The ASA Evidence Standard

A Public Reference for Evidence Scoring in the Treadstone 71 Adversary Scoring Architecture

White Paper · ASA Evidence Standard v0.3 · part of the Treadstone 71 Adversary Scoring Architecture (T71-ASA) · Q2 2026 free public release · document v0.3.2

BLUF

The ASA Evidence Standard scores an intelligence artifact across 30 factors grouped into nine clusters and returns three numbers: a T-Score for evidence quality, a D-Score for adversary deception likelihood, and an F-Score for forecast confidence on the adversary's most-likely achieved endgame. The standard preserves the NATO Admiralty source-and-information grades as two of its 30 factors and adds 28 factors built for the cyber, AI, and cognitive-warfare conditions analysts work in today. The tri-score separates two judgments the older single-grade systems collapse: how good the evidence looks, and whether an adversary built the evidence to look that way.

The ASA Evidence Standard is the evidence layer of the Treadstone 71 Adversary Scoring Architecture. It supplies the source grades that anchor every TAI sub-index score, and it separates verified fact from adversary claim across the catalog. This reference publishes the standard's structure. The precise within-cluster weights, the scoring formulas, and the T71 Cognitive Forensics Suite stay with the paid tiers and the inaugural release.

Chapter 1 — Why the Standard Exists

1.1 The evidence-scoring problem in 2026

Analysts work with evidence the 1940s Admiralty Code authors never imagined. A typical week touches Telegram channels run by pseudonymous milbloggers, leaked code carrying cultural-language artifacts, AI-generated video of named officials, coordinated inauthentic behavior networks running across five platforms at once, financial flows routed through six jurisdictions, and intercepts that arrive machine-translated, the translation choices shaping the read. The artifact set that lands on an analyst's screen each morning carries layers of curation the Admiralty Code, built for naval HUMINT debriefs in 1942, cannot describe.

The gap matters operationally. An analyst who grades a manipulated artifact as B-2 — Usually Reliable, Probably True — signals high confidence to the customer. The customer decides on that confidence. When the artifact turns out to be the visible part of an adversary's coordinated deception plan, the customer's decision runs against a picture the adversary built. The analytic chain failed not because the analyst missed the deception — the Admiralty Code does not score deception — but because the standard lacked the vocabulary for the question.

1.2 The Admiralty Code and its limits

The Admiralty Code, codified in NATO STANAG 2022 and folded into AJP-2.1, grades source reliability on an A-F scale and information credibility on a 1-6 scale. The two grades served well for decades of HUMINT-led

collection. The standard carries them forward as Factor 1 and Factor 8, which preserves backward compatibility and lets a customer read an ASA score against the rating language already in the workflow.

The two grades do not, on their own, address synthetic media, cryptographic provenance, coordinated inauthentic behavior, analyst bias disclosure, or adversary deception operating against the analytic question. The ASA Evidence Standard keeps the lineage and adds the missing dimensions.

1.3 The cyber, AI, and cognitive-warfare gap

Three pressures define the modern evidence problem. Cyber artifacts arrive with technical chains of custody that an analyst can verify or fail to verify. AI synthetic media arrives indistinguishable from authentic media to the unaided eye. Cognitive-warfare artifacts arrive shaped by coordinated influence operations that target the analyst's read as much as the public's. The ASA Evidence Standard scores all three, and it scores the analyst's own tradecraft fit alongside them.

Chapter 2 — The Architecture

2.1 Design principles

The standard rests on four principles. Backward compatibility keeps the Admiralty grades in place. Dual scales serve both human reading and machine integration. Cluster grouping lets an analyst score related factors as a coherent set. Auditability means a customer traces any score back through factors and clusters to named open-source citations.

2.2 Dual scales and five named bands

Every factor scores on two parallel scales. The 0-6 scale preserves Admiralty granularity. The 0-100 scale supplies operational resolution for tooling. Five bands carry consistent names and thresholds across factors.

