

The Mind Domain Gets an Engine

Adversary use of AI in cognitive warfare, influence operations and hybrid warfare, and the enterprise AI that now sits on the battlefield

DATE 12 October 2026
AUTHOR Treadstone 71
STATUS Final

Copyright 2026 Treadstone 71. Reproduction without written permission is prohibited.

Document control

Date	12 October 2026
Author	Treadstone 71
Status	Final

Contents

Who 10
What 10
So What 10
Why Now 10
Impacts So Far 11
Outlook 11
Recommendations and Opportunities 12
Omissions in the Record 12

1 Scope, method and assumptions 12
1.1 The question and the frame 12
1.2 Priority intelligence requirements 13
1.3 Method stack 13
1.4 Analytic stance 14
1.5 Bias disclosure and analytic integrity 14
1.6 Key Assumptions Check 16

2 Key Judgments 16

3 Iran 21
3.1 Three propaganda bodies, one operating model 21
3.2 The same bodies watch their own people 22
3.3 The loud networks and the attribution wall 23
3.4 The opposition mirror 23
3.5 What Iranian state media say 24
3.6 Adversary reaction read (ACS) 25
3.7 Iranian audiences (HTIM) 25
3.8 What the record shows on Eitaa, Bale and Rubika 26

4 Russia 27
4.1 State media run AI inside the newsroom 27
4.2 The chatbot as audience 27
4.3 Storm-1516 and the Baltic states 29
4.4 CopyCop and the Firebird data center 29
4.5 Self-hosted models that no lab sees 30
4.6 Matryoshka against the German state elections 31
4.7 Contractors, small operators and the monitoring state 32
4.8 What the counts say 33
4.9 What the Kremlin says 34
4.10 Adversary reaction read (ACS) 35
4.11 Russian-theater audiences (HTIM) 36
4.12 Russia as a thief of model capability, on an empty record 36

5 China 36
5.1 Security bureaus run AI as analyst staff 36
5.2 Influence work for hire, and small reach 38
5.3 Taiwan's claims about hired firms 39
5.4 What the PLA writes 41
5.5 Capability theft 42
5.6 The espionage line 43
5.7 Domestic rules beside foreign operations 44
5.8 Adversary reaction read (ACS) 44

Copyright 2026 Treadstone 71 - www.treadstone71.com

5.9 Two rows the record leaves thin 45
6 North Korea, hired hands and proxies 45
6.1 Persona factories 45
6.2 Kimsuky builds decoys with an agent and leaves local tooling behind 46
6.3 Money and enterprise AI as a target 47
6.4 What Pyongyang says 48
6.5 The market for hired hands 48
6.6 STEMPLES Plus across the four states 50
6.7 Adversary reaction read (ACS) 51
6.8 Two North Korea rows the record leaves thin 52
7 Enterprise AI as audience, target and channel 52
7.1 The gateway as target 53
7.2 Accounts, models and research as prizes 55
7.3 Assistants and agents as a channel 55
7.4 Intrusion at the speed of an agent, and fraud by video call 56
7.5 The counterweights 58
7.6 The CARM decision test 58
7.7 Adversary reaction read (ACS) 58
7.8 Three rows the enterprise record leaves thin 59
8 Scoring and prioritization 60
8.1 The ATCRI ranked threat ledger 60
8.2 CWIA and CWC for impact and counter-influence priority 61
8.3 Four audiences under HTIM 62
8.4 One decision under CARM 63
8.5 What the scores leave out 63
9 What governments and allies say, and what they leave out 64
9.1 The American assessment goes quiet 64
9.2 Cyber agencies warn about speed 66
9.3 European services count influence 67
9.4 Alliance and think-tank voices 69
9.5 What the record leaves out 69
9.6 The Gulf states and Singapore 70
10 Patterns, trends and tendencies 71
10.1 What the evidence measures 71
10.2 Who reports on whom 72
10.3 When the evidence appeared 72
10.4 How claims lose evidence as they climb 73
10.5 How stable the ATCRI ranks are 74
10.6 Tendencies across the four states 75
10.7 The outside view 76
10.8 What the patterns change 78
11 Analytic tests 78
11.1 Analysis of Competing Hypotheses 78
11.2 Hypothesis Evolution Tracking 82
11.3 Dynamic Bayesian Narrative Analysis 83
11.4 Devil's advocacy 84
11.5 The effect ceiling 84
11.6 The trap map 85
11.7 Reporter dependence 86
11.8 Cliff distance 87

Copyright 2026 Treadstone 71 - www.treadstone71.com

11.9 Hindcast audit ... 87
12 Outlook in measured baselines, resolvable paths, indicators and dated forecasts ... 88
12.1 How the outlook is built ... 89
12.2 Measured baselines ... 90
12.3 Five paths and how each resolves ... 94
12.4 Drivers read through STEMPLES Plus as a forward scan ... 96
12.5 Cones of plausibility on measured indices ... 97
12.6 Indicators of change ... 98
12.7 Which indicators matter most ... 102
12.8 What the indicators do to the outlook ... 103
12.9 Actors and the tilt of the drivers ... 105
12.10 Scored futures by adversary, vector and date ... 106
12.11 Scored futures by target region ... 110
12.12 Ten dated forecasts ... 111
12.13 Momentum-Reversal Analysis ... 113
12.14 Prerequisites for paths B and C ... 113
12.15 Wildcards, premortem and robust actions ... 114
12.16 What moves the headline ... 115
12.17 What breaks the picture, confidence and review dates ... 117
12.18 The election window: 9 October to 9 November ... 119
13 So what for readers ... 126
13.1 For the enterprise that runs AI ... 126
13.2 For the buyer of counter-influence work ... 126
13.3 For analysts and editors ... 126
13.4 Decision calendar ... 127
13.5 A regional read for operators outside the United States ... 127
13.6 Who cares, what to decide and what it costs ... 130
14 Questions answered and unanswered ... 131
14.1 What the report now answers ... 131
14.2 What the report answers in part ... 132
14.3 What the report does not answer ... 133
14.4 Which questions matter most ... 135
15 Evidence limits ... 135
16 Evidence ledger ... 136
16.1 Index coding sheet ... 141
16.2 Rows near a band edge ... 146
16.3 The adversary record ... 147
16.4 Sources read after the ledger froze ... 160
17 Keywords and wrap-up ... 164
Keywords ... 164
Wrap-up ... 164
Glossary and abbreviations ... 166

Copyright 2026 Treadstone 71 - www.treadstone71.com

List of figures

Figure 1. ATCRI ranked threat ledger. Bar length shows raw risk. Pale bars mark confidence below 70. 61
Figure 2. Ledger rows by the highest rung of evidence each reaches. 71
Figure 3. Ledger rows by originator and theater, with mean T-Score. 72
Figure 4. Dated ledger rows by month of publication. 73
Figure 5. Band midpoint of the report's judgments by attribution rung. 74
Figure 6. ATCRI rank under 20,000 random one-point changes to every rating. 75
Figure 7. Stress test of the hypothesis posterior under twelve test conditions. 82
Figure 8. Posterior probability of each hypothesis across the evidence stages. 83
Figure 9. Logic of the outlook: what is counted, what is modeled, what is judged. 89
Figure 10. Measured baselines: two shares with intervals, and ATVA breadth by actor. 92
Figure 11. The four named paths on two uncertainties. Path E sits outside the grid. 95
Figure 12. Cones for the local-model share and the provider-telemetry share, trailing 12-month windows. 97
Figure 13. Breadth of documented vectors by actor: counted today, projected with the forward grid. 98
Figure 14. Chance each indicator fires under each path. 100
Figure 15. Information gain of each indicator about the five paths, read alone. 102
Figure 16. Joint information of the best watch list against the sum of single gains. 103
Figure 17. Path shares after the watch list resolves, across 30,000 simulated worlds. 104
Figure 18. Four updating storylines, each tempered at 0.5, by review date. 105
Figure 19. Adversary Threat Vector Analysis: where the record documents each vector. Sums score Documented as 2 and Partial as 1. 107
Figure 20. Chance of new public documentation by 31 December 2027, by actor and vector. 109
Figure 21. Path A and Path B shares as each input moves across its tested range. 116
Figure 22. Resolution calendar: indicators and forecasts by check date. 118
Figure 23. Forecast window: phases, dated checks and cumulative odds. 121
Figure 24. Calls ranked by probability and colored by confidence. 123
Figure 25. Ledger rows by T-Score and D-Score. 140

Copyright 2026 Treadstone 71 - www.treadstone71.com

List of tables

Table 1. 1.2 Priority intelligence requirements....13
Table 2. 1.3 Method stack....13
Table 3. 1.6 Key Assumptions Check....16
Table 4. 2 Key Judgments....17
Table 5. 2 Key Judgments....21
Table 6. Four chatbot probes of Russian content, with the denominator and the definition of a hit....28
Table 7. 6.6 STEMPLES Plus across the four states....50
Table 8. 7 Enterprise AI as audience, target and channel....53
Table 9. 8.1 The ATCRI ranked threat ledger....60
Table 10. 8.2 CWIA and CWC for impact and counter-influence priority....62
Table 11. 8.2 CWIA and CWC for impact and counter-influence priority....62
Table 12. 8.3 Four audiences under HTIM....63
Table 13. 8.4 One decision under CARM....63
Table 14. 9.3 European services count influence....68
Table 15. 11.1 Analysis of Competing Hypotheses....79
Table 16. 11.1 Analysis of Competing Hypotheses....80
Table 17. 11.2 Hypothesis Evolution Tracking....82
Table 18. 11.3 Dynamic Bayesian Narrative Analysis....83
Table 19. 11.5 The effect ceiling....84
Table 20. 11.6 The trap map....86
Table 21. Seven single-origin cases, relay items found and independent sources at case level....86
Table 22. Weakest supporting row per Moderate judgment, and the gaps of 24 that push it under the floor....87
Table 23. Brier score by half-year window from 1 October 2024....88
Table 24. Outlook assumptions and what follows if each fails....89
Table 25. Measured baselines with Wilson 95 percent intervals....91
Table 26. Provider and local shares under nine row sets....94
Table 27. Five paths, their resolution rules, shares and bands....95
Table 28. STEMPLES Plus forward scan, one driver per factor....96
Table 29. Visibility of each route by path, the chance the route sees an event that occurs....98
Table 30. Indicators of change, check dates, the chance each fires, and what each does to the path shares (points)....99
Table 31. Indicator chances against the hindcast, four half-year windows from October 2024....101
Table 32. Path shares before and after the full watch list resolves (percent)....104
Table 33. Four patterns of signals and the path shares that follow (percent, A B C D E)....104
Table 34. Driver ratings by state and path, scores 0 to 4....105
Table 35. Tilt by state: the state's modeled share divided by the aggregate share (1.00 = no tilt)....106
Table 36. ATVA grid, status today....107
Table 37. The eight least-evidenced cells....108
Table 38. Adjustments to the forward grid, with the basis for each....109

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 39. Forward documentation chance for selected cells, by date (percent).....110
Table 40. Chance of a public report by 31 December 2027, by target region and actor (percent, with Kent band).....111
Table 41. Ten dated forecasts with point probabilities.....111
Table 42. Momentum-Reversal reads, conditional on the disruption.....113
Table 43. Prerequisites for Path B and the chance of each.....113
Table 44. Six wildcards, chances by 31 December 2027, and where each lands among the paths.....114
Table 45. Premortem causes, first tells and counters.....115
Table 46. Stable actions: ratings, expected payoff and rank stability.....115
Table 47. Path shares under changed inputs (percent, A B C D E).....116
Table 48. Decision triggers, first moves and planning dates (Treadstone 71 planning figures).....118
Table 49. Review dates, what resolves and what to rerun.....118
Table 50. Base rates from four cycles.....120
Table 51. Phases, predictions and breakpoints.....120
Table 52. Scenarios to 9 November (S). Probabilities sum to 100. Confidence is Low for all four.....122
Table 53. Dated calls to 9 November (F).....122
Table 54. Estimates to 9 November (E).....124
Table 55. Levers on the result (forecast).....124
Table 56. Likely frames by outcome (P).....125
Table 57. Signposts that move the calls.....125
Table 58. Decisions by date.....125
Table 59. 13.1 For the enterprise that runs AI.....126
Table 60. Decision calendar for readers, in date order.....127
Table 61. 13.5 A regional read for operators outside the United States.....128
Table 62. Effort model, main units and assumed times (minimum, most likely, maximum).....130
Table 63. 13.6 Who cares, what to decide and what it costs.....131
Table 64. 14.1 What the report now answers.....131
Table 65. 14.2 What the report answers in part.....132
Table 66. 14.3 What the report does not answer.....133
Table 67. 15 Evidence limits.....135
Table 68. 16 Evidence ledger.....136
Table 69. 16.1 Index coding sheet.....141
Table 70. 16.1 Index coding sheet.....145
Table 71. 16.1 Index coding sheet.....145
Table 72. Scored rows within five points of an edge.....146
Table 73. Statements by adversaries and their channels, 1 Feb 2023 to 9 Oct 2026.....147
Table 74. Sources read after the ledger froze, unscored.....161
Table 75. Glossary of terms, methods and abbreviations.....166

Assessment of 6 October 2026 | By Treadstone 71

Three states treat AI as standing staff inside their influence and security work. Iranian propaganda institutions, Russian state media and Chinese security bureaus use commercial and local models to draft, translate, run personas, profile targets and write reports. North Korean units use models for fake identities and lure

documents. Lower cost and higher output are our inference. No source measures either. Belief change goes unmeasured too. Reach stays small in most documented cases, and nothing in the record ties AI to the largest audience counts. The sharper change sits at the edges: chatbots as audiences, AI projects as targets, stolen model capability, and AI supply chains as the way in. Labs and vendors that sell the models supply almost all public evidence. Read it as a map of what they see. The map covers only part of what adversaries do. Counting the ledger sizes the blind spot loosely. Local models appear in 2 of the 16 rows from non-provider sources, with an interval of 3 to 36 percent. One of the two is Google's short, hedged finding on a Chinese intrusion actor that hosts its own model in victim clouds. Adding 21 candidate rows and two slices from February to August brings in 18 non-provider rows, since 5 of the 23 added rows carry the provider flag. That gives 4 of 34, or 12 percent, with an interval of 5 to 27. Fourteen more reports from the four older months add ten non-provider rows, none showing an actor running its own model, and give 4 of 44, or 9 percent (4 to 21). No Western government source in the record reports adversary use of local models at all.

Who

Four states and their hired hands. Iran: the Islamic Culture and Communications Organization (ICCO), the Islamic Propaganda Office of Khorasan Razavi, and the Bina Cultural Observatory. Russia: state media pipelines at Sputnik and RIA Novosti, plus two networks: CopyCop, which Recorded Future also tracks as Storm-1516, and Matryoshka. Researchers tie the CopyCop network to the GRU. China: municipal public security and state security bureaus, a law-enforcement-linked cyber unit, and private firms that sell influence to provincial clients. North Korea: Kimsuky and the IT-worker networks. Each actor follows its own incentives. The evidence shows no joint plan. Shared use of Western commercial AI creates a common pattern. No source shows a common command.

On the receiving end sit firms that build, buy or host AI, and the audiences those firms depend on: customers, regulators, investors and staff.

What

AI moved from experiment to staff work. Anthropic documents Iranian bodies using Claude to write soft-war doctrine, persona systems and amplification schedules. The same report traces Russian state-media output to Claude-written articles and broadcast tickers. Chinese police units, it finds, run sentiment-monitoring work on a model.

Four other lines run beside that staff work. Chatbots became an audience. Researchers report that a Russian network floods the web with content that AI systems then ingest. AI projects became a target. Recorded Future reports that the CopyCop network ran impersonation videos against an Armenian AI facility ahead of its opening. Model capability became a prize. Anthropic reports that Chinese labs drain frontier models through fraudulent accounts. AI supply chains became an entry point. The FBI says a criminal group poisoned the LiteLLM gateway that holds API keys for many providers.

So What

Volume is cheap. Belief is not. The public record counts accounts, posts, videos and pipelines. Effect almost never gets counted. A defender who budgets against volume chases a cost curve. Measuring effect shows whether any of it lands.

The enterprise exposure is sharper than the influence story. Any organization that runs AI holds something adversaries treat as an audience, a target and a channel at the same moment. One owner must hold all three.

Why Now

Four clocks run at once. The war with Iran began on 28 February 2026. Iranian outlets and IRGC voices describe the contest as cognitive war, as Mehr, ISNA-sourced and Asriran items show. Midterms fall on 3 November 2026. The foreign malign influence center sits in a refocus, and the 18 March threat assessment carries no AI-influence language. Anthropic's 10 September report names units and labs. The 8 September advisory from NSA, CISA and FBI warns of industrial-scale distillation by China-based AI companies. Six firms appear in its body, with nothing larger than billions of tokens across millions of exchanges. Its authors judge the activity likely proceeded with Chinese government awareness and state no basis. Russia signed Federal Law 243-FZ on 26 July 2026. That law is a framework law with no penalties. General effect came on 1 September. Its sovereign and national model classes start on 1 March 2027. Part of that clock is the publication calendar. Of 61 dated ledger rows, 25 fall in September and October 2026. From February to August the count runs about five rows a month. September's 13

rows come from five reports. On a report basis September matches that pace. Providers cut their cases finer in September, and that explains the cluster better than a surge in operations.

Impacts So Far

Documented reach is mostly small. OpenAI reports "no evidence of meaningful breakout beyond its own activity" for a PRC-linked operation that targeted American debates on AI, and it rates that activity Category One, confined to one platform. ISD finds that Matryoshka made basic errors and kept its reach limited in Germany. Two cases cut the other way. OpenAI rates the Russian-run Burke Institute operation at the lower end of Category Three on the Breakout Scale, with some indications of breakout to authentic audiences. Its significance, the report adds, lies less in the audience than in the infrastructure. Anthropic reports a Category Four case, tracked as GTG-04001, in which a Russian-speaking operator ran daily content on Radio Lengo Songo, an FM station in the Central African Republic. Each rating is the reporter's own.

The largest audience figures belong to networks where no source shows that AI made the reach. Two pro-Iran X networks pulled in over a billion views during the first month of the war. ISD finds no evidence that either has state backing. A joint report from the EEAS and Ukraine's counter-disinformation center, a belligerent body, counts 1.39 billion views across 244,000 publications on Ukraine's EU accession, as the EEAS page announcing it states. The figure rests on that announcement page alone. No split of AI-made content appears inside the count.

The real losses sit elsewhere. Five ledger rows record loss on the enterprise side and one on the influence side. Anthropic counts over 16 million exchanges and about 24,000 fraudulent accounts for three Chinese labs in its February report, and its September report counts more, led by over 151 million exchanges for Alibaba. Those are provider counts and no government counted them. Judgment KJ5 leaves every volume count out of its basis. Only Anthropic holds them, and the advisory body carries no per-firm count. Reports put the exposure of the malicious LiteLLM releases at about three hours. Wiz puts LiteLLM in 36 percent of the cloud environments in its data. Anthropic reports that Chinese and Iranian security units turned AI on real people to widen their surveillance. The Iranian and Chinese cases each rest on Anthropic alone, so the judgment steps down once under the Section 2 rule, and Section 3.2 rates the Iranian case Roughly even chance. Recorded Future counts about 1.6 million combined views for the Firebird impersonation videos. Mandiant says plainly that AI did not directly cause the breaches of 2025. No ledger row tests belief or behavior change, and no source shows an election outcome that AI changed.

