AI-Generated Evidence Admissibility:

A Formal Classification Framework for Synthetic Content

Vincent Gonzalez

Independent Scholar

modulign.org · VinceGonzalez@me.com

2026

Target: Harvard Journal of Law and Technology · Stanford Technology Law Review · Yale Journal of Law and Technology

---

Abstract

Courts in the United States and internationally are now regularly confronted with evidence alleged to be AI-generated, AI-enhanced, or of uncertain synthetic origin — and they lack a formal framework adequate to the problem. Existing evidentiary doctrine evaluates authenticity through procedural compliance: chain-of-custody documentation, expert testimony, and forensic analysis of content properties. These mechanisms were designed for physical and digital evidence that has an observable causal history. They are structurally inadequate for synthetic content, whose causal history traces to a generative model's parameters rather than to any observable state of affairs.

This paper argues that the problem is not one of labeling or detection but of classification architecture — the formal system by which evidence is categorized according to its epistemic relationship to observable reality. Current governance frameworks, including proposed Federal Rule of Evidence 707, the EU AI Act, and the C2PA technical standard, treat synthetic origin as a property that can be attached to content through metadata. This paper proposes an alternative approach: a formal classification grammar in which synthetic origin is encoded in the content's address structure — its identity within the classification system — rather than appended to it as a separable tag.

The paper instantiates this approach through the Modulign Standard, a Dimensional Address Grammar for Observable Reality (DAG-OR) whose VR/·:SYN classification signature functions as a formal epistemic primitive. Within the address grammar, content whose causal history traces to a generative model rather than to an observable state of affairs receives an address — VR/·:SYN — that is structurally inseparable from its identity in the system. This paper demonstrates that content so classified cannot satisfy the authentication predicate of Federal Rule of Evidence 901(b)(9) as observational evidence about the reality it depicts — not because any rule prohibits its admission, but because its address structure formally encodes the absence of the causal chain that observational authentication requires. The content remains admissible as evidence of the artifact's own existence and provenance — the distinction is between what the content depicts and what the content is.

The paper further distinguishes AI-generation cases from AI-enhancement cases, a distinction courts have frequently conflated, and proposes a formal evidentiary framework that resolves both categories. It addresses the constitutional implications for criminal proceedings under the Confrontation Clause and Due Process Clause, identifies the separation-of-powers considerations inherent in judicial reliance on a private classification standard, engages the enforcement and data-protection dimensions of the EU AI Act, and proposes an adoption pathway. The Modulign Standard is a published specification; it has not yet been independently validated, adopted by a standards body, or tested in adversarial proceedings. The paper's

claim is that the formal architecture is logically sound and that the adoption and validation pathway is feasible — not that the infrastructure is already operational.

1. The Evidentiary Problem

The admission of AI-generated content into legal proceedings is no longer a theoretical problem. Courts have confronted it in multiple jurisdictions, applying a range of existing evidentiary doctrines without a formal framework adequate to the questions it raises.

1.1 The Case Record

In United States v. Reffitt (No. 21-cr-32, D.D.C. 2022), defense counsel argued that prosecution evidence of the defendant's participation in the January 6 Capitol events could be a deepfake — deploying evidentiary uncertainty around synthetic content as a reasonable-doubt strategy against content that was observationally grounded. The court had no formal standard for evaluating this claim beyond the traditional authentication requirements of FRE 901, which do not specifically address the possibility of generative synthesis.

In State v. Puloka (Wash. Super. Ct. 2024), a court analyzed AI-enhanced bystander video under the Frye standard for scientific evidence, FRE 702 (expert testimony), and Rules 401 and 403 (relevance and prejudice), ultimately denying admission. The court applied frameworks designed for expert scientific testimony to a question that is fundamentally about the causal history of the content itself — whether the enhancement altered what the original recording captured.

In Huang v. Tesla (No. 2:22-cv-02846, C.D. Cal.), defense counsel raised the possibility that video evidence of a party's statements could be AI-generated, again without a formal standard for adjudicating the claim.