Band	0-6	0-100	Sherman Kent gloss	ICD 203
Apex	6	90-100	Almost certain (93%+)	High
Mature	5	70-89	Highly likely (75-85%)	High / Moderate
Operational	3-4	45-69	Likely / Roughly even (40-65%)	Moderate
Developing	1-2	20-44	Unlikely (15-30%)	Low
Nascent	0	0-19	Highly unlikely (<15%)	Low

The Apex through Nascent naming aligns with the broader T71 capability-maturity vocabulary used across the six T71 Standard pillars, which keeps a customer's reading consistent across products. The band thresholds match the Admiralty lineage mapping in the published scoring workbook — A/1 grades land in the Apex band at 90-100, and B/2 grades land in the Mature band at 70-89, where most vendor CTI reporting sits. The Sherman Kent mapping draws on Kent's 1964 *Studies in Intelligence* paper, “Words of Estimative Probability,” using the central values of Kent's ranges with adjustments for 2026 operational reality.

2.3 The seven evidence-quality clusters

Twenty evidence-quality factors distribute across seven clusters. Each cluster groups factors that share an analytic dimension and that an analyst scores together.

Cluster	Factors	What it scores
C1 Source Credibility	F1 Source Reliability (Admiralty A-F), F2 Source Authority, F3 Source Independence, F4 Source Access/Proximity	The source, independent of content
C2 Information Validity	F5 Corroboration, F6 Internal + External Consistency, F7 Specificity/Granularity, F8 Information Credibility (Admiralty 1-6)	The content, independent of source
C3 Relevance	F9 Topical Relevance, F10 Temporal Relevance/Currency	Fit to the standing requirement and the decision deadline
C4 Cyber/Technical Integrity	F11 Chain of Custody/Provenance, F12 Technical Verifiability, F13 Metadata Integrity	The technical chain of custody
C5 AI/Synthetic Media Risk	F14 AI-Generation Probability, F15 Provenance Standard Adherence, F16 Manipulation Forensics	Synthetic-media risk (inverse-scored: higher score, lower risk)
C6 Cognitive/Influence-Op Risk	F17 Coordinated Inauthentic Behavior Indicators, F18 Propaganda Technique + BEND Markers	Influence-operation risk (inverse-scored)
C7 Analytic Tradecraft Fit	F19 Bias Disclosure, F20 Confidence Calibration + Reproducibility	The analyst's own work

F1 preserves the Admiralty A-F grade. F8 preserves the Admiralty 1-6 grade. The two carry the lineage; the other 18 evidence-quality factors carry the modern dimensions. C5 and C6 score inverse — a clean artifact with no synthesis signal and no coordinated-inauthentic-behavior signal earns a high score.

2.4 The T-Score

The seven evidence-quality clusters integrate through a weighted mean into the T-Score on the 0-100 scale. A high T-Score indicates strong, well-corroborated evidence with a clean technical chain, low synthetic-media risk, low influence-operation risk, and disclosed analyst tradecraft. The standard publishes the cluster structure here; the precise weights stay with the paid release and carry forward to empirical recalibration at v1.0.

Chapter 3 — The Deception and Forecast Extension

3.1 The deception-detection gap

Evidence quality and deception likelihood are different questions. A high-quality artifact set shaped by a coordinated adversary plan produces high evidence confidence and dangerous false comfort under any single-grade standard. Barton Whaley's principle holds: deception runs more often than analysts detect it. The standard answers the deception question on its own track rather than folding it into evidence quality.

3.2 Cluster C8 — Counterdeception Detection (the D-Score)

Cluster 8 outputs the D-Score. The score expresses the analyst's estimate that adversary denial and deception operate against the analytic question. A high D-Score warns that the artifact set arrives shaped by an adversary

plan. A low D-Score does not certify the absence of deception. The base score sits at 25 — the Developing band, “unlikely but plausible” — and adjusts upward on observed indicators.