Outlook

The baseline still leads through 2027, and it leads thinly: AI becomes ordinary staff equipment for state cognitive operations, effect stays unproven, and the loudest numbers stay with networks that no source ties to AI output. Path A stands at 39 percent, Unlikely (15 to <45). A plurality, short of a majority. The shift to local models stands at 30 percent, Unlikely (15 to <45). That 30 is the model's chance for the path. No one measured the share. The measured share of local-model rows is 2 of 16 non-provider rows (3 to 36 percent), or 4 of 34 (12 percent) with the extended set. A poisoned-machine era, with measured and confirmed distortion of chatbot answers or enterprise retrieval, takes 18 percent, Unlikely (15 to <45). Shock events with comparison evidence take 5 percent, Highly unlikely (5 to <15). Another 8 percent of futures cannot resolve. Confidence in every share is Low.

The order of A and B does not hold as a finding. Three coding choices decide how far apart they sit. The 4-of-34 count drops the local flag on the OpenAI February slice (row RC-CN-6) by convention. Code that row local, as the rule treats the Google row EN-3, and the count reads 5 of 35, since the row also leaves the provider group and the base grows from 34 to 35. A then reads 38.1 and B 36.0. Restore the first coding of EN-3 and A rises to 50

Copyright 2026 Treadstone 71 - www.treadstone71.com

percent while B falls to 19, so that row sets the size of A's lead and leaves the order alone. Lower the rule's local-share threshold from 30 percent to about 25 percent and B overtakes A. Adding the 21 candidate rows leaves A ahead at 43 against 31. Section 12.16 tests each input.

Silence from the labs does not mean calm. When only the provider-route indicators stay silent, A falls from 39 to 20 percent and B rises from 30 to 50. That result rests on judged visibility figures, and the record holds cases that cut against them. A year with no signal at all leaves 15 percent of futures unresolved. China leans toward the local shift. Russia leans toward the poisoned machine. Iran leans toward the shock. The eight best indicators carry 83 percent of the information in all sixteen active ones. They start with the measured local share, a state surveillance-unit disclosure and a chatbot audit. A hindcast over up to four half-year windows per indicator scores the judged indicator chances at a Brier of 0.14, against 0.25 for a coin flip. Read the score as a floor. The test leaks. Section 12 sets out the counted baselines, the cones, the scored futures by adversary, vector, region and date, and ten dated forecasts. A forecast window from 9 October to 9 November, five days after the results go public, fills Section 12.18, with scenarios, twenty dated calls and the levers adversaries pull on the result.

Recommendations and Opportunities

Enterprises: name one owner for AI exposure across influence, supply chain and agent permissions, and inventory every gateway and credential store, by 31 October; strip tool permissions from agents until each has a reason to hold them, and trace what feeds the retrieval layer, by 30 November; grade every AI-generated analytic output on the ASA risk clusters, add live-video checks to remote hiring, and require a callback on every payment instruction that arrives by video call, by 31 December; and rehearse an impersonation attack against the next AI facility or product launch. Operators in Europe, Australia, Singapore and the Gulf find the regional read in Section 13.5.

Section 1.5 discloses the house commercial interest.

Omissions in the Record

Four omissions mark the government record. The record holds no government measure of belief or behavior change from a live operation. Nor does any Western government source report observed adversary use of local or self-hosted models. The public pages of seven executive bodies, including Justice, State and the White House, hold no item from 1 September onward on AI-enabled influence aimed at the midterms, as of 7 October. A Department of War directive dated 22 September orders Cyber Command and three agencies, NSA, DIA and NGA, to work on foreign election threats, and it names no AI. CISA's own 2026 election plan carries no AI text and gives the agency no foreign-influence task. Almost no government source separates fact from adversary claim. Valdai and PLA Daily show the pattern.

A Public Attribution Series dossier on the midterms is dated 12 October. In it, each verified fact stays out of any sentence that carries an adversary claim.

1 Scope, method and assumptions

1.1 The question and the frame

Treadstone 71 received a direct question: do we hold reports, raw intelligence or opinion on how adversary groups weaponize AI, or exploit the new exposures that enterprise AI creates? The assessment date is 6 October 2026, and items dated 7 October carry their date. Older material sits in the ledger too: of 61 dated rows, 25 date from September or October 2026, 30 from February through August, one from January and 5 from earlier. We

grade every row under the ASA Evidence Standard. An extended set adds 21 candidate rows from February to August 2026 and 14 reports from October 2025 to January 2026. Section 12.2 sets them beside the ledger. The ledger counts leave them out.

The frame is cognitive warfare, influence operations and hybrid warfare. Threat intelligence asks which tool broke which system. Our question is what an adversary wants a population or a decision maker to believe and do, and how AI changes the price, the speed and the reach of getting there. The Cognitive Warfare founder's brief states the test: cyber operations attack systems, and cognitive operations attack the people who decide. Technical items enter only where they feed a belief, a decision or a capability that shapes one.

1.2 Priority intelligence requirements

The Priority Intelligence Process Plus ties each requirement to a decision. Six requirements drive the product.

Table 1. *1.2 Priority intelligence requirements*

PIR	Requirement	Decision it drives
1	How do Iran, Russia, China and North Korea use AI in cognitive and influence operations today, and at what maturity?	Direction of AI Tradecraft Adoption (ATA) movement in the next TAI cycle
2	What effect do those operations produce on belief and behavior?	Whether a client funds counter-influence work, and how it measures the result
3	Which parts of the enterprise AI stack do adversaries treat as audience, target or channel?	Who owns AI exposure inside a firm
4	How do adversaries respond to takedown and exposure?	Where collection sits in 2027
5	What do governments and allies say, and what do they leave out?	How far a client relies on public attribution
6	Which observation changes our mind?	The review triggers in Section 12

1.3 Method stack

The Adaptive Cyber Intelligence Lifecycle frames the product. Action lives in Section 13, and the tripwires and the review dates hold the evolve step.

Table 2. *1.3 Method stack*

Framework	Where it runs	Output
ASA Evidence Standard, Public Edition 1.2 (DOI 10.5281/zenodo.23137881)	Section 16 and every Level 3 item	Anchor grade, T-Score, D-Score, F-Score and a product sentence per item
Sherman Kent bands, half-open	Every estimative judgment	Band with numeric range
STEMPLES Plus	Sections 3 to 6, with the comparison in 6.6 and the forward scan in 12.4	Twelve-factor environment read per actor, and twelve forward drivers
Cultural Nexus and HTIM	Sections 3 and 4	Terrain Vulnerability Index and HTIM Index for four audiences
CWIA and CWC	Section 8	Impact verdicts and Counter-Influence Priority for eight campaigns
ATCRI	Section 8	Ranked threat ledger of ten threats
ACS and CARM	Sections 3 to 7, with the decision test in 7.6	Adversary reaction reads and a decision verdict
Advanced SATs: DBNA, HET, MRA, ATVA	Sections 11 and 12	Narrative posteriors, hypothesis tracking, pivot forecasts, vector map
Foundational SATs: ACH, Key Assumptions Check, Devil's Advocacy	Sections 1.6 and 11	Hypothesis scores, assumption table, counter-case

Framework	Where it runs	Output
Decoy's Dilemma	Section 11.6	Reflexive-control trap map
ASA Measurement of Effect	Section 11.5	Design-strength ceiling on every effect claim
Reporter-Dependence and Echo Mapping, Threshold and Cliff Sensitivity Analysis, Forecast Calibration and Hindcast Audit	Sections 11.7 to 11.9	Relay counts for single-origin cases, distance of each confidence from its floor, and the strength of the hindcast lead
Indicators of Change, cone of plausibility, strategic foresight analysis, forward ATVA, prerequisite trees, premortem	Sections 12.2 to 12.17	Three counted indices with intervals, five paths with resolution rules, twenty indicators with likelihoods by route and path, cones on counted indices, scored futures by adversary, vector, region and date, action stress test
Pattern analysis, outside view and question audit	Sections 10 and 14	Evidence ladder, reporter and month profiles, attribution ladder, rank stability, reference-class base rate, questions answered and open

Six of the Treadstone 71 Advanced SATs (MRA, HET, ACS, ATVA, HTIM and DBNA), taken from Tradecraft at Machine Speed, run here from their published glosses.

1.4 Analytic stance

Five rules govern the reading.

Emergence framing. Each actor is a node with its own capability and incentive. Parallel behavior across Iran, Russia, China and North Korea reflects shared tools and shared incentives. A joint plan does not explain the pattern. No ledger source shows one.

Presence versus persuasion. The ledger separates evidence that an operation exists from evidence that it moved anyone. Mostly, the 2026 record proves presence.

The attribution wall. Open sources prove the work product, the channels and the institutional footprints. They do not prove who gave the order. Where a judgment crosses that wall, the text says so. The band drops.

Fact and claim never share sentences. Verified facts and adversary assertions each carry their own sentence and their own grade. Every judgment carries a Kent band. Confidence appears separately, using the ASA T-Score band: High for Strong, Moderate for Adequate, Low for Weak. One override applies. Confidence drops one step when a judgment rests on an absence of evidence or on our own inference. The Basis column says so.

Single-reporter step-down. Anthropic rows step down once where a judgment rests on them alone, as Section 1.5 describes. The same step applies where only the reporter holds the telemetry or the environment: RU-6 on the CopyCop self-hosted inference stack, RU-8 on Telegram monitoring bots, and KP-3 on the local-model tooling that Genians found on a Kimsuky host. Nothing shows the North Korean stack in use against a victim, and its one source frames the find as capability building. RU-6 carries partial outside support. A Treasury release of December 2024 ties a server the GRU paid for, built to run generative AI, to another body's name, and a French preprint supports a Llama 3 attribution for the network's text. Neither shows the hardware, so the step stands. Findings on published artifacts, such as the opencode decoy files, stay exempt, because a reader tests them against the same files.

1.5 Bias disclosure and analytic integrity

Anthropic's 10 September report supplies the heaviest evidence in the product: rows IR-1, IR-2, IR-8, RU-1, CN-1, CN-7, CN-10, PX-1 and EN-12. The report page shows no date. Reuters, in a digest dated 11 September, reports the release as Thursday, which is 10 September. One lab with a commercial interest in its own enforcement

record supplies the heaviest evidence. That conflict sits on the table. Three controls apply. Anthropic-only rows grade B2, never A. Row CN-7 grades B2 as well. Beside Anthropic's report there stands a US government advisory, but the advisory body names six firms, gives no volume beyond billions of tokens and states no basis for its awareness judgment. Its references cite Anthropic and Google, so its independence stays unsettled. A judgment that rests on Anthropic alone drops one band step and says so on the page, as in Sections 3.2, 3.4, 5.6 and 6.5. Corroboration, where it exists, gets cited. The ACH in Section 11.1 drops every Anthropic-derived row, the advisory row included. H1 keeps its lead in each removal test. Under one recoding test it loses the ACH lead to H3. Section 11.1 reports that test.

Anthropic, OpenAI, Google and Microsoft each gain standing when they publish enforcement results. The ledger records that incentive as evidence row E12 and weighs it. Treadstone 71 holds the same incentive. The house sells analysis and training, and findings that raise the urgency of AI exposure favor those sales. Weigh every recommendation accordingly.

ASA's five-point rule sends any score within five points of a band edge to a second reader. Twenty-seven of the 48 scored ledger rows sit five points or fewer from a band edge at 40, 60 or 80. A five-point margin carries little precision. KJ4 prints one band lower because its two Strong rows (EN-1, EN-6) sit at the edge. KJ10 prints Low because EN-5 reads Weak on the ledger itself. All T, D and F scores stay unvalidated expert ratings. ASA 1.2 calls its own scores a structured expert rating with a documented method, and it says no validation numbers exist yet.

Copyright 2026 Treadstone 71 - www.treadstone71.com

1.6 Key Assumptions Check

Table 3. 1.6 Key Assumptions Check

#	Assumption	Status	If wrong
A1	Lab and vendor reports sample adversary behavior fairly	Doubtful. Labs see only their own users and publish what they choose	Every prevalence figure shifts. Displacement to local models stays invisible
A2	Vendor attribution to named state bodies holds	Supported for institutions that name themselves in prompts and footers. Weaker for unit-level and chain-of-command claims	Attribution bands drop one step
A3	Reach equals effect	Unsupported. No reviewed source measures belief change. Reach bounds effect from above, so the ACH reads small reach as evidence against a rise in effect and nothing more. Section 11.1 tests the matrix with those rows neutral	The whole effect judgment (KJ2) moves toward H2
A4	"AI-enabled" is a stable category across sources	Doubtful. The EEAS 27 percent is a share of detected incidents, and the CrowdStrike 89 percent is a growth rate in attacks by adversaries it tags AI-enabled. Three annexes to the EEAS report give a functional wording and no coding rule. CrowdStrike's report body defines no AI-enabled adversary and prints no count or baseline	Cross-source comparisons lose meaning
A5	Government silence means absence	Unsupported. The foreign influence center sits in a refocus	Absence of US attribution tells us about US capacity, not adversary activity
A6	The analytic tool is neutral	Unsupported. See 1.5	Controls in 1.5 limit the damage
A7	The three ACH hypotheses compete as rivals	Doubtful. They overlap, and H3 holds beside H1 or H2	The posteriors read as an analytic index and no more, and the five paths in Section 12 do not overlap, since rules in a fixed order assign each future to one path
A8	The record's calendar tracks adversary behavior	Doubtful. 25 of 61 dated rows fall in September and October 2026, the weeks when labs and agencies published	The "Why Now" clocks partly measure the publication calendar

2 Key Judgments

Every estimate prints with a Kent band for likelihood and a separate confidence rating. The ASA T-Score of the weakest supporting row sets confidence. Strong evidence earns High, Adequate earns Moderate, and Weak evidence or a single belligerent source earns Low. Adequate evidence caps likelihood at Likely. Only judgments built on Strong rows print higher. Where a judgment depends on a gap in the record or on our own reasoning, confidence steps down once. The Basis column marks each case.

One rule bands absence. A judgment that calls an act or effect real while the record holds no measure of it takes Roughly even chance when a working path exists and nothing shows use at scale. Unlikely fits where open sources show alignment and no order, and confidence steps down once. KJ3a, KJ5a, KJ9a, KJ11 and KJ12 follow the rule.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 4. 2 Key Judgments

#	Judgment	Band	Confidence	Basis
KJ1	Iranian, Russian and Chinese state bodies run AI as standing staff for content, personas and monitoring. North Korean units run it for fake identities and lure documents	Likely (55 to <85)	Moderate	Sections 3 to 6. IR-1, RU-1, CN-1, KP-1, KP-3. CN-1 is the weakest row, at T 60 with an unrounded 60.3. That print stands on a cliff, since a second analyst who scores CN-1 one point lower moves the confidence from Moderate to Low
KJ2	The documented reach of state operations stays small in most cases	Likely (55 to <85)	Moderate	Section 11.1. RU-7, RU-11, CN-2. Two independent sources report limited reach: ISD and OpenAI. In the base case H1 reads 84.5 and H3 reads 11.3, so the two rivals that agree on small reach hold 95.8 together. Twelve stress tests put H1 alone between 57.2 and 87.7, median 72.8. Two cases cut the other way. OpenAI rates the Burke Institute operation at the lower end of Category Three, with some indications of breakout to authentic audiences. Anthropic reports a Category Four case, GTG-04001, with daily content on Radio Lengo Songo, an FM station. OpenAI's report of 8 October, after the counting window, rates a Russian false-front operation Category Five, its first, and an Iranian placement operation Category Four. Each rating is the reporter's own. The Adequate cap holds the print at Likely
KJ2a	AI cut the cost and raised the output of state operations	Likely (55 to <85)	Low	Section 11.1. Inference from staff-use findings (IR-1, RU-1, CN-1). No row measures cost or output. Override applies
KJ3	Russian state-linked content sits in the Common Crawl training source	Likely (55 to <85)	Moderate	Section 4.2. RU-2
KJ3a	That content reaches chatbot answers through retrieval and changes what users believe, at scale	Roughly even chance (45 to <55)	Low	Section 4.2. RU-3 is a model, and its rate comes from ISD's count of chatbot answers that cite a state-attributed source: nearly 18 percent of 300 test responses, 11 percent for neutral prompts, two links from the Pravda network. Citation is no belief measure and no user-exposure rate. No memorization test, no belief measure
KJ4	Criminal groups compromise AI gateways, provider keys and agent frameworks now, and one compromise reaches many keys	Likely (55 to <85)	Moderate	Section 7.1. EN-1 (Level 3) grades at T 80 and EN-6 at T 81, each Strong and each within five points of the 80 edge, so the edge rule sets the print one band lower. The FBI FLASH lists the Trivy, KICS, LiteLLM and Telnyx SDK compromises. EN-6 (Nx and Salesloft, 2025) is precedent for the path and no AI gateway. Highly likely, High returns only if a second analyst's ratings put each row at 80 or more
KJ4a	Our ratings place those gateways and frameworks first among ten threats on defender risk	Likely (55 to <85)	Low	Section 7.1. An ordering set by our own ATCRI ratings, so it says nothing about what adversaries value. EN-1

Copyright 2026 Treadstone 71 - www.treadstone71.com

#	Judgment	Band	Confidence	Basis
KJ5	Chinese labs distilled US frontier models at industrial scale through fraudulent accounts	Likely (55 to <85)	Moderate	Section 5.5. CN-7, with a US government advisory (AA26-251A). The advisory names six firms: DeepSeek, Moonshot, Alibaba, MiniMax, StepFun and Z.AI. Its footnotes lead back to Anthropic and Google, so independence from Anthropic stays unconfirmed. Volume counts stay out of the basis. Anthropic alone holds the February figures of over 16 million exchanges and about 24,000 accounts and the larger September ones. The advisory gives only "billions of tokens" and "millions of exchanges" in aggregate. CN-7 grades B2 at T 64, Adequate, because the advisory states no basis for its awareness judgment
KJ5a	The Chinese government knew of the distillation	Roughly even chance (45 to <55)	Low	Section 5.5. The advisory judges the activity "likely with Chinese government awareness" and shows no basis, no source and no confidence level anywhere in its text. China's foreign ministry reply of 9 September names no advisory and no company, and it rebuts no technical claim
KJ6	Russian, Chinese and North Korean operators run work on models Western labs cannot see, so lab reports undercount adversary activity. KJ14 bounds the size	Likely (55 to <85)	Low	Sections 4.5, 5.6 and 6.2. RU-6, KP-3, EN-3. Only Recorded Future, Genians and Google hold the telemetry, so the single-reporter step-down applies. Anthropic's own report adds two cases: a China-based actor ran a self-hosted model beside Claude (GTG-17002), and a Mali surveillance platform built with Claude's help ran on local models (GTG-50027). Nobody saw hardware. RU-6 is a self-hosted inference stack, and the KP-3 stack has not been shown in use against a victim. The measured share is 2 of 16 non-provider rows (3 to 36 percent), or 4 of 34 (12 percent) with the extended set. PROMPTSTEAL calls a hosted Qwen model through the Hugging Face API, and PromptLock, which ESET says ran a model locally through Ollama, was an NYU research prototype. Neither shows an actor running its own hardware. One Western government release, the US Treasury's of 31 December 2024, says a GRU affiliate built a server to host generative AI tools, and it names no model. Section 12.2 bounds the size only loosely

