [3], and extended to technical and specialised knowledge in *Kumho Tire Co. v. Carmichael*, 526 U.S. 137 (1999) [4]. The 2000 amendment of FRE 702 codified the trilogy's core principles.

4.2. Structural Limitations in the Forensic AI Context

The decisions discussed in this Section are selected to illustrate recurring structural patterns in how U.S. courts have handled forensic AI under Daubert, rather than to provide an exhaustive survey. *State v. Loomis* and the *Chubbs–Foley–Pickett* line are treated as leading appellate authorities on, respectively, algorithmic risk scoring and the discoverability of probabilistic-genotyping source code; *United States v. Gissantaner* and *United States v. Miller* are treated as illustrative of recurring patterns (error-rate review without AI-specific disaggregation), and the recharacterisation of facial-recognition output as a non-testimonial investigative lead. The claim advanced is therefore about a doctrinal pattern, not about the frequency of any single holding.

The testability criterion proves difficult to operationalise for machine learning systems. A deep learning model's "methodology" is not a discrete, falsifiable hypothesis; its testing consists of performance benchmarking against curated datasets, a process that can demonstrate accuracy on the test set without establishing generalisability to real-world operational conditions.

The peer-review criterion is systematically undermined by vendor trade-secret claims, and courts have responded inconsistently. In *People v. Superior Court (Chubbs)* (Cal. Ct. App. 2015) [14] and *Commonwealth v. Foley*, 38 A.3d 882 (Pa. Super. Ct. 2012) [15], courts declined to compel probabilistic genotyping developers to produce source code. More recently, *State v. Pickett*, 466 N.J. Super. 270 (App. Div. 2021) [16] marked a significant doctrinal shift: the New Jersey Appellate Division compelled TrueAllele's developer to disclose its source code to the defence under a protective order. Together, *Chubbs*, *Foley*, and *Pickett* reveal the structural incoherence of relying on adversarial challenge to enforce transparency in the absence of a statutory disclosure floor.

The error rate criterion is applied inconsistently and without AI-specific disaggregation. In *United States v. Gissantaner*, 990 F.3d 457 (6th Cir. 2021) [17], the Sixth Circuit upheld the admissibility of STRmix probabilistic genotyping evidence, reversing the district court's exclusion and finding the software sufficiently reliable notwithstanding defence challenges to its handling of low-template DNA and contributor thresholds, without requiring error rates disaggregated by profile complexity, mixture contributor number, or demographic group.

Courts have also admitted FRT-assisted evidence without resolving whether FRT's cross-demographic accuracy disparities render it unreliable under Daubert. In *United States v. Miller* (E.D. Va. 2021) [18] and subsequent unreported decisions, FRT was treated as an investigative lead "confirmed" by a human detective, a framing that circumvents Daubert scrutiny by recharacterising the AI output as non-testimonial, despite the fact that the human identification was causally downstream of and cognitively anchored to the AI match.

4.3. The COMPAS Problem: Opacity, Due Process, and the Limits of Judicial Gatekeeping

In *State v. Loomis*, 881 N.W.2d 749 (Wis. 2016) [19], the Wisconsin Supreme Court upheld the use of a COMPAS score in sentencing, rejecting the defendant's due process argument on three grounds: the score was not the determinative sentencing factor; the defendant received the score and a general description of the instrument; and the defendant had the opportunity to challenge the score through expert testimony.

Each rationale attracts sustained scholarly criticism. The "not determinative" reasoning ignores cognitive anchoring effects, the well-documented tendency of decision-makers to calibrate subsequent judgments to numeric anchors even where formally instructed otherwise [20]. The "general description" reasoning conflates procedural disclosure with substantive contestability: receiving a score and a marketing description is not equivalent to having access to the data, weights, and validation studies needed to meaningfully challenge the output. The "expert testimony" reasoning assumes that defence counsel has the resources and technical expertise to retain a machine learning expert, an assumption empirically implausible in the vast majority of cases and systematically disadvantaging indigent defendants.

