# The ASA Evidence Standard

## Grading evidence, deception, and endgame confidence in intelligence work

Public Edition 1.2. First published September 15, 2026. Revised October 4, 2026. Treadstone 71 LLC, Half Moon Bay, California. License: CC BY-NC-ND 4.0.

Cite as: Treadstone 71 LLC. The ASA Evidence Standard, Public Edition 1.2. Half Moon Bay, CA: Treadstone 71 LLC, October 4, 2026. https://www.treadstone71.com/asa-evidence-standard. All versions: https://doi.org/10.5281/zenodo.22662241. Supersedes Public Edition 1.1, Zenodo version DOI 10.5281/zenodo.23087062.

**Public-safety note.** The edition in hand discloses the standard at the cluster and factor level, with public glosses, scoring anchors, grading rules, bands and actions, a deception catalog, and four worked examples. Equal cluster weights are the method of the Public Edition. The internal edition adjusts the weights, adds per-factor scoring bands, and holds the calibration record. Each place where the internal edition adds detail carries the tag [internal edition].

**Version history.**

| Version | Date | Change |
|---|---|---|
| 1.0 | September 15, 2026 (printed). Deposited September 8, 2026 | First public edition. Three tearlines and a 30-technique reference. Zenodo record 22662242, version v1. |
| 1.1 | October 1, 2026 | Four tearlines. The 30-technique reference becomes a crosswalk to the Advanced Structured Analytic Techniques, 66 techniques in nine families. Section 2.8 on technical artifacts. Table 5 wording for a deception score of 85 or more aligned to a one-band drop. Cluster scores added to the Section 6.4 affidavit. Corrections to Sections 1, 2.2 to 2.4, 6.2, and 6.3, to references and links, and to license scope. Zenodo version 3, DOI 10.5281/zenodo.23087062. |
| 1.2 | October 4, 2026 | Section 2.9 states how clusters C8 and C9 produce the D-Score and the F-Score. Table 5A fixes the executive words for each score. Version 1.0 row corrected to the deposited file. Citation points to the all-versions DOI. Section 12 names the license, CC BY-NC-ND 4.0, and bars commercial reuse and derivative works. |

## Tearlines

The paper is cut at four tearlines. Read the tearline that matches the desk.

Table A. Tearlines.

| Tearline | Reader | Sections | Advanced Structured Analytic Technique families |
|---|---|---|---|
| Tearline 1, Strategic | Commanders, boards, contracting officers | Executive summary; What changes for the reader; 2.5 Bands and actions; 10 Adoption; 13 Requests | Threat Calibration; Strategic Foresight |
| Tearline 2, Operational | Desk leads, collection managers, program owners | 1 The two-axis code; 2.6 Scoring effort; 7 AI-augmented work; 8 Where the standard runs | Counterintelligence; Cognitive Denial and Deception; Offensive Tradecraft |
| Tearline 3, Tactical | Analysts and threat hunters grading items | 2.1 to 2.4 Anchor, factors, tri-score, anchors; 2.9 D and F; 3 Rules; 4 Probability language; 5 Deception catalog; 6 Worked examples | Adversary Cognition; Cognitive Warfare |
| Tearline 4, Technical | Detection engineers, forensic examiners, review boards, methodologists, counsel | 2.7 Dependencies; 2.8 Grading technical artifacts; 2.9 Engine; 9 Validation and limits; 11 Compliance; 12 Use and license | Cyber and AI Disruption; Convergent Deception |

Each section heading below carries its tearline in brackets.

## Tactical and technical intelligence

Tactical cyber intelligence follows the adversary's decisions and conduct during an engagement. Analysts ask how the operator works. Technical cyber intelligence follows the machine record: artifacts, forensic signatures, and telemetry produced during or after execution. Analysts ask what ran, and where. The two tearlines grade different evidence, and the standard keeps them apart.

Mixing them reduces analysis to indicator tracking. A hash match proves a file ran. Who typed the commands, and what the operator does next, stays unanswered. Technical indicators spoil fast, because a recompiled binary or a rotated proxy changes the signature and leaves the adversary's skill untouched. Tactical analysis rebuilds operator conduct, and operator conduct changes only when the operator retrains.

Table B. Tactical and technical intelligence compared.

| Dimension | Tactical (Tearline 3) | Technical (Tearline 4) |
|---|---|---|
| Object graded | Operator behavior, tradecraft, maneuver sequence | Machine artifacts, forensic residue, code |
| Question answered | How does the adversary conduct the operation? | What asset or code ran on the endpoint or network? |

| Dimension | Tactical (Tearline 3) | Technical (Tearline 4) |
|---|---|---|
| Typical evidence | Tactics, techniques, and procedures (TTPs), command sequences, movement patterns, evasion habits | Hashes, IP addresses, registry keys, YARA signature rules, compiled payloads, command-and-control beacon intervals |
| Shelf life | Medium to long. Changing human tradecraft takes retraining | Hours to days. A recompile or proxy rotation changes the signature |
| Reader | Incident commanders, threat hunters, collection managers | Detection engineers, security operations center analysts, automated endpoint and log-correlation rules |
| Method | Behavioral profiling, intent reading, stylometry of operator scripts, forecasting | Signature extraction, reverse engineering, static and dynamic tracing |
| Where ASA grades it | C1 to C3 and C6 to C9 | C4 and C5, and the signature side of F23 |
| Advanced Structured Analytic Technique families | Adversary Cognition; Cognitive Warfare | Cyber and AI Disruption; Convergent Deception |

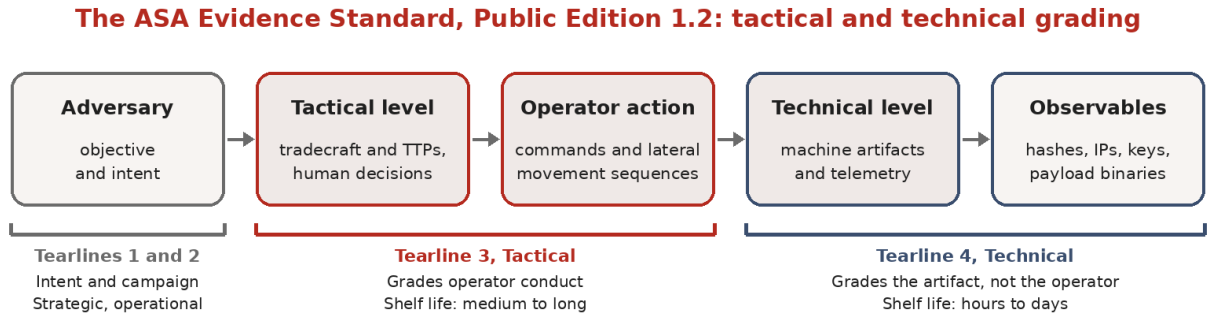



*Figure A. Where tactical grading ends and technical grading begins.*

Grades follow the evidence type. A technical grade attaches to the artifact. Judgments about operator intent, motive, or endgame earn their own grade in Tearline 3, from tactical evidence. Section 2.8 states the rule for the grader. A forecast of the operator's next move, built on behavioral dependencies seen in similar campaigns, prints a Section 4 band, and the T-Score on the behavioral evidence sets the cap.

## Executive summary [Tearline 1]

The ASA Evidence Standard grades an intelligence artifact across 30 factors in 9 clusters and returns three separate numbers. Evidence quality gets the T-Score. Deception likelihood, meaning the probability an adversary shaped the evidence, gets the D-Score. Endgame confidence gets the F-Score. ASA stands for Adversary Scoring Architecture, the Treadstone 71 engine beneath the Adversary Index. Its target is a documented failure of the two-axis Admiralty

code, which lets a well-built denial and a corroborated attribution sit at the same grade and carries no check for deception, self-interest, sourcing type, or synthetic media. Beneath the new factors, the classic Admiralty scales stay in place as the entry anchor, source reliability A through F and information credibility 1 through 6, each printed as its own table with the NATO STANAG 2511 definitions and never averaged. Each score carries a band, and each band carries an action: the T-Score caps the likelihood a product prints, the D-Score holds a decision, and the F-Score decides whether an endgame gets named. Self-serving claims grade no higher than fairly reliable. Verified fact and adversary claim never share a sentence. An assessment without a confidence marker is a draft, not a product.

## What changes for the reader [Tearline 1]

For a commander or a board, the standard prices deception and self-interest into the grade before money moves. Under the two-axis code, adversary-built evidence enters the decision at full credit. ASA flags shaped evidence on its own axis with the D-Score, and the F-Score names the endgame to plan against instead of the endgame the adversary announced. Section 6.1 grades the source behind the 2002 mobile-weapons estimate and shows the judgment the standard permits: Unlikely, with a decision hold.

For an analyst, the standard replaces a gut read with a fixed scoring path. Minutes cover a routine item and an hour covers an item that drives a decision. What comes out is a number a second reader reproduces.

For a review board or a litigator, the standard leaves a one-page ledger. The ledger runs from the headline judgment back to a named source, a grade on each axis, the deception markers observed, and the analyst's own bias disclosure. A finding survives the post-mortem because the chain is on paper.

## 1. The two-axis code and where it fails [Tearline 2]

The Admiralty code, codified in NATO STANAG 2511 (Reference 2) and carried in Allied Joint Publication AJP-2.1 (Reference 1), grades a source on a letter scale from A to F and grades a report on a number scale from 1 to 6. A report tagged B2 tells a reader the source usually proves reliable and the content is probably true. The pairing served decades of human-source collection. Five weaknesses weigh more now than when the British Admiralty's Naval Intelligence Division built the code during World War II (Reference 4).