In Mata v. Avianca, Inc. (No. 22-cv-1461, S.D.N.Y. 2023), an attorney submitted a brief containing case citations fabricated by ChatGPT — synthetic content presented as though it had a real referent. The court sanctioned counsel under professional responsibility doctrine. The case illustrates that the problem extends beyond audiovisual deepfakes to any content type whose apparent referent may not exist.

These cases share a structural feature: courts are being asked to make admissibility determinations about the epistemic status of content — whether it has an authentic causal connection to observable reality — using evidentiary doctrines designed for different questions. The courts were not failing; they were reasoning carefully with the tools available. But those tools were designed for evidence whose causal history is, in principle, traceable. They do not generalize to synthetic content whose causal history traces to model parameters rather than to observable states of affairs. The result is inconsistency across jurisdictions and analytical frameworks that fit the problem imperfectly.

1.2 The Regulatory Gap

The Advisory Committee on Evidence Rules released proposed Rule 707 for public comment in August 2025. As drafted, Rule 707 applies to evidence that the proponent acknowledges was "generated or substantially assisted" by AI, requiring notice and foundational information supporting authentication. The rule does not create a formal classification mechanism, does not provide a standard for resolving disputes about whether content is AI-generated, and does not apply to the contested-authenticity scenario — the deepfake defense, the "this video might be synthetic" argument — that drives the most urgent cases.

The Advisory Committee Notes indicate that the rule addresses the disclosure problem rather than the detection or classification problem, and that contested-authenticity cases are expected to be resolved under

existing authentication doctrine. But existing authentication doctrine is precisely what the case record reveals to be inadequate for this category of dispute. This paper argues that the gap is not in the rules themselves but in the classification infrastructure those rules presuppose. The evidentiary doctrines are adequate; what is missing is a formal classification standard that activates them for synthetic content.

2. The Foundational Error in Current Frameworks

Every major regulatory and governance framework currently addressing AI-generated content treats synthetic content as a labeling problem. The EU AI Act (Regulation (EU) 2024/1689), Article 50, requires providers of AI systems that generate synthetic content to ensure that outputs are marked in a machine-readable format and detectable as artificially generated. The NIST AI Risk Management Framework (AI RMF 1.0, 2023) identifies content provenance and labeling as key measures for trustworthy AI. C2PA (Coalition for Content Provenance and Authenticity) provides a technical standard for cryptographically signed provenance manifests that record what tools produced a piece of content.

These frameworks share a foundational assumption: that the difference between observed and generated content is a property that can be attached to content — through a label, a watermark, a manifest.

The foundational error is this: the difference between observed and generated content is not a property of the content. It is a fact about the content's causal history. A deepfake of a crime scene is not a photograph with incorrect content. It is a categorically different kind of object — one whose causal history traces to a generative model's parameters, not to the physical properties of any real scene. No label changes this. A disclosed deepfake is still a deepfake. A watermarked synthetic voice recording is still a synthetic voice recording. The label describes the content's provenance; it does not alter the content's epistemic status.

This distinction matters because the evidentiary weight of observational evidence derives precisely from its causal history. Courts have long recognized this: authentication requirements under FRE 901 ask not whether evidence looks real, but whether it was produced by a process that accurately captures reality. The process matters because the process is what connects the evidence to the fact. An AI system generating a video does not engage the observable features of any real-world state of affairs. Its output therefore cannot carry the epistemic weight of evidence that did.

The governance frameworks address the disclosure dimension of this problem competently. They do not address the classification dimension — the question of how synthetic content should be formally categorized within an evidentiary system — because they lack a formal primitive that encodes epistemic status rather than merely recording provenance metadata. Labels and classification are complementary; neither substitutes for the other.