4.4. State-Level Divergence and the Absence of a Federal AI Evidence Standard

In the absence of a federal AI evidence statute, state courts have developed divergent approaches to forensic AI admissibility. Illinois' Artificial Intelligence Video Interview Act (2020) [21] and New York City's Local Law 144 (2021) impose AI transparency requirements in employment contexts without extending analogous protections to the criminal justice domain. The resulting patchwork means that a defendant's ability to challenge AI-generated evidence depends largely on the accident of jurisdiction, an outcome arbitrary from both due process and equal protection perspectives.

5. Comparative Analysis: Divergence, Convergence, and the Harmonisation Opportunity

5.1. Fundamental Structural Divergence

The most fundamental divergence between the EU AI Act and the Daubert framework is temporal and institutional. The Act operates prospectively and administratively, imposing requirements on providers before and during deployment, enforced by National Competent Authorities with investigatory and sanctioning powers; compliance is a precondition of market access. Daubert operates reactively and judicially: triggered at the point of admission in individual proceedings, administered by generalist judges with typically limited AI literacy, and resolved without access to the technical documentation the Act mandates.

These structural differences play out across concrete governance dimensions. On disclosure, the Act imposes mandatory technical documentation (Article 11) and automatic logging (Article 12), whereas under Daubert disclosure is triggered only by adversarial challenge and is routinely defeated by trade-secret shields. On bias, the Act requires dataset examination and ongoing monitoring (Articles 10 and 72), while Daubert contains no explicit bias rule and places the burden on the challenging party. On human oversight, the Act mandates oversight by design (Article 14), whereas FRE 702 imposes no structural requirement. And on error rates, the Act requires a system-level performance metric ex ante, while Daubert treats the error rate as one non-dispositive factor, typically assessed only in the aggregate.

The divergence widens across the remaining dimensions. The Act prohibits certain uses outright, most notably real-time remote biometric identification, permitted only under three narrow exceptions and subject to prior judicial authorization; whereas U.S. law offers no AI-

specific statutory prohibition, reaching such practices only through the Fourth Amendment. On territorial reach, the Act delivers single-market harmonisation across 27 Member States, while the U.S. position remains state-by-state absent any federal AI-evidence statute. Finally, the Act establishes continuous post-deployment governance (provider monitoring under Article 72, incident reporting under Article 73, and a parallel deployer duty under Article 26(5)); Daubert imposes no post-admission monitoring and revisits an admitted output only on appeal.

5.2. Convergence: Transparency as a Shared Core Value

Despite their structural differences, both frameworks converge on transparency as a necessary condition for the legitimate use of forensic AI. The Act's Articles 11–13 transparency obligations and Daubert's peer-review and methodology-disclosure requirements reflect the same epistemic commitment: that AI-generated claims must be derivable from and contestable on the basis of articulable, accessible information about the processes that produced them.

5.3. Comparative Gap Analysis

A final dimension, sanctions, warrants particular emphasis, because the EU AI Act applies a two-tier penalty structure. Under Article 99(4), non-compliance with Chapter III high-risk AI obligations attracts fines of up to €15 million or 3% of total worldwide annual turnover. Under Article 99(3), deployment of a prohibited AI practice under Article 5 attracts a higher penalty of up to €35 million or 7% of global annual turnover. Article 101 governs fines exclusively for providers of general-purpose AI models and is not the relevant provision for the high-risk forensic AI systems examined here. On the U.S. side, there is no analogous monetary exposure: the only sanction is evidentiary exclusion, and, for unconstitutional collection, the Fourth Amendment exclusionary rule, with no statutory fine imposed on developers.

5.4. The Mutual Reinforcement Opportunity

The EU AI Act and Daubert are not merely parallel regimes, they are potentially mutually reinforcing. EU conformity assessment produces exactly the technical documentation, bias-assessment results, and performance metrics that Daubert courts currently lack, while Daubert's adversarial mechanism supplies the case-specific human accountability that the Act's oversight requirements envision but do not fully operationalise at the point of evidentiary deployment.