**Failure one:** the axes collapse in practice. STANAG 2511 requires reliability and credibility to be judged independently, and AJP-2.1 repeats the rule at paragraph 3.33 (Reference 1). Analysts do not hold them apart. Mandel, Irwin, Dhami, and Budescu ran two experiments with professional intelligence analysts, 25 in the first and 27 in the second after screening (Reference 4), across a six-by-six grid of reliability and credibility levels. As inconsistency between the two cues rose,

intra-analyst reliability, inter-analyst agreement, and effective cue use all fell. Directionality of the scale terms led analysts to lump the first three levels of each scale together and the fourth and fifth together. Classification level, the third cue tested, produced no statistically significant effect on accuracy judgments. A 2025 study in Judgment and Decision Making (Reference 5), with 74 analysts and 175 nonexperts across two experiments, found judgment reliability highest when the two cues were moderately consistent and lower when they were maximally inconsistent or maximally consistent. Under inconsistent cues, 44 percent of analysts in the first experiment and 66.7 percent in the second used neither cue as a reliable predictor of accuracy.

**Failure two:** F6 becomes a dumping ground. The code assigns F to a source "which has not been used in the past" and 6 to an item "which provides no basis for comparison with any known behaviour pattern of a target." A new source reporting a new event lands at F6 by the letter of the standard. Open-source and cyber collection produce new sources daily. The highest-uncertainty bin fills with material that deserves the closest handling.

**Failure three:** no treatment of self-interest. The code has no field for a source that pleads its own cause. A ministry denying an operation and a neutral technical report grade on the same letter, with nothing in the method that caps the denial.

**Failure four:** no line between technical and human sourcing. Irwin and Mandel (Reference 3) note the code draws no distinction between a human source and a sensor, or between a primary source and a relay. A cryptographically verifiable artifact and a secondhand rumor carry the same grade.

**Failure five:** no denial-and-deception check. Nothing in the two axes asks whether the adversary built the artifact to be found. Whaley's Stratagem (Reference 6), first issued at MIT in 1969 and reissued by Artech House in 2007, tabulated more than one hundred cases of surprise from 1914 to 1968 and found that planned deception seldom failed to produce surprise. A high grade on well-made adversary material signals false comfort to the customer.

Deception runs more often than analysts detect it.

## 2. What the standard adds [Tearlines 1 to 4]

### 2.1 The entry anchor [Tearline 3]

The standard keeps the Admiralty scales at the front door and preserves the STANAG 2511 wording. AJP-2.1 Edition B Version 3 prints the same terms in Table 3.1. The long definitions below come from STANAG 2511, first edition 2003, which Reference 4 calls defunct. Source reliability appears as one table. Information credibility appears as a second table. The two grades print side by side and never merge.

Table 1. Source reliability (NATO STANAG 2511).

| Grade | Term | Definition |
|---|---|---|
| A | Completely reliable | Refers to a tried and trusted source which can be depended upon with confidence. |
| B | Usually reliable | Refers to a source which has been successful in the past but for which there is still some element of doubt in a particular case. |
| C | Fairly reliable | Refers to a source which has occasionally been used in the past and upon which some degree of confidence can be based. |
| D | Not usually reliable | Refers to a source which has been used in the past but has proved more often than not unreliable. |
| E | Unreliable | Refers to a source which has been used in the past and has proved unworthy of any confidence. |
| F | Reliability cannot be judged | Refers to a source which has not been used in the past. |

Table 2. Information credibility (NATO STANAG 2511).

| Grade | Term | Definition |
|---|---|---|
| 1 | Confirmed by other sources | If it can be stated with certainty that the reported information originates from another source than the already existing information on the same subject, it is classified as "confirmed by other sources." |
| 2 | Probably true | If the independence of the source of any item cannot be guaranteed, but if, from the quantity and quality of previous reports its likelihood is nevertheless regarded as sufficiently established. |
| 3 | Possibly true | If, despite there being insufficient confirmation to establish any higher degree of likelihood, a freshly reported item does not conflict with the previously reported behaviour pattern of the target. |
| 4 | Doubtful | An item of information which tends to conflict with the previously reported or established behaviour pattern of an intelligence target. |
| 5 | Improbable | An item of information which positively contradicts previously reported information or conflicts with the established behaviour pattern of an intelligence target in a marked degree. |
| 6 | Truth cannot be judged | Any freshly reported item of information which provides no basis for comparison with any known behaviour pattern of a target. |

## 2.2 Nine clusters, thirty factors [Tearline 3]

Behind the anchor sit 30 factors in 9 clusters. Fifteen evidence-quality factors across five clusters, C1 through C4 and C7, roll up to the T-Score. Five risk factors in two clusters, C5 and C6, set a cap on the T-Score. Cluster C8 produces the D-Score and cluster C9 the F-Score. Factor 1 carries the Admiralty A to F grade and Factor 8 carries the Admiralty 1 to 6 grade, so a unit reads an ASA result against the language already in its workflow. Each factor scores 0 to 100 on the anchors in Section 2.4.

Table 3 gives the public gloss for each factor, the question the analyst answers. Clusters C4 and C5 carry added grading rules in Section 2.8 [Tearline 4]. The internal rubric adds per-factor evidence thresholds [internal edition].

**Table 3. Clusters and factors with public glosses.**

| Cluster | Factor | Public gloss |
|---|---|---|
| C1 Source Credibility | F1 Source Reliability (Admiralty A to F) | The source's track record, graded on Table 1. |
| C1 | F2 Source Authority | Does the source hold a position to know the matter reported? |
| C1 | F3 Source Independence | Does the source stand apart from the parties with a stake in the report? |
| C1 | F4 Source Access and Proximity | How close did the source sit to the event, in time and in place? |
| C2 Information Validity | F5 Corroboration | How many independent sources agree, and how independent are they? |
| C2 | F6 Internal and External Consistency | Does the item hold together, and does it fit what is already established? |
| C2 | F7 Specificity and Granularity | Does the item name checkable times, places, and amounts? |
| C2 | F8 Information Credibility (Admiralty 1 to 6) | The content grade on Table 2. |
| C3 Relevance | F9 Topical Relevance | Does the item answer the requirement on the desk? |
| C3 | F10 Temporal Currency | Is the item current against the decision deadline? |
| C4 Cyber and Technical Integrity | F11 Chain of Custody and Provenance | Is the path from collection to the analyst documented and unbroken? |
| C4 | F12 Technical Verifiability | Does a second party reproduce the technical finding from the artifact? |
| C4 | F13 Metadata Integrity | Do timestamps, headers, hashes, and file properties agree with the claimed origin? |
| C5 AI and Synthetic Media Risk (risk flag) | F14 AI-Generation Probability | How likely is the artifact machine-generated in whole or in part? |
| C5 | F15 Provenance Standard Absence | Does the artifact lack a content-credential or provenance record, or fail validation? |

| Cluster | Factor | Public gloss |
|---|---|---|
| C5 | F16 Manipulation Forensics | Do forensic checks find editing, splicing, or synthesis? |
| C6 Cognitive and Influence-Op Risk (risk flag) | F17 Coordinated Inauthentic Behavior Indicators | Do account networks, timing, and amplification patterns show coordination? |
| C6 | F18 Propaganda Technique and BEND Markers | Does the item carry named propaganda techniques or BEND maneuver signatures? |
| C7 Analytic Tradecraft Fit | F19 Bias Disclosure | Did the analyst record assumptions, gaps, and tool use, including AI assistance? |
| C7 | F20 Confidence Calibration and Reproducibility | Does a second analyst reach the same grade from the same ledger? |
| C8 Counterdeception (D-Score) | F21 Deception Indicator Density | How many catalog markers appear, weighted by severity? |
| C8 | F22 Cross-INT Discrepancy | Do other disciplines contradict the item? |
| C8 | F23 Signature Anomaly Index | Does the item's technical or linguistic signature depart from the actor's known pattern? |
| C8 | F24 Cover Story Plausibility | Does the item's benign explanation survive a hostile reading? |
| C8 | F25 Action-Narrative Coherence Gap | Does what the actor says match what the actor does? |
| C9 Endgame Forecasting (F-Score) | F26 Target Set Convergence | Are observed targets converging on one objective? |
| C9 | F27 Capability-Intent Alignment | Does demonstrated capability match the stated or inferred intent? |
| C9 | F28 Timeline Compression | Are intervals between observed steps shortening? |
| C9 | F29 Cross-Domain Coordination | Do cyber, kinetic, economic, and narrative lines move together? |
| C9 | F30 Endgame Path Probability | Which of the three endgame hypotheses does the evidence favor? |

BEND refers to the sixteen influence maneuvers in Carley's social-cybersecurity framework, four positive and four negative for narrative and four of each for network structure: back, build, bridge, boost; engage, explain, excite, enhance; neutralize, negate, narrow, neglect; dismiss, distort, dismay, distract. Reference 15 gives the source.

### 2.3 The tri-score [Tearline 3]

Three numbers come out of one assessment. Each runs from 0 to 100, and the D-Score starts at its floor of 25. Figure 1 shows the path from item to product sentence.