3. The VR/·:SYN Primitive: Epistemic Status as Address Structure

3.1 The Classification Architecture

The Modulign Standard introduces the VR/·:SYN classification signature as a formal epistemic primitive for synthetic content within its Dimensional Address Grammar for Observable Reality (DAG-OR). The Standard assigns every classified observation a structured address encoding its realm of origin, medium of observation, and other dimensional coordinates. Observational content captured in the physical world receives a physical realm designation: SR/ (surface reality), AT/ (atmospheric), AQ/ (aquatic), OR/

(orbital). The realm designation is not a tag — it is the content's address within the classification system. Content without an address does not exist in the system.

The VR/·:SYN signature encodes two things simultaneously. First, VR/ (virtual realm): the content has no physical locus. It is not constituted at any spatio-temporal coordinate. Assigning it a physical realm would assert a physical existence it does not have, creating a contradiction with the Standard's foundational definitions. Second, :SYN (synthetic medium): the observation was generated rather than captured. No sensor, instrument, or human observer engaged the observable features of a real-world state of affairs to produce it. Together, VR/·:SYN encodes the absence of causal grounding in observable reality.

3.2 The Classification Rule and the Semantic Claim

The VR/·:SYN primitive embodies a biconditional whose two directions do different work and must be distinguished. The left-to-right direction — content is synthetic, therefore it must receive VR/·:SYN — is a classification rule, a design decision encoded in the Standard. It mandates that synthetic content receive the VR/·:SYN address because no alternative address is available without contradiction: synthetic content cannot be assigned a physical realm without asserting a physical existence it does not have.

The right-to-left direction — content bears VR/·:SYN, therefore it is classified as synthetic — is a semantic claim. It holds within the formal system because VR/·:SYN is the only address that synthetic content can receive without contradiction, and because no non-synthetic content can receive it without falsifying the observation record. The address and the epistemic status are coextensive within the grammar, not because of an arbitrary stipulation, but because the grammar's dimensional constraints force the coextension.

3.3 Structural Inseparability, Not Metaphysical Constitution

In the philosophy of language, constitutive encoding means that the encoding makes something the case — a marriage certificate constitutively creates a legal marriage; a photograph of a wedding merely records one. VR/·:SYN does not make content synthetic. The generative model did that. What VR/·:SYN does is embed the classification in the content's formal identity within the system, such that the classification cannot be separated from the content without destroying the content's address — and therefore its existence in the system.

This is structurally different from a label. A provenance label — a C2PA manifest, an Article 50 watermark — is attached to content after the content exists as an independent object. The label is separable: it can be removed, overwritten, or challenged independently of the content. VR/·:SYN is not separable in this way. Within the Modulign classification system, VR/·:SYN is the content's address. Removing the classification does not produce unlabeled content; it produces unaddressed content — content that does not exist in the system. The append-only architecture of the Observation Registry ensures that a subsequent reclassification would create a contradiction record in the Relation Registry rather than replacing the original entry.

The precise claim is not that VR/·:SYN metaphysically constitutes synthetic origin. The claim is that within the address grammar, VR/·:SYN is structurally inseparable from the content's identity — making it categorically more robust than a label, which is structurally separable. This distinction — between structural inseparability and metaphysical constitution — is what differentiates the address-grammar approach from the labeling approach, and it is sufficient to carry the evidentiary argument that follows.

3.4 The Evidentiary Consequence

The evidentiary consequence follows from the Address-Theoretic Knowledge formulation (ATK) developed in Gonzalez (2026a). Under ATK, knowledge of observable reality requires that the justificatory procedure engage the observable features causally produced by the truth-making state of affairs. Synthetic content, by definition, was not produced by a causal process that engaged the observable features of any

real-world state of affairs. A deepfake of a crime scene was not produced by the crime scene. A synthetic voice recording was not produced by the person whose voice it mimics.

Under FRE 901(b)(9), authentication requires evidence describing "a process or system" and showing that "the process or system produces an accurate result." The process that produces VR/·:SYN content — a generative model — does not produce accurate results about any real-world state of affairs. It produces content that may resemble real-world states of affairs, but resemblance is not causal connection, and it is causal connection that authentication under 901(b)(9) evaluates.