6. The Forensic AI Evidence Standard (FAES): A Proposed Framework

6.1. Rationale and Design Principles

The Forensic AI Evidence Standard (FAES) responds to the five structural governance failures identified in Section 2. It is not designed to replace either the EU AI Act or Daubert; it operates as a purpose-built evidentiary overlay that translates the Act's compliance architecture into the epistemic vocabulary courts require, supplements Daubert's reliability inquiry with AI-specific operational criteria, and establishes a cross-jurisdictional floor for forensic AI evidence in participating jurisdictions.

FAES rests on five design principles: *disclosure primacy*, every evidentiary AI claim must be accompanied by sufficient information to assess its epistemic basis; *accountability anchoring* (every claim must be endorsed by an identified human expert bearing professional and legal responsibility); *computational continuity* (the chain of custody between raw input and evidentiary output must be reconstructable from an immutable log); *harmonised recognition* (systems validated under one major regulatory regime should not require duplicative validation under another for admissibility purposes); and *proportional calibration* (the stringency with which the four pillars apply scales with the evidentiary role and decisional weight of the AI output).

The calibration principle operationalises the taxonomy introduced in Section 2.1. Where an AI output is offered as substantive or near-determinative evidence, all four pillars apply in full. Where it functions only as a corroborating tool or an investigative lead, a reduced but non-zero subset applies: a Technical Transparency Disclosure and a Human Oversight Attestation remain mandatory, while the full Computational Audit Trail obligation attaches if and when the output is later relied upon as substantive proof. Crucially, the human-oversight concern does not diminish as the AI output recedes from the centre of the case; it intensifies. As the critique of *State v. Loomis* in Section 4.3 shows, the cognitive-anchoring effect means that an output formally treated as "only one factor" may nonetheless drive the outcome. Calibration, therefore, lowers the documentary burden for genuinely low-stakes uses without licensing courts to relax scrutiny of how an AI output influences the human judgement built upon it; the weight a court assigns to AI-assisted evidence offered as one factor among many should be expressly conditioned on the adequacy of the Section 6.4 attestation and on whether the anchoring risk has been addressed.

6.2. Pillar 1 — Technical Transparency Disclosure (TTD)

Any party seeking to introduce AI-generated evidence must produce, as a precondition of admissibility, a Technical Transparency Disclosure (TTD) comprising four mandatory components.

Component A is a plain-language system description: an accessible account of the AI system's purpose, training methodology, input modalities, and output representation, at a level of technical detail sufficient for a competent forensic expert to evaluate its appropriateness for the evidentiary task.

Component B is a disaggregated performance statement: the system's validated error rates (false positive and false negative) presented separately for each relevant demographic subgroup (at minimum: sex, age group, and where statistically reliable, racial or ethnic group), each evidence quality category, and each operational condition category relevant to the deployment context. Aggregate performance metrics without subgroup disaggregation do not satisfy this requirement.

Component C is a version confirmation: a declaration identifying the precise version of the AI system used to produce the evidence, with documentation confirming that this version was the subject of the validated performance data disclosed in Component B. Where the operational version differs in any respect (training data updates, architectural modifications, threshold recalibrations, or software patches), this must be disclosed together with any available data on the effect of those differences on performance.

Component D is a limitations disclosure: a documented inventory of all known failure modes, operational constraints, and unresolved technical concerns, including unmitigated demographic performance gaps, sensitivity to degradation in evidence quality, and conditions under which the system's output should not be relied upon.

The TTD draws directly on the Act's Article 11 technical-documentation mandate, operationalised for the evidentiary context. It simultaneously satisfies Daubert's peer-review and error-rate requirements by giving courts the technical specificity that an independent scientific reviewer would require, without depending on whether the vendor's proprietary methodology has been voluntarily made public.

6.3. Pillar 2 — Computational Audit Trail (CAT)

For any specific item of AI-generated evidence, the producing party must be able to produce, on demand, an immutable, time-stamped log of all processing steps applied to the raw input data to produce the evidentiary output. The CAT must encompass: data ingestion and preprocessing transformations; model version identification and configuration parameters active at the time of processing; intermediate processing steps to the extent technically accessible; the raw output of the AI system prior to any post-processing threshold application; and the threshold or scoring rule applied to convert raw output to the evidentiary claim.

