

The Human Attestation Protocol: Confrontation Clause Compliance for Automated Classifications in DAG-OR

Vincent Gonzalez

Independent Scholar · ORCID 0009-0004-7283-8598

modulign.org · VinceGonzalez@me.com

This work is licensed under a Creative Commons Attribution 4.0 International License (CC BY 4.0). <https://creativecommons.org/licenses/by/4.0/>. Modulign™ and MGN™ are trademarks of Vincent Gonzalez and are not licensed under CC BY 4.0.

Abstract

When an automated Modulign classification is offered as evidence in a criminal proceeding, the Confrontation Clause of the Sixth Amendment requires that the defendant be able to confront the witness against them. An algorithm cannot be cross-examined. This paper specifies the Human Attestation Protocol (HAP): the formal procedure by which a human witness provides the testimony necessary to satisfy *Crawford v. Washington* (541 U.S. 36, 2004), *Melendez-Diaz v. Massachusetts* (557 U.S. 305, 2009), and *Bullcoming v. New Mexico* (564 U.S. 647, 2011) when automated classifications are offered in criminal proceedings. The protocol identifies three mandatory human roles—the Certifier, the Operator, and the Interpreter—specifies what each must attest to, maps each attestation to the constitutional and evidentiary requirements it satisfies, and defines the documentation that must accompany the testimony. The protocol also addresses the special case of fully autonomous (:A-level) classifiers and the limitations of the human attestation pathway for black-box systems.

Keywords: Confrontation Clause, *Crawford*, *Melendez-Diaz*, *Bullcoming*, human attestation, automated evidence, expert testimony, chain of custody, forensic classification

1. Introduction: The Constitutional Problem

The Sixth Amendment provides: “In all criminal prosecutions, the accused shall enjoy the right . . . to be confronted with the witnesses against him.” In *Crawford v. Washington* (2004), the Supreme Court held that this right applies to all testimonial statements—statements made with the primary purpose of establishing facts relevant to a criminal prosecution. If a statement is testimonial, the declarant must be available for cross-examination, or the statement is inadmissible.

A Modulign classification, when offered to prove that an observation occurred as classified, is a testimonial statement. It asserts facts. It is prepared for use in a proceeding. It falls within *Crawford*'s scope. When the classification is produced by a human classifier, *Crawford* is straightforward: the classifier testifies. When the classification is produced by an automated classifier—certified at :T, :E, or :A under the %AUT framework [7, 17]—*Crawford* presents a structural problem. The automated system cannot testify. It cannot be cross-examined. It cannot respond to challenges about its reasoning, its biases, its failure modes, or its competence on the specific input at issue.

The synthesis paper [16] identified this problem and specified a scoping condition: automated classifications in criminal proceedings require human attestation. This paper specifies what that attestation requires.

2. The Three-Role Framework

The Human Attestation Protocol identifies three mandatory human roles. Each corresponds to a distinct aspect of the automated classification that the defense has a right to challenge. The roles may be filled by the same person if that person has the knowledge and qualifications to attest to all three aspects. They may not be filled by a person who lacks the relevant knowledge—a surrogate who can describe the process in general but was not involved in the specific classification is insufficient under Bullcoming.

2.1 The Certifier

Who: the person who evaluated the automated system and granted its %AUT certification. Attests to: the system's identity (make, model, version, training data provenance); the certification level and its basis (evaluation conducted, test set used, error rates measured, domain restrictions); for :E systems, the characterized error profile; for :A systems, the OOD detection capability and scope; and the certification's current status. Constitutional basis: Daubert admissibility requires a known error rate and evidence that the technique has been tested. Evidentiary basis: FRE 702 (expert testimony) and FRE 901(b)(9).

2.2 The Operator

Who: the person responsible for the automated system's operation at the time the specific classification at issue was produced. Attests to: that the system was functioning within certified parameters at the time of classification; the input data and its chain of custody prior to classification; the operational environment and system logs; and for :A systems, whether the OOD detector triggered during or near the classification. Constitutional basis: Bullcoming requires that the person who actually conducted the test—not a surrogate—testify. The Operator is the DAG-OR analog to the laboratory technician who ran the forensic instrument. Evidentiary basis: FRE 901(b)(9). Note: FRE 803(6) (business records exception) does not apply because under Melendez-Diaz, forensic reports prepared for litigation are testimonial regardless of their business-record character.