VR/·:SYN content therefore cannot satisfy the 901(b)(9) authentication predicate as observational evidence about the reality it depicts. This is not a per se exclusion. The content remains admissible as evidence of the artifact's own existence and provenance — to prove that a particular synthetic artifact was created, that it has certain properties, that it was distributed in a certain way. The distinction is between the content as a window onto reality (inadmissible, because the window has no causal connection to the scene) and the content as an object in its own right (admissible, subject to standard relevance and authentication requirements).

4. The Critical Distinction: AI-Generation Versus AI-Enhancement

Courts have frequently conflated two structurally distinct categories of AI-related evidence, applying analytical frameworks suited to one category to disputes that belong to the other. The Modulign architecture resolves this conflation by encoding the distinction formally.

Category	Definition	Modulign Classification	Evidentiary Question
A — AI-Generation	Content wholly synthetic; no real-world observation in causal history	VR/·:SYN (mandatory)	Whether content has any causal connection to observable reality
B — AI-Enhancement	Real observational content processed by AI after capture	SR/ original + →DRV→ derivation record; ^CONT flag if disputed	Whether enhancement altered an existing causal chain in ways affecting reliability

4.1 Category A: AI-Generation Cases

Reffitt, Huang, and Mata raise Category A questions. In each, the claim — explicit or implicit — is that content with no causal connection to observable reality has been presented as though it has one. Under the proposed framework, genuine AI-generation cases are resolved by reference to the classification record. Content classified under the ^EVID protocol carries a VR/·:SYN address that is cryptographically signed, append-only, and machine-verifiable. Content that lacks any classification record is epistemically unverified. The appropriate judicial response is not to presume authenticity; it is to treat the absence of classification as an evidentiary gap that the proponent must fill through other means.

4.2 Category B: AI-Enhancement Cases

Puloka is a Category B case: bystander video AI-enhanced to clarify footage. The original content retains its observational grounding — it was produced by a real causal process engaging real observable features. The question is whether the enhancement altered what the original captured. Under the proposed

framework, the original content retains its SR/ realm designation. The enhancement is documented in a →DRV→ (derivation/provenance) relation record linking the enhanced version's identifier to the original's identifier with a timestamped, observer-annotated processing record. If the enhancement is disputed, a ^CONT (contested) meta flag is applied. The classification history is permanent and auditable.

4.3 Why the Distinction Matters

AI-generation cases raise a question about the content's fundamental epistemic status — whether it has any causal connection to observable reality. AI-enhancement cases raise a question about the integrity of a causal chain that exists — whether the enhancement altered the content in ways that affect its evidentiary reliability. These are different questions requiring different analytical frameworks. Applying FRE 702 expert-testimony analysis to a Category A dispute misframes the issue. The question is not whether the generative model is reliable; it is whether content produced by a generative model has any causal connection to the reality it depicts. The formal distinction between Category A and Category B prevents both analytical errors by encoding the relevant difference at the classification level.

5. The ^PROV / ^EVID Certification Architecture

The Modulign Standard introduces a certification distinction that is legally significant and has no precise equivalent in existing evidentiary frameworks.

Certification	Produced By	Evidentiary Status
^PROV (Provisional)	AI system self-reporting its own output; classification without full CDP; observer below evidence-grade threshold	Useful for disclosure and compliance; not evidence-grade for legal proceedings
^EVID (Evidence-Grade)	Certified classifier applying the full eight-step CDP, or post-generation observer satisfying evidence-grade threshold with two-expert consensus (:E×2)	Admissible as evidence of the artifact's existence, classification, and provenance; cannot serve as SR/ observational evidence about depicted reality

The governance implication is precise. Requiring AI systems to self-classify their output — as the EU AI Act's Article 50 effectively does — creates ^PROV records: useful disclosure, but not evidence-grade. Requiring certified human classifiers to verify that output creates ^EVID records: formally documented, independently verifiable, and capable of supporting evidentiary authentication. Both are valuable. Neither substitutes for the other. Policy frameworks must specify which they require and in what context.