#	Judgment	Band	Confidence	Basis
KJ7	In the cases we read, single subscribers and small vendors deliver influence and surveillance programs that once needed a staffed unit, for states with thin capacity	Roughly even chance (45 to <55)	Low	Sections 5.3 and 6.5. PX-1 stands on Anthropic alone, so the band drops one step. The CN-12 release page names no state buyer. The 399-page document set behind it names PLA Unit 31040 in a 2020 procurement letter, and in its own scripts and product table GoLaxy names state bodies as deployment sites and users. Those are the firm's claims. GoLaxy claims a team of about 50 on the GoPro system and about 900 in the firm, so it is no single-subscriber case. One step down takes PX-1 at T 72 from Moderate to Low. Section 5.3 holds the NSB tasking chain (CN-6, C4, T 49, Weak) at the same band, and that Weak row sets Low by itself
KJ8	The thin American record on AI-enabled influence reflects a changed posture and thinner capacity, and no source shows adversary operations falling	Likely (55 to <85)	Low	Section 9.1. GV-1 grades Strong, which sets High. The judgment rests on a gap in the record and on our own causal reasoning, so confidence steps down twice, to Low. No executive-agency statement from 2026 on AI-enabled influence aimed at the midterms exists in the record. The CISA election plan of 24 September runs 13 pages with no text on AI, deepfakes or disinformation and no foreign-influence role for CISA. Its infrastructure scope alone explains that silence, so the plan supports the null and leaves the cause open. No such statement appears on the White House, State, Justice, FBI, DHS or Treasury pages as of 7 October either. The Department of War memorandum of 22 September, on election threats from abroad, names no AI. A letter of 2 October from Senator Warner and Representative Himes requests a classified briefing round by 31 October plus public unclassified updates, and it carries no AI wording
KJ8a	European bodies attribute, count and sanction influence work that American bodies leave out	Likely (55 to <85)	Moderate	Section 9.3. RU-9 grades B3 at T 60, Adequate, which caps the print at Likely and sets Moderate. The row sits on a cliff, at an unrounded 59.7, and one point off RU-9 sends the confidence down to Low. GV-4 stays at Level 1. The EEAS fourth report carries three annexes on sanctions, law enforcement and digital regulation. None sets out how incidents are coded, and the prevalence measure carries no coder agreement. No source measures belief change
KJ9	Adversaries abuse AI coding assistants through planted configuration files in the wild	Likely (55 to <85)	Moderate	Section 7.3. EN-3. Google does not state the scale

#	Judgment	Band	Confidence	Basis
KJ9a	Adversaries exploit agent and MCP permission flaws in the wild at scale	Roughly even chance (45 to <55)	Low	Section 7.3. EN-2, EN-10. Research proves the path and nothing proves use at scale. The Flowise flaw (CVSS 10.0) shows honeypot detections from one IP address, per a VulnCheck researcher's post, with no victim named. Override applies
KJ10	Threat actors run credential theft and multi-victim intrusion with agentic AI at a pace that shrinks defender warning time	Likely (55 to <85)	Low	Section 7.4. Rows EN-3, EN-5 and EN-12 carry it. EN-5 grades B3 at T 53, a Weak row by the ledger's own scale, and the weakest supporting row sets confidence
KJ11	Iranian state direction of the two pro-Iran networks that drew more than one billion views	Unlikely (15 to <45)	Low	Section 3.3. IR-7. Open sources show alignment and no order. Override applies
KJ12	Deepfake-enabled payment fraud against finance teams stays a live exposure	Roughly even chance (45 to <55)	Low	Section 7.4. EN-13 sits at Level 2 with anchor B2, T 65 and an Adequate band, which gives Moderate. One case from January 2024 stands behind it, about 33 months old at the assessment date, and a single case steps the print down once, to Low. Arup's accounts book a 25.4 million pound exceptional item and say no systems were compromised. Hong Kong police put the transfer at HK$200 million
KJ13	Through 2027 the leading path, AI as ordinary staff equipment with no measured effect, no counted majority shift to local models and no confirmed chatbot distortion, holds 39 percent. A shift to local models holds 30. Most futures lie elsewhere	Unlikely (15 to <45)	Low	Section 12.3. A model of counted baselines and judged inputs. Rules assign each future to one path. The 30 is the model's chance for Path B and no measured share. Path A leads in 18 of 22 sensitivity runs in 12.16 and trails Path B in 4, and the order of the two is not separable. Recode one OpenAI row, or lower the rule threshold to 25 percent, and the lead closes or flips
KJ14	Local-model use stands below one third of state-linked AI operations today, and lab reports undercount adversary activity by an amount the counts bound only loosely	Likely (55 to <85)	Low	Section 12.2. DLS reads 2 of 16 rows, interval 3 to 36 percent, and 4 of 34 rows, interval 5 to 27, once the extended set joins. The latent share has a median near 23 percent, 10th to 90th 8 to 46, and a 73 percent chance of sitting under one third. That 73 percent holds at detectability 0.5 and falls to 42 percent at 0.25, so the print leans on the middle setting. RU-6, KP-3, EN-3. The Adequate cap holds the print at Likely, and the single-reporter step-down applies

Three things hold the set together.

The strongest judgment sits on the enterprise side. KJ4 prints at Likely, the band the edge rule leaves it. The FBI, Palo Alto Unit 42 and Wiz documented the LiteLLM compromise from separate vantage points, and the FBI has no product to sell. Two Strong rows behind it sit at the band edge, which holds the print below Highly likely.

The weakest concern effect and attribution. KJ3a, KJ5a, KJ9a and KJ11 either measure nothing or rest on a source with an interest, and they stay on the page because each marks a place where a reader otherwise assumes more.

The set carries a conflict. Anthropic supplies the heaviest evidence for KJ1, KJ5, KJ7 and KJ10. Section 11.1 tests the effect finding without every Anthropic row and without every vendor-telemetry row. H1 stays in front on the posterior in every test. H3 takes the ACH lead when three coding choices change. The two rivals agree that documented reach stays small, which is KJ2. They disagree on cause. That disagreement holds KJ2a at Low. The table below applies the same removal to the judgments themselves.

Table 5. *2 Key Judgments*

Judgment	With Anthropic	Without Anthropic rows	What carries it
KJ1	Likely, Moderate	Russia and China hold at Likely on other sources. The staff-work claim for state newsrooms and security bureaus leaves with RU-1 and CN-1. Iran falls to Roughly even chance, Low	Russia: RU-4, RU-6, RU-8. China: CN-2, CN-3. Iran: IR-5, plus the IR-3 and IR-4 statements. North Korea: KP-1, KP-3
KJ5	Likely, Moderate	Likely, Low. The advisory and Google's report of distillation against its own models remain, and the advisory cites Anthropic. Anthropic's account of the named labs and its counts leave	CN-7 advisory (names six firms, gives no per-company count, cites Anthropic and Google), EN-3
KJ7	Roughly even chance, Low	Unlikely, Low	PX-1 is the only row on hired hands who serve state clients. CN-6 carries Taiwan's own claim of state tasking, graded C4. The CN-12 release page shows no state buyer, and the document set behind it holds a 2020 procurement letter that names PLA Unit 31040 and deployment and user claims, state bodies named, that are GoLaxy's own
KJ10	Likely, Low	Likely, Low	EN-3, EN-5 (B3) and GV-2. The espionage example leaves with EN-12

3 Iran

Iran runs the most open case in the record. Its operators named their own institutions to the model. Iranian outlets say that Iran is the target of a cognitive war. Two of them, from Mehr News, appear in Section 3.5. Anthropic's record shows Iranian bodies running one.

3.1 Three propaganda bodies, one operating model

Anthropic removed three Iranian state-aligned accounts. Each traced to a named institution. The first account came from the Islamic Culture and Communications Organization (ICCO), which sits under the Ministry of Culture and Islamic Guidance. Claude produced a ministerial influence portfolio for it. Project codes ran A-01 through B-04. Claude also drew up a funeral and succession plan for the Supreme Leader. The second came from the Islamic Propaganda Office of Khorasan Razavi in Mashhad. Claude wrote doctrine for a content factory the staff call "Manjanegh" (Catapult), and persona-tailored content for a paid campaign. Distribution ran through Eitaa, Bale and Rubika, plus X, Instagram and Telegram, with paid amplification from IRGC-affiliated channels. Anthropic names two IRGC-affiliated channels, @hamyane_sepah and @moghavematnews_iran, and shows what Claude produced. Its table gives no platform for either handle. Public pages dated 7 October show @hamyane_sepah as a real Eitaa channel with 21.9K followers. The handle @moghavematnews_iran is a six-follower channel-sale page on Eitaa. On Bale the same handle is a war-news channel. Bale lists 155 members, and the preview holds ten posts, all from 14 March, none naming AI. The report does not show how much of the Eitaa content came from the model. Reach goes unreported. A third body is the Islamic Propaganda Organization with its Bina Cultural Observatory. There Claude repackaged IRGC spokesperson communiqués, serialized a war-related messaging campaign and laundered material through a think-tank front. Anthropic's report rates the

operation Category Three, per the Breakout Scale: multiple platforms, with content seen on IRGC-aligned Eitaa channels.

Iran blocks access to Claude. The operators got in anyway, with VPNs and foreign phone numbers. They still named their locations, institutions and roles in conversation. In their own words, the program was a non-military plan to shape opinion at home and abroad, a "soft war" or "cognitive warfare" effort.

Assessment: the three bodies run AI as standing staff for doctrine, personas and amplification timing. Likely (55 to <85), Moderate confidence. The Anthropic row grades B2 with T 75. That caps the band at Likely. Recorded Future's Iran AI playbook, published on 16 July before Anthropic's report, judges that AI sped up Iranian operations and found its greatest use in information warfare. That backs the class of use. Insikt documents no propaganda body that runs AI as routine staff work, so the three named bodies stay on Anthropic's telemetry. Most-Likely Achieved Endgame: the bodies keep AI inside routine staff work and shift part of the load to home-built or open-weight models as Western access closes. Likely, F-Score 62. Alternatives: a sovereign Iranian model pipeline by year-end 2027, Unlikely (15 to <45); abandonment of AI under security pressure, Almost no chance (1 to <5).

Iran's score on the AI Tradecraft Adoption sub-index in TAI moves up with the finding. Anthropic's indicator table already names the formations, channels and codenames a Public Attribution Series dossier needs.

3.2 The same bodies watch their own people

The influence work sits next to a domestic control program. Anthropic reports that it banned 16 accounts run by two linked units, which it links at high confidence to Iranian paramilitary and domestic security agencies. One was a seven-department organization with provincial offices. That unit claimed to hold an identity database of Iranian nationals and to have surveilled and profiled 6,388 Iranians in a single year, and it ran social-network analysis over 155,216 tweets. The two units built tools for one central system that the report calls Arman. A Qom-based provincial unit used Claude as its engineering department. Its flagship product, a Firefox extension called "al-Najm al-thāqib," went into production to harvest user identities from major social networks. Anthropic says the units answered taskings from senior government officials. Its report adds that they vetted candidates for official posts. The agencies stay unnamed, and Reuters, in its digest of 11 September, calls them unidentified.

Anthropic also describes a separate actor that used free Claude accounts across 16 single-operator organizations to build malware, a delivery pipeline and a phishing portal for domestic targets, and another Iran-nexus actor that built an automated open-source profiling harness aimed at Israeli government and non-government individuals and at Jewish diaspora organizations. In a joint alert of 15 September, the FBI, the UK NCSC and the Dutch AIVD describe the older, non-AI side of the same repression work against dissidents, activists and journalists abroad.

Assessment: Iranian security units use AI to scale surveillance of domestic and diaspora audiences. Roughly even chance (45 to <55), Low confidence. Anthropic alone reports the AI work (row IR-2, B2), so confidence steps down once, from Moderate to Low (Section 1.4). The record holds relays of Anthropic's text and no statement from any Iranian unit. BBC Persian, RFE/RL, IranWire and Iran International restate the report, and none adds telemetry. Google's tracker of 8 September shows the Iranian actor APT42 using Gemini for research and lures, which backs the class of state use of commercial models and says nothing about the Iranian units. The FBI alert in IR-9 carries no AI element. A Cultural Nexus read explains incentive. Iran scores high on measure 01, information control and media trust. Measure 06, surveillance acceptance, runs high as well. A tool that cuts the cost of watching people finds a ready buyer.

Copyright 2026 Treadstone 71 - www.treadstone71.com

3.3 The loud networks and the attribution wall

Researchers at ISD report that since the February war began, Iranian embassy and official X accounts have treated "shitposting," including AI-generated video, as statecraft. ISD compared about 150 diplomatic and government accounts over two 50-day windows. Views rose from 55 million to 896 million. Recorded Future rounds that figure up to nearly one billion. The accounts are overt. ISD does not say how much of their content is AI-made. Recorded Future writes that attribution of the viral AI videos to the Iranian state remains uncertain, and it names state outlets that redistribute them as the strongest link. A separate ISD study finds two coordinated pro-Iran X networks drew more than one billion views in the war's opening month with false, misleading and AI-generated war content. BRICS4CLICKS had 23 accounts. They took over 772 million. Verified4War had 18 accounts. They took over 370 million. Iranian and Russian diplomatic accounts shared their posts. ISD states it "has no further evidence that either network is state-backed."

Assessment: Iranian state direction of the two networks. Unlikely (15 to <45), Low confidence. Shared posts show aligned interests. No order shows. The case shows where open sources stop: they prove the content and the sharing pattern, and nothing in the ledger proves command. CWIA scores the two networks Elevated (70.0), the highest impact score in the record. Reach alone drives it.

For Treadstone 71, the TAI rubric scores executed operations and ignores stated intent. Unattributed volume leaves Narrative Volume for Iran where it stood. The Cognitive Warfare Monitor carries the networks as a tracked narrative with a Claim grade of C and no state label.

3.4 The opposition mirror

Anthropic reports that it also disrupted a distributed network aligned with the MEK and the NCRI. In its account, the network ran a shared AI agent that held persistent memory and produced content on a schedule across several actors. Operators cloned a real activist's voice from private messages and held live conversations as that person with contacts inside Iran. Psychographic dossiers on named people in Iran, arrest histories included, followed. Undisclosed synthetic "ordinary Iranians" and historical-figure deepfakes went out, and affiliated media lost their watermarks.

The accused organization answered. In a statement of 13 September, the PMOI says Anthropic never contacted it or the NCRI, and it denies systematic use of AI. The PMOI concedes that one account used AI for about a month and was closed, and it says that account redistributed public positions. That text quotes Anthropic's own caveat that the report found no "visible sign of coordination." As the accused party's account, it is reliable only as to what was said. Between the two sides, the one concrete fact, a single closed account, stays undisputed.

Each side uses the tools.

Assessment: opposition-aligned networks and state bodies now use the same class of tools, including live impersonation. Roughly even chance (45 to <55), Low confidence. Anthropic alone reports the network (row IR-8, B2). The record holds relays and no second report. Only one text adds wording of its own, and its outlet is hostile to the MEK. That text carries no telemetry. The PMOI reply is the accused party's account and adds none either. Confidence sits at Low. Every public claim about "what Iranians think" now carries a source-risk flag from either side. The ASA C6 cluster, cognitive and influence-operation risk, applies to nearly every Persian-language political artifact on social media.

3.5 What Iranian state media say

Two Mehr News pages use the phrase cognitive war. The first is an item timed 01:08 on 7 October about a night rally in Arak during Police Week, held the evening before. Sardar Farzad Tajareh, police commander of Markazi Province, is the speaker. He says that cohesion between the people, the police and the armed forces forms «سدی نفوذناپذیر در برابر جنگ شناختی و هجمه‌های رسانه‌ای دشمن», "an impenetrable barrier against cognitive war and the enemy's media onslaught." The enemy, having failed in the physical, military and security fields, now turns to perceptual war and false narratives, he adds, and he urges media literacy. The page mentions no AI. A commander speaks for his province. Alone, the item signals tone.

The second is a Mehr commentary dated 14 September on the Anthropic report. Its English headline reads "Claim of Iran's use of Claude AI to attack American forces." The body relays Anthropic's finding of three accounts tied to Iran-aligned actors and does not deny it. Most of the text covers the Yemeni missile-software cell and other states. Anthropic's three Iranian influence accounts describe no attack on American forces. A separate Iran-nexus case in the same report (GTG-30005) sits closer to the headline. The report files it under military reconnaissance and says the actor had Claude build targeting handbooks on US naval forces from open sources, among them personnel rosters scraped from photo captions, ship and aircraft transponder identifiers and satellite-imagery query scripts. Research toward targeting is the claim, and no attack appears in it. The commentary argues that the line between use and misuse follows American power. Claude's place on US classified networks, which Anthropic announced in February 2026 under a 2025 Defense Department contract worth up to 200 million dollars, anchors the argument, and the text cites reports on the Maven system as speeding target ranking against Iran. None of those reports is named. The conclusion carries the two quoted phrases: «جنگ الگوریتمی آغاز شده است», "the algorithmic war has begun," and «زیرساخت شناختی جنگ», "cognitive infrastructure of war." They describe frontier models as a general military layer. Sovereign frontier-model capacity, secure compute and defensive agents are the commentary's remedy. Nothing in the text discusses Iranian influence operations or presents cognitive war as something done to Iran.

ISNA's item of 2 April, headed "Amazon's cloud business in Bahrain damaged," relays the Financial Times on damage to Amazon cloud operations after what Bahrain's interior ministry called an Iranian attack. One clause says the report came a day after the IRGC threatened US firms in the Middle East, Microsoft, Google and Apple among them, as legitimate targets. The inference puts the threat on or about 31 March. Three firms are named, and no AI content appears. Digiato, an Iranian technology site, carried the threat under a byline on 1 April at 11:23. The article says the IRGC told staff of US firms in the region to leave their workplaces and stay a kilometer away. From 20:00 Tehran time that day, for every assassination in Iran, a corresponding unit of those firms was to be destroyed. The page prints 18 names as the list the IRGC published: Cisco, HP, Intel, Oracle, Microsoft, Apple, Google, Meta, IBM, Dell, Palantir, Nvidia, JPMorgan, Tesla, General Electric, Spire Solution, G42 and Boeing. Digiato does not reproduce the notice itself. ISNA's item runs on 2 April, places the Financial Times report on Wednesday, 1 April, and says the Bahraini announcement came earlier that day. If so, the fire was reported before Digiato's 20:00 deadline. No primary text settles the date of the threat or the order of threat and strike. Euronews and MIT Sloan Management Review Middle East report an IRGC statement from the same weeks that names US technology firms as targets. Their dates run from 12 March to 1 April and their lists run from 17 to 18 firms. Recorded Future reads the IRGC statements as naming AI firms only as targets of strikes. Proxy outlets relay further claims in that vein. PalToday on 4 April relayed an IRGC claim of a strike on US forces and on an AI company's site in the UAE. Al-Masirah English on 17 July carried a claimed hit on an AI center in Bahrain. On 7 October Al-Ahed relayed a brigadier's threat against "American AI" in the region. Each is an adversary claim. We verified none.

Copyright 2026 Treadstone 71 - www.treadstone71.com

The frame also reaches command level. Etemad Online, crediting ISNA, carries a speech by an adviser to the IRGC commander-in-chief at a 9 Dey rally in Gorgan. The page prints no year. His remark on the 12-day war points to 30 December 2025. The enemy cannot mount a military attack, he says, and has "focused on cognitive war." Its methods are forgetting, distortion, discrediting and censorship, plus polarization and pressure on hope. A reprint on Asriran dates an earlier line to February 2023, on a stamp that reads poorly. Maj. Gen. Hossein Salami, then IRGC commander-in-chief, told Basij content producers that "today's war is in cyberspace" and that the enemy writes every narrative it is able to write. Neither item mentions AI. The two items put the cognitive frame in IRGC command speech before the Anthropic report appeared.

Two more Persian items carry the argument from other offices. A Mehr guest note of 29 July, by the head of international affairs at the judiciary's lawyers' center, asks whether Palantir, Anduril and Shield AI are partners in war crimes. He grants that no international court has attributed any firm's role in a specific attack on Iran. His argument rests on the Rome Statute and the Nuremberg industrial cases. Company officers face liability for aiding crimes, he says, and a responsibility gap will widen. On 8 October the head of the Majlis Research Center, in an Eghtesad Online item, rebuts a story that parliament sat closed during the war. In cognitive war, he says, the enemy need not invent a narrative, because Iranians sometimes hand over its raw material. The item has no AI content.