Five factors compose Cluster 8, covering deception-marker density, cross-INT discrepancy, signature anomaly against the documented adversary baseline, and the remaining counterdeception signals. The deception-marker factor scores against a named catalog: 11 Treadstone 71 cyber and physical deception techniques — Mimicking, Inventing, Decoying, Double Play, Double Bluff, Red Flagging, Paltering, Negative Spin, Feint Demonstration, Masking, and Dazzling — and seven Soviet active-measures patterns — Dezinformatsiya, Kombinatsiya, Provokatsiya, Fabrikatsiya, Diversiya, Proniknovenniye, and the agent-of-influence pattern. The cross-INT factor measures variance across independent intelligence disciplines reporting on the same event: ambiguity-type deception produces measurable divergence when different sensors collect on different fabricated alternatives. The marker severity weights and the within-cluster weighting stay with the paid standard and Appendix B.

3.3 Cluster C9 — Adversary Endgame Forecasting (the F-Score)

Cluster 9 outputs the F-Score, paired with three named endgame hypotheses. The cluster forecasts what the adversary plans, which targets the activity converges on, when execution will fail, and what outcome the adversary actually wants from the current position.

Five factors compose Cluster 9. Target Set Convergence measures how many independent indicators across disciplines point to a common target class — real preparation concentrates, feints scatter. Capability-Intent Alignment scores the intersection of what the adversary can execute and what it intends to execute, since capability alone produces false positives and intent alone produces false negatives. Timeline Compression measures whether preparatory signals accelerate toward execution or decay toward cancellation. Cross-Domain Coordination measures synchronization across the AI, cyber, physical, and cognitive domains. The fifth factor and the within-cluster weighting stay with the paid standard.

3.4 The tri-score output and the triple endgame

The standard returns three scores per assessment plus named endgame hypotheses.

The T-Score scores evidence quality from clusters C1 through C7. The D-Score scores deception likelihood from cluster C8 with its floor of 25. The F-Score scores forecast confidence from cluster C9. A cross-score interpretation matrix guides the analyst's posture when the three combine.

T	D	F	Analyst posture
High	Low	High	Publish with high confidence; standard tradecraft
High	High	High	Treat artifacts as adversary product; counter-forecast against the deception hypothesis
High	Low	Low	Solid evidence, unclear endgame; collect on the adversary's value structure
Low	High	High	Likely a known adversary hand, poorly evidenced; build collection on alternate streams
Low	Low	High	Forecast outpaces evidence; check C9 inputs for over-fitting
Low	High	Low	Worst case — poor evidence shaped by deception; treat as adversary-fed
Low	Low	Low	Insufficient; defer publication

T	D	F	Analyst posture
High	High	Low	Strong evidence, clear deception signal, unclear endgame; surface the deceptive purpose

The named endgames preserve three forecasts every adversary assessment must hold apart: the Stated Endgame the adversary declares, the Optimal Endgame available from the realized position, and the Most-Likely Achieved Endgame friction will actually produce. Russian operations in 2022 supply the case — the Stated Endgame ran Kyiv decapitation, the Optimal Endgame from the realized position became Donbas consolidation, and the achieved outcome ran between the two. Collapsing the three into one estimate hides the structure a customer decision rests on.

Chapter 4 — The Cognitive Forensics Suite (Held)

The Treadstone 71 Cognitive Forensics Suite extends the standard with 15 novel detection methods across strategic, operational, and tactical bands — among them doctrinal drift detection, counterfactual inference from negative space, decision-cycle topology mapping, and cross-persona entropy scoring. The suite shipped with the July 15, 2026, inaugural release inside the paid tiers; this free reference does not document the method mechanics.

Two methods sit exclusive to the Executive Desk tier — M9 (CRT) and M15 (CAVI) — because they require offensive operational integration no commercial vendor maintains. The two release only under a named-personnel mutual non-disclosure agreement.

Chapter 5 — Worked Example

Consider an open-source artifact: a video, posted to a pseudonymous Telegram channel, showing a named official appearing to announce a policy reversal. An analyst scores it under the standard.