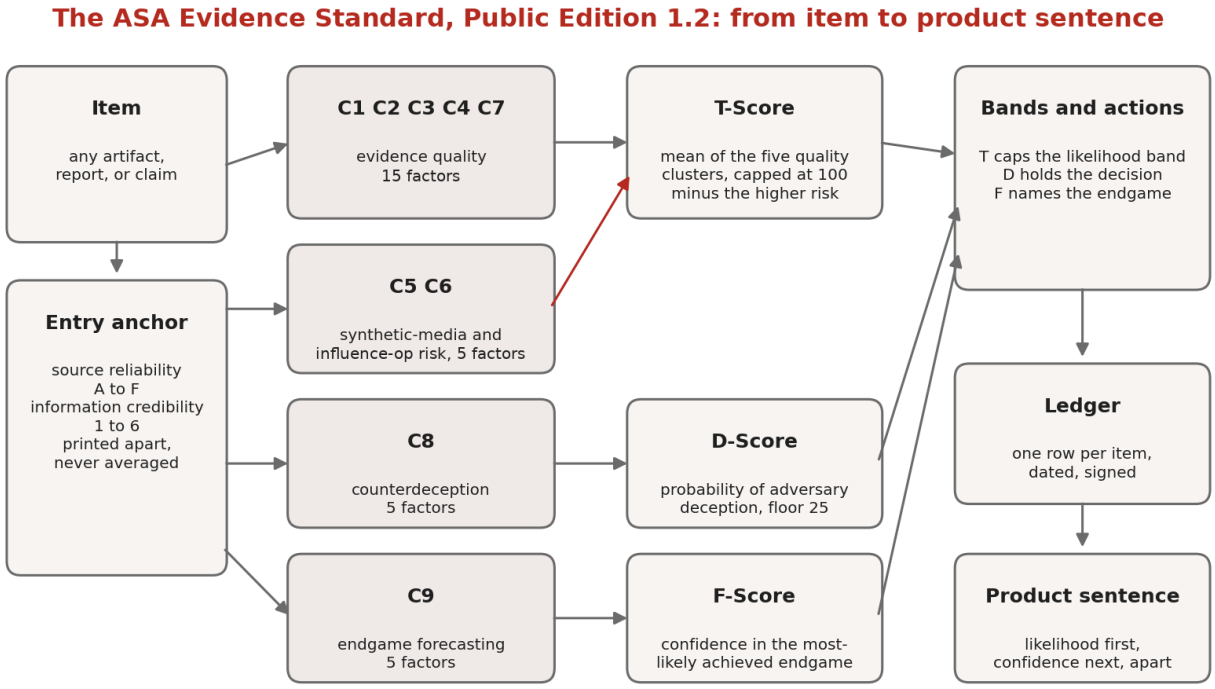



*Figure 1. From item to product sentence.*

The T-Score is the Public Edition method in one line: the mean of the applicable quality clusters, C1, C2, C3, C4, and C7, capped at 100 minus the higher of the two risk scores. Each cluster score is the mean of its factor scores. Where a cluster does not apply to an item, such as C4 on a purely human-source report, it is marked not applicable and left out of the mean. The risk clusters C5 and C6 score in the risk direction, 0 for no risk and 100 for confirmed risk, and they never add to the T-Score. They only pull it down. No artifact earns a bonus for being clean. A synthetic or coordinated artifact pays a penalty in proportion to the risk. The internal edition refines the cluster weights within the same structure [internal edition].

The D-Score is a probability, expressed on the Kent-style spectrum in Section 4, that the item set carries adversary deception directed at the analytic question. Cluster C8 produces it. Its floor is 25, inside the Unlikely band, and it rises on observed indicators. A low D-Score never certifies the absence of deception; it records that no indicator fired. Product sentences print the D-Score with its Kent-style term, as in "D-Score 78, deception Likely." Section 2.9 states how C8 and C9 produce the D-Score and the F-Score.

The F-Score is confidence, 0 to 100, in one hypothesis only: the Most-Likely Achieved Endgame. Cluster C9 produces it and holds two further hypotheses apart, the Stated Endgame the adversary declares and the Optimal Endgame available from the realized position. Each of the three carries its own probability in the ledger, and the product names an endgame only when the F-Score band in Section 2.5 permits. Russian operations in 2022 show the split. Moscow's stated objective was

a decapitation strike on Kyiv. From the realized position, the optimal objective was consolidation in the Donbas. What Russia achieved fell between the two.

A forecasting score sits inside an evidence standard for one reason. Endgame confidence is only as good as the graded evidence beneath it. The F-Score is the confidence output of graded evidence applied to the three-hypothesis frame, not a separate forecasting method.

The three scores never merge. A high T-Score beside a high D-Score is the case the two-axis code cannot express: good-looking evidence an adversary built.

### 2.4 Scoring anchors [Tearline 3]

Factors take a 0 to 100 score from one of two anchor sets. Quality factors, F1 through F13, F19, F20, and F26 through F30, use the quality anchors. Risk factors, F14 through F18 and F21 through F25, use the risk anchors. Factors 1 and 8 map from the Admiralty grade: A and 1 score 90, B and 2 score 70, C and 3 score 50, D and 4 score 30, E and 5 score 10, and F and 6 score 50 with a not-yet-judged flag.

**Table 4. Scoring anchors.**

| Band | Quality factors | Risk factors |
|---|---|---|
| 0 to 20 | Absent or contradicted | No indicator observed |
| 21 to 40 | Weak, a single indication | One weak indicator |
| 41 to 60 | Partial, indications with gaps | One strong indicator or two weak |
| 61 to 80 | Satisfied on the item's own terms | Two or more indicators that agree |
| 81 to 100 | Satisfied and independently corroborated | Confirmed by a second discipline or an admission |

Per-factor thresholds sit in the internal rubric [internal edition].

### 2.5 Bands and actions [Tearline 1]

A score without an action is a label. Every score lands in a band, and the band binds the product.

**Table 5. Bands and actions.**

| Score | Band | Action the band requires |
|---|---|---|
| T 80 to 100 | Strong | Any likelihood band on the spectrum is available to the product |
| T 60 to <80 | Adequate | Likelihood capped at Likely; nothing prints above 85 percent |
| T 40 to <60 | Weak | Likelihood capped at Roughly even chance; the product states the evidence gap |

| Score | Band | Action the band requires |
|---|---|---|
| T 0 to <40 | Insufficient | No forward judgment from the item; the product reports the gap and tasks collection |
| D 95 to 100 | Deception Almost certain | The decision that rests on the item holds; counter-deception collection is tasked; the product carries a deception warning line; a second reader signs; the likelihood cap drops one band |
| D 85 to <95 | Deception Highly likely | Same as above |
| D 55 to <85 | Deception Likely | Deception warning printed; second reader mandatory; the likelihood cap drops one band |
| D 45 to <55 | Deception Roughly even | Note printed in the product and the ledger |
| D 25 to <45 | Deception Unlikely, the floor band | Ledger note only |
| F 85 to 100 | Endgame named, Highly likely or above | Plan against the Most-Likely Achieved Endgame |
| F 55 to <85 | Endgame named, Likely | Name it; print the two alternatives with their probabilities |
| F 45 to <55 | Endgame open | Present three hypotheses; name none |
| F 0 to <45 | Endgame not assessed | The product says so |

The caps compound. A T-Score of 65 permits Likely. Add a D-Score of 70 on the same item and the cap drops one band, to Roughly even chance. The product then prints Roughly even chance and a deception warning, and a second reader signs before release.

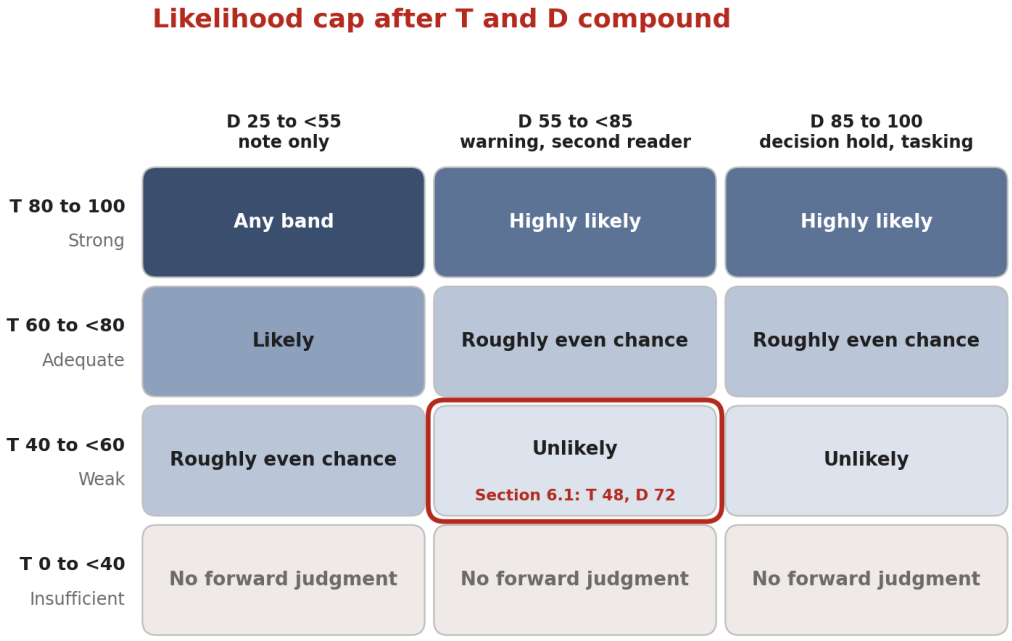



*Figure 2. Likelihood caps after T and D compound. Table 5 sets the rules.*

**Table 5A. Executive words for the three scores.**

| Score | Executive word | Table 5 band it stands for |
|---|---|---|
| T | Strong | T 80 to 100, Strong |
| T | Solid | T 60 to <80, Adequate |
| T | Moderate | T 40 to <60, Weak |
| T | Not published | T 0 to <40, Insufficient: no forward judgment |
| D | Low | D 25 to <45, deception Unlikely |
| D | Guarded | D 45 to <55, deception Roughly even |
| D | Elevated | D 55 to 100, deception Likely or above; a D of 85 or more also holds the decision |
| F | Highly likely | F 85 to 100, endgame named |
| F | Likely | F 55 to <85, endgame named with two alternatives |
| F | Even chance | F 45 to <55, endgame open |
| F | Not assessed | F 0 to <45 |

The CWTR Tier 3 Executive Desk Edition prints each score as a word. Table 5A fixes what each word means. A word carries the cap and the routing of its band, and the ledger keeps the number.