A proponent who produces content bearing an ^EVID VR/·:SYN classification has formally acknowledged synthetic origin in a cryptographically non-repudiable, machine-verifiable form. A proponent who produces content lacking any classification record has produced epistemically unverified content. The court need not rely solely on the proponent's testimony about the content's origin — the classification record either exists and is verifiable, or it does not.

6. Constitutional Dimensions

The proposed framework operates within constitutional constraints that must be identified, even where this paper does not fully resolve them.

6.1 The Confrontation Clause

The Sixth Amendment guarantees criminal defendants the right to confront the witnesses against them. Crawford v. Washington (541 U.S. 36, 2004) held that testimonial statements of absent witnesses may be admitted only where the declarant is unavailable and the defendant has had a prior opportunity to cross-examine. Melendez-Diaz v. Massachusetts (557 U.S. 305, 2009) extended this principle to forensic laboratory reports, holding that their authors must be available for cross-examination when those reports are introduced in criminal proceedings.

When a Modulign ^EVID classification is introduced in a criminal proceeding — as evidence that content is synthetic — the classification functions as an expert assertion about the content's causal history. Under Melendez-Diaz, the certified classifier who produced the ^EVID record must be available for confrontation if the classification is introduced against a criminal defendant. The framework accommodates this: the ^EVID protocol requires a named, credentialed observer whose identity is recorded in the classification record. That person is the witness subject to confrontation. Courts routinely manage confrontation requirements for forensic analysts and chain-of-custody witnesses; this creates no novel constitutional problem.

6.2 Due Process and Burden-Shifting

The paper proposes that content lacking any classification record be treated as epistemically unverified, with the burden of demonstrating observational grounding falling on the proponent. The due process implications differ depending on context.

In civil proceedings, burden-shifting is a standard procedural tool. In criminal proceedings, the asymmetry matters. When the prosecution introduces evidence and the defense argues it might be AI-generated, placing the burden on the prosecution to demonstrate observational grounding is consistent with the beyond-a-reasonable-doubt standard established in In re Winship (397 U.S. 358, 1970). When the defense introduces evidence and the prosecution contests its authenticity, the appropriate resolution is to treat the burden as one of production rather than persuasion: the defense must produce some evidence of observational grounding, but the ultimate burden of proving unreliability remains with the party challenging it. This mirrors the existing framework under FRE 901(a), which requires only that the proponent produce "evidence sufficient to support a finding" of authenticity.

6.3 Separation of Powers and Private Standards

The proposed framework relies on a privately developed classification standard. Courts routinely rely on ASTM standards, NIST guidelines, ISO certifications, and C2PA manifests; the Daubert framework explicitly contemplates judicial reliance on methodologies developed outside the judicial system, subject to the court's gatekeeping function. Three conditions make such reliance legitimate: the standard must be publicly available and independently auditable; it must be subject to adversarial testing; and judicial reliance must be on the methodology, not the standard's conclusions. The Modulign Standard is published on Zenodo under an open-access license (DOI: 10.5281/zenodo.19348703), freely available for inspection, critique, and independent implementation. Adversarial testing is identified in Section 10 as a necessary step before judicial adoption. The court remains the ultimate gatekeeper; a Modulign classification is evidence to be evaluated, not a determination to be deferred to.

7. International and EU Dimensions

7.1 The EU AI Act: Disclosure and Classification

Article 50 of the EU AI Act requires providers of AI systems that generate synthetic content to mark outputs in a machine-readable format detectable as artificially generated. The Modulign VR/·:SYN primitive proposes a formal classification architecture for that same content. Both are necessary; neither substitutes for the other.

Three dimensions require attention. First, enforcement architecture: Article 50 obligations fall on providers and are enforced through national market surveillance authorities under administrative law. The Modulign classification framework, as proposed for evidentiary use, would be enforced by courts under evidentiary law. A jurisdiction implementing both must specify how they interact — whether an Article 50 disclosure satisfies a court's classification requirement, whether courts can require classification beyond what Article 50 mandates, and how conflicts between the two regimes are resolved.