The CAT is the computational analogue of the physical chain of custody, and its absence from current forensic AI practice is a structural deficiency with no counterpart in any other forensic domain. The Act's Article 12 automatic-logging obligations supply the regulatory mandate; FAES operationalises that infrastructure as an evidentiary precondition. Where a provider cannot produce a CAT because the system does not maintain sufficient logs, the evidence should be inadmissible, a rule that incentivises investment in audit-capable architectures.

6.4. Pillar 3 — Human Oversight Attestation (HOA)

AI-generated evidence submitted under FAES must be accompanied by a Human Oversight Attestation (HOA): a signed declaration by a qualified human expert in the relevant forensic domain who has undertaken a substantive review of the AI system's output in the specific evidentiary context and attests to four matters:

(i) that the system's output is within the validated operational parameters documented in the TTD for evidence of the type and quality presented;

(ii) that no anomalous conditions (system alerts, unusual confidence patterns, evidence quality flags) were detected or ignored;

(iii) that the expert has independently assessed the reliability of the output in light of all known limitations disclosed in the TTD; and

(iv) that the expert accepts professional and legal responsibility for the evidentiary claim being advanced on the basis of the AI system's output.

The HOA operationalises the EU AI Act's Article 14 human oversight mandate at the point of evidentiary deployment, together with Article 26(2)'s requirement that deployers assign oversight to competent, trained persons. It also transforms Daubert's expert-testimony mechanism into a vehicle for meaningful AI accountability: the named expert who signs the HOA is the person whom defence counsel can cross-examine, and whose qualifications, affiliations, potential conflicts, and methodology choices are placed before the trier of fact.

A structural objection arises when the attesting expert is a deployer-side analyst (for example, a forensic examiner employed by a police agency) who is denied access to the provider's proprietary internals by the very trade-secret claims that FAES is designed to overcome. The conflict is resolved by the framework's precondition structure rather than by a carve-out. Because the TTD is itself a precondition of admissibility, a provider that withholds from the attester the information needed to make the four attestations simply renders the evidence inadmissible; the incentive therefore runs toward disclosure rather than concealment. Where legitimate confidentiality interests are genuinely implicated, the mechanism modelled in *State v. Pickett* (compelled disclosure to the defence under a protective order) supplies a template for reconciling the attester's and the opponent's access with the provider's commercial interests, without permitting trade secrecy to defeat the attestation altogether.

6.5. Pillar 4 — Cross-Jurisdictional Mutual Recognition (CMR)

FAES proposes a Mutual Recognition Protocol under which a forensic AI system that has obtained a high-risk conformity assessment certificate under the EU AI Act (including the technical documentation, dataset governance, and human oversight design required by Articles 9–14, alongside the post-market monitoring obligations of Article 72) is accorded a rebuttable presumption of reliability for the purpose of Daubert admissibility analysis in U.S. federal criminal proceedings, and vice versa where reciprocal U.S. federal certification procedures are established. The presumption is rebuttable on evidence that the system's performance in the specific operational deployment context differs materially from its certified performance profile.

To address the procedural questions that a presumption of this kind leaves open, FAES specifies the following mechanism. The proponent bears the initial burden of production, satisfied by tendering the EU high-risk conformity certificate together with the matching Technical Transparency Disclosure and Computational Audit Trail for the specific evidential item. The

burden then shifts to the challenger, who may rebut the presumption by making a threshold showing that the system's performance in the specific operational deployment context differs materially from its certified profile, for example, that the evidence falls outside the validated parameters disclosed under Pillar 1, or that an intervening version change is undocumented under Component C. A sufficient threshold showing triggers a reliability hearing at which the proponent must establish admissibility under the governing Rule 702 standard; the 2023 amendment to Rule 702 makes explicit that each admissibility requirement must be established by a preponderance of the evidence, and the certificate is evidence going to, but not dispositive of, that determination. Consistent with *General Electric Co. v. Joiner*, the trial court's ruling on the presumption and on any rebuttal is reviewed on appeal for abuse of discretion. The presumption thus streamlines admissibility for certified systems without displacing the court's gatekeeping responsibility or the opponent's opportunity to contest case-specific reliability.