## 2.6 Scoring effort [Tearline 2]

Nobody scores 30 factors on every item, and the standard does not ask for it. Three levels of effort apply. Level 1 is the anchor only. Routine items receive the two Admiralty grades from Tables 1 and 2 and enter the ledger. Time: minutes. Level 2 adds three clusters. Items that bear on a judgment receive C1, C2, and C8 in addition to the anchor. Time: under thirty minutes. Level 3 is the full thirty. Items that drive a decision, and every item in the TAI catalog, receive all nine clusters and the full tri-score. Time: about an hour per item.

Any score within five points of a band boundary in Table 5 goes to a second reader before publication.

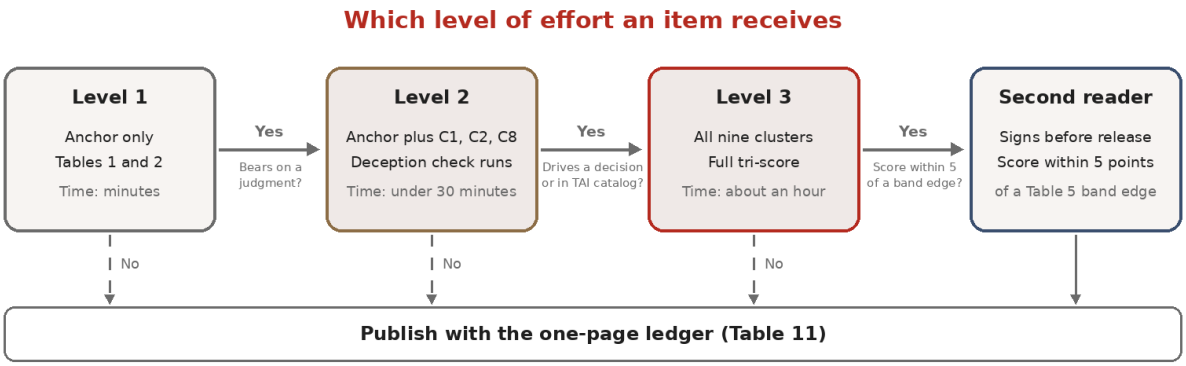



*Figure 3. Routing an item to Level 1, 2, or 3.*

## 2.7 Dependencies between the scores [Tearline 4]

The three scores print separately. They are not statistically independent, and the standard says so rather than pretend otherwise. F3 source independence and F5 corroboration feed the T-Score and also inform the C8 judgments on F22 cross-INT discrepancy and F24 cover-story plausibility. A weak source lowers T and raises D from the same observation. Because the F-Score rests on evidence T has already graded, a low T-Score pulls F down through the shared inputs. Each observation that fed more than one score is marked in the ledger. A reviewer who sees a low T and a high D on the same item reads them as one weakness seen twice, not two.

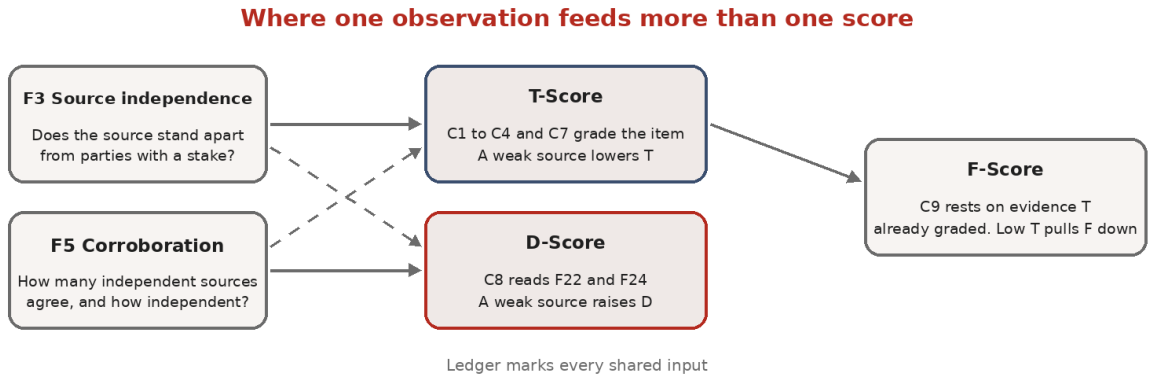



*Figure 4. One observation, three scores.*

## 2.8 Grading technical artifacts [Tearline 4]

Technical artifacts take the full cluster treatment when the item drives a decision. Cluster C4 grades the artifact. C5 asks whether a machine built or altered it, and Section 7 governs AI output. Four practices apply to C4 and its neighbors.

**Hash at collection.** F11 requires a hash and a timestamp at acquisition, plus an unbroken record from that point to the analyst. An artifact with no acquisition record scores below the top band on F11.

**Reproduce before relying.** F12 asks a second examiner to extract the finding from the artifact, such as a configuration, a beacon interval, or cipher material. A finding that one examiner alone reproduces carries that single-examiner condition in the product text.

**Age the indicator.** Technical indicators spoil in hours to days. F10 grades currency against that clock, and the ledger carries the collection time. Rule 9 triggers a re-grade when infrastructure rotates or a sample recompiles.

**Keep the grade on the artifact.** A technical grade supports a statement about the file, the address, or the cipher material. Statements about intent, tradecraft, or endgame draw on tactical evidence and earn their own grade. Where technical evidence alone supports a statement about intent, the product names that single-discipline condition.

## 2.9 Computing the D-Score and the F-Score [Tearlines 3 and 4]

The two scores follow from graded factors, with no weights beyond the equal weights of the Public Edition.

**The D-Score.** Factors 21 to 25 carry risk-direction scores on the Table 4 anchors, and the grader examines at least three. D equals the larger of 25 and their mean, rounded half up. A factor left unexamined drops out, while one examined and found clean scores 0 to 20 and stays in. Fewer than three examined factors print D at its floor with the note "check incomplete."

**The F-Score.** Factors 26 to 30 score on the quality anchors, again with three examined at minimum. F is their mean with the same rounding. A factor drawn only from items below T 40 drops out, since Table 5 allows no forward judgment from those items. Fewer than three usable factors print "Endgame not assessed."

**The endgame ledger.** F is the probability of the Most-Likely Achieved Endgame. Alongside it, the ledger carries probabilities for the Stated Endgame and the Optimal Endgame. No hypothesis outranks the Most-Likely Achieved Endgame; when one does, the grader relabels them. Probabilities need not sum to 100, because the hypotheses overlap.

**One limit.** Averaging dilutes one confirmed indicator. When a grader holds a confirmed indicator, an admission or a second-discipline finding, the product text records it and the item goes to a second reader. The score itself stays the mean.

**Automation.** The reference engine applies the rules of Sections 2.3 to 2.9, Rules 1 and 7, the T-Score cap, Table 5, and the five-point routing test, and it reproduces every score printed in Section 6 [Tearline 4].

## 3. Rules of grading [Tearline 3]

Nine rules govern every graded item.

**Rule 1.** Self-serving claims cap at fairly reliable. A source speaking in its own cause grades no higher than C on Table 1, so a ministry denying its own operation never reaches A or B.

**Rule 2.** Admissions against interest are exempt from the cap. A source conceding a fact that damages its own position grades on the ordinary scales, and the ledger records the against-interest condition. Evidence law, ICD 206 sourcing practice (Reference 8), and every debriefer treat an admission against interest as stronger than a self-serving claim. The cap exists for the denial, not the confession. On January 11, 2020, after three days of denial, Iran's armed forces general staff stated that its air defense had shot down Ukraine International Airlines flight PS752 by mistake (Reference 17), and the IRGC aerospace commander accepted responsibility on state television. The statement runs against the speaker's own cause and grades C1: C on Table 1, given the three days of denial that preceded it, and 1 on Table 2, confirmed by independent video and by the Ukrainian and Canadian findings. In the ledger the against-interest condition is marked, and the product prints the shootdown as Almost certain.

**Rule 3.** Verified fact and adversary claim never share a sentence. A finished product states what independent evidence confirms in one sentence and states what the adversary asserts in another, each with its own grade.

**Rule 4.** Every judgment carries a band or a ranking. Attribution, motive, effect, and future implication each carry a probability band from Section 4 or a source ranking from the anchor tables. An assessment without a confidence marker is a draft, not a product.

**Rule 5.** The denial-and-deception check runs on every item at Level 2 and above. Factor 21 scores marker density against the catalog in Section 5.

**Rule 6.** Technical and human sourcing grade on their own terms. Cluster C4 scores a technical artifact on chain of custody, reproducibility, and metadata integrity. A verifiable artifact and a relayed rumor do not land on the same grade by default.

**Rule 7.** Single-source items carry a corroboration penalty. Factor 5 scores cross-source agreement. A single uncorroborated source cannot reach the top band on C2, and the product states the single-source condition in the text.

**Rule 8.** Conflicts between axes are recorded, not reconciled. A strong source carrying doubtful content stays recorded as strong-source, doubtful-content, with each grade visible. The method never splits the difference into one blended number.

