

MODULIGN STANDARD

Automated Classifier Certification Framework

Proposed Appendix to v3.1 Specification

Vincent Gonzalez · Independent Scholar
Draft 1.0 · 2026

AUT-o · Purpose and Regulatory Context

The Modulign Standard v3.0 permits automated systems (%AUT) to produce classifications, including ^EVID-grade classifications, provided domain competence requirements are met. Version 3.0 does not specify how an automated system achieves or demonstrates competence certification. This appendix establishes that framework.

MANDATORY STATUS BY JURISDICTION

NON-OPTIONAL under the EU AI Act (in force 2024–2026):
Any high-risk AI system producing evidence used in legal proceedings must meet transparency and explainability requirements.
%AUT classifiers producing ^EVID in EU jurisdictions are high-risk.

INCREASINGLY REQUIRED in U.S. federal proceedings:
Following emerging precedent on algorithmic evidence admissibility, automated forensic classifiers face heightened Daubert scrutiny.
This framework satisfies those requirements by design.

This appendix must be published and implemented BEFORE any automated classifier produces ^EVID-grade evidence in legal proceedings.

AUT-1 · Definitions

Automated Modulign Classifier (AMC): Any software system, machine learning model, computer vision pipeline, or sensor array that produces Modulign classification codes without per-classification human review.

Human-in-the-Loop Classifier (HILC): An automated system whose outputs are reviewed and confirmed by a human observer before emission. HILCs are classified as %HYB, not %AUT.

Certification Authority (CA): An entity holding Modulign Authority (:A) status in the relevant domain, designated to evaluate AMC certification submissions.

Operational Domain: The specific Modulign domain(s) for which an AMC seeks certification. Certification is domain-specific. An AMC certified for TRN is not certified for SEC without separate evaluation.

AUT-2 · Competence Levels for Automated Systems

The %AUT competence hierarchy follows the same three levels as %HUM, with modified eligibility criteria reflecting the auditability requirements of automated systems:

Level	Human Requirement	%AUT Requirement
:T — Trained	Applied Decision Protocol under supervision at least once	Validated against labeled dataset: ≥1,000 observations per domain, ≥85% segment-level accuracy
:E — Expert	Domain specialist with demonstrated classification accuracy	≥10,000 labeled observations per domain, ≥95% segment-level accuracy, independently audited by CA
:A — Authority	Registry-certified domain authority	≥50,000 observations, ≥99% accuracy, independently audited, explainability documentation accepted, continuous monitoring required

AUT-3 · Explainability Requirements

All %AUT classifiers producing ^EVID-grade classifications must provide segment-level reasoning documentation for each classification. This is not optional. Black-box classifications are not eligible for :E or :A certification regardless of accuracy metrics.

AUT-3a — Decision Trace

For each of the eight Decision Protocol steps, the AMC must output a structured record of the observable features that drove the classification at that step.

DECISION TRACE REQUIREMENT

For machine learning systems:

Attention mapping, saliency analysis, or equivalent interpretability technique producing human-readable output per CDP step.

Minimum required trace fields per step:

- Step identifier (1–8)
- Primary observable features considered
- Decision output and alternatives considered
- Feature contribution scores

A system that produces a code string without a decision trace is NOT eligible for :E or :A certification.

AUT-3b — Uncertainty Quantification

The AMC must output a confidence score at the segment level — not only at the overall classification level. Segment-level confidence is required for all four mandatory segments and for any optional segment included in the classification.

SEGMENT-LEVEL CONFIDENCE

Required for: DOM · SUB · LOCUS (G1–G4) · NOD

Required for any optional segment included in code

Confidence aggregation method must be:

- Documented in the certification submission
- Reproducible given the same inputs
- Consistent with the weighted scoring formula of CDP Step 8

AUT-3c — Out-of-Distribution Detection

The AMC must include a mechanism for detecting observations outside its training distribution. Any classification where the OOD score exceeds the threshold established during certification must be automatically assigned ^PROV.

OOD DETECTION — MANDATORY

Classification of out-of-distribution observations as

high-confidence ^EVID is a certification VIOLATION.

OOD detection must be:

- Active on every production classification
- Calibrated during certification evaluation
- Logged in the classification's JSON reasoning record

OOD score field required in every AMC JSON output.