2.3 The Interpreter

Who: the person who reviews the specific classification's protocol traceability record and can explain in court how the automated system applied the CDP to produce the specific segment values at issue. Attests to: the protocol traceability record for the specific classification; whether the classification is consistent with the CDP; and the Interpreter's own assessment of whether the classification is correct—not merely that the system produced it, but that the Interpreter, as a qualified human classifier, endorses or identifies concerns with the result. Constitutional basis: the Interpreter provides what the algorithm cannot—a human mind that has examined the specific classification and can be questioned about it. Evidentiary basis: FRE 702 and FRE 703.

3. Attestation Documentation

3.1 Certifier's Documentation Package

Document	Content
Certification Report	Complete evaluation record for the automated system's %AUT certification: test set, methodology, results, error characterization, scope limitations, OOD test results (:A only)
Certification Date and Status	Current certification, expiration, and recertification history
Training Data Provenance	Description of training data, sources, and known biases or limitations
Specification Version	Version of the Modulign Standard against which the system was certified

3.2 Operator's Documentation Package

Document	Content
System Log	Operational log for the classification run producing the GSI-ID at issue: timestamps, input data identifiers, system state, error or anomaly flags
Input Chain of Custody	Documentation of how input data reached the automated system—provenance from original observation to system input
Configuration Record	System configuration at time of classification: version, parameters, active domain restrictions
OOD Detection Log (:A only)	OOD detector output for the classification at issue: whether it flagged the input, and action taken

3.3 Interpreter's Documentation Package

Document	Content
Protocol Traceability Record	Step-by-step trace of the CDP as executed by the automated system for the specific classification
Interpreter's Analysis	Written assessment: consistency with CDP, whether a human classifier would reach the same result, any concerns or qualifications
Correction History	Correction Chain for the GSI-ID at issue, if any corrections exist

4. Testimony Requirements

4.1 All Three Roles Must Testify

Under Bullcoming, surrogate testimony is insufficient. Under Melendez-Diaz, the analyst who prepared the report must appear. The HAP extends this to all three roles: the Certifier, the Operator, and the Interpreter must each testify in person (or by constitutionally adequate means—live video testimony where permitted). If one person fills multiple roles, they may testify in all three

capacities, but the testimony must cover all three attestation areas. Omitting any of the three creates a gap the defense is entitled to exploit.

4.2 The Defense's Right to the Record

The defense is entitled to the complete documentation package for all three roles. This is not a discretionary disclosure—it is a constitutional right. The traceability record, the certification report, the system logs, and the Interpreter's analysis are all materials necessary for meaningful cross-examination. Withholding any of them is a potential due process violation. For classifications produced by government-operated nodes, *Brady v. Maryland* requires disclosure of any material favorable to the defense, including error characterization data and OOD detection results suggesting the input was near the boundary of the system's competence.

4.3 Expert Qualification

The Certifier and the Interpreter must be qualified as expert witnesses under FRE 702. The Operator is a fact witness—testifying to what happened during the classification run, not offering opinions about system reliability—unless their testimony extends to system design, in which case expert qualification may be required.

5. Black-Box Limitations

5.1 The Explainability Ceiling

Protocol traceability can show which CDP step produced which output, but it may not be able to explain why the network's internal computations led to that output. The Interpreter can say: “At CDP Step 2, the system assigned §DOM = BIO based on these input features.” The Interpreter cannot necessarily say why the system assigned BIO if the system's internal feature representations are opaque. The Interpreter must disclose the ceiling: “I can tell you what the system did. I can tell you what input features were associated with the decision. I cannot fully explain the internal computation that produced the decision from those features.” The court then decides the weight. The explainability ceiling does not automatically render the classification inadmissible—it affects weight, not admissibility, if the Daubert foundation is otherwise satisfied.

5.2 The :A-Level Special Case

A :A-level autonomous classifier operates without human supervision—the Operator may not have been present or aware of the specific classification when it was produced. For :A-level systems, the Operator attests to the system's operational configuration and monitoring status, not direct supervision. The Operator's testimony is: “The system was configured as follows, was operating within certified parameters as verified by automated monitoring, and produced the classification autonomously. No anomalies were detected by the monitoring system or the OOD detector at or near the time of classification.” This is a weaker attestation than for supervised systems. The defense is entitled to argue that the weaker attestation warrants less weight. The HAP does not claim to make autonomous classifications as well-attested as supervised ones. It claims to provide the constitutionally minimum attestation for admissibility.