**Rule 9.** Grades are re-opened, not overwritten. An item is re-graded when new information touches any factor, when a forecast it supports resolves, at each TAI release cycle, and on customer request. The ledger keeps every prior grade with its date. When a re-grade moves an item across a band in Table 5, the product cites the current grade and the prior one in the same line. Example 6.4 shows an item moving from F6 to E5 when attribution lands.

## 4. Probability language [Tearline 3]

House products use a Kent-style spectrum with half-open intervals. Kent's 1964 paper (Reference 9) proposed fixed numeric ranges behind estimative words; the terms below are house terms on that model, not Kent's original vocabulary. Half-open intervals put every probability value in exactly one band with no overlap at a boundary. The spectrum runs from 1 to 99 percent. Values of 0 and 100 are reserved for confirmed fact and confirmed falsehood, which are not estimates. Closed ICD 203 bands do not appear in house products.

**Table 6. House Kent-style spectrum.**

| Term | Interval |
|---|---|
| Almost certain | 95 to 99 percent |
| Highly likely | 85 to <95 percent |
| Likely | 55 to <85 percent |
| Roughly even chance | 45 to <55 percent |
| Unlikely | 15 to <45 percent |
| Highly unlikely | 5 to <15 percent |
| Almost no chance | 1 to <5 percent |

A band attaches to each forward-looking element of a judgment. Attribution takes a band or a source ranking. Motive carries a band. Effect carries a band. Future implication carries a band. The band answers how likely an event or a state of affairs is, within the cap the T-Score and D-Score set in Table 5.

Confidence is a different quantity. A confidence level describes the strength of the evidence base behind the judgment. ICD 203 (Reference 7) directs that a confidence level and a likelihood never combine in one sentence, and the standard holds to the rule. A finished product prints the likelihood first and the confidence in the sentence that follows, and the T-Score is the number behind the confidence statement.

## 5. The deception catalog [Tearline 3]

Factor 21 scores an item against a fixed catalog. Eleven techniques come from Treadstone 71 deception doctrine. Seven patterns come from Soviet active-measures practice, which Russian services and their imitators still run. One line each.

**Table 7. Deception catalog.**

| Entry | Definition |
|---|---|
| Mimicking | The item copies the form of a legitimate source or artifact to borrow its credibility. |
| Inventing | The item presents an event, actor, or document that does not exist. |
| Decoying | The item draws attention toward a false target so a real one passes unwatched. |
| Double Play | The item is built to be discovered as a deception, so the real operation behind it is dismissed as more of the same. |
| Double Bluff | The item tells the truth in a form the target is expected to disbelieve. |
| Red Flagging | The item plants a false attribution marker, such as a foreign-language string or a borrowed tool, to point at the wrong actor. |
| Paltering | The item uses true statements to create a false impression. |
| Negative Spin | The item reframes an adversary success as failure or an adversary failure as success. |
| Feint Demonstration | The item shows visible activity in one place to draw a response away from another. |
| Masking | The item hides the true signature under noise or under a benign cover activity. |
| Dazzling | The item floods the channel with variants so the true signal cannot be isolated. |
| Dezinformatsiya | Fabricated or altered information released through a channel built to look independent of the state. |
| Kombinatsiya | Several operations run in combination so that each covers for the others. |

| Entry | Definition |
|---|---|
| Provokatsiya | An action staged to provoke a target response that serves the operator. |
| Fabrikatsiya | A forged document or record introduced as authentic. |
| Diversiya | A diversionary act that draws resources and attention from the main effort. |
| Proniknovenniye | Penetration of a target organization by a controlled person. |
| Agent of influence | A person in a position of trust who shapes a target's decisions on behalf of the operator without disclosing the relationship. |

The internal rubric carries the severity weight for each entry [internal edition].

## 6. Worked examples [Tearline 3]

Four items follow. The first is a retrospective on the case that produced the current United States sourcing directives, and the other three are current public cases. Each shows the anchor grade, the cluster scores under the Public Edition method, the tri-score, the band and action from Table 5, and the sentence a finished product prints. All four use public sources and none is offered as a new finding.

### 6.1 A retrospective on the source behind the 2002 mobile-weapons estimate

From January 2000 to September 2001, Defense human intelligence (HUMINT) disseminated roughly 100 reports from an Iraqi defector in Germany, code-named Curveball, describing mobile biological-weapons production facilities. The reports reached United States agencies through the German liaison service. United States officers met the source once, in May 2000. Two further sources appeared to corroborate: one asylum seeker who recanted in October 2003, and one source brought forward by the Iraqi National Congress. In 2002 a German intelligence officer told the CIA's European division chief that his service "was not sure whether Curveball was actually telling the truth." The Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction, the Robb-Silberman Commission, reported in March 2005 (Reference 11) that those doubts never reached Secretary Powell before his February 5, 2003 address to the United Nations, and that the errors stemmed from poor tradecraft and poor management. In October 2002 the National Intelligence Estimate carried its biological-weapons judgments at high confidence. The CIA withdrew the reporting in 2004.

**Anchor grade at the time, under the standard:** C4. The source sought asylum and resettlement on the strength of his reporting, which puts Rule 1 in force and holds reliability at C. Credibility sits at 4, Doubtful: the content conflicted with the handling service's own doubts and with later imagery of the named site.

**Cluster scores:** C1 40, with independence low through a liaison relay and access claimed but never tested; C2 35, with one meeting, one recanting corroborator, one interested corroborator,

and Rule 7 in force; C3 90; C4 not applicable, no technical artifact; C7 25, no bias disclosure and no reproducible path to the raw reporting. C5 risk 0, no synthetic media in 2002. C6 risk 35, exile-group amplification of the corroborating source. Cap: 65. T-Score: the mean of 40, 35, 90, and 25 is 48, under the cap, so T is 48, Weak. D-Score 72, deception Likely, driven by F22 cross-INT discrepancy, F24 cover-story plausibility, and Inventing markers on production-facility details the site later failed to support. F-Score not assessed: the item is a source report, and the endgame frame applies to the actor, not the source.

**Table 5 action:** T Weak caps the likelihood at Roughly even chance; D Likely drops the cap one band, to Unlikely, and holds the decision.

**Finished-product sentences under the standard:** "Iraq operates mobile biological-weapons production units, Unlikely, 15 to <45 percent. The evidence base supports only low confidence, T-Score 48, single liaison-relayed source, anchor grade C4. Deception Likely, D-Score 72. Decision hold: independent access to the source and imagery of the site named in the reporting are required before any judgment above Unlikely prints."

The estimate that shipped in October 2002 printed high confidence. Applied to the same file, the standard permits Unlikely and a hold.

### 6.2 A multi-agency attribution assessment

On May 24, 2023, Microsoft Threat Intelligence named the actor Volt Typhoon (Reference 13) and assessed with moderate confidence that the campaign was building capability to disrupt communications infrastructure between the United States and Asia in a future crisis. Eight months later, on February 7, 2024, CISA, the NSA, and the FBI published joint advisory AA24-038A (Reference 12) with seven co-authors: the Department of Energy, the Environmental Protection Agency, the Transportation Security Administration, and the cyber agencies of Australia, Canada, the United Kingdom, and New Zealand. The advisory assesses with high confidence that the actor pre-positioned on information-technology networks to enable lateral movement to operational-technology assets, and reports footholds held in some victim environments for at least five years. Living-off-the-land technique, valid-account use, and named indicators appear in the same document.

**Anchor grade:** B2. A multi-agency government body with a track record sits at B, Usually reliable. Content matches Microsoft's independent May 2023 reporting and prior reporting. The ten authoring agencies share observations and count as one government source, so independence stays unconfirmed and the content sits at 2, Probably true, not 1. Reliability does not reach A because attribution rests on behavioral and technical inference rather than a confession.

**Cluster scores:** C1 85, C2 82, C3 90, C4 80, C7 78. C5 risk 5, C6 risk 8. Cap: 92. T-Score: the mean of the five is 83 and the cap does not bind, so T is 83, Strong. D-Score 30, deception Unlikely, the floor plus a small allowance for a false-flag operation on the technical signature. F-

Score 60, Likely, for the Most-Likely Achieved Endgame: access maintained as an option, with disruption executed only in a crisis over Taiwan. Disruption at a time of Beijing's choosing, the Optimal Endgame, sits at 25, Unlikely. The Stated Endgame is the denial graded in 6.3, at 3, Almost no chance.

Action from Table 5: T Strong permits any band; D Unlikely adds a ledger note only; F Likely names the endgame and prints the alternatives. A second reader signs: T 83 sits three points above the Strong edge at 80, and F 60 sits five points above the Likely edge at 55.

**Finished-product sentences:** "A PRC state-sponsored actor pre-positioned on United States infrastructure networks for disruptive effect, Highly likely, 85 to <95 percent. Confidence in the evidence base is high, T-Score 83, anchor grade B2, D-Score 30. Most-Likely Achieved Endgame: access held as an option, disruption reserved for a Taiwan crisis, Likely, F-Score 60; alternatives, disruption at Beijing's chosen time, Unlikely; no operation, Almost no chance."

### 6.3 An adversary claim

On October 15, 2024, the PRC Ministry of Foreign Affairs spokesperson (Reference 14) stated that Volt Typhoon is an international ransomware group and that United States agencies inserted foreign-language strings into code to misdirect attribution toward China. Chinese state outlets carried National Computer Virus Emergency Response Center reports making the same claim, including a July 8, 2024 report tying the activity to a criminal group named Dark Power.

**Anchor grade:** C5. Rule 1 caps reliability at C because the source speaks in its own cause, denying an operation attributed to its own state. Credibility sits at 5, Improbable, because the content positively contradicts corroborated multi-source reporting. No admission against interest appears, so Rule 2 does not apply.