AUT-4 · Certification Process

Step 1 — Submission Package

The AMC operator submits the following to a Modulign Certification Authority:

1. System description: architecture, training data sources, training methodology, operational domains
2. Validation dataset: independently compiled, not used in training, minimum size per AUT-2
3. Accuracy report: segment-level accuracy across all mandatory and claimed optional segments
4. Explainability documentation: decision trace methodology, uncertainty quantification method, OOD detection mechanism
5. Operational parameters: hardware requirements, inference latency, throughput capacity, documented failure modes
6. Jurisdiction compliance statement: list of jurisdictions for ^EVID use, confirmation of applicable AI accountability law compliance

Step 2 — Independent Audit

The Certification Authority conducts an independent evaluation:

- Re-runs AMC against a held-out evaluation dataset not provided by the operator
- Reviews decision traces for a random sample of ≥ 500 classifications
- Evaluates OOD detection on a deliberately adversarial observation set
- Confirms explainability output is human-interpretable and adequate for cross-examination

Step 3 — Certification Decision

Decision	Meaning	Action Required
Full certification	AMC meets all requirements at claimed competence level	Valid 12 months. Publish in registry.
Conditional certification	AMC meets requirements at lower competence level only	Specified conditions for upgrade. Publish with conditions.
Certification denied	AMC fails one or more requirements	Documented grounds provided. Operator may remediate and resubmit.

Certification decisions are published in the Modulign registry. Operators may appeal to the Standards Advisory Council.

Step 4 — Continuous Monitoring

:E and :A certified AMCs are subject to continuous monitoring requirements:

- Monthly accuracy audit: random sample of ≥ 100 production classifications, reviewed by %HUM·[DOM]:E or higher
- Immediate re-certification required if: segment accuracy drops below threshold
- Immediate re-certification required if: AMC is retrained on new data
- Immediate re-certification required if: AMC deployed in new jurisdiction with different evidentiary requirements
- Immediate re-certification required if: ^EVID classification challenged in proceedings citing AMC methodology

AUT-5 · Prohibited Practices

The following practices void %AUT certification and render all ^EVID classifications produced during the violation period retroactively ^PROV.

CERTIFICATION-VOIDING PROHIBITED PRACTICES

- P1 Producing ^EVID classifications without segment-level decision traces
- P2 Classifying OOD observations as high-confidence ^EVID without flagging
- P3 Retraining AMC on data that includes observations it previously classified (circular training)
- P4 Operating AMC outside its certified domain without separate certification
- P5 Suppressing OOD detection results to maintain apparent high confidence
- P6 Claiming ^EVID eligibility for %HYB systems where human review is nominal rather than substantive

AUT-6 · Legal Deployment Notice

Any proceeding that intends to present %AUT ^EVID classifications as evidence must, upon request, produce the following:

7. The AMC's current certification record from the Modulign registry
8. The segment-level decision trace for the specific classification at issue
9. The OOD score for the specific classification at issue
10. The most recent monthly accuracy audit report

DEFAULT ON FAILURE TO PRODUCE

Failure to produce these documents on request does not automatically invalidate the classification.

The court may instruct that the ^EVID status be treated as ^VFYD for purposes of the proceeding:

- authenticity maintained
- non-accidentality guarantee removed

This preserves the evidentiary record while appropriately limiting the epistemic claim made on behalf of the classification.

AUT-7 · Relationship to Existing AI Governance

Framework	AUT Framework Status	Notes
EU AI Act	Compliant by design	AMCs producing legal evidence = high-risk. Explainability (AUT-3) + accuracy auditing (AUT-4) satisfy Art. 9–15.
NIST AI RMF	Aligned	Decision trace = Explainability. OOD detection = Reliability. Monthly audits = Monitoring.
ISO/IEC 42001	Compatible	Certification process maps to conformity assessment requirements.
U.S. Daubert	Satisfies	Documented methodology, reproducible, known error rate (confidence score + accuracy audit), auditable application.
ACPO Digital Evidence	Extends	AUT-3 decision trace extends ACPO's integrity principle to automated classifiers.

MODULIGN™ · %AUT Certification Framework · Draft 1.0 · 2026

Proposed Appendix — Modulign Standard v3.1

© 2026 Vincent Gonzalez · All Rights Reserved