Each statement above is a claim. Rule 1 caps it at C. The ledger rows IR-3, IR-4 and IR-10 grade C3, C4 and C4, and none supports a judgment about facts. IR-3 and IR-4 record what the regime says about cognitive war and about AI. The third row, IR-10, records a threat to named firms with no AI element. Mehr's commentary denies nothing in Anthropic's finding and argues for sovereign models. No motive inference rests on it. Rule 1 holds the Mehr rows at Weak, with T-Scores of 54 and 48. Section 16.3 lists the further items read in the original, each coded Level 1 and left unscored.

3.6 Adversary reaction read (ACS)

The Adversarial Cognitive Simulator, set on a nation-state intelligence apparatus with the trigger "takedown of a tool and public exposure," returns the following for Iran.

Worldview. The apparatus treats information as a front in a wider war and uses "soft war" and "explanatory jihad" as organizing terms. Fears: loss of Western tool access, exposure of named institutions, and leaks of sensitive plans such as the succession material. Biases: confirmation bias on enemy intent, and satisficing on tool quality. Likely actions: rotate accounts and phone numbers, shift work to domestic or open-weight models, hire contractors to hold the tool relationship, and answer exposure with denial and reframing. Four outlets that covered the Anthropic report record no official Iranian response. CARM pathways: deny, reframe, relocate. Miscalculation risk: over-trust in weaker domestic models, and a second leak through a less guarded channel.

3.7 Iranian audiences (HTIM)

The Human Terrain Influence Mapping run on Iranian domestic audiences returns a TVI of 7.0, a PES of 6.0 and an HTIM Index of 66.0, confidence 72. Section 8 shows the inputs. The score reflects high demographic and cultural resonance for state narratives, high channel saturation on the walled-garden platforms, and a middling behavior-shift velocity.

Copyright 2026 Treadstone 71 - www.treadstone71.com

3.8 What the record shows on Eitaa, Bale and Rubika

The three Iranian platforms carry the state content that Anthropic's report traces. Public pages on Eitaa and Bale, dated 7 October, hold what the record shows. The record holds no public Rubika channel page for either handle. That is the largest blind spot left in the Iran section.

Eitaa's @hamyane_sepah carries the title «حامیان سپاه پاسداران», "Supporters of the IRGC," and a follower count of 21.9K. The channel page of 7 October holds 12 posts, numbered 73383 to 73394 and stamped across about two hours. They relay, among other items, statements by the Houthi leader, Yemeni military claims, a Financial Times line on Hormuz and two videos, plus one medical advertisement. Twelve posts stamped between 7 September and 2 October carry the Persian words for AI in the channel's visible view, which is truncated. Five of the 12 pass on a claim that someone else used AI. They carry the Wall Street Journal on Iran using a US model against warships, the Financial Times on Yemenis using Claude to build missiles, unnamed Hebrew sources on Iran and Yemen, a Majlis commission head quizzing an AI tool on how often Trump had claimed victory, and Trump on consulting AI before the Maduro arrest. Three more relay AI statements by Musk, Bessent and Trump, one relays a Defense News item on AI-equipped drones, one uses the term as slang, and one shows no AI wording in its text. One post, dated 27 September, carries a video that the channel itself captions as AI-made images of the first moments of the search for Nasrallah. That is the only AI-generated media item on the channel page, and the channel labels it.

The Bale preview of @moghavematnews_iran, titled «مقاومت نیوز», "Resistance News," shows ten posts, all stamped 14 March and about 207 days old on 7 October. None mentions AI, and the preview shows image placeholders in place of media. The Eitaa page for that name, titled as a buy-and-sell channel with six followers, shows text posts dated 19 September to 31 October 2025. Two offer fake members and one links to a channel-sales group. Nothing on it ties the page to the IRGC. On 7 October @hamyane_sepah stood live on Eitaa only, and @moghavematnews_iran stood live on Bale and on Eitaa as a sales page, with no deletion showing. For @hamyane_sepah, Bale answers that the requested conversation does not exist. Anthropic lists a private Eitaa channel as Monjaneq, which reads as a second spelling of the factory name Manjanegh (Catapult) in Section 3.1, and its address loads only a contact stub, so a deleted channel and a private one look alike.

A second snapshot of @hamyane_sepah, taken on 9 October, holds about three dozen posts stamped 16 and 17 Mehr, which convert to 8 and 9 October. Counters now show 27.8 thousand photos and 19.2 thousand videos beside the 21.9K followers. The bio promises monitoring of resistance groups, and one admin handle is published. No post carries AI, cognitive-war, internet-control or platform wording. Houthi and Yemeni strike claims, Hormuz warnings and diplomatic relays fill the page. Two relays touch the information war. One passes on a claim that the director of United Against Nuclear Iran said Iran's psychological war had drained American morale for the war. The other carries a major general's list of naval assets, electronic warfare among them. Forwarded advertisements for a skin cream, a diabetes scheme and an instant loan sit among the war posts. That fits a monetized supporters' channel and does not settle who runs it. Every view counter reads one, which looks like a capture artifact.

Public research on automated or AI-driven amplification across the three apps is a documented null. Searches in English and Persian to 7 October and Filterwatch's report of 6 March find none. That report lists Bale, Eitaa, Rubika and Soroush Plus as the domestic apps that officials urged people to join after the wartime shutdown. Rubika belongs to Hamrah-e Aval, a telecom with IRGC links. IranWire reported in August 2021 that Rubika created thousands of fake accounts in users' names, with verification ticks. That case predates AI. The Open Technology Fund's Security Lab audited all three apps in December 2023 and October 2024. None offers end-to-end encryption, and the servers see which sites users reach. Eitaa and Rubika run on Telegram code, and Bale is

Copyright 2026 Treadstone 71 - www.treadstone71.com

a fork of Actor. The audit ties Eitaa to a University of Qom incubation center and Bale to Sadad and Bank Melli. Rubika's own documents claim AI image analysis for hijab detection, and the audit finds no AI analysis of message content in the app code. The audit predates the window and bears on platform visibility only. No 2025 or 2026 research from Citizen Lab, OONI, ISD, Recorded Future or Check Point names bot or AI amplification on any of the three platforms.

The limits are plain. A single snapshot dated 7 October forms the record: a dozen posts from a single channel across about two hours, a truncated view, and a Bale preview whose newest post is nearly seven months old. Classification reads post text. Photos and video sit outside it, so a result of no AI claim is no finding that the media are authentic. English glosses of Persian text are working glosses and no translation of record. Logged-in content, Rubika and the Monjaneq channel sit outside the record. One snapshot says nothing about how the channels amplify content, and the null on research stays where it was.

4 Russia

Russia shows the widest spread of AI use in the record. State media run models as staff. Networks that researchers tie to the GRU forge brands and clone voices. A third line targets chatbots themselves. The Kremlin talks about sovereignty and says nothing about influence work.

4.1 State media run AI inside the newsroom

Anthropic banned accounts tied to two Russian editorial pipelines (GTG-24015). In the first, a contractor with links to Russia pulled content from Telegram channels and had Claude write Latin American Spanish articles for Sputnik en Español and for a Telegram channel with the handle @ATodaPotencia. A Sputnik Mundo presenter and producer watched the editorial work. The contractor also fed the output to a Telegram channel that presents itself as independent and recasts Kremlin-aligned narratives as local commentary. Local voice. Kremlin script.

In the second, an employee of a Russian state-owned media outlet used Claude to build live-broadcast material: tickers, screen captions, voiceover scripts and short headlines. Russian newswires, the SVR and the Defense Ministry supplied the inputs. Newswire in. Broadcast out. In at least one confirmed case the material reached Russian airwaves.

Anthropic's report names Sputnik Moldova, RIA Novosti, Sputnik en Español and Sputnik Africa as outlets that published Claude-written content. Anthropic states high confidence that the output reached state-owned and state-funded media. The actors hid their identities and their access path.

Assessment: Russian state media use a commercial model as newsroom staff, through contractors and through employees. Likely (55 to <85), Moderate confidence. Row RU-1 grades B2 with T 76, so the cap holds the band at Likely. Anthropic names the outlets. Their published content is open to checking, so the finding rests on more than the vendor's word. OpenAI's operation in Section 4.7 repeats the working pattern from a second vendor.

The Russia line of the AI Tradecraft Adoption sub-index in TAI moves on the same finding. In the Cognitive Warfare Monitor, the RU track gains a named-outlet entry with a vendor attribution grade of B2.

4.2 The chatbot as audience

Russia writes some content for machines. The DFRLab audit of 8 April found content from the Pravda network, Glassbridge and RT inside Common Crawl, a major AI training source. A Llama 3.1 405B Base model reproduced some of that content nearly verbatim in some cases. Nobody yet tests Pravda memorization properly, DFRLab

states, because most Pravda content post-dates the model cutoff. DFRLab splits the problem in two. Training-data poisoning changes what models know. Retrieval-layer poisoning alters what a model finds when it searches. The lab judges that the network "almost certainly" intends to shape the responses of certain large language models. Analysts at the American Sunlight Project, which coined the term LLM grooming, estimate at least 504,060 Russia-influenced, Ukraine-related chatbot responses per month in the EU. Its estimate applies a rate of roughly 18 percent from ISD. ISD ran 300 test responses from four chatbots in five languages, and nearly 18 percent cited a Russian state-attributed source. Neutral prompts drew 11 percent, biased ones 18 and malicious ones 24. Only two links came from Pravda. ISD says it cannot tell whether grooming happens at scale. Its test measured no belief and no click. The project applies the blended rate to ordinary news queries, where ISD's neutral rate was 11 percent, and it calls a citation an influenced response. Its authors call the result a conservative modeled baseline. No judgment rests on it. The rate still feeds the modeled chatbot-exposure inputs in Section 8. So the CWIA, CWC and ATCRI entries for chatbot poisoning inherit the doubt.

The figure is an estimate. Published probes do not measure the same thing, and the table sets their definitions side by side.

Table 6. *Four chatbot probes of Russian content, with the denominator and the definition of a hit.*

Study	Date	Chatbots and prompts	Responses	What counts as a hit	Headline
NewsGuard	6 March 2025	10 chatbots, 15 false narratives, 3 prompt styles	450	The chatbot repeats the false claim, even with a caution. Pravda links count apart	33.55 percent repeated a claim. 56 of 450 (12.4 percent) cited Pravda, 7 of 10 chatbots among them
Harvard Kennedy School Misinformation Review	15 October 2025, fieldwork 22 April 2025	4 chatbots, 13 prompts, 2 locations	416	A response confirms a known false claim, or it cites a Pravda domain	21 (5 percent) supported disinformation. 34 (8 percent) cited Pravda, 30 of them while cautioning or debunking
ISD	27 October 2025	4 chatbots, 5 topics, 5 languages, 3 query types	300	A response cites a Russian state-attributed source under the EEAS typology	Nearly 18 percent. Two cited links came from Pravda
DFRLab and Check First	12 March 2025	4 chatbots, a simple test	Not stated	Content sourced to Pravda domains with no warning	No percentage. None of the tested chatbots warned or refused

The numbers do not conflict. They count different things. NewsGuard counts a repeated claim from any source. The Harvard team reads that as a conflation of repetition and citation. ISD counts any Russian state-attributed source and not only the Pravda network. Each rate belongs to a prompt set its authors chose, and none measures what a user sees. The Harvard authors attribute most Pravda citations to data voids and warn that 416 responses limit the power of the test.

NewsGuard published two audits on the same question. One tested Claude on 20 pro-Kremlin prompts and found false claims repeated on three. The other counted 906 false reports that spoof 163 outlets in 21 countries across the Matryoshka, Storm-1516 and Doppelganger operations. NewsGuard sells a competing product and audits Anthropic, and the 906 count is its own. Neither audit enters the ledger. Nor does either carry weight in any judgment.

Assessment one: Russian state-linked content sits in Common Crawl, which feeds major AI models. Likely (55 to <85), Moderate confidence. Row RU-2 grades B2.

Assessment two: that content reaches chatbot answers through retrieval and changes what users believe, at scale. Roughly even chance (45 to <55), Low confidence. No probe of memorized Pravda text has run on a model

Copyright 2026 Treadstone 71 - www.treadstone71.com

whose training cutoff post-dates the network's expansion, though retrieval-style probes have. Modeling produced the exposure figure. Row RU-3 grades C4, T 53 and D 25, because the project relabels ISD's citation rate as influence. At ISD's neutral-prompt rate of 11 percent, the project's own inputs yield roughly 308,000 responses monthly in place of 504,060. No reviewed source measures a change in belief. Impact scores follow the evidence: Elevated on the CWIA at 50.0 with confidence 50, and a Counter-Influence Priority of 176.0, ranking fifth of eight.

For Treadstone 71, chatbot-as-audience belongs in the Cognitive Warfare Tradecraft Reference as a maneuver with two exemplars, retrieval poisoning and training poisoning, and a falsifier: a memorization test that comes back clean.

4.3 Storm-1516 and the Baltic states

DFRLab records four Storm-1516 campaigns between 30 July and 17 August. One used X's Grok Imagine to turn a photo of a Latvian soldier into a short video. A photo became a video. Two campaigns used cloned sites, one of EUobserver and one of Defense Post. A third put a Euronews logo on a video, and DFRLab found no clone of the Euronews site. The Clemson Media Forensics Hub spotted the activity first. DFRLab writes that the campaigns show patterns characteristic of those attributed to the GRU. Recorded Future writes the name of the same network as "CopyCop (also known as Storm-1516)." The report counts one network, and Sections 4.3 to 4.5 give three views of it.

One narrative told Latvians that only 25 percent of conscripts report for duty. Latvian media put attendance at medical exams near two-thirds. The two figures count different things, a report for duty against attendance at an examination. So the rebuttal contradicts the claim only by implication.

The tool outweighs the video. A Russian operation picked a Western commercial generator, owned by a US company, to make its clip.

Assessment one: Storm-1516 adds AI-generated video to forged outlet branding in its campaigns. Likely (55 to <85), Moderate confidence.

Assessment two: the GRU directs the pattern. Roughly even chance (45 to <55), Moderate confidence. DFRLab matches patterns to earlier GRU attributions and ties none of the four campaigns to the GRU itself. Open sources do not show command, so the direction call ranks below the pattern call. Row RU-4 grades B2, T 72, D 25. Only the fourth campaign used Grok Imagine, and DFRLab counts 1,651 accounts and 275 mentions. Scores: Monitor on the CWIA (42.5, confidence 80) and Counter-Influence Priority 228.0, rank 3 of 8.

For Treadstone 71, Storm-1516 belongs in the Monitor RU track as a repeat entry with a velocity note. A fifth campaign next quarter makes it a Public Attribution Series candidate.

4.4 CopyCop and the Firebird data center

Enterprise AI became influence terrain in the summer. In a post of 18 August, Recorded Future Insikt reports that three impersonation videos "very likely" targeted the joint US and Armenian-backed Firebird AI data center in Armenia before it opened. Two videos posed as TechCrunch and Gizmodo reports. The third posed as an Iranian military voice. Insikt ties the videos to CopyCop campaign infrastructure. DFRLab's Storm-1516 in Section 4.3 is that same network. Combined views of the three videos pass 1.6 million, across data from 24 June to 13 July.

The videos told viewers that the site sits on an earthquake risk, that the local grid will fail under its load, and that Iran will strike it. Insikt records each of those statements as fabricated.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Look at the audience. Armenians who live near the site. Investors who weigh the project. Officials who signed off on it. The target was a building. Every person who decides its fate made up the audience.

Assessment: CopyCop points influence work at AI infrastructure projects to raise doubt among investors and local publics. Likely (55 to <85), Moderate confidence. Row RU-5 grades B2, Level 3, T 68, D 25. The grade covers Insikt's analysis of the videos and the infrastructure links. Clips themselves stay out of the T-Score, because video evidence carries ASA cluster C5 risk.

Of the eight campaigns scored in Section 8, Firebird takes the top Counter-Influence Priority at 280.0. The CWIA reads Elevated (52.5, confidence 80). Size does not drive that rank. Impersonation of trusted outlets against a named facility, with a plausible trigger, drives it. The Armenian audience scores an HTIM Index of 62.0 at confidence 75, highest of the three Russian-theater audiences. Political resonance scores 8 of 10 and channel saturation 6.

For Treadstone 71, Firebird goes into the Sovereign AI Architecture Audit as a training case, and the video set goes to the Tradecraft Reference as an exemplar of brand impersonation against infrastructure. Enterprise clients that open an AI facility get a rehearsal of the attack before launch. Section 13 turns that into action.

4.5 Self-hosted models that no lab sees

Insikt's earlier CopyCop report, dated 18 September 2025 on its page and 17 September in the PDF, says the same network built at least 200 new fictional media sites and passed 300 for the year, including a German set. The report states that CopyCop runs self-hosted, uncensored large language models based on Meta's Llama 3. Insikt cites the Washington Post and the US Treasury for GRU sponsorship of John Mark Dougan's language-model servers. Dated 23 October 2024, the Post article rests on more than 150 Russian documents that a European intelligence service obtained and passed to the paper. The documents are unpublished. Payments reached Dougan's Moscow bank account from April 2022, the Post reports, from a man whom two European security officials name as a GRU officer. The Post says the server launched in the summer of 2024, after his sites ran into trouble reaching Western AI systems. Dougan denies the payments and calls the documents fabricated, Dugin denies any GRU partnership, and Korovin dismisses the claims. The Post describes a server meant to generate its own AI content and end reliance on Western technology. Llama, self-hosted and CopyCop appear nowhere in its text, so the language-model gloss belongs to Insikt alone. Insikt's own evidence is an Open WebUI login page on a domain tied to Dougan, seen in March 2025, a Python script that calls Ollama, shown in a French television interview in June 2025, and knowledge-cutoff artifacts in published articles. Insikt names two Llama 3 fine-tunes as the likely models.

Two outside items touch the claim. A US Treasury release of 31 December 2024 says the Center for Geopolitical Expertise, a Moscow affiliate of the GRU, built a server for generative AI tools to avoid foreign web hosts. The GRU paid for the server, and for part of the rent on the apartment that houses it. Treasury names the center, not CopyCop. The Post drew on documents from a European intelligence service, and Treasury's account perhaps rests on related intelligence. We cannot tell, so the two count as one origin. A French academic test of 2,646 articles from the campaign, posted to arXiv on 16 August 2026, supports Insikt's Llama 3 attribution and hints at Mistral models as well. The test reads text and shows no hardware. CrowdStrike's Global Threat Report says a pro-Russia propagandist used AI to build media websites and videos for campaigns aimed at US and German elections across 2024 and 2025. The report names no actor, no operation, no count and no audience, so it gives no second source for the CopyCop findings.

Dougan has spoken on the point himself. NewsGuard reports that he told a Moscow roundtable on 27 January 2025 that a single server in his home writes almost 90,000 articles a month, and that an open, uncensored Llama

Copyright 2026 Treadstone 71 - www.treadstone71.com

trained on a Russian view is the kind of model to use. Our fetch of the NewsGuard page failed with an access error, so the text rests on search excerpts, and we saw no video. His statement is a Level 1 item and does not tie CopyCop output to that server.

Sber's AI unit, publicly known as the owner of the ai-forever organization, shows what a Russian developer publishes in the open. The organization's Hugging Face page of 9 October lists 122 models and 44 datasets, among them a family of text-diffusion models of 0.6 to 4 billion parameters, and the newest entries are hours old. That shows availability and says nothing about any operator's use.

Lab and vendor reports in Section 4 sample their own users. A network that runs its own weights leaves no account to ban. Nothing to read. Two Russian-language searches returned tutorials, Ollama security checklists and coverage of an Ollama vulnerability, and no report of an adversary running a local model. The null rests on two searches and proves little. Section 6 carries a Kimsuky finding that points the same way on the North Korean side, from one source. On the Chinese side, a Google sentence in Section 5 on suspected UNC6508 activity points that way, again from one source.