Second, definitional alignment: the AI Act defines AI-generated content through the functional characteristics of the system that produced it (Article 3, read with Article 50). The Modulign Standard defines synthetic content through its address grammar — the absence of a causal chain to observable reality. These definitions are likely coextensive in practice but are not identical in structure. Edge cases — AI systems that process real sensor data in ways that blur the generation/enhancement boundary — may be classified differently under the two frameworks. A full harmonization analysis is necessary before cross-jurisdictional implementation.

Third, data protection: the Modulign classification architecture records observer identities, competence annotations, and classification decisions in an append-only ledger. Under GDPR Articles 17 and 21, data subjects have rights to erasure and to object to processing. An append-only ledger recording personal data of classifiers may conflict with these rights. The GDPR's exceptions for legal obligation (Article 6(1)(c)) and the establishment, exercise, or defense of legal claims (Article 9(2)(f)) provide a basis for processing, but this tension must be addressed in any EU implementation.

7.2 Cross-Border Evidentiary Harmonization

If a US court requires Modulign ^EVID classification and an EU court requires Article 50 disclosure, content crossing jurisdictions must satisfy both. A classification standard intended for international use must treat cross-border compatibility as a first-order requirement. The Modulign Standard's open-access publication and jurisdiction-neutral address grammar are designed with this in mind, but specific harmonization work has not yet been completed.

8. Application to Proposed Rule 707

The Advisory Committee's proposed Rule 707 represents a significant and welcome step. The Modulign framework does not replace it — it provides classification infrastructure that could support a more comprehensive version.

Rule 707 establishes a notice requirement and a foundation requirement for AI-generated evidence whose synthetic origin is acknowledged. This addresses the transparency dimension of the problem: parties and courts should know when evidence was generated by AI, and the proponent should be prepared to explain how and why.

The rule does not apply to the contested-authenticity scenario — the case where one party introduces evidence and the opposing party claims it may be AI-generated. These are the cases where the evidentiary framework is most urgently needed.

A revised Rule 707, or a successor rule, could incorporate three elements from the proposed framework. First, a formal classification requirement: AI-generated content offered in evidence would be required to bear a classification record satisfying specified protocol requirements. Second, a treatment for unclassified content: content lacking any classification record would be treated as epistemically unverified for authentication purposes, with the burden of demonstrating observational grounding falling on the proponent. Third, a formal distinction between Category A and Category B cases: the rule would specify different analytical frameworks for AI-generation disputes and AI-enhancement disputes, preventing the doctrinal conflation the case record reveals.

9. Implications for Expert Testimony Under Daubert

The proposed classification architecture has direct implications for expert testimony foundations under Daubert v. Merrell Dow Pharmaceuticals (509 U.S. 579, 1993). Daubert requires that expert testimony be based on sufficient facts or data, be the product of reliable principles and methods, and reflect a reliable application of those principles and methods to the facts of the case. The Modulign CDP is designed as precisely such a documented, reproducible method.

An expert witness who has produced a Modulign ^EVID classification under the CDP can support testimony with: a documented eight-step classification procedure whose reliability is specified in a published standard; a confidence score with a documented baseline; an observer competence annotation mapping to existing forensic certification structures; and a cryptographically signed classification record that is independently verifiable.

A critical qualification: Daubert requires not only that a methodology be specified and documented, but that it be testable, have a known or potential error rate, be subject to peer review and publication, and enjoy general acceptance in the relevant scientific community. The Modulign Standard has been published and is subject to peer review. Its error rate has not been empirically established. Its general acceptance is limited to date. These are necessary steps on the path to Daubert qualification, not obstacles that defeat it. But they must be completed before the Standard can be presented to a court as having satisfied all Daubert factors.

10. The Adoption Pathway

A classification standard has value only to the extent it is adopted. The proposed framework requires engagement from multiple stakeholders, and this paper must be transparent about what adoption requires.