**Cluster scores:** C1 25, C2 15, C3 70, C4 30, C7 55. C5 risk 15, a real statement rather than a synthetic artifact. C6 risk 70, coordinated state-media amplification. Cap: 30. T-Score: the five average 39; the cap of 30 binds, so T is 30, Insufficient. D-Score 78, deception Likely, driven by Fabrikatsiya in the cited reports, Red Flagging in the foreign-language-string claim, and F22 cross-INT discrepancy against the technical record. Trellix and ThreatMon, whose research the reports cited, rejected the connection the reports drew (Reference 19), which reinforces the read. F-Score 70, Likely. The Most-Likely Achieved Endgame is attribution-muddying in third-country audiences, as against the Stated Endgame of technical correction at 5.

**Table 5 action:** T Insufficient permits no forward judgment from the item; D Likely prints a deception warning and requires a second reader.

**Finished-product sentences:** "Chinese state media claims the actor is a criminal ransomware group and that Washington planted foreign-language strings to misdirect attribution. The claim is recorded as adversary assertion, anchor grade C5, T-Score 30, and supports no judgment. Deception Likely, D-Score 78. The claim does not revise the attribution in 6.2."

### 6.4 An influence-operation artifact

On September 4, 2024, the United States Department of Justice announced the seizure of 32 internet domains used in the Russian influence campaign known as Doppelganger (Reference 16). The unsealed affidavit describes the Social Design Agency, Structura National Technology, and ANO Dialog operating under direction of the Russian Presidential Administration, cybersquatting domains that mimic legitimate news outlets, running fabricated influencers and fake social-media profiles, and using AI-generated content to drive traffic to the spoofed sites. Graded here is an article on one such domain, styled as a report from a major Western newspaper.

Two ledger rows apply. Row one grades the article as evidence for its own claim, and row two grades the affidavit as evidence for attribution of the article. Keeping the rows apart is Rule 3 at work.

**Row one, the article.** Anchor grade on first encounter: F6. The domain had no track record and the article offered no basis for comparison, the Failure-two case the two-axis code produces. Anchor grade after attribution to the SDA network, under Rule 9: E5. A network with a documented record of fabrication sits at E, Unreliable. Content that contradicts the genuine outlet's reporting sits at 5, Improbable. Cluster scores after attribution: C1 10, C2 15, C3 75, C4 20, C7 60. C5 risk 60, since the affidavit documents AI-generated content across the campaign without proving it for the single article. C6 risk 90, coordinated inauthentic amplification documented. Cap: 10. T-Score: the five average 36; the cap of 10 binds, so T is 10, Insufficient. D-Score 90, deception Highly likely, driven by Mimicking, Dezinformatsiya, and Fabrikatsiya markers at high density.

**Row two, the affidavit.** Anchor grade: B1. A federal court filing under penalty of perjury sits at B, and its content is confirmed by the domain seizures and the Treasury designations, 1. Cluster scores: C1 86, C2 91, C3 85, C4 90, C7 88. C5 risk 0, C6 risk 5. Cap: 95. T-Score: the five average 88; the cap of 95 leaves T at 88, Strong. D-Score 25, the floor.

F-Score 70, Likely, on the campaign's Most-Likely Achieved Endgame: measurable erosion of support for Ukraine in targeted audiences, short of the electoral effect the operators sought. The Stated Endgame, drawn from the affidavit, influence over election outcomes, sits at 40, Unlikely.

Action from Table 5: row one permits no judgment from the article and carries a deception warning; row two permits any band for the attribution.

**Finished-product sentences:** "A spoofed news article on a cybersquatted domain carried a fabricated report, anchor grade E5 after attribution, T-Score 10, and supports no judgment on its own claim. Deception Highly likely, D-Score 90. Its attribution to the Russian Doppelganger network is Almost certain, 95 to 99 percent, given the unsealed affidavit and the domain seizure. The evidence base for the attribution supports high confidence, T-Score 88."

Under the two-axis code, the four items carry four letter-number pairs and nothing else. With the standard applied, three of the four carry a deception score, two of the four forbid a forward judgment, and one of the four, the retrospective, changes the sentence a government printed.

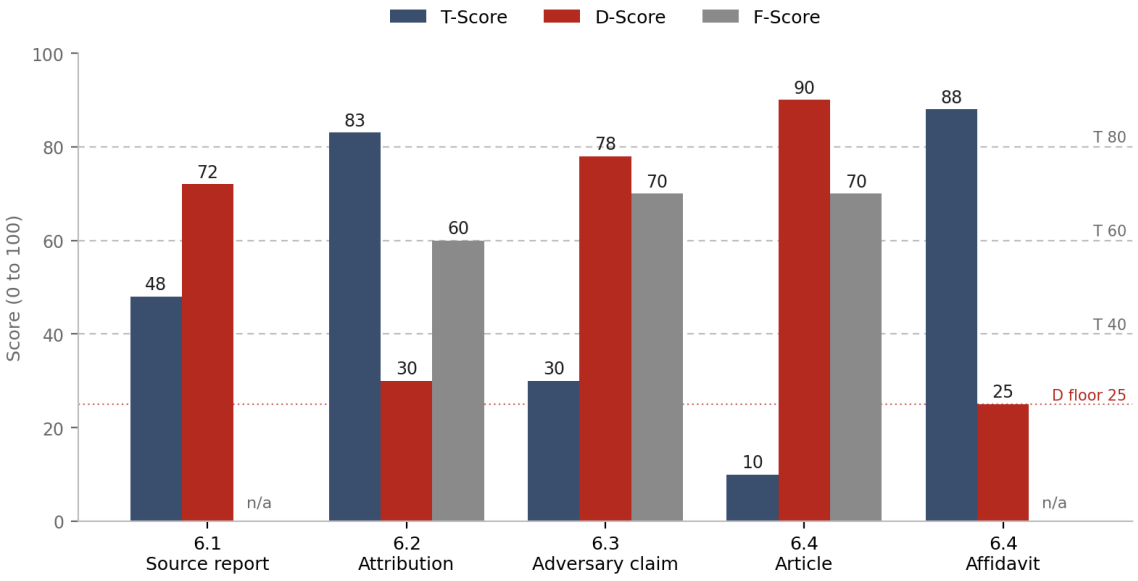



*Figure 5. Tri-scores for the worked examples. The F-Score shown for 6.4 belongs to the campaign, not the article. Sources: Section 6 grading.*

## 7. The standard inside AI-augmented work [Tearline 2]

Treadstone 71 runs AI inside its tradecraft, and the standard grades AI the same way it grades any other source. Four provisions apply.

A model's output is a source and enters the ledger with a reliability grade on Table 1. Where a model or pipeline has no track record on the question at hand, it starts at F and earns a letter over time against resolved outcomes. No model's assertion carries an A on the vendor's name alone.

Factor 19 requires disclosure of AI assistance. The ledger records the model, the version, the prompt or query, and the parameters used to produce or pre-score an item. Factor 20 requires reproducibility, so a second analyst with the same inputs reaches the same grade.

Cluster C5 grades synthetic inputs, whether they arrive from an adversary or from the analyst's own tooling. An AI-generated summary of a source is graded as a derived artifact, one step removed from the primary, and never substitutes for the primary in the ledger. A model never scores F14 on its own output; a human or a second tool does.

Any factor takes a model pre-score. A human signs the ledger. The signature attests that a person read the primary sources behind every Level 3 grade.

## 8. Where the standard runs [Tearline 2]

**Inside TAI.** The standard is the evidence layer of the Adversary Scoring Architecture beneath the Treadstone 71 Adversary Index. Every sub-index score rests on artifacts the standard has graded. TAI's inaugural release shipped July 15, 2026. Its next release lands October 5, 2026 and carries the TAI Q3 refresh and the inaugural Cognitive Warfare Tradecraft Reference (CWTR). Crowd-forecast values are authoritative in TAI. The standard governs the evidence grading beneath them, and each TAI profile carries a falsifier, the one observable that moves its score, which is the resolvable forecast the validation program in Section 9 scores against.

**STEMPLES Plus.** The framework scores an adversary across Social, Technical, Economic, Military, Political, Legal, Educational, and Security dimensions, plus the Plus factors of demographics, religion, and psychological terrain. Under each input, the standard grades the evidence so a reader traces a domain score to a graded source.

**Advanced Analytic Dominance techniques.** The Advanced Structured Analytic Techniques catalog holds 66 techniques in nine families (Reference 18). Section 8.1 maps each family to the grades it consumes. Table A places each family in the tearline whose reader uses it first. Two classic methods, analysis of competing hypotheses and the assumptions check, live in Reference 10 and sit outside the 66. ASA grades the evidence those techniques take as input. No technique in the catalog replaces a grade, and no grade replaces a technique.

**Analytic writing and training.** The firm's analytic-writing products carry the fact-versus-claim rule and the confidence-band rule into finished prose. Training at cyberinteltrainingcenter.com teaches the standard as part of intelligence tradecraft.

### 8.1 Crosswalk to the Advanced Structured Analytic Techniques

The catalog compiles the Advanced Structured Analytic Techniques Practitioner's Manual. Nine families cover threat calibration, adversary cognition, cyber and AI disruption, cognitive warfare, counterintelligence and economic intelligence, strategic foresight, cognitive denial and deception, convergent deception on the hybrid battlefield, and offensive tradecraft analytics. Table 8 shows where ASA grades enter each family.