The T-Score runs moderate. Source credibility (C1) sits low — a pseudonymous channel with no institutional authority and unclear proximity. Information validity (C2) is mixed — the content is specific but lacks independent corroboration. Cyber/technical integrity (C4) is weak — no cryptographic provenance, inconsistent metadata. AI/synthetic-media risk (C5) is the decisive cluster — detector consensus flags a moderate-to-high probability of synthesis, and no C2PA credential is present. The integrated T-Score lands in the Operational band: the evidence is real enough to act on cautiously but far from clean.

The D-Score runs high. Two deception markers register — a Red Flagging pattern and a Fabrikatsiya pattern — and a cross-INT discrepancy appears, since signals and human reporting do not corroborate the announcement. The D-Score rises well above its floor of 25 into the Mature band.

The F-Score runs moderate. Target Set Convergence is weak — the artifact does not concentrate on a coherent target class — and Timeline Compression shows no acceleration. The named Most-Likely Achieved Endgame reads as a narrative-shaping operation rather than a genuine policy signal.

The tri-score reads T-moderate, D-high, F-moderate. The cross-score matrix points the analyst to the deception posture: treat the artifact as probable adversary product, do not publish the policy reversal as fact, and counter-

forecast against the narrative-shaping hypothesis. A single Admiralty grade would have recorded “B-3, Usually Reliable, Possibly True” and lost the deception signal entirely.

Chapter 6 — How the Standard Fits TAI, and How to Engage

The ASA Evidence Standard is the evidence layer of the five-layer T71-ASA architecture. Every TAI sub-index score rests on artifacts the standard has graded, and the tri-score discipline keeps verified fact separate from adversary claim across the 42-actor catalog. The Universal Indicator Framework supplies the indicators, the Cultural Nexus layer supplies the context, and the ASA Evidence Standard supplies the evidence grades that make the composite auditable.

The standard publishes here as a free reference because an open evidence standard is the point — a customer evaluates the logic and reproduces a finding rather than trusting a black box.

Chapter 7 — Limitations and Roadmap

The standard carries known limitations. The within-cluster weights rest on expert-judgment calibration pending pilot empirical validation; the v1.0 release will publish revised weights from observed performance across the pilot corpus. Deception scoring measures the likelihood of denial and deception, not its certainty — a low D-Score never certifies the absence of an adversary plan. Forecast scoring depends on indicator coverage; thin collection produces thin forecasts, and the standard marks that thinness rather than hiding it.

The roadmap ran from this v0.3 public reference through the July 15, 2026, inaugural release, which activated the full standard across the paid tiers. The v1.0 release publishes empirically recalibrated weights.

Direct engagement runs through www.treadstone71.com, www.cybershafarat.com, and www.cyberinteltrainingcenter.com.

Document control

Field	Value
Document title	The ASA Evidence Standard — A Public Reference
Architecture	Treadstone 71 Adversary Scoring Architecture (T71-ASA)
Standard	ASA Evidence Standard — 30 factors, 9 clusters, tri-score (T/D/F)
Document version	v0.3.2 — August 13, 2026 (supersedes v0.3.1 of July 22, 2026)
v0.3.2 change note	Reconciles the inaugural release date to the July 15, 2026 ship date and records the October 5, 2026 release (TAI Q3 refresh plus CWTR inaugural). Standard version unchanged at v0.3; no structural change.
v0.3.1 change note	Reconciled the Section 2.2 band thresholds to the published scoring workbook (Apex 90-100; Mature 70-89). Standard version unchanged at v0.3; no structural change.
Information cutoff	June 26, 2026
Classification	Free public release — Tier 0

Field	Value
Held for paid tiers	Within-cluster weights, scoring formulas, Appendix B marker weights, the Cognitive Forensics Suite
Inaugural release	July 15, 2026
Next release	October 5, 2026 — TAI Q3 refresh plus CWTR inaugural
Author	Treadstone 71 Editorial Production Team

End of document.