Assessment: Russian operators run at least one large influence network on self-hosted open-weight models that Western labs cannot observe. Likely (55 to <85), Low confidence. Recorded Future holds the evidence for the self-hosted models: login pages, a script and model artifacts. No source shows CopyCop hardware itself. The accurate phrase is a self-hosted inference stack. Treasury and the Post give partial support on the funding, and they count as one origin. The arXiv test gives partial support on the model family. None of three covers the hosting. The single-reporter step-down stays. Row RU-6 grades B2, T 65 and D 25, five points above the Weak edge. The consequence reaches past CopyCop. Every prevalence figure in the lab reports undercounts that activity by an amount we cannot measure. The limit sits in Assumption A1 (Section 1.6).

The record also holds a counterweight. Beyond the CopyCop network, every Russian state-nexus case in the record runs on a commercial Western model: Claude in Anthropic's GTG-20006 (Section 7.4) and GTG-04001, and Gemini for SANDWORM RELIC. DFRLab records the same network on Grok Imagine in one of four campaigns. Self-hosting is one network's practice. The record shows no Russian pattern.

Most-Likely Achieved Endgame: Russian operators keep the commercial-model workflow wherever it works, add self-hosted models for part of the volume and the sensitive work, and thin the lab record. Likely (55 to <85), F-Score 60. Alternatives: a deliberate poisoning push against chatbot retrieval at scale, Unlikely (15 to <45); a full pivot to a sovereign Russian model stack by the end of 2027, Unlikely (15 to <45).

4.6 Matryoshka against the German state elections

ISD found 269 posts on X between 24 June and 1 September that targeted German state and municipal elections. About 40 percent of the operation's videos seem to use AI voiceovers. Cameo clips of celebrities, including Sarah Jessica Parker and Mike Tyson, paired real faces with cloned voices. The operation also pushed fabricated allegations under the tag #StopMerzToday. ISD reports basic errors, limited reach.

Assessment: Matryoshka used AI voiceovers in about 40 percent of its videos, and its reach in Germany stayed limited. Likely (55 to <85), Moderate confidence. Row RU-7 grades B2, T 76. ISD's report of weak craft and thin audiences supports the Efficiency Amplifier hypothesis in the ACH (Section 11.1) and weighs against the Capability Shift hypothesis. Matryoshka scores Monitor at 30.0, lowest of the eight campaigns. High media trust and strong institutional filters hold the German electorate at an HTIM Index of 46.0, confidence 75, with a Behavior score of 3.

Copyright 2026 Treadstone 71 - www.treadstone71.com

For Treadstone 71, the case works as a falsifier against any client brief that treats AI volume as a measure of threat. The Cognitive Warfare Monitor carries it with Velocity set to low.

4.7 Contractors, small operators and the monitoring state

OpenAI banned Russia-origin accounts on 25 August. The operators prompted in Russian and produced English posts and comments for a front called the International Burke Institute. They built a "Burke Sovereignty Index" and used VPNs. ChatGPT received orders to hide any linguistic clues that the authors were Russian. OpenAI judges the reach modest: typical posts drew low views and the official accounts few subscribers, though the Telegram channels counted 10,000 to 20,000 followers each. The vendor's Brookings Breakout Scale rating falls in Category Three, at its low edge: multiple platforms, with some signs that the content reached authentic audiences. OpenAI puts the significance less in the audience reached than in the infrastructure built.

The front described itself as an Israel-based expert community. That description is the operators' claim. OpenAI's finding of Russian-language prompts is a fact from the vendor's logs. Of 36 sampled articles, 34 were copied from elsewhere, OpenAI says. Row RU-11 grades B2, T 66, D 25.

Anthropic reports a wider-reach case. In GTG-04001 a Russian-speaking operator in Bangui ran a daily content operation through Radio Lengo Songo, a 98.9 FM station that an outside investigation says Wagner created in 2017, alongside RT, Sputnik Afrique and TASS. The operator, Anthropic says, acted as the local media coordinator for Politology, an influence branch that Anthropic assesses to have come under SVR control in late 2023. On the Breakout Scale the operation rates Category Four, with content broadcast daily on the station, amplified on Telegram and carried by local outlets. The report notes that the widest authentic reach in its influence cases came where state media outlets, FM radio among them, carried the content.

Google's Threat Intelligence Group adds two Russian items in its 8 September report. SANDWORM RELIC, also tracked as APT44, used Gemini to script automated password spraying, host profiling and workflow automation against Ukraine. UNC5792, a Russia-based cluster, integrated AI models into Telegram monitoring bots to flag content of interest to Russian authorities, in Google's words. The detail describes experiments: the group used AI to learn how to wire in API keys, to sort messages as suspicious or neutral and to write structured reports, and Google says those activities triggered Gemini's safety responses. Prompt telemetry is the probable evidence, not a sighting of deployed bots. The page does not state its evidence basis. A public service announcement of 26 June from the FBI and CISA ties UNC5792 to Russian intelligence. That text covers phishing of messenger users. Bots and AI go unmentioned. Google's wording on API credential integration points to a hosted model. The first is a cyber operation. That item enters the report only as proof that a unit Google ties to the GRU treats a commercial model as routine equipment. Item two belongs in the domestic control picture next to Iran and China.

Five Russian vendors publish AI counts, and none names Russian state AI use. BI.ZONE's "Dark AI" report of 4 June, 58 pages in Russian, studies 7,477 records pulled between March 2025 and April 2026 from Telegram and seven shadow forums, and its disclaimer says more than 7,400. According to BI.ZONE, 77 percent of the analyzed messages concern bypassing the limits of public models, 22 percent concern uncensored models, and 1 percent cover everything else: malware, attack stages, AI agents and integrated tools. Mentions rose from December 2025 and peaked in January 2026, and BI.ZONE links the spikes to model releases. In its own tests, model-written code drew flags from 13 of 72 and 12 of 71 scanning engines, and BI.ZONE says the models built no working tool without a skilled developer. One forum thread recommends pairing Claude with less restricted or locally hosted solutions. A forum user gave that advice, and it shows no operator running a model. Across 58 pages the report names no state, no intelligence service, no APT group and no government. The count covers messages, not

Copyright 2026 Treadstone 71 - www.treadstone71.com

attacks. BI.ZONE sells AI defense products. A BI.ZONE Threat Intelligence item that CNews carried on 1 October describes two September offers of infostealers with built-in AI victim profiling. Each is a criminal ware, and the article does not say whether a vendor hosts the model or the buyer runs it. In its paper "AI in 2026", Positive Technologies reports confirmed AI use in 10 percent of MITRE ATT&CK techniques, up from 5 percent in 2024, calls malware with AI modules extremely rare, and names no Russian state actor. Its study of 7 October, "Trends in the Shadow Market of Cyber Threats", runs 32 pages in Russian. The actors are criminal. No state and no APT group is named, and APT groups appear only in general terms. The TITAN ransomware group advertises on its leak site an AI module that, in TITAN's words, ranks stolen files and works out a ransom. Positive Technologies repeats the claim and does not test it. The study judges the AI in the XIRA RAT advert probably marketing. KAIDO RAT, which drives a Claude command-line tool as a command channel and targets Brazilian financial firms, also appears in the study. No source for the item appears on the page, and no incident count backs it. In tone the two papers differ. The January paper treats malware with AI modules as extremely rare, while the October conclusion has malware shifting toward autonomous agents, on evidence of adverts and leak-site text. On LAMEHUG, the paper says the malware called Qwen through the Hugging Face API and gives the operators no state attribution. CrowdStrike, a Western vendor, ties LAMEHUG to FANCY BEAR in its annual threat report for 2026 and judges that the malware gained nothing meaningful in effectiveness, a case of "experimentation rather than operational reliance." Kaspersky adds attack counts at press-release depth. Solar's July count of 417 million web events comes from a sample of several dozen companies, and Solar's analyst infers AI use without showing it. Solar is state-owned. F6's global report leaves out Russia and Belarus, so it cannot speak to Russian state use. Every AI case the vendors name is a criminal actor, an unnamed group or an actor attacking Russia. Kaspersky's Securelist item of 3 July on Armored Likho, an actor of unstated origin that attacks Russian government and power targets, infers AI-written loaders from verbose code comments and emoji bullets. BI.ZONE's March item on a werewolf cluster makes the same inference for another group, from detailed code comments. Neither rests on provider telemetry. Western vendors name Russian state-nexus actors on commercial models, and the Russian vendors do not. That asymmetry matters more than any count, though absence in Russian vendor output is expected whatever the facts and carries little weight. Each vendor sells AI defense or detection. None enters the ledger.

Assessment: Russia-origin influence operators and a cyber unit tied by Google to the GRU use US commercial models as everyday tools, and the influence operators mostly get small reach. Two cases break that rule: OpenAI's Burke operation at the lower end of Category Three, and Anthropic's GTG-04001 at Category Four, where an FM station carried the content. Likely (55 to <85), Moderate confidence. A second assessment: a Russian cluster builds AI-assisted monitoring of Telegram for state clients. Likely (55 to <85), Low confidence. Google alone holds the telemetry. Confidence steps down once. Row RU-8 grades B2, T 76.

4.8 What the counts say

Government counts need a careful read.

The EEAS fourth report on foreign information manipulation and interference, dated 12 March 2026, counts 27 percent of 2025 incidents as AI-involved. Three annexes deal with sanctions, law enforcement and digital-regulation tools. Not one explains how incidents are coded for AI. The nearest definition is the summary's line that one in four detected incidents involved AI tools used to produce or distribute content. Figure 11 sorts AI use by function, as content creation and distribution. Of 540 incidents the EEAS detected in 2025, 147 make the 27 percent, up from 41 cases in 2024. The report calls its set a selective, time-bound sample built partly from public reporting or independent analytical attribution. Russia takes 29 percent of incidents by attribution and China 6 percent, and 65 percent stay unattributed. No coding rule is published. CrowdStrike prints its 89 percent in the

foreword, the introduction and the overview of the 2026 Global Threat Report. Nothing in the body defines an AI-enabled adversary, counts adversaries or attacks, or sets a baseline. CrowdStrike's figure is a yearly growth rate for attacks launched by adversaries that CrowdStrike tags AI-enabled, with no denominator and no unit. The two figures measure different things, one a share of influence incidents and one a growth rate in intrusion activity. They cannot be compared, averaged or trended together. Section 1.6 lists the mismatch as assumption A4. The wording and existence of the EEAS figures grade A2, taken from the EEAS itself. Row RU-9 grades the measure B3, because it lacks a coding rule and a coder-agreement figure, rests on a selective sample partly drawn from public reporting, and counts channels two ways. The row stays at Level 2, T 60 and D 25. T 60 is 59.7 before rounding, and one point lower reads Weak.

The joint EEAS and Ukraine Centre for Countering Disinformation report of 23 June counts 244,000 publications that drew 1.39 billion views on Ukraine's EU accession between January 2025 and April 2026. About 80 incidents targeted EU audiences. Ukraine's center is a belligerent body, and the counts are its own. The center says the "ZOV" network holds 68 sites, 40 Telegram channels and 40 VK pages, and the report describes "AI-enabled mass production of content." Row RU-10 grades C3 with T 51, Weak, and D 27, Unlikely. No method for the counts appears on the page.

On 23 September the same centre reported a new series of AI videos in "street interview" format in central Kyiv, which it places on Russian-controlled TikTok accounts and which push "peace at any price" («нову серію маніпулятивних ШІ-відео у форматі вуличних інтерв'ю в центрі Києва»). Earlier items cover AI fakes about women's mobilization and outraged veterans. The attribution to Russia belongs to the center alone. Row RU-13 grades C3. The row carries no weight.

Assessment: Russia uses AI to mass-produce content aimed at EU audiences. Likely (55 to <85), Moderate confidence, because independent sources, ISD and DFRLab, show the same behavior in Sections 4.3 and 4.6. The size of that output, and the share it holds of all Russian operations, stays unknown.

4.9 What the Kremlin says

At the Valdai plenary on 1 October, Putin called AI the most obvious new dilemma. He said experts warn more often that AI is able to bring everyone to a common denominator through «кем-то настроенных алгоритмов», "algorithms set up by someone." His next line read «Такой цифровой тоталитаризм будет таким же тупиковым, мрачным и антигуманным», "Such digital totalitarianism will prove just as dead-end, gloomy and inhumane." The warning belongs to unnamed experts. No state is named, and the text does not say Western. He asked how increasingly digitalized states secure sovereignty, as an open question about states in general and not about AI models. On AI in security he said progress cannot stop, a human must take the final decision, and the limits are a question of the survival of humanity. The words cognitive, information and cyber never occur in the Kremlin transcript, which carries the speech. He says nothing about state use of AI in influence operations. The Kremlin published a list of instructions (Pr-2434) on 29 September, and Rossiyskaya Gazeta reported it on 30 September. Under the list, the government must organize «централизованный сбор, обработку и предоставление научно-технической информации», "centralized collection, processing and provision of scientific and technical information," to train sovereign and/or national foundation models, with a report due on 1 March 2027. The same article lists two further instructions: amend legislation if needed, and give scientists access to software built on Russian AI technologies, with a decision on funding. A Rossiyskaya Gazeta ticker of 9 October carries a headline that infrastructure of a Yandex data center in Kaluga was damaged after a drone attack. Only the headline is captured. Federal Law 243-FZ, "On supporting the development of AI technologies," passed the Duma on 8 July, passed the Federation Council on 17 July and was signed on 26 July. Article 1 of the official text, on publication.pravo.gov.ru, lets later acts set separate rules for large models used in defense, state

security and public order. The law says nothing about information or influence operations. Large foundation models of at least one billion parameters fall under the law. A sovereign model needs a Russian developer, full reproducibility including training, inference and storage in Russian-owned data centers, and a Government-run confirmation of conformity with Russian law and traditional spiritual and moral values. Components developed in Russia or abroad, other large foundation models among them, are permitted in a national model under open licenses, with the same data-center and values tests. The government will set the cases in which use is limited to sovereign or national models. Article 9 gives sites with over 500,000 daily Russian users a duty to let users label AI output. Anyone who builds audio or visual material with a large model gains the possibility of adding a notice, in a format agreed with the model provider. No mandatory watermark follows. Article 11 sets no penalties. General effect began on 1 September 2026, the sovereign and national classes and the labeling article start on 1 March 2027, and existing systems with data held in Russia run to 1 September 2032. Mishustin said on 6 October, in remarks posted on government.ru, that the government will define which sensitive sectors run only sovereign or national models. He said the Max messenger has 130 million registered users, a count of registrations and not of active users. The decree meant to name those sectors has no published text. A working draft has reached the press, and a law-firm digest of one Kommersant report is the only account of it in the record. The draft carries no weight.

Every item above is state signaling or a state outlet's account of policy. Rule 1 caps RU-12 at C3. The law is a policy motive for interest in model capability and no evidence of misuse. Policy items carry no ledger row. They support a judgment about doctrine and none about conduct.

Public Telegram windows of 6 and 7 October cover six channels, each with about 20 posts: the foreign ministry spokesperson Zakharova (461,000 subscribers), Solovyov Live, Rybar, Readovka, Mash and the Ministry of Digital Development. The Digital Development ministry's window is stale and runs to late March. No claim of Russian AI use appears in the windows. One Solovyov post repeats a claim, unverified, that Washington has renamed AI "superintelligence." In Zakharova's channel, the foreign ministry presents AI fakes as a Ukrainian and Western product. The ministry announced the forum "Dialogue about Fakes 4.0" for 22 September, where Lavrov was to speak. No speech text exists in the record. Putin's greeting to the same forum calls AI and its linked threats a serious challenge in infocommunications. The greeting is silent on Russian state use. That null result measures a window and says nothing about their activity. Wagner-linked channels, NoName057(16) and CARR sit outside the sample.

Assessment: Russian leaders frame AI as a sovereignty and domestic-control problem. Roughly even chance (45 to <55), Low confidence. The Valdai transcript, the 29 September instruction list, the law's official text, Mishustin's remarks and Putin's greeting name no state use of AI in influence operations. Putin's greeting and Mishustin's remarks treat AI deception as a threat to Russia. Texts of other speakers at the forum are absent.

We infer that a closed model stack inside Russia shrinks what Western labs see. Roughly even chance (45 to <55), Low confidence. The reasoning is ours. Nothing in the ledger shows the stack at work. The law's national class lets foreign open-license components in, so a national model need not be domestic throughout.

Rows AO-RU-01 to 03 in Section 16.3 list the Russian statements read in the original, with the sources read after the ledger froze in Section 16.4.

4.10 Adversary reaction read (ACS)

The Adversarial Cognitive Simulator, run on a Russian intelligence-linked influence apparatus under the trigger "public attribution of a state-media pipeline and a Western-tool takedown," gives the read below.

Worldview. Information works as terrain in a standing confrontation with Western institutions. Reflexive control shapes how it designs operations: feed an opponent the inputs that lead the opponent to choose what the sender wants. Fears: a ban on Western tools, exposure of named contractors and state-media staff, and a sovereign-model project that fails to match Western capability. Biases: mirror imaging of Western institutions, and trust in volume as proof of effect. ISD's German findings fit the second bias.

Likely actions: rotate contractors and accounts, shift volume work onto self-hosted models, as CopyCop already does, answer attribution with denial and counter-accusation, and reuse forged outlet branding because the tactic keeps working. CARM pathways: deny, accuse back, relocate. Miscalculation risk: the apparatus counts reach as persuasion and spends more on content that moves no one. A second leak through contractors who hold the tool relationship adds a further risk.

4.11 Russian-theater audiences (HTIM)

Three audiences, three scores. The German electorate returns a TVI of 5.33, a PES of 3.5 and an HTIM Index of 46.0. Latvian conscripts and Baltic audiences return 6.67, 4.5 and 58.0. The Armenian public and the AI-investment community return 6.67, 5.5 and 62.0. Section 8 holds the inputs. The ordering matches the record. Audiences with the most to lose from a stalled project and the least trust in outside institutions score highest. The audience with the strongest institutional filters scores lowest.

4.12 Russia as a thief of model capability, on an empty record

Anthropic's February and September reports attribute every distillation campaign to Chinese labs. Google reports distillation by private companies and researchers and names no Russian actor. Mandiant notes theft of AI data in extortion cases in the second quarter, with no Russian attribution. The public record holds no sign of Russian state distillation or weight theft. One case comes near: GTG-50020, a Russian-speaking, financially motivated actor. That actor injected prompts into an AI vendor's evaluation sandbox and took production API keys for several providers. The campaign reached about 30 AI companies in about four days. Its stated goal was access to a pre-release Claude model, and it never gained that access. The case shows a criminal after capability, and it shows neither weight theft nor a state act. Anthropic is target and reporter. The null covers open sources and vendor reports, and the vendors observe only their own users. That null grades C3. The only Russian-state link to model capability is the sovereign-model law in Section 4.9, which is a motive and no act.

5 China

China presents two records that matter more than its influence work. Security bureaus use AI to watch their own people and the diaspora. Chinese labs take US model capability. Influence operations come third. Their reach stays small, and the PLA's own writing shows more ambition than the evidence shows practice.

5.1 Security bureaus run AI as analyst staff

Anthropic banned three clusters of accounts that it ties to PRC-aligned actors in security, public security and public-opinion work. Each used Claude in place of a staffed team. Three clusters, one pattern. The labels differ by cluster, and the sections below keep them apart.