6. International Applicability

6.1 The Confrontation Right Beyond the Sixth Amendment

The HAP is specified with reference to U.S. constitutional law because the Crawford/Melendez-Diaz/Bullcoming jurisprudence is the most developed on confrontation rights in forensic evidence contexts. The right to confront witnesses is not unique to the U.S.: Article 6(3)(d) of the European Convention on Human Rights, Article 14(3)(e) of the International Covenant on Civil and Political Rights, and Article 67(1)(e) of the Rome Statute all provide equivalent rights. The HAP's three-role framework is designed to be jurisdiction-agnostic in structure. The specific evidentiary rules are U.S.-specific; the constitutional logic—the defendant has a right to confront a human who can answer for the classification—applies wherever confrontation rights exist.

6.2 Civil Law Jurisdictions

In civil law jurisdictions, the court is responsible for fact-finding and expert testimony differs from adversarial proceedings. The HAP applies with adaptation: the Certifier, Operator, and Interpreter may be called as court-appointed experts rather than party-called witnesses, and questioning is conducted by the judge. The three-role structure and documentation requirements remain applicable because the epistemic need—a human who can explain and answer for the automated classification—is identical regardless of procedural framework.

7. Open Problems

The multiple-person cost: the HAP may require three separate witnesses for a single automated classification. In proceedings with many classified evidence items, this multiplies witness costs significantly. The constitutional answer is clear—the defendant's rights are not subject to cost-benefit analysis—but practical pressure to consolidate roles or streamline testimony will be significant. Legacy classifications: the current pipeline has produced over 3.79 million classifications without a designated Certifier, Operator, or Interpreter at classification time. If any of these are offered in a legal proceeding, the HAP must be applied retroactively. Experts can review and testify about past events, but this is weaker than contemporaneous attestation. The AI witness question: the HAP assumes AI systems cannot be witnesses, which is currently true as a matter of law. If legal systems evolve to permit AI testimony, the HAP would require revision. The protocol is designed for the current legal landscape.

8. Conclusion

An algorithm can classify. An algorithm cannot testify. The Human Attestation Protocol bridges this gap—not by pretending the algorithm can speak for itself, but by specifying exactly which humans must speak for it, what they must say, and what documentation they must provide. The three-role framework—Certifier, Operator, Interpreter—maps onto the constitutional requirements established in Crawford, Melendez-Diaz, and Bullcoming with sufficient specificity that a forensic laboratory, a node operator, or a prosecutor can implement it today.

The protocol is conservative. It errs on the side of more testimony, more documentation, and more disclosure. This conservatism is deliberate. The first automated Modulign classification offered in

a criminal proceeding will be scrutinized intensely. The protocol must survive that scrutiny—not by clever argument, but by having specified in advance exactly what the Constitution requires.

References

- [1] Gonzalez, V. (2026). *The Modulign Standard v3.0*. Zenodo. <https://doi.org/10.5281/zenodo.19348703>
- [7] Gonzalez, V. (2026). *%AUT Certification Framework (Modulign v3.1 Appendix)*. Zenodo. <https://doi.org/10.5281/zenodo.19559602>
- [10] Gonzalez, V. (2026). *Chain of Custody as Embedded Property: Formalising $\wedge$ EVID in DAG-OR*. Zenodo. <https://doi.org/10.5281/zenodo.19642385>
- [16] Gonzalez, V. (2026). *Modulign: Architecture, Implementation, and Applications of a Dimensional Address Grammar for Observable Reality*. Zenodo. <https://doi.org/10.5281/zenodo.19726994>
- [17] Gonzalez, V. (2026). *The Epistemology of Observation: A Formal Certification Framework for Human and Automated Classifiers in DAG-OR*. Zenodo. <https://doi.org/10.5281/zenodo.19726226>
- [19] Gonzalez, V. (2026). *The Modulign Correction Protocol: Formal Specification for Append-Only Error Resolution in DAG-OR*. Zenodo. <https://doi.org/10.5281/zenodo.19726401>