10.1 AI Providers

AI providers would need to integrate VR/·:SYN classification at the point of content generation — either by adopting the Modulign address grammar internally or by maintaining a translation layer between their internal systems and the Modulign grammar. The incentive structure, however, aligns with existing compliance pressures. AI providers are already preparing for Article 50 obligations under the EU AI Act and for C2PA integration. A Modulign classification module that outputs VR/·:SYN addresses alongside C2PA manifests and Article 50 disclosures would satisfy all three requirements through a single integration.

10.2 Forensic Laboratories and Certifiers

The ^EVID protocol requires certified classifiers whose competence has been verified against the Standard's observer requirements. The question of who certifies the certifiers must not be circular. The proposed pathway mirrors existing forensic certification structures: an independent accreditation body — analogous to the American Board of Criminalistics or ISO/IEC 17025 accreditation for forensic laboratories — would certify classifiers against the Standard's published competence requirements. The Standard defines the requirements; the accreditation body verifies that individuals meet them. This is the same structure used for forensic DNA analysts, latent print examiners, and digital forensic practitioners.

10.3 Courts

Judicial adoption is the final step, not the first. Courts will not — and should not — adopt a classification standard that has not been independently validated, implemented by forensic practitioners, and tested in adversarial proceedings. The adoption pathway is: (1) publication and open review, completed; (2) independent peer review, in progress; (3) pilot implementation by forensic laboratories and AI providers; (4) empirical validation, establishing error rates and testing reproducibility; (5) adversarial testing in proceedings where opposing counsel can challenge the methodology; (6) standards body engagement with NIST, ISO, or equivalent; and (7) judicial adoption by courts applying Daubert or Frye after steps 1–6 are completed. This paper proposes the formal architecture and argues that it is logically sound. It does not claim that the architecture has completed this pathway.

10.4 Licensing and Intellectual Property

A classification standard proposed for use in judicial proceedings must be openly available. Proprietary control over the epistemic infrastructure of the legal system is unacceptable — as the experience with proprietary forensic tools has demonstrated. The Modulign Standard is published under an open-access license. Any party may implement it, audit it, critique it, or build upon it without license fees or restrictions.

11. The Detection Problem and Why It Does Not Defeat the Framework

Detection is the question of verification: given content of unknown provenance, can we determine whether it is synthetic? This is genuinely hard and getting harder. Classification is a different question: given content we produced, can we classify it correctly? For content producers — AI companies, individuals using generative tools — classification at the point of production is trivial. The VR/·:SYN primitive addresses the classification side of the problem, not the detection side. Content that reaches a court without a verifiable classification record is unclassified. The appropriate judicial response is to treat it as epistemically unverified — not as presumptively authentic, and not as presumptively synthetic. The burden falls on the proponent to establish observational grounding through other means.

11.1 The "All Digital Processing" Objection

A defense objection now appearing in litigation argues: "All digital processing modifies content. If modification introduces causal discontinuity, then all digital evidence is epistemically suspect." Digital processing that preserves an unbroken derivation chain from the original observational record — documented in the Relation Registry with timestamped, observer-annotated processing records — provides evidence of causal continuity with observable reality. The derivation record documents the processing steps through which the original observational content was transformed, allowing a court to evaluate whether those steps preserved or compromised evidentiary reliability. A VR/·:SYN output with no observational ancestor in the derivation chain is not the same object as SR/ content with a documented processing history.

The distinction is formal, auditable, and resistant to the objection because it rests on the documented history of the content, not on the binary question of whether any processing occurred.

12. Limitations

This paper proposes a formal framework and argues for its logical soundness. It does not claim that the framework is ready for immediate judicial adoption.

First, the Modulign Standard is a published specification. It has not been independently validated through empirical testing, adopted by a recognized standards body, or subjected to adversarial scrutiny in legal proceedings. The adoption pathway described in Section 10 identifies the steps required to address this limitation.