GTG-14020 ran a religious affairs intelligence operation. Anthropic calls the actor PRC government-aligned. The label rests on inference from targeting that maps onto United Front, State Security and former religious-affairs priorities. No organ is named. The page gives no confidence level. Dossiers in Chinese covered religious leaders and diaspora figures across Asia, with Catholic, Tibetan Buddhist, Falun Gong and Taiwanese Christian

Copyright 2026 Treadstone 71 - www.treadstone71.com

communities as targets. GTG-14022 ran what the Chinese party-state calls 舆情, public opinion monitoring. Its operator told Claude to produce restricted briefings, 舆情简报, that cataloged dissidents, activists, ethnic minority communities, diaspora groups and foreign media as threats to political stability. Anthropic assesses with medium confidence that the actor was a contractor working for government clients, likely connected to the state security or united front ecosystem. The page does not call it a state organ. GTG-14021 gives the most detail. Anthropic ties the cluster to a municipal cyber police unit and finds three levels of actor: an individual cyber police officer, a police academy student and a bureau of several people. The cluster used government document templates for "situational awareness" briefings, sensitivity tier taxonomies and stability-maintenance vocabulary. Operators ran Claude Code with custom surveillance skills. They queried a government surveillance database. An internal AI usage manual codified one prompt formula for distribution inside the bureau. Anthropic names Hong Kong pro-democracy figures, organizers of Tiananmen commemorations, Uyghur advocacy groups and international human rights organizations as the transnational repression targets. The vendor connects one operator to a municipal public security detective who studies at a police academy. Anthropic places the state security bureau, likely, in Zhejiang. Neither link rises above medium confidence, and the identification of the cyber police officer carries low confidence. Device clocks showed UTC+8, whatever the exit node.

A fourth China-based cluster sits on the same page and outside our ledger. GTG-14010 is a contractor, at low confidence, working to PRC state security collection priorities on Uyghurs in Syria, with bureau-level government clients for surveillance bids.

A staffed analyst team becomes a prompt formula. The formula spreads by memo.

Anthropic's report is the source for all three clusters. Identifiers GTG-14020, GTG-14021 and GTG-14022 sit on the page, with the unit labels given above. The page carries no publication date. Reuters, writing on 11 September, calls the release a Thursday event, which puts it on 10 September 2026, the date we use. Its case-by-case relay frames each item as Anthropic's account, calls the religious-surveillance actors China-linked, lists Catholic cardinals, Taiwanese Christian leaders, Tibetan Buddhists, dissidents and activists among the targets, and repeats the low-confidence label on the Uyghur-in-Syria contractor. Reuters also records that China's foreign ministry said it had no awareness of the report and opposed "distortion of facts and smears." A second vendor sees the pattern. OpenAI's February "Cyber Special Operations" report records an account tied to Chinese law enforcement that asked ChatGPT to polish status reports on 网络特战, cyber special operations, and links the activity to Spamouflage and to harassment of the @whyyoutouzhele account and of Safeguard Defenders. ChatGPT refused a plan that targeted Japan's prime minister. In that report a user pasted status reports that said provincial teams experimented with locally deployed DeepSeek-R1, Qwen2.5 and YOLOv8. That is an actor's claim, and OpenAI says it was unable to confirm it. OpenAI's own page summary lists local models among the strategy's features, while the body treats them as the user's claim. The June report repeats that Chinese cyber special operations favor locally deployed Chinese open-weight models. No case accompanies the point. June puts the point in OpenAI's own voice, "as we reported in February", and drops the February caveat that the user's claim went unconfirmed. OpenAI adds that it cannot say what drove the operators to American tools. No June actor ran a local model, and the repeat brings no new evidence.

Assessment: PRC-aligned actors at a municipal police unit, a state security bureau and a contractor run AI as standing analyst staff for domestic sentiment surveillance and diaspora targeting. Likely (55 to <85), Moderate confidence. Row CN-1 carries B2 at T 60, and row CN-2 corroborates at T 64. Unrounded, the CN-1 score is 60.3, and KJ1 leans on it as its weakest row. A reading one point lower moves KJ1 from Moderate to Low confidence. Anthropic rates its bureau-level links at medium or low and gives no confidence for GTG-14020's unit, so we hold attribution to a named bureau a step under the pattern judgment. No outside text names a city police bureau, a state security bureau or a contractor in AI-enabled surveillance. ASPI's December 2025 study of Chinese AI and

human rights supports the domestic picture and names no cluster. OpenAI's reports and Google's tracker of 8 September show other PRC actors using commercial models. That corroborates the class at B2 and leaves each cluster on Anthropic alone. The ATCRI ranks AI-run surveillance and transnational repression second of ten (raw risk 441, confidence 80).

A Cultural Nexus read shows why demand runs high. Measures 01 and 06 of the nexus, information control and surveillance acceptance, run high for China. A cheaper way to monitor meets little resistance at home.

The finding lifts China in the AI Tradecraft Adoption sub-index of TAI. A Public Attribution Series dossier gains a second Chinese formation type, the municipal cyber police unit.

5.2 Influence work for hire, and small reach

Private firms carry part of the Chinese influence load. OpenAI's 10 June report on PRC-linked work around American AI debates describes two case studies, "Data Center Bandwagon" and "Tech and Tariffs." For the first, OpenAI judges the operators likely part of a social media team at a private Chinese technology firm that works for provincial government clients. The PDF edition of the report carries that judgment. For "Tech and Tariffs" the report points elsewhere. OpenAI sees public-security indicators and activity aligned with CCP interests, yet it leaves the operators' institution unestablished. False claims that attackers had compromised ChatGPT user data came from a separate set of X accounts, and none of them posted model-written text. OpenAI ties them to "Tech and Tariffs" by behavior and calls the tie likely. Impact stayed low. The report finds "no evidence of meaningful breakout beyond its own activity" and rates the activity Category One on the Breakout Scale and names one platform only. Small reach, again.

Graphika reports that Tumblr accounts it assesses with high confidence as Spamouflage published AI-generated bilingual articles, in English and Tibetan, on the Central Tibetan Administration election. A set of 103 coordinated X accounts posted AI-generated cartoons. Graphika writes that AI tools "most likely enabled the production of bilingual content at the volume we observed." The article is dated 4 March 2026, and its high-confidence Spamouflage label covers the Tumblr cluster. On the 103 X accounts, Graphika says they act in coordination, and that Spamouflage-linked accounts spread a poster that was likely AI-generated. The text states flatly that the Tumblr articles were AI-generated and infers that AI enabled the volume. Graphika names no model or vendor.

Two more items set the limit. Graphika's July report on Spamouflage and manipulated event flyers, aimed at a Safeguard Defenders event and Tibet and Uyghur protests, carries no AI element in its public summary. Meta's H2 2026 adversarial threat report says that every coordinated inauthentic behavior network it disrupted in the report incorporated some form of generative AI that Meta was able to identify. That is a finding on networks Meta chose to take down. So the sample is selected. Meta titles the edition Second Half 2026. The cover date is August 2026, and the report says its investigations ran primarily in the first half of 2026, so the title names the half of publication. Its case studies for the half name no China-origin network. Two China-based surveillance-for-hire firms appear by name, Surfilter and Sinovatio. Meta ties neither to the state. Meta's cumulative counts since it began reporting stand at 53 networks from Russia, 41 from Iran and 14 from China. They span years, and we do not weigh them against the half's cases. Row CN-11 stands as the AI-prevalence row for Meta and not as a China row. AI shows up almost everywhere.

Philippine and Japanese election material adds two cases. A DFRLab report of 30 June finds 50 Spamouflage profiles with AI-generated images on Philippine Facebook pages, beside a separate Philippine state-aligned network that DFRLab ties, at moderate confidence, to an army unit. Japan's February election drew press counts of 3,000 and 394 accounts with simplified-Chinese traits and 59 AI images. Nothing ties them to the Chinese state. Nikkei judged the influence limited. NHK found no clear evidence of foreign manipulation.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Microsoft's threat analysis center lists no East Asia report after April 2024, a documented null. Unit 42 and ASPI show no 2026 item on PRC AI influence. Index and HTML pages from CVERC, Qianxin, 360 and NSFOCUS show no US or Taiwan AI operation. CVERC's annual situation report of September 2026, 127 pages, was read in full. Its body names no US or Taiwan operation and holds no passage on cognitive warfare. One vendor table in it codes a group as North American, APT-C-78 in 2025 only, with no entity named and no evidence. Its AI-abuse section covers fraud and nothing political. A preface says the text is no official government position. The report is a joint product of CVERC with Antiy, 360, NSFOCUS, Qianxin and Huaan Yunke, so the null now reaches their contributions. Sangfor and Venustech pages and the separate report bodies of those firms sit outside the sample. No HTIM Index exists for a Chinese audience, because the record names no audience for a PRC operation in enough detail to score.

Assessment: PRC-aligned influence operators use AI for translation and volume, and the audience stays small. Likely (55 to <85), Moderate confidence. Rows CN-2, CN-3 and CN-11 grade B2 at T 64, 64 and 63, each sitting no more than five points over the Weak line. Row CN-11 adds only the prevalence point. Scores track the evidence. Spamouflage on the Tibetan election rates Monitor at 35.0, with a counter-influence rank score of 88.8. The US AI-debate operation rates Monitor at 32.5, with a rank score of 69.6, the lowest priority of the eight campaigns in Section 8.

For Treadstone 71, the pair backs a standing rule for the Cognitive Warfare Monitor. The CN track records AI presence in each entry. Velocity scores from measured reach, never from account counts.

5.3 Taiwan's claims about hired firms

Taiwan's National Security Bureau makes the broadest claim in the China record. More than 45,000 sets of inauthentic accounts operated in 2025 and spread 2.314 million disinformation items, the NSB asserts. The bureau also asserts that the Cyberspace Administration of China, the United Front Work Department and the PLA Cyberspace Force tasked firms named Zhongkedianji, Beijing Xingguang and OneSight to use generative AI with more than 10,000 bot accounts. Its list continues. iFlytek and Magic Data built a Taiwan-accent voice database. GoLaxy and Meiya Pico crawled and profiled targets. China North Industries Group developed AI video generation and delivery systems.

Each line is a claim. Taiwan is party to the dispute. The nine-page English PDF prints no date, and the bureau's site lists the release as 11 January 2026. Row CN-6 grades C4 with T 49, Weak, and D 27. The record holds no PRC response.

One independent item touches that ecosystem. The Vanderbilt Wicked Problems Lab, in its September 2025 work on GoLaxy documents, describes a state-aligned firm that builds precision profiles and runs AI-driven propaganda. Vanderbilt's news piece of 25 August 2025 says the documents run to nearly 400 pages and show profiles of thousands of US political figures, congressional leaders among them. Press coverage had put forward counts of more than 2,000 figures and at least 117 members of Congress, and neither Vanderbilt page carries them. The release page shows no date and lists two New York Times items from early August 2025. Vanderbilt's content server hands out a document set of 399 pages and 112.9 MB, and all 399 pages are now read. Papers inside date from November 2020 to March 2024. The slide behind the press figures sits on page 7. Its heading names political-figure data on M, an abbreviation the set leaves unexpanded, and section content reads it as the United States. One sentence beneath it lists the 117th Congress, 2022 congressional candidates, the establishment, Trump supporters, right-wing forces at 4000+, hundreds of NGO accounts, and other political figures and opinion leaders at 2000+. A sample table of right-wing figures follows. The ordinal 117 names a

Congress and counts no members. Other figures and opinion leaders make up the 2000+ pool. No page in the set carries a count of members of the 117th Congress.

The set holds more than one kind of paper. Company decks present GoLaxy, trading as Zhongke Tianji, as the commercial arm of a Chinese Academy of Sciences institute, and they name its users as party, government and military. One bid result, dated 9 November 2020, names PLA Unit 31040 in a small procurement of an entity-extraction tool. Bid prices and scores print on the page. No proof of tasking follows from it. A 2021 sales sheet lists 17 projects with provincial public-security, regional and institute buyers. Two stand at sale complete, and neither is an influence product. Minutes of 11 May 2023 and 19 October 2023 and a run of open-topic papers do not name GoLaxy as a party. They link to the firm only through the release's grouping, and the report treats them with that caution.

The firm's own words state intent and claim capability. Two versions of a script for visiting leaders differ. The first, on page 25, says parts of the GoPro system "have been deployed" at the State Council's Hong Kong and Macao Affairs Office, the Ministry of State Security, the Joint Staff intelligence bureau and front-line strategic support units. Added support covers opinion guidance on Hong Kong, the Taiwan Strait and the US election. The second, on pages 121 to 124, says the team has no national project or funding yet. A claim of several hundred human-like bot accounts on Twitter, Facebook and Telegram follows, with a platform for automatic batch posting and commenting. The team hopes to work with the Cyberspace Affairs Office, the Propaganda Department, a Military Commission technology body and State Security. Its goal is "East Wind" over West Wind, and it admits a gap to full service of the overseas opinion struggle. The firm pitches each claim, and no evidence comes with either.

A rotated photograph on page 379 shows a status table of seven products with a main-user column. The Cyberspace Administration's office is the user of three, among them bot detection and algorithm-flaw analysis, two at the fielded stage. Two more, at acceptance, go to the Commission's science and technology office: an intelligent propaganda assistance system and an account-cultivation platform. A user printed as "311" takes a sixth. The Taiwan Affairs Office is the user of an election-situation analysis system, still in research. As a vendor's own list, it shows no date and no origin. A 2021 plan on pages 381 to 391 states the GoPro purpose as targeted propaganda overseas and says the core user group was still to be settled in 2022 and 2023.

Other pages show design and some product. A weekly Taiwan opinion report, signed 1 December 2023, reviews the week of 25 November camp by camp ahead of the January 2024 election and counts 634,448 items. The report ends by advising "our side" to mobilize a cyber army for cognitive operations. No addressee and no AI tool is named. An open-topic paper from the institute side asks for 100 digital identities and 1,000 virtual persons on Facebook, Twitter and TikTok, in Taiwan Mandarin and Hokkien. A deck on account nurturing claims to beat platform bot detection on twelve platforms, and one screen shows 3,692 accounts beside a claimed pool of more than 860,000 monitored accounts. For a Hong Kong case, the script picks 3,000-plus persons and 100-plus opinion leaders out of 9.66 million monitored Twitter users and 180,000 Hong Kong users. The requirements specification of 6 March 2024 is an analytic monitoring product, with text analysis, translation, sentiment and clustering, and it asks for no persona or generation feature.

The AI language runs from claimed to specified. Generation and rewriting tools are listed as present modules. A December 2022 paper asks for generated text, images and digital humans. Minutes of October 2023 ask for fuller, more human-like accounts, long-term pre-positioned Taiwan and Lithuania topics, a Twitter spread experiment on a non-political technology topic, and an exploration of large models. One proposal says current capability cannot support opinion warfare at scale. No prompt log and no deepfake of a named person appears. The papers show a firm that sells and specifies, and a Taiwan product delivered to an unnamed "our side." They

Copyright 2026 Treadstone 71 - www.treadstone71.com

show no use against any named person. They leave NSB tasking unconfirmed, since page 379 lists the Cyberspace Administration against defensive tools only and the United Front Work Department and the PLA Cyberspace Force do not appear in it. Row CN-12 stays at B3, Level 1, unscored. Section 16.3 lists the pages by item.

Assessment one: Chinese firms sell AI profiling and influence services to state clients. Likely (55 to <85), Moderate confidence. OpenAI's provincial-client finding carries the judgment. The Vanderbilt release page, which relays GoLaxy's own papers, points the same way and adds no weight. Behind that page, the documents show a firm that names state users in its own scripts and table, one small PLA procurement and no contract for an influence product. State tasking stays unconfirmed.

Assessment two: Chinese party and military bodies direct named firms to run AI influence campaigns against Taiwan. Roughly even chance (45 to <55), Low confidence. The NSB stands alone on the tasking chain, and the one outside item on the ecosystem confirms neither the chain nor its absence. In the documents, a Cyberspace Administration office uses defensive tools and a Military Commission body uses propaganda tools. They name neither of the NSB's other two taskers.

5.4 What the PLA writes

PLA publications state the doctrine in plain terms. A 13 April PLA Daily essay by the PLA Academy of Military Science describes "AI+cognitive attack," in the Chinese phrase «AI 赋能的认知打击，推动全域毁伤从物理域向认知域深度延伸», "AI-enabled cognitive strikes drive all-domain destruction deep from the physical domain into the cognitive domain." Three more items reach the record at second hand. PLA Daily ran a 22 January essay on cognitive electronic warfare. Xinhua's Globe magazine ran a 10 June piece by a National Defense University professor on cognitive-domain operations and «制脑权», "brain control." Xinhua Liaowang carried a 3 August line, «认知域更成为不战而屈人之兵的新战场», "the cognitive domain becomes a new battlefield for subduing the enemy without fighting." The two Xinhua quotes carry the same limit.

The 13 April essay also asserts that the United States and Israel ran AI cognitive offensives against Iran. That assertion is a single sentence. The sentence names no tool, no platform, no date and no source. That same essay cites unnamed press reports that Claude and Maven served an operation it calls «史诗怒火», and that AI working from outdated intelligence caused the Minab school strike. Neither claim carries a source, and the essay shows PLA authors citing Claude by name. The assertion is a PLA-side claim, untested, and it gets its own row under Rule 3. Row CN-5 grades C6 with T 42, Weak, and D 45, Roughly even chance, since we hold no documented null.

The originator of the 14 September Hong Kong Free Press item is an article by Chen Yixin, China's minister of state security, in the magazine China Cyberspace, issue 8, released on 13 September. Sina Finance hosts the text. He lists six risks to AI security, names "Claude Mythos", "GPT-5.5-Cyber" and the open-source tool "Lobster", and says hostile forces abuse deepfake audio and video and AI text to batch-produce political rumors. According to The Record, he did not accuse any of the named systems of being used in attacks on China. On the Hong Kong Free Press page, the item runs as AFP wire copy, posted on 14 September. AFP names the two models and does not name Lobster or use the word deepfake. Its copy says the models pose possible risks to China's information infrastructure. That article is a PRC official's claim, so Rule 1 caps it. Its text shows what Beijing says it fears. As evidence of a hostile operation, it counts for nothing.

Chinese-language and Russian-language Telegram channels in the sample hold commentary only, and no primary material on PRC AI operations appears in them. The foreign-affairs column index of PLA Daily for 6 August to 24 September shows no title on AI or cognition.

Copyright 2026 Treadstone 71 - www.treadstone71.com

The exact phrase 认知域作战 does not occur in the essay, and the report quotes only terms that do. Row CN-4 grades C2 with T 63 and lands in the Adequate band, three points above the edge. The rows prove that PLA authors write about AI-enabled cognitive attack. They prove nothing about practice.

Assessment: the PLA runs AI-enabled cognitive operations today. Roughly even chance (45 to <55), Low confidence. Doctrine and the surveillance record in Section 5.1 point the same direction. No source shows a PLA operation.

Rows AO-CN-01 to 31 in Section 16.3 list the GoLaxy and CVERC statements by page, each at Level 1.

5.5 Capability theft

The largest volume in the record sits here, and Anthropic counted it. Provider counts supply the figures below. No government has audited them. Anthropic's February report counted over 16 million exchanges and about 24,000 fraudulent accounts across three Chinese labs, DeepSeek, Moonshot and MiniMax. Its September report counts more. Anthropic puts Alibaba at over 151 million exchanges, the largest distillation attack it says it has measured, Moonshot at over 23 million and DeepSeek at over 12.1 million. For Zhipu, branded outside China as Z.ai, it counts 273 fraudulent accounts over ten days and 770,609 exchanges through the cleaning step. Anthropic reports that Zhipu captured Claude's reasoning traces, replayed the traces through Claude to clean them, and trained its GLM models on the result. The report adds that Zhipu targeted Claude Opus 4.6 mostly to grade the responses of another US frontier model. SenseTime and others appear in it as well. On 8 September the NSA, CISA and FBI issued advisory AA26-251A on industrial-scale distillation by China-based AI companies. The advisory names six firms: DeepSeek, Moonshot AI, Alibaba Group, MiniMax, StepFun and Z.AI. Activity runs back to at least late 2024. Volume appears only as billions of tokens across millions of exchanges, with no count for any firm and no 151 million figure. The agencies judge the activity likely with Chinese government awareness, and the text states no basis. NSA's press release names no company, gives no volume and makes no awareness statement. The release says the operations spread across providers to avoid single-point detection.