CMR is not unprecedented: mutual recognition of regulatory equivalence is a well-established instrument in international trade law, financial services regulation, and pharmaceutical approval. The Budapest Convention on Cybercrime's mutual legal assistance framework [22] provides a procedural template adaptable for this purpose.

Taken together, the four pillars map onto anchors in both frameworks. Pillar 1 (TTD) draws on the Act's Article 11 technical-documentation and Article 13 transparency obligations and discharges Daubert's peer-review and methodology-disclosure expectations. Pillar 2 (CAT) rests on the Article 12 automatic-logging, Article 16(d) provider record-keeping, and Article 26(6) deployer log-retention duties, functioning as the computational analogue of the physical chain of custody and of FRE 901 authentication. Pillar 3 (HOA) operationalises Article 14 human oversight and the Article 26(2) deployer competence duty, and connects to expert testimony under FRE 702. Pillar 4 (CMR) rests on the Articles 9–14 conformity obligations and Article 72 post-market monitoring, and on the U.S. side on FRE 702(b)'s "sufficient facts or data" requirement together with the statutory recognition proposed under the Federal Forensic AI Evidence Act (Section 6.6).

6.6. Pillar Application to Generative Systems

The large-language-model deployments flagged in Section 2.1 require specific treatment because generative systems strain the premise of the Computational Audit Trail. A CAT presupposes that the chain between input and output can be reconstructed from an immutable log; however, a non-deterministic generative model may produce different outputs from identical inputs, so logging alone does not guarantee reproducibility.

FAES addresses this in two ways. First, for any generative output offered in evidence, the CAT must additionally record the model version, the complete prompt, the sampling parameters that govern non-determinism (including the temperature and, where the system exposes one, the random seed), and the verbatim output, so that the conditions of generation are fully documented even where the output cannot be regenerated bit-for-bit. Second, where exact reproducibility cannot be assured, that irreproducibility is itself a reliability limitation that must be disclosed under Pillar 1 and that constrains the evidentiary role the output may properly play. Applying the proportional-calibration principle of Section 6.1, a non-reproducible generative output should be confined to investigative or triage functions and should not, absent independent corroboration, be admitted as substantive or near-determinative evidence. This treatment closes the gap between the generative use cases introduced in Section 2.1 and the audit-trail infrastructure that Pillars 2 and 3 depend on.

6.7. Institutional Implementation

FAES requires implementation at three institutional levels. At the legislative level, the United States Congress should enact a Federal Forensic AI Evidence Act establishing FAES as the minimum standard for AI-generated evidence in federal criminal proceedings, preempting inconsistent state-level approaches, and authorising the Federal Rules Advisory Committee to amend FRE 702 to incorporate FAES-specific criteria.

At the regulatory level, NIST, whose AI Risk Management Framework 1.0 [23] provides a compatible technical vocabulary, should publish a Forensic AI Profile under the AI RMF addressing risk categories, performance metrics, and transparency requirements for forensic AI systems, and the National Institute of Justice should condition grant funding for AI-based forensic tools on the publication of TTD-compliant performance studies.

At the judicial level, the Federal Judicial Center should develop a Forensic AI Bench Book, modelled on its existing reference guides for DNA evidence and statistical analysis, giving federal judges decision trees for evaluating FAES compliance, including checklists for TTD review, CAT verification, and HOA adequacy.

7. Conclusion

The governance of AI-generated forensic evidence stands at a critical inflection point. The EU AI Act has created, for the first time, a comprehensive ex ante architecture that ensures forensic AI systems meet a meaningful floor of technical quality, transparency, and human accountability before producing outputs that may determine individual liberty. The Daubert framework, though institutionally entrenched and procedurally flexible, was designed for human expert witnesses and openly published methodologies, a world that proprietary machine learning forensics has rendered structurally inadequate for its gatekeeping function.