Second, the case law survey covers the most prominent cases but is not exhaustive. AI-generated evidence disputes are proliferating across jurisdictions, and a comprehensive empirical survey of judicial treatment would strengthen the problem identification.

Third, the paper's central claim — that structural inseparability within an address grammar is categorically more robust than labeling — depends on the assumption that the address grammar is correctly designed and consistently implemented. Independent audit and adversarial testing are the appropriate responses.

Fourth, the framework addresses the classification dimension of the AI-generated evidence problem. It does not address the detection dimension — the challenge of identifying synthetic content when the producer has not classified it. Detection tools, forensic analysis, and witness testimony remain necessary complements.

Fifth, the framework's international applicability depends on harmonization work that has not yet been completed, particularly the alignment of the Modulign grammar's definitions with the EU AI Act's definitions and with emerging treaty language.

13. Conclusion

The evidentiary challenge posed by AI-generated content is not principally a technological problem. It is an epistemological one: courts lack a formal framework for evaluating what kind of epistemic object a given piece of content is, and what evidentiary weight it can bear as a consequence. Proposed Rule 707, the EU AI Act, and existing forensic frameworks address the disclosure dimension without addressing the classification dimension: the formal encoding of epistemic status within an evidentiary system.

This paper proposes a classification architecture — instantiated in the Modulign Standard — that resolves this gap through a formal epistemic primitive: the VR/·:SYN classification signature. Within the address grammar, VR/·:SYN is structurally inseparable from the content's formal identity. Content so classified cannot satisfy the FRE 901(b)(9) authentication predicate as observational evidence about the reality it depicts, because its address structure formally encodes the absence of the causal chain that observational authentication requires. The content remains admissible as evidence of the artifact's own existence and provenance.

The framework formally distinguishes AI-generation from AI-enhancement — a distinction courts need and current doctrine does not provide. It specifies a certification architecture (^PROV / ^EVID) that separates disclosure-grade from evidence-grade classification. It operates within the constitutional

constraints of the Confrontation Clause and Due Process Clause. It complements rather than replaces the EU AI Act's disclosure requirements. And it identifies a concrete adoption pathway through existing institutional channels.

> *The formal architecture is sound. The specification is published and openly available. What remains is the empirical and institutional work of validation, adoption, and adversarial testing. That work can now proceed on a defined foundation.*

References

Advisory Committee on Evidence Rules. 2025. Proposed Amendment to Federal Rules of Evidence: Rule 707. Committee on Rules of Practice and Procedure, Judicial Conference of the United States.

C2PA (Coalition for Content Provenance and Authenticity). 2023. C2PA Technical Specification v1.3. Joint Development Foundation.

Crawford v. Washington, 541 U.S. 36 (2004).

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

European Union. 2024. Regulation (EU) 2024/1689 of the European Parliament and of the Council (AI Act). Official Journal of the European Union.

Federal Rules of Evidence, Rules 401, 403, 702, 901.

Gonzalez, V. 2026a. Modulign Standard. Zenodo. https://doi.org/10.5281/zenodo.19348703

Gonzalez, V. 2026b. The Formal Logic of Modulign. Zenodo. https://doi.org/10.5281/zenodo.19350847

Huang v. Tesla, No. 2:22-cv-02846 (C.D. Cal.).

In re Winship, 397 U.S. 358 (1970).

Mata v. Avianca, Inc., No. 22-cv-1461 (S.D.N.Y. 2023).

Melendez-Diaz v. Massachusetts, 557 U.S. 305 (2009).

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

State v. Puloka, Wash. Super. Ct. (2024).

United States v. Reffitt, No. 21-cr-32 (D.D.C. 2022).

Disclosure

The author is the sole developer of the Modulign Standard. The Standard is published under an open-access license and is freely available without license fees or restrictions. The author has no commercial interest in classification services, certification services, or compliance tools built on the Standard. This paper proposes the Standard as a formal framework for academic and judicial evaluation, not as a commercial product.