Asked about the advisory by AFP on 9 September, the foreign ministry spokesperson gave a generic reply. The ministry's page, timed 17:24 that day, is an excerpt of five answers, and carries the reply in the two languages. She said Chinese AI rests on self-reliance and open cooperation, and in the Chinese words «希望美方…不要对中国进行不实指责和抹黑», "we hope the US side will not make false accusations and smears against China." The reply names no advisory, no company and no distillation. Nor does the reply rebut any technical claim. The rejection is an adversary assertion. Row CN-8 grades C5, since the reply gives nothing to test. The question dates the advisory to 8 September, which independently matches the NSA release.

OpenAI's post of 30 September, "Disrupting a coordinated model-distillation campaign," attributes a core cluster to individuals associated with Moonshot AI, with activity from 1 July, spikes on 24 and 25 July of 16,000 requests from more than 4,000 users, and more than 15,000 users disrupted by 28 July. Those are OpenAI's counts, from a provider. No ledger row carries the post, and no judgment rests on it.

Distillation matters to the cognitive frame because the stolen capability returns as a product. A distilled model carries the capability of the US original and the rules of the lab that trained the copy. NIST's CAISI found in September 2025 that DeepSeek R1-0528 followed hijacking instructions 12 times as often as US models and answered 94 percent of jailbreak requests. The CNAS "Red Lines" report of 12 June says Chinese AI systems "enable more effective censorship, surveillance, influence campaigns, and espionage" in the cognitive domain. A chapter of the Estonian Foreign Intelligence Service report of 10 February, titled "Chinese artificial intelligence distorts perceptions," calls DeepSeek "a channel for promoting Chinese propaganda in the West." Row CN-9

combines the three at T 72. The CAISI and CNAS figures stay unchecked against primary text. CAISI's evaluation of DeepSeek V4 Pro on 1 May 2026 has nothing on narratives or hijacking.

Assessment one: Chinese labs distilled US frontier models at industrial scale through fraudulent accounts. Likely (55 to <85), Moderate confidence, at the top of the band. Anthropic and a US government advisory point the same way. The advisory lists Anthropic and Google among its references, so independence stays uncertain. Row CN-7 grades B2, because the advisory gives its awareness judgment no stated basis and cites Anthropic and Google. Its T-Score of 64, Adequate, keeps the print at Likely. ATCRI ranks model theft and distillation third of ten (raw risk 432, confidence 87).

Assessment two: the Chinese government knew of the distillation. Roughly even chance (45 to <55), Low confidence. The advisory says the activity was likely with Chinese government awareness and shows no source, evidence or confidence level for the line. Beijing's reply addresses no claim. The record holds no independent proof of awareness.

For Treadstone 71, distillation raises China's AI Tradecraft Adoption score more than any influence item in the record. Treadstone 71 should also add model-provenance questions to its Sovereign AI Architecture Audit casework, because a client that fine-tunes a Chinese-built open-weight model inherits the behavior CAISI, CNAS and the Estonian service describe.

5.6 The espionage line

Anthropic's November 2025 report on an AI-orchestrated espionage campaign assessed with high confidence that a Chinese state-sponsored group manipulated Claude Code into trying to infiltrate roughly 30 targets: large technology companies, financial institutions, chemical manufacturers and government agencies. Few succeeded. The September 2026 report describes that campaign as suspected state-sponsored.

A second case, GTG-10007, sits in the September report. The two differ. GTG-10007 is an exploit foundry run by Chinese-speaking operators likely in Changsha, Hunan. Two are identified as undergraduate students. Several used Claude as the engineering and orchestration layer of an offensive program, with reconnaissance against foreign government networks in the Middle East, Europe and Southeast Asia. Anthropic counts roughly fifty targeted organizations and says hands-on intrusions fell exclusively on victims in China. The report calls the program a sustained espionage operation and names no state organ. Reuters, relaying the report, repeats the Changsha location, the two Hunan university students, the figure of about 50 organizations that includes foreign government networks, and the workflows run with limited human supervision. The detail that hands-on intrusions fell on victims in China rests on Anthropic alone.

Google's tracker of 8 September (EN-3) adds a different fact. UNC6508 is a PRC-nexus actor that targets academic, medical and military research institutions, with proprietary AI research as a focus. The tracker says Google has also observed suspected UNC6508 activity that compromises cloud environments to deploy local LLM infrastructure. Google gives the reason as avoiding commercial AI API monitoring while using the victim's compute. The lead sentence carries the word suspected. Two sentences follow without it: the avoidance reason, and a line that the group continues to research how to set up and use AI tools, open models run locally among them. The research line plausibly rests on the actor's queries to Gemini. Every other mention relays the tracker or sits on a vendor profile page. Its text names no model, no victim count, no date and no confidence beyond the word suspected, and it does not say which of Google's three telemetry sources saw the activity. Google's report of 15 June 2026 on the same actor's medical-research campaign carries no AI-model finding. So it cannot support the claim. No lab sees that use on its own platform. The case is thin. One Chinese actor on self-hosted models is

the single documented case of its kind among the non-provider rows in the ledger. That case carries more weight in the outlook of Section 12 than any other row.

Assessment: a Chinese state-linked intrusion actor hosts its own open-weight model inside victim clouds. Likely (55 to <85), Low confidence. Row EN-3 grades B2 and scores T 68. One reporter holds the evidence and states no basis. So confidence steps down once. The wording "suspected" leaves the credibility of the specific claim at 3.

Cyber operations enter the report here because they feed decisions. Stolen government and corporate material tells Beijing what officials plan. What officials plan shapes what Beijing chooses to say to them.

Assessment: PRC state-sponsored actors use AI to orchestrate intrusion campaigns at scale, with limited success. Roughly even chance (45 to <55), Low confidence. Anthropic alone reports the two cases, so confidence steps down once, from Moderate to Low. When the November 2025 case came out, researchers questioned its autonomy claim and noted that Anthropic published no indicators of compromise (BleepingComputer, 14 November 2025). The state attribution belongs to that case, and Anthropic's own later wording softens it. Row CN-10 grades B2 at T 60 and D 25, and the T score is 59.75 before rounding. The row blends the November 2025 campaign and GTG-10007, which Rule 8 forbids, so a split into CN-10a and CN-10b, each scored apart, is pending. Mandiant adds a counterweight in M-Trends 2026, and Section 7.5 quotes it.

5.7 Domestic rules beside foreign operations

Several Chinese state items sit outside the ledger because they describe domestic regulation and no operations. China's cyberspace regulator released its AI Safety Governance Framework 3.0 on 14 September. Mandatory labels for AI-generated content have applied since 1 September 2025, and China Daily reports that on 28 April the CAC summoned ByteDance over labeling failures on CapCut, Catbox and Dreamina. The regulator's September notices report 124 generative AI services filed in July and August, 1,112 filed in all, and a clean-up campaign that has removed 5.61 million items and handled 49,000 accounts. Those are CAC's own counts. Five September CAC notices mention no foreign AI-enabled influence. CVERC's report of September 2026 adds detail on the same regime. Four departments issued labeling rules in March 2025, in force on 1 September 2025, backed by a mandatory national standard of February 2025. The regulator announced app takedowns and summons over missing labels in November 2025. In force since 1 January 2026, the amended Cybersecurity Law adds a clause on AI ethics, risk monitoring and safety supervision, and raises fines to as much as 10 million yuan for the gravest harm to critical infrastructure. The same report urges a channel with the United States and Europe for mutual recognition of deepfake labels and cross-border AI incident reporting. That is a Chinese state center's statement of intent, with no counterpart action on record. A joint white paper from CVERC, Tianjin University, Tianjin Normal University and a Tianjin firm tests eight Chinese large models on 38 datasets and 668,820 items. The paper runs 177 pages with no author, date or version. Its agent tests run four coding agents, Claude Code among them, on two Chinese models only, so the scores do not rank those agents. The political-values axis holds 4.67 percent of the items. We use it for existence and structure only. China requires labels on AI content at home while its security organs use AI to watch the same population. Two facts sit in the record. The record does not reconcile them.

5.8 Adversary reaction read (ACS)

Fed the trigger "public advisory naming labs for distillation and a vendor takedown of bureau accounts," the Adversarial Cognitive Simulator returns the read below for a Chinese security and intelligence apparatus.

Worldview. To the apparatus the cognitive domain is a battlefield where victory comes without fighting. The apparatus treats foreign AI as a tool to borrow and a threat to answer. Fears: a cut-off from Western models,

Copyright 2026 Treadstone 71 - www.treadstone71.com

sanctions on named labs, and exposure of bureau staff. Biases: confidence in scale and in state coordination of private firms, and a habit of reading US attribution as political theater.

Likely actions: deny through the foreign ministry, counter-accuse (the ministry already calls the United States "the leading source of instability in cyberspace"), reroute access through gray-market intermediaries, and shift bureau workloads to domestic models such as DeepSeek and Qwen. CARM pathways: deny, accuse back, reroute. Miscalculation risk: a domestic model that leaks bureau data, and a distilled model that inherits safeguards too weak for a sensitive state workload.

5.9 Two rows the record leaves thin

China as a target of chatbot work. No source in the record describes a PRC operation that poisons chatbots or retrieval. The sources on hand are the Graphika article, the OpenAI June report, the Meta report and the Estonian chapter, and none targets the question. So the row stays open and is no documented null. The Estonian service goes the other way and says Chinese models distort perceptions among their users. That is a government assessment of a channel and no test result.

China and local models. One case carries the finding: Google's suspected UNC6508 activity. Anthropic's report holds a second case. In GTG-17002 a China-based defense researcher built electronic-warfare software with Claude, retargeted the default scenario to 12 sites in Taiwan, and ran a self-hosted model on an internal network beside Claude. Reuters, relaying Anthropic, says account information tied the actor to Chinese research institutions, among them the Academy of Military Sciences. OpenAI's pasted-report claim is a third item and stays unconfirmed. The counterweight is large. OpenAI's two June operations used ChatGPT. Anthropic's GTG-10007 used Claude. Google's BASIN CASTLE, RAVINE CASTLE and an unnamed pentest-framework group used Gemini. The record holds no report of a PLA or Ministry of State Security actor running Qwen or DeepSeek locally. Open-weight Chinese models with refusals removed circulate on Hugging Face, and Chinese developer forums discuss local deployment and gray-market API relays. That shows availability, not use. Two Chinese-language searches, on 本地大模型 and on self-built models, returned relays of Anthropic's report and defender write-ups, and no report of an actor's own model. The actor names, 本地部署 DeepSeek and GoLaxy self-hosting sit outside that sample.

6 North Korea, hired hands and proxies

The record shows North Korea running deception operations. AI makes the deception cheaper. The target is a hiring manager, a software maintainer or a researcher at a desk. Around North Korea sits a market of contractors, vendors and single subscribers who sell the same tools to states that cannot build them. Proxy movements stay almost silent.

6.1 Persona factories

Microsoft's 6 March report names Jasper Sleet and Coral Sleet, the second formerly tracked as Storm-1877, as North Korean actors behind fraudulent remote IT-worker employment. Microsoft finds that the actors use generative AI to produce culturally matched name lists and email address formats, pull requirements from job postings and tailor fake identities to the role. Jasper Sleet used the Faceswap application to place workers' faces on stolen identity documents and reused one AI-generated profile photo across several personas. The actors ran voice-changing software in interviews to hide accents. AI wrote their resumes and letters. Microsoft adds that Coral Sleet shows rapid capability growth from AI-assisted iterative development of malware components, and it calls agentic use early experimentation. No local-model claim for any North Korean actor appears in the post.

CrowdStrike's Global Threat Report of 24 February 2026 says FAMOUS CHOLLIMA used several AI tools across its fake-employment work. The report prints an 89 percent rise in attacks credited to adversaries it labels AI-enabled, and nowhere does it define the label or give a count to divide by. Row EN-14 scores that report alone and overlaps KP-2, so the two rows count once. Section 1.6 lists the definition problem as assumption A4. Okta's 11 February report, "The North Korean on your payroll," covers the same scheme from the employer's side, and an undated Okta post covers generative AI services behind North Korean IT-contracting scams. Okta rests its attribution on US authorities and third parties.

The human network sits behind it. US courts sentenced facilitators on 20 March, 15 April and 6 May. The three releases show no AI element. A 31 July alert from eleven governments, published by the State Department, addresses the IT-worker scheme. Its wording on AI and real-time deepfakes is weaker than the press accounts of it, and the State text governs. South Korea's intelligence service put out a white paper on 8 June that, per an Asia Today article, covers North Korean use of AI in cyber operations. No text of the white paper itself is in the record. The source is a party to the conflict. That item is a statement. 38 North's 27 February analysis, by an outside analyst, reaches a similar view across the cyber, economic and military domains.

Assessment: North Korea uses AI to scale fake identities that pass hiring screens. Likely (55 to <85), Moderate confidence. Rows KP-1 and KP-2 grade B2, at T 77 and T 61. KP-2 sits just over the Weak line, and a point lower moves the call to Low confidence. The ATCRI ranks AI-built personas and insider access eighth of ten (raw risk 324, confidence 75).

The audience here is a person. Enterprise hiring is the terrain. A firm that uses video interviews and resume screening has already agreed to the contest, and the adversary uses its own AI for the response.

The finding lifts North Korea on the AI Tradecraft Adoption sub-index of TAI and gives the KP track of the Cognitive Warfare Monitor a standing narrative, Insider by Persona, with Microsoft and Okta as the origin grades.

6.2 Kimsuky builds decoys with an agent and leaves local tooling behind

Genians reported on 7 September that Kimsuky used the AI agent "opencode" to create decoy documents. Thirteen LNK files dated 11 to 19 August 2026 carried 29 decoys, 11 of them unique, Genians found. PDF metadata lists the creator as "opencode," the author as "anonymous" and a "(temporary value)" placeholder. The campaign used GitHub personal-access-token LNK attacks.

A second Genians report, dated 10 August, holds the local-model material. The report lists traces on an actor host: Ollama, a GPT4All LocalDocs installer and database, Msty Studio, and the LLamaSharp package with CUDA support, together with Microsoft Semantic Kernel packages and a faster-whisper archive. Cursor AI also appears, used to draft a requirements file and a GitHub command-and-control test. Genians names no model weights, shows no model output and finds no evidence of training. "Local" here means installed tooling and databases, not a model whose behavior anyone has seen. Key logs reconstruct under the South Korean Dubeolsik layout, while retained spellings ("싸이트," "로출되였는지") read as North Korean. Genians notes that actors are able to use the South Korean layout. The layout settles nothing. An Astrill VPN and the Arirang manufacturer appear among clues the report calls insufficient. Genians reports a workflow in which Korean text went through Google Translate into ChatGPT prompts, and it was unable to confirm that a paste into an AI service took place. If that path ran, it ran through the cloud. Genians assesses the actor as "in the phase of accumulating technologies and capabilities to integrate AI across its overall attack operations."

Keyboard layout and spellings conflict. Nobody has resolved the conflict. The two reports must not be merged: the 7 September report holds the opencode metadata and no keyboard, spelling or local-model analysis, and the

10 August report holds the rest. Genians' attribution to Kimsuky and to Operation GitPower rests on a public builder, header patterns from a February 2024 report and a Fortinet report of April 2026. That call is an assessment. No confidence word attaches to it.

Assessment one: Kimsuky uses AI agents to produce lure documents at speed. Likely (55 to <85), Moderate confidence. Row KP-3 grades B2, T 72.

Assessment two: Kimsuky operators work with local model tooling that no commercial lab sees. Likely (55 to <85), Low confidence. Only Genians saw the environments, so confidence steps down once. The stack has not been shown running against a victim, and no second source reports it. Press relays of 10 to 12 August add no telemetry, and no S2W, NSHC or KISA text appears in the record. Row KP-3 grades B2 at T 72, with credibility 2 for the artifacts and 3 for the inference about the operator. That row covers the 7 September page. The 10 August page, which carries the local-tooling finding, has no ledger row of its own. Evidence stops at tooling and databases, with no weights or output. Of the three, the reading "Kimsuky runs its own model" has the least support. Other vendors (Microsoft, CrowdStrike, Google's 11 May report, which shows APT45 sending thousands of prompts to hosted models, and AhnLab's August roundups of 18 and 30 September) report no North Korean self-hosted model. The ledger keeps the local flag for KP-3 because Genians observed the tooling. Section 12.16 shows what the headline does if the flag goes.

The CopyCop finding in Section 4.5 sits at a firmer evidence level. Recorded Future reports model artifacts and a script on gated hosts, and reads them as a self-hosted stack. Genians reports tooling and databases on a Kimsuky host and no model. Neither is visible to the labs. A US Treasury account of a GRU-funded AI server in Moscow adds partial outside support on the Russian side. Nothing outside Genians backs it.

6.3 Money and enterprise AI as a target

Chainalysis reports that $387 million was stolen from Bitget on 24 September. The firm attributes the theft to North Korea and says it pushes North Korean crypto theft in 2026 past $1 billion. Chainalysis used its own AI tooling to trace the funds. That counts as defender AI. The firm claims no North Korean AI use in the theft. A second Chainalysis item reports blockchain dead-drop attacks up 420 percent in twelve months, with Iran and North Korea among the leading states, and the firm infers a link to AI coding tools. The inference belongs to the vendor.

CrowdStrike's 3 August Threat Hunting Report says the North Korea-nexus actor STARDUST CHOLLIMA injected a malicious npm package into at least 131 trusted Mastra AI framework packages in June. The report puts npm at 87 percent of first-half 2026 software-registry threats. CrowdStrike's blog states no basis and no confidence level for the Mastra attribution. Microsoft's 17 June post on the same compromise assesses Sapphire Sleet with high confidence, on infrastructure and post-compromise technique overlap, and counts 140 or more packages. The package counts differ by source (131, 140 or more, 144 in some press), and we report the difference without reconciling it. Each vendor uses a different actor name, and neither states the overlap in its text. Public output from the FBI, CISA, the NCSC and ASD carries nothing on Mastra. The null proves nothing about what the agencies hold.

An AI framework package is enterprise AI used as an entry point. Section 7 takes the pattern up with LiteLLM and the MCP finding.

None of the three items shows North Korea using AI against a population. They show it using the AI economy as a place to steal and to hide.

Copyright 2026 Treadstone 71 - www.treadstone71.com

6.4 What Pyongyang says

The sample holds the Rodong Sinmun date pages from 1 January to 3 October, more than 276 pages checked by title, and KCNA's front pages for 1 to 7 October, 33 in all. A KCNA Watch snapshot of 7 October shows 152 articles from the last day and 1,361,217 in the archive, and none of its five newest items mentions AI or cyber topics. Every item below is a statement by the regime's own press.

Rodong of 5 June, page 4, calls AI one of the core industries of the Ninth Party Congress and a driver of the fourth industrial revolution. The framing is civil and economic. NK Insider reads that item as a turn to military AI, and the reading is its own inference. A Rodong explainer titled "The Brain of Modern Industry: Artificial Intelligence Technology" ran on 19 July, page 5. In Rodong of 17 September, a KCNA report covers a national AI conference of 15 and 16 September, with lectures on large language models. On 19 September, page 2, Kim Jong Un calls for raising the AI level of offensive and defensive combat weapon systems. The text confirms NK News's relay and concerns weapon systems, not factory processes. NK News frames the same visit as a push for AI and automation upgrades at munitions plants. Its report dates the inspection to 16 to 18 September, places the plants under the Second Economic Commission, which oversees defense manufacturing, and reads some photos as 600 mm rocket launchers. NK News also notes an inspection of suicide drones. The two readings overlap and differ in emphasis. KCNA of 3 October reports a strategic weapon launching drill in a statement attributed to Kim Yo Jong. The statement says the missile is able to change its flight orbit at low altitude and that AI was introduced into it. KCNA owns the claim, and nothing outside it verifies the claim. Only that item among the 33 front pages mentions AI.