This paper has argued that the two regimes are not irreconcilable competitors but mutually reinforcing frameworks that, appropriately harmonised, can address the five governance failures identified in Section 2: opacity, auditability deficits, bias amplification, versioning problems, and cross-jurisdictional fragmentation. The Forensic AI Evidence Standard operationalises that harmonisation across legislative, regulatory, and judicial levels, calibrated to the evidentiary weight of the AI output and extended to the generative systems now entering forensic practice.

The stakes are not abstract: every day in which AI systems produce forensic evidence that cannot be audited, challenged, or meaningfully attributed to a responsible human expert is a day in which the due process guarantees of democratic legal systems are rendered partially fictitious. The RELAF community is well positioned to advance the scholarly, technical, and institutional work needed to close this gap, work for which the Forensic AI Evidence Standard is a starting point, not a conclusion.

Declaration on Generative AI

During the preparation of this work, the author used Claude to format, check grammar, and check spelling. After using this tool, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

[1] European Parliament and Council. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act) and amending certain Union legislative acts. Technical Report L 2024/1689, Official Journal of the European Union, 2024.

[2] Daubert v. Merrell Dow Pharmaceuticals, Inc., 509 U.S. 579, 1993. United States Supreme Court.

[3] General Electric Co. v. Joiner, 522 U.S. 136, 1997. United States Supreme Court.

[4] Kumho Tire Co. v. Carmichael, 526 U.S. 137, 1999. United States Supreme Court.

[5] J. Buolamwini and T. Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Proceedings of the 1st Conference on Fairness, Accountability, and Transparency (FAT* 2018), volume 81 of Proceedings of Machine Learning Research, pages 77–91, 2018.

[6] K. Hill. Wrongfully accused by an algorithm. The New York Times, June 24 2020.

[7] D. Taylor, J. Buckleton, and I. Evett. Testing likelihood ratios produced from complex DNA profiles. Science and Justice, 55(5):379–384, 2015. doi:10.1016/j.scijus.2015.04.001.

[8] R. Wexler. Life, liberty, and trade secrets: Intellectual property in the criminal justice system. Stanford Law Review, 70(5):1343–1429, 2018.

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

[10] C. Haskins. How ShotSpotter works — and why its critics are calling for a ban. VICE News, July 20 2021.

[11] R. Simmons. Big data, machine judges, and the legitimacy of the criminal justice system. UC Davis Law Review, 52(3):1067–1126, 2019.

[12] C. Rudin. Stop explaining black box machine learning models for high stakes decisions and use interpretable models instead. Nature Machine Intelligence, 1(5):206–215, 2019. doi:10.1038/s42256-019-0048-x.

[13] K. Lum and W. Isaac. To predict and serve? Significance, 13(5):14–19, 2016. doi:10.1111/j.1740-9713.2016.00960.x.

[14] People v. Superior Court (Chubbs), No. B257612, 2015. California Court of Appeal.

[15] Commonwealth v. Foley, 38 A.3d 882, 2012. Pennsylvania Superior Court.

[16] State v. Pickett, 466 N.J. Super. 270 (App. Div.), 2021. New Jersey Superior Court, Appellate Division.

[17] United States v. Gissantaner, 990 F.3d 457, 2021. United States Court of Appeals for the Sixth Circuit.

[18] United States v. Miller, No. 19-cr-00651, 2021. United States District Court for the Eastern District of Virginia.

[19] State v. Loomis, 881 N.W.2d 749, 2016. Wisconsin Supreme Court.

[20] J. Dressel and H. Farid. The accuracy, fairness, and limits of predicting recidivism. Science Advances, 4(1):eaao5580, 2018. doi:10.1126/sciadv.aao5580.

[21] Illinois Artificial Intelligence Video Interview Act, 820 ILCS 42/1 et seq., 2020. State of Illinois.

[22] Council of Europe. Convention on Cybercrime (Budapest Convention), ETS No. 185, November 23 2001.

[23] National Institute of Standards and Technology. Artificial Intelligence Risk Management Framework (AI RMF 1.0). Technical Report NIST AI 100-1, U.S. Department of Commerce, Gaithersburg, MD, January 2023.