**Table 8. Where ASA grades enter the Advanced Structured Analytic Techniques.**

| Family | Techniques (count) | ASA input or check |
|---|---|---|
| Threat Calibration and Risk Indexing [Tearline 1] | 6 | The T-Score sets the weight an item carries in a ranking or an update. A D-Score of 55 or higher sends the item to a second reader before it moves a ranking. |
| Adversary Cognition and Decision Modeling [Tearline 3] | 9 | Cluster C9 supplies the three endgame hypotheses that Adversarial Cognitive Simulation and Minimal Viable Hypothesis Testing test. Hypothesis Evolution Tracking keeps the audit trail that Rule 9 requires. Decompositional Uncertainty Mapping tiers read against Table 9. |

| Family | Techniques (count) | ASA input or check |
|---|---|---|
| Cyber and AI Disruption [Tearline 4] | 7 | Synthetic Influence Payload Dissection output feeds cluster C5, factors F14 to F16. Recursive Malicious Code Attribution findings grade in cluster C4, on reproducibility under F12. |
| Cognitive Warfare [Tearline 3] | 8 | Narrative Convergence Analysis, Stigmergic Coordination Detection, Emergence-versus-Direction Discrimination, and Influence Infrastructure Mapping supply the coordination indicators behind F17 and F18. Perception Discrepancy Analysis and Cognitive Warfare Impact Assessment outputs grade as derived artifacts, one step from the primary evidence. |
| Counterintelligence, Insider Threat, and Economic Intelligence [Tearline 2] | 4 | Counterintelligence Deception Diagnostics feed F21 marker density and F22 cross-INT discrepancy. |
| Strategic Foresight [Tearline 1] | 4 | Every forward element of a scenario carries a Section 4 band. The F-Score decides whether a product names an endgame. |
| Cognitive Denial and Deception [Tearline 2] | 6 | Planning family. ASA grades the evidence about target cognition that Cognitive Denial Surface Mapping rates. Where Friendly Cognitive OPSEC Mapping finds friendly exposure, the D-Score tests whether an observed item exploits it. |
| Convergent Deception on the Hybrid Battlefield [Tearline 4] | 9 | Provenance-Layer Deception Analysis pairs with F15, provenance standard absence. Synthetic-Reality Baseline Drift Tracking protects the baseline that F5 and F14 score against, because synthetic content in reference corpora inflates confidence. |
| Offensive Tradecraft Analytics [Tearline 2] | 13 | Planning family. ASA grades the evidence behind effect and attribution forecasts. Attribution-Management Analysis tests an attribution posture and Blowback and Escalation Forecasting forecasts blowback, and each prints a Section 4 band on every forecast element. |

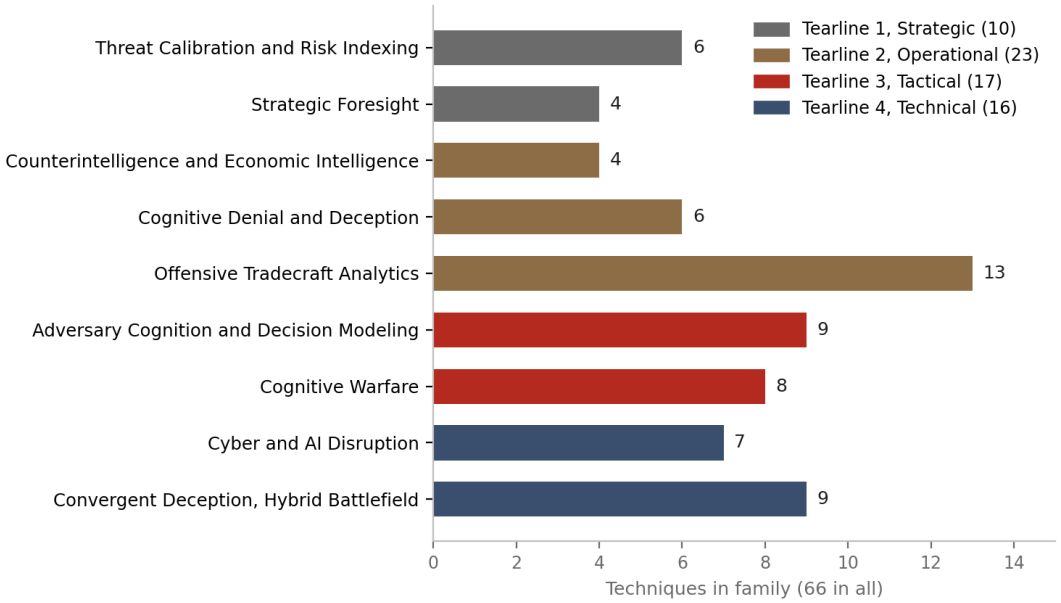



*Figure 6. Advanced Structured Analytic Techniques by family and tearline. Counts sum to 66 (Reference 18).*

The catalog chains six techniques in the Integrated Decision-Support Workspace: Adversarial Cognitive Simulation, Decompositional Uncertainty Mapping, Adaptive Threat Calibration and Risk Indexing, Cyber Adversary Reflex Mapping, Capability Resilience Score, and Hypothesis Evolution Tracking. Its handoff specifications call for evidence tags at the first handoff and source-quality tagging on every Hypothesis Evolution Tracking entry. ASA supplies the grade behind each tag. Rule 9 and Hypothesis Evolution Tracking meet at one point. The ledger tracks grade history and the tracking log holds hypothesis history, so a re-grade that crosses a Table 5 band belongs in each record.

The catalog grades each piece of evidence on a five-tier scale: Verified, Corroborated, Asserted, Adversary claim, and Inferred. Tiers attach to evidence, not to sources, just as ASA grades attach to items. Table 9 reads the tiers against Tables 1 and 2.

**Table 9. Catalog source-quality tiers read against the ASA anchor.**

| Catalog tier | Catalog meaning | ASA treatment |
|---|---|---|
| Verified | Confirmed through multiple independent observation sources | Credibility 1 on Table 2. Reliability grades on Table 1. Independence scores in F3 and F5. |
| Corroborated | Multiple sources, not fully independent, or partial verification | Credibility 2. The STANAG 2511 wording for 2 already covers a source of unconfirmed independence. |
| Asserted | One credible source, no independent corroboration | Credibility 3 at most. Rule 7 applies and the product states the single-source condition. |

| Catalog tier | Catalog meaning | ASA treatment |
|---|---|---|
| Adversary claim | Source controlled by the adversary or advancing its position | Rule 1 caps reliability at C. Credibility grades on the content. Rule 2 exempts an admission against interest. |
| Inferred | Analyst reasoning with no direct evidence | No Table 1 grade. The ledger records a judgment with a Section 4 band. The item never counts as evidence in C1 or C2. |

The two systems differ at one point: the catalog lists defendant statements in counterintelligence cases as adversary claims, while Rule 2 grades a statement against the speaker's own interest under Tables 1 and 2 as usual.

Appendix C of the catalog prints a different probability spectrum. Its ranges share endpoints at 25 and 55 percent, and it gives Probable and Likely the same range. Section 4 governs any product that applies ASA. A product that prints a catalog term prints the range beside it. Table 10 shows where each catalog term lands.

**Table 10. Catalog probability terms read against the house spectrum.**

| Catalog term | Catalog range | House term for that range |
|---|---|---|
| Almost certain | 90 to 99 percent | Highly likely for 90 to <95. Almost certain for 95 to 99. |
| Very likely | 75 to 89 percent | Likely for 75 to <85. Highly likely for 85 to 89. |
| Probable, Likely | 55 to 74 percent | Likely |
| Even chance | 45 to 55 percent | Roughly even chance for 45 to <55. Likely at 55. |
| Unlikely | 25 to 44 percent | Unlikely |
| Very unlikely | 11 to 25 percent | Highly unlikely for 11 to <15. Unlikely for 15 to 25. |
| Almost certainly not | 1 to 10 percent | Almost no chance for 1 to <5. Highly unlikely for 5 to 10. |

## 9. Validation and limits [Tearline 4]

**What has been exercised.** The standard graded the evidence base for every actor in the TAI v2.1 catalog, 42 actors and 7 Iranian-Axis annexes, ahead of the July 15, 2026 inaugural release, and runs again for the October 5, 2026 refresh.

**What has not been measured.** Treadstone 71 has not published inter-rater reliability data, calibration data against resolved forecasts, or error rates for the standard.

Until those numbers exist, the tri-score is a structured expert rating with a documented method, not a validated instrument.

**The burden problem.** Mandel and colleagues found that inconsistent cues reduce analyst agreement. Thirty factors present more cues than two. The standard answers the finding in three

ways. Level 1 and Level 2 keep most items to a handful of cues. Rule 8 records conflicting grades instead of forcing a blend. The five-point second-reader rule catches the boundary cases where disagreement concentrates. Whether the answer holds is an empirical question the validation program below is built to test.

**The validation program.** Three measurements are scheduled. First, inter-rater reliability on a fixed set of 40 graded items, scored blind by three raters per item, reported as agreement per factor and per cluster. Raters are analysts certified in Treadstone 71 tradecraft methods, drawn from the firm's certification cohorts and from client units under mutual non-disclosure agreement. Second, calibration of F-Scores against the falsifiers in the TAI catalog as they resolve across two release cycles. Third, error analysis on D-Scores against cases where deception was later confirmed or ruled out, with the 6.1 retrospective as the first entry. Results publish with the release that follows completion.

## 10. Adoption [Tearline 1]

An intelligence unit adopts the standard in five steps.

1. Set the entry anchor. Grade each source A through F and each report 1 through 6 on Tables 1 and 2, held apart.
2. Choose the level. Route routine items to Level 1, judgment-bearing items to Level 2, and decision-driving items to Level 3.
3. Apply the rules. Cap self-serving claims, exempt admissions against interest, split fact from claim, attach a band to every judgment, and run the deception check.
4. Read the bands. Take the action Table 5 requires before a sentence prints.
5. Publish with the ledger. Attach the one-page ledger and send any case that sits five points or less from a band edge to a second reader.

Table 11 is the ledger. One row per item, one column per field. The sample entry shows the 6.3 item.

**Table 11. One-page evidence ledger.**

| Field | Entry for item 6.3 |
|---|---|
| Item | PRC MFA spokesperson statement, October 15, 2024 |
| Source | Ministry of Foreign Affairs of the PRC, via state media |
| Reliability (A to F) | C, capped under Rule 1 |
| Credibility (1 to 6) | 5 |
| Level | 3 |
| Clusters not applicable | None |
| T-Score and band | 30, Insufficient, cap 30 from C6 risk 70 |

| Field | Entry for item 6.3 |
|---|---|
| D-Score and band | 78, deception Likely |
| F-Score and endgame | 70, Likely; Most-Likely Achieved Endgame: attribution-muddying |
| Deception markers | Fabrikatsiya, Red Flagging |
| Cross-INT discrepancy | Contradicts AA24-038A and vendor technical reporting |
| Corroboration | None independent |
| Against-interest condition | No |
| Shared inputs | F3 and F5 fed T and D |
| Action taken | No forward judgment; deception warning printed |
| AI assistance | None in grading; disclosure line blank |
| Analyst bias disclosure | Prior familiarity with the 6.2 attribution recorded |
| Second reader | Required; initials and date |
| Re-grade history | Initial grade; no prior entries |
| Product sentence | See 6.3 |

A review board reads down the ledger and sees the chain behind each judgment.

## 11. Relationship to ICD 203, ICD 206, and AJP-2.1 [Tearline 4]

**Table 12. Compliance mapping.**

| Standard element | How the ASA Evidence Standard satisfies it |
|---|---|
| ICD 203: describe quality and credibility of underlying sources | Entry anchor plus C1 and C2 grade source and content on published scales |
| ICD 203: express and explain uncertainties | Kent-style band attaches to every judgment, within the Table 5 caps |
| ICD 203: distinguish information from assumptions and judgments | Rule 3 keeps verified fact and adversary claim in separate sentences |
| ICD 203: incorporate analysis of alternatives | C9 holds Stated, Optimal, and Most-Likely Achieved endgames apart with separate probabilities |
| ICD 203: confidence and likelihood stated separately | Section 4; T-Score backs the confidence sentence |
| ICD 206: source descriptors, including factors that affect quality such as denial and deception | C8 D-Score grades deception on its own axis; F3 and F4 grade independence and access |
| ICD 206: source summary statement | Table 11 ledger functions as the source summary |
| AJP-2.1 / STANAG 2511: reliability and credibility rated independently | Tables 1 and 2 print separately; Rule 8 forbids blending |

| Standard element | How the ASA Evidence Standard satisfies it |
|---|---|
| AJP-2.1: evaluation of information before inclusion in intelligence | Level 1 grading precedes ledger entry for every item |

The house spectrum differs from the closed ICD 203 bands, which run almost no chance 1 to 5 percent, very unlikely 5 to 20, unlikely 20 to 45, roughly even chance 45 to 55, likely 55 to 80, very likely 80 to 95, and almost certain 95 to 99. Every house product states which spectrum it uses, and a product prepared for a United States Intelligence Community customer prints the ICD 203 term beside the house term on request.

## 12. Use, citation, and license [Tearline 4]

Public Edition 1.2 is published under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International license (CC BY-NC-ND 4.0). The license does not permit commercial reuse or derivative works. Each needs written permission from the firm.

The firm grants one set of permissions without fee: citation, quotation with attribution, and adoption for internal analytic use by any intelligence unit, agency, or enterprise. Adoption for internal use includes reproduction of Tables 1 through 7, 11, and 12 in internal ledgers and training material with the citation line intact.

Reproduction of the standard in a commercial product, resale, incorporation into a paid service, translation or adaptation, and any instrument or software that implements the scoring rules all require a written license from Treadstone 71 LLC. Ingestion of the standard into AI training corpora, embedding indexes, or retrieval pipelines requires a written license. The firm reserves text and data mining rights under EU Directive 2019/790, Article 4.

The ASA Evidence Standard, STEMPLES Plus, The Adversary Index, TAI, and Advanced Analytic Dominance are trademarks of Treadstone 71 LLC.

## 13. The ledger template and the internal edition [Tearline 1]

The one-page ledger in Table 11 is available as a fillable template. Request it at info@treadstone71.com with the subject line "ASA ledger."

The internal edition, with refined cluster weights, per-factor thresholds, severity weights, and the calibration record, is available to TAI Tier 3 subscribers and to training clients under mutual non-disclosure agreement. Send the request to the same address.

Treadstone 71 LLC. Half Moon Bay, California. +1 424 234 3629. info@treadstone71.com. treadstone71.com. cyberinteltrainingcenter.com.

## References


1. NATO Standardization Office. AJP-2.1, Allied Joint Doctrine for Intelligence Procedures, Edition B Version 3, May 2022, Table 3.1 and paragraph 3.33. https://www.jedi-sec.us/downloads/MISC_PDF/AJP-2.1%20EDB%20V3%20E.PDF
2. STANAG 2511 source reliability and information credibility definitions, working-group transcription. https://eturwg.c4i.gmu.edu/?q=node%2F128
3. Irwin, D., and Mandel, D. R. Critical review of the Admiralty code. https://www.blockint.nl/intel-analysis/critical-review-of-the-admiralty-code/
4. Mandel, D. R., Irwin, D., Dhami, M. K., and Budescu, D. V. "Meta-informational cue inconsistency and judgment of information accuracy: Spotlight on intelligence analysis." Journal of Behavioral Decision Making 36, no. 3 (2023): e2307. https://doi.org/10.1002/bdm.2307
5. Kelly, M. O., Budescu, D. V., Dhami, M. K., and Mandel, D. R. "The effect of source reliability and information credibility on judgments of information quality in intelligence analysis." Judgment and Decision Making 20 (2025). https://www.cambridge.org/core/journals/judgment-and-decision-making/article/effect-of-source-reliability-and-information-credibility-on-judgments-of-information-quality-in-intelligence-analysis/E67548E8010A47345C3439D45D9EC6B3
6. Whaley, B. Stratagem: Deception and Surprise in War. MIT Center for International Studies, 1969; Artech House, 2007. https://dspace.mit.edu/handle/1721.1/81857
7. Office of the Director of National Intelligence. ICD 203, Analytic Standards, January 2, 2015. https://www.intelligence.gov/assets/documents/intelligence-community-directives/ICD_203.pdf
8. Office of the Director of National Intelligence. ICD 206, Sourcing Requirements for Disseminated Analytic Products. https://irp.fas.org/dni/icd/icd-206.pdf
9. Kent, S. "Words of Estimative Probability." Studies in Intelligence 8, no. 4 (1964). https://www.cia.gov/resources/csi/studies-in-intelligence/archives/vol-8-no-4/words-of-estimative-probability
10. Heuer, R. J., and Pherson, R. H. Structured Analytic Techniques for Intelligence Analysis. CQ Press. https://books.google.com/books/about/Structured_Analytic_Techniques_for_Intel.html?id=Js1w15Q7X4gC
11. Commission on the Intelligence Capabilities of the United States Regarding Weapons of Mass Destruction. Report to the President, March 31, 2005, Chapter One, Case Study: Iraq. https://www.govinfo.gov/content/pkg/GPO-WMD/pdf/GPO-WMD-1-6.pdf
12. CISA, NSA, FBI, and partners. Joint Cybersecurity Advisory AA24-038A, PRC State-Sponsored Actors Compromise and Maintain Persistent Access to U.S. Critical

Infrastructure, February 7, 2024. https://www.cisa.gov/news-events/cybersecurity-advisories/aa24-038a

13. Microsoft Threat Intelligence. Volt Typhoon targets US critical infrastructure with living-off-the-land techniques, May 24, 2023. https://www.microsoft.com/en-us/security/blog/2023/05/24/volt-typhoon-targets-us-critical-infrastructure-with-living-off-the-land-techniques/
14. Ministry of Foreign Affairs of the People's Republic of China. Spokesperson's remarks on Volt Typhoon, October 15, 2024. https://www.mfa.gov.cn/eng/wjb/zzjg_663340/jks_665232/jkxw_665234/202410/t20241015_11507722.html
15. Carley, K. M. "Social cybersecurity: an emerging science." Computational and Mathematical Organization Theory 26 (2020): 365 to 381. https://link.springer.com/article/10.1007/s10588-020-09322-9
16. United States Department of Justice, Eastern District of Pennsylvania. Justice Department Disrupts Covert Russian Government-Sponsored Foreign Malign Influence Operation, September 4, 2024. https://www.justice.gov/usao-edpa/pr/justice-department-disrupts-covert-russian-government-sponsored-foreign-malign
17. BBC News. Iran plane crash: Ukrainian jet was "unintentionally" shot down, January 11, 2020. https://www.bbc.com/news/world-middle-east-51073621
18. Bardin, J. S. Advanced Structured Analytic Techniques: Consolidated Catalog, from the Advanced Structured Analytic Techniques Practitioner's Manual. Sixty-six techniques across nine families. East 71st Tradecraft Publishing, manuscript compiled June 2026.
19. Recorded Future News. Beijing accused of misusing Western research to claim Volt Typhoon is a ransomware group. The Record, July 2024. https://therecord.media/china-accused-misusing-western-cybersecurity-research-volt-typhoon