The sample holds no 2026 AI statement from North Korean agencies. A KCNA item of 19 September 2025 carries the title "Respected Comrade Kim Jong Un Guides Work of Unmanned Aeronautical Technology Complex" and states a program the regime says it runs. The Rodong pages for 18 and 19 September 2025 carry no matching title. That statement is an adversary claim. Row KP-6 grades C3 and carries no weight on capability. The 2026 items show how the regime frames AI, as an industrial and military priority, and show nothing about its capability. They carry no ledger rows. The sample holds selected Rodong bodies only, and Voice of Korea, Naenara and DPRK Today sit outside it.

Silence in the open source says nothing about activity. The null result measures the sample. What Pyongyang chooses to publish enters the measure too.

Rows AO-KP-01 to 03 and AO-PX-01 to 11 in Section 16.3 list the Korean-language pages and the proxy statements read in the original.

6.5 The market for hired hands

Anthropic's September report describes influence and surveillance programs that single people or small firms ran for others. Nine influence cases came from Russia, Iran, Turkey, the Gulf, South Asia, Africa and Europe, and the actors included governments, state media, private firms that sell influence, domestic political operators and one opposition movement in exile.

Four cases show the market. GTG-54006 is a single actor in Gaibandha District, Bangladesh, who rotated through 29 Claude accounts and drove a custom program named "fake_news_3.py" that generated fixed batches of 15 headlines, three detailed fabricated stories and 15 image prompts for a pro-Awami League operation aimed at rural Bangladesh. Anthropic found no proof that the party itself directed or funded the network. In a second case, GTG-84002, Anthropic describes a UAE-directed influence operation that targeted the Muslim Brotherhood, the Sudan conflict and UN accountability mechanisms, and says it linked the work with high confidence to UAE government officials. A likely Bamako-based independent consultant, tracked as GTG-50027,

Copyright 2026 Treadstone 71 - www.treadstone71.com

used Claude as the main engineering workforce for "Lakana 360," a population-scale surveillance platform for Mali's state intelligence service that monitors roughly 25 million SIM cards on all three national mobile operators. Anthropic states that the platform ran fully on premises using local models, and that its account bans did not touch the deployed product. The fourth case, GTG-54009, involves an entity named S2T Unlocking Cyberspace, which open-source research suggests is an Israeli-Singaporean commercial intelligence vendor, profiling social media accounts of Iranian and Persian Gulf-based users.

One consultant built an interception platform. A lone actor ran a fake-news plant. The price of a program dropped to a subscription.

Assessment: AI lets a single subscriber or a small vendor deliver programs that once needed a staffed unit. States with thin capacity buy them. Roughly even chance (45 to <55), Low confidence. Anthropic alone reports the programs, so the single-source rule drops confidence one step. Row PX-1 grades B2, T 72. Anthropic's attribution of the UAE direction carries the vendor's high confidence, and we did not test it.

Proxy movements stay quiet in the open source. The evidence is thin. Public Telegram windows in the sample hold 100 recent posts each for Al-Manar English, Al-Masirah English and a Hezbollah-aligned channel, 80 for Al-Ahed, 89 for Palestine Today and 62 for Quds News Network, with keyword searches of their history. Eight more windows cover Al-Manar's Arabic channel (771 posts), Al-Masirah's satellite channel (600 posts scanned for keywords, back to 5 October), the channel of the Houthi-run Saba agency in Sanaa (668 posts), the channel of the Aden government's Saba agency (100 posts), three channels that call themselves Qassam's official media (97 to 99 posts each) and Al-Aqsa TV's channel, whose newest post dates from May 2023. The windows hold post text only. No post text announces AI-generated media, and none claims to have made content with AI. The AI items relay or discuss other parties' AI. Among them, Al-Masirah English carried a claim of its own on 16 November 2025 that a spy network used advanced AI, and we did not test it. Al-Masirah's satellite channel and Sanaa Saba accuse Saudi-aligned media of false narratives in the window and name no AI. The 600-post windows cover 2.5 to 6.5 days. One Qassam-named channel reaches farthest, back to 18 February at 97 posts. A window says nothing beyond itself. One Quds News Network post relays a +972 Magazine claim that Israel funds networks to shape AI chatbot answers on Gaza. The post is secondhand. That claim is an adversary assertion of unknown merit. An IDF statement of 1 October rebuts a documentary on AI in targeting and shows nothing about proxy media.

The Qassam Brigades' own Telegram page (8,261 subscribers) offers a stub and no public preview. Three channels that claim to be its official media carry the same posts and a recurring notice, which the channels themselves state, that fake pages and accounts impersonate Qassam. Which channel is official stays unknown. A Qassam-labeled clip, synthetic or not, proves nothing about who made it. Home pages for Al-Masirah (masirahtv.net, where almasirah.net.ye redirects) and Sanaa Saba (saba.ye) form one text snapshot taken on 7 and 8 October. The text carries no AI claim and no AI-made media. Saba's one AI term sits in a market wire on a falling Nikkei index. Each page runs Houthi-side accusations against the Saudi side, a Saudi media campaign among them, and none names a technology. A deputy information minister says Yemeni strength defeated a Saudi military, media and political attack. The Houthi leader's speech of 7 October, on the third anniversary of the 2023 attack, calls on Turkey and Pakistan to resist Saudi plots, says Saudi Arabia works under the American umbrella and alleges a secret Saudi-Israeli alliance. No AI appears in it. The pages also show an open Saudi-Houthi confrontation, with Riyadh and Abha strike claims, which the report does not otherwise cover. Those are adversary claims and stay untested. The Aden agency's sabanew.net held 68 headlines in the snapshot and carried no Yemeni AI actor. That agency is the adversary-side body for Houthi purposes, and its headlines are statements by the Aden government and the Saudi-led coalition.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Four items are statements by proxy-aligned or interested parties. None is verified fact about reality. An Al-Manar item is Naim Qassem, the Hezbollah secretary general, speaking to the Al-Ahed news website on 23 June 2025 and telling it to keep pace with modern technology and AI. A Hamas-adjacent outlet, Shehab, promoted an AI-produced film on 7 October 2026 and named no producer. Arabi21 (29 December 2025) described AI-produced scenes in Hamas material, and RT Arabic called a Hezbollah video AI-made on a date the record does not confirm. RT is a Russian state outlet and carries low weight. Independent fact-checkers supply the firmer items: AFP Fact Check Arabic (20 May 2026) on an AI image in the regional conflict, Kashif (24 November 2025) on an AI image of the Qassam spokesman, and Arij (3 to 4 March 2026) on a video a detector scored 94.6 percent likely AI-made, a probabilistic score and not proof. Arij and Raseef22 (21 August 2026) describe a coordinated, AI-enabled campaign of fabricated videos aimed at Hezbollah, operator unknown. That campaign targets a proxy and is not a proxy product. Houthi fabrication allegations in our set come from the Yemeni government side and Saudi-aligned press. CNN Arabic and DW show recirculated old footage on pro-Houthi accounts, not AI and not leadership direction.

Anthropic's report adds one Houthi-region case, GTG-87001: a cell in northern Yemen used Claude Code in place of software engineers on guidance and control software for three weapons programs. Anthropic reports a failed field test of a guided rocket, and no evidence of an operational device. The report also states that the actors had already built an offline simulation toolkit that does not rely on Claude. A toolkit is no self-hosted model. The record holds no Houthi source on any of it.

Row PX-2 stays at F6. Fourteen Telegram windows holding 62 to 771 posts each, and the two Houthi home pages as one-day text snapshots make a thin sample. Text alone enters the sample, Qassam authorship stays unproven, Al-Aqsa's channel is stale and four websites sit outside it. The secondary items above are outlet characterizations that we do not score. That row records a gap in the sample. Nothing about activity follows from it.

One more Iranian-linked front sits here. Check Point's 2021 report puts the Moses Staff channel at t.me/moses_staff. A Handala leak-site post of 7 March 2026, relayed by CTIWatch, gives t.me/HANDALA_HPR2. Telegram offers no public preview for either, so the record holds no content. No handle for Abraham's Ax appears. KELA and Outpost24 say Handala channels were shut and reopened under new names. Group-IB and SOCRadar give no handles.

6.6 STEMPLES Plus across the four states

The STEMPLES Plus frame reads each state's environment through twelve factors. Each cell below comes from the record in Sections 3 to 6. A cell marked "not observed" means the record holds nothing, which measures the record as much as the state.

Table 7. 6.6 STEMPLES Plus across the four states

Factor	Iran	Russia	China	North Korea
Social	State propaganda bodies and paramilitary units adopt a Western tool despite the access block	Contractors and state-media staff share the workflow	Bureau staff share one prompt formula through an internal manual	Facilitator networks in the US and an IT-worker diaspora
Technical	Claude as engineering department; a Firefox harvester shipped to production	One network, CopyCop or Storm-1516: a self-hosted inference stack on Llama 3 models, per one vendor, and Grok Imagine video in one campaign	Claude Code with custom skills; gray-market distillation; a suspected local open-weight model in victim clouds (UNC6508)	Local model tooling on one host (Ollama, GPT4All, Msty); opencode; Cursor

Copyright 2026 Treadstone 71 - www.treadstone71.com

Factor	Iran	Russia	China	North Korea
Economic	Paid amplification through IRGC-affiliated channels	Moscow spent about $150 million on the Moldova influence campaign, per the Estonian FIS (an agency statement)	Private firms sell influence to provincial clients	Crypto theft above $1 billion in 2026
Military	IRGC messaging and paramilitary units	Storm-1516 and APT44, which researchers tie to the GRU	PLA writing on AI-enabled cognitive attack	Raising the AI level of combat weapon systems, per Rodong of 19 September (statement); AI in a missile, per KCNA (claim); vendor-named clusters
Political	"Soft war" and "cognitive war" in regime language	Sovereign and national model law from March 2027	AI Safety Governance Framework 3.0	AI framed as an industrial and military priority in state media (statement)
Legal	Claude access blocked inside Iran	Sovereign-model rules ahead	Mandatory AI content labels since 1 September 2025	US sentencing of facilitators; 31 July alert from eleven governments
Educational	Not observed	Not observed	Police academy student among actors; NDU professor writes on "brain control"	Not observed
Security (internal)	Surveillance units take taskings from officials	Telegram monitoring bots (Google only); Max neuroprofile ad tests	Stability maintenance and sentiment monitoring	Not observed
Religion	Institutions named for Islam; religious frame for propaganda	Not observed	Catholic, Tibetan Buddhist, Falun Gong and Taiwanese Christian communities as targets	Not observed
Demographics	Persian-speaking domestic and diaspora audiences	Moldovan, Latin American, African and EU audiences	Diaspora across Asia; Tibetan and Uyghur communities	Western remote-hiring markets
Social psychology	"Manufactured grassroots" hashtags such as #IranStands	Forged outlet branding as a trust lever	Stability-maintenance vocabulary and fear of instability	Interview trust, defeated by voice changers
Strategic	Hold the regime narrative at home and abroad	Shape chatbots and doubt about AI investment	Acquire capability and control the cognitive domain	Raise revenue and gain insider access

The table shows three patterns. China and Iran show the densest internal-security entries, and each relies on bureaucratic prompt practice. Russia shows the densest information-ecosystem entries, with chatbots and forged brands as terrain. North Korea leaves three of twelve cells empty and fills the rest from remote-hiring, revenue and state-media evidence, the last of which shows intent and no capability. A client that models all four states with one profile loses the differences that decide the defense.

6.7 Adversary reaction read (ACS)

The Adversarial Cognitive Simulator, aimed at a North Korean cyber and revenue apparatus with the trigger "exposure of a persona network and enforcement against facilitators," returns the read below.

Worldview. Revenue and access come first. Influence work ranks low. Fears: loss of hiring pipelines, seizure of facilitator laptops, and exposure of identity stock. Biases: reliance on volume of applications, and confidence that employers will not check. Likely actions: rotate personas and facilitators, move more work to local models,

and shift from IT-worker roles to contractor and open-source maintainer roles. CARM pathways: relocate, conceal, replace. Miscalculation risk: reused AI-generated photos across personas, as Microsoft documents, link identities that the actor wants kept apart.

Most-Likely Achieved Endgame: persona fraud keeps scaling through AI-built identities while the actor adds local models for part of its tooling and stays partly below the lab record. Likely (55 to <85), F-Score 62. Alternatives: an AI-enabled social engineering breakout against a major open-source maintainer, Unlikely (15 to <45); a collapse of the facilitator network under prosecution, Unlikely (15 to <45).

6.8 Two North Korea rows the record leaves thin

Chatbots as audience. English and Korean searches for North Korean targeting of chatbots or their training data return no 2026 source that shows it. The Russian Pravda-network pattern in Section 4 does not appear in any North Korean material in the record. That row stays a documented null at 4/6/10. The number is a prior, not a finding.

Local and self-hosted models. The best evidence is the Genians tooling traces in Section 6.2, at trace level. Microsoft, CrowdStrike and Google report hosted use only. That row stays at the evidence level of installed tooling, with no weights, output or training shown. Nobody has shown the tooling turned on a victim. Genians alone reports it, and press relays supply nothing more. Repository and model results on Kimsuky, GitPower and Ollama terms show no item tied to the cluster, and Korean press of 9 to 12 August restated the Genians report and brought no analysis of its own. The account names in the Genians 7 September report are not in the record, so no account-level check exists.

7 Enterprise AI as audience, target and channel

A firm that runs AI systems holds three exposures at once. Adversaries read its chatbots and retrieval layers as an audience, its keys, models and accounts as targets, and its assistants and agents as a channel for instructions. Where one team owns only one of the three, the other two sit open.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 8. 7 Enterprise AI as audience, target and channel

Role	Exposure	Strongest evidence	Best band
Target	LLM gateway and provider keys	TeamPCP and the LiteLLM release (EN-1); Nx and Salesloft precedents (EN-6)	Criminal compromise is documented (FBI FLASH). The ATCRI places the threat first of ten for defender risk, and no band applies to what adversaries value
Target	Who stands behind AI-stack compromise	TeamPCP is criminal (EN-1). Microsoft assesses Sapphire Sleet with high confidence and CrowdStrike names STARDUST CHOLLIMA for the Mastra injection (KP-2). The NCSC links about 75 percent of UK critical infrastructure incidents to states and does not isolate AI (GV-2)	Likely (55 to <85), Moderate confidence, for a state-nexus actor behind the Mastra injection
Target	AI accounts, model logic, research	GTIG on stolen accounts, distillation and extortion (EN-3); Anthropic GTG-50020 (EN-12)	Likely (55 to <85)
Channel	Assistants and agents	GTIG on DUSTMAKER (EN-3); MCP research (EN-2); NSA note (GV-6)	Likely (55 to <85), Moderate confidence, for GTIG's account that one criminal group built assistant abuse into its malware, scale unstated. Roughly even chance (45 to <55), Low confidence, for exploitation of agent and MCP flaws at scale, since no public proof of use exists
Channel	Agent-run intrusion	GTIG on a six-hour campaign (EN-3); Gambit (EN-5)	Likely (55 to <85)
Audience	Chatbots and retrieval	Pravda network (RU-2, RU-3)	Presence of Pravda content in a training source is documented (RU-2). Probes report retrieval hits under different definitions (Section 4.2). Evidence on effect is absent, and no band applies
Audience	People around an AI project	Firebird (RU-5)	Documented in one case, attributed by Recorded Future (RU-5)
Target	Finance staff and payment approval	Arup deepfake video-call fraud (EN-13)	The 2024 fraud is documented. Present exposure: Roughly even chance (45 to <55), Low confidence, with no 2026 case in the ledger

7.1 The gateway as target

The FBI's 2 July FLASH lists TeamPCP's 2026 compromises of Trivy, KICS, LiteLLM and the Telnyx SDK. TeamPCP stole cloud tokens, SSH keys and Kubernetes secrets. Extortion ran on the side. The FBI calls the group a cyber criminal group and names no state.

Palo Alto Unit 42 is reported to record the LiteLLM step, and no copy of its page text is on hand. Malicious releases 1.82.7 and 1.82.8 went live after attackers used PyPI tokens stolen in the earlier Trivy compromise, and the payload swept environment variables for LLM API keys. LiteLLM sits between a firm and its model providers. One gateway holds many keys. Reports put the exposure at about three hours, and vendor timestamps differ. Neither the duration nor the Wiz estimate that LiteLLM sits in 36 percent of cloud environments traced to a primary page, so each stays a reported figure. Downstream, Mercor's own notice, dated 25 June, confirms that compromised LiteLLM versions hit it in late March. Mercor says it worked with Mandiant and Latacora, and it reports very limited customer impact and no employee data affected. Lapsus$ claimed the breach on its leak site and posted samples. TechCrunch reviewed that sample on 31 March: Slack data, ticketing data and two videos.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Mercor declined to say whether the incident links to Lapsus. Neither the volume nor the Lapsus$ role enters the record.

Google's Threat Intelligence Group tracks the group as UNC6780 and reports a series of large-scale open-source supply chain compromises since March. A federal indictment of Ruben Ian Thomson, reportedly unsealed in the Northern District of California on 27 August over spring TeamPCP attacks, stands in the record as a press report. The record holds no indictment text and no Justice Department page. Anthropic's September report separately says multiple actors compromised AI wrapper services' LiteLLM deployments and used prompt injection to pull production API keys from their cloud containers. That is Anthropic's account of a different path, and it sits apart from the TeamPCP releases.

Two 2025 precedents show the path is old. Nx s1ngularity published malicious packages for four hours after an npm token theft, and the post-install script tried local AI tools, among them Claude and Gemini, then uploaded the results to a public GitHub repository. Salesloft Drift used stolen OAuth tokens from a chat-agent integration to export Salesforce data in bulk. Row EN-6 grades A2, T 81.

Two vendors attribute the June compromise of Mastra AI framework packages to a North Korea-nexus actor (Section 6.3, row KP-2). The mechanism is dependency injection: a malicious npm package, easy-day-js, added to the packages. No cited source describes prompt injection in the case. CrowdStrike's 3 August blog says STARDUST CHOLLIMA injected a malicious npm package as a dependency into at least 131 Mastra packages, and it states no basis. That blog is the primary for the count of 131. The full threat hunting report sits behind a registration form, and the evidence base holds no copy. Microsoft counts 140 or more and cites infrastructure and post-compromise technique overlap at high confidence. The two vendors use different actor names and perhaps share community indicators. They count as two assessments of one cluster, not two proofs. No government attribution for Mastra appears in the FBI IC3 listing, CISA's North Korea advisories page and bulletin search, an NCSC supply-chain blog, ASD's alert listing or CCCS's alert list. That null covers listings and a few pages, not the sites in full. The nearest government text is the joint advisory of 22 September from ASD, the FBI, DC3, Germany's BND and BfV and other agencies on North Korean WaterPlum npm malware. No Mastra package appears in it. The FBI names no state in the TeamPCP cases. Attribution stays vendor-only.

Assessment one. The record documents the compromise itself: the FBI lists LiteLLM among TeamPCP's 2026 compromises, and Mercor confirms a hit. Beyond that case, the judgment is that criminal groups reach AI gateways, provider keys and agent frameworks as a pattern, and that one compromise reaches many keys. Likely (55 to <85), Moderate confidence. Row EN-1 grades Level 3 at T 80, and row EN-6 at A2 and T 81. Each sits at the Strong edge, and the edge rule drops the print by one band. The FBI FLASH carries the call. Unit 42 and Wiz reach the record as reports. Highly likely returns if a second analyst rates EN-1 and EN-6 at 80 or more.

Assessment two: the ATCRI ranks AI supply-chain compromise first of ten threats for defender risk (raw risk 576, confidence 85). The rank follows from ratings we set by hand, and it measures the risk a defender meets. Ledger rows say nothing of what adversaries value, so no claim about their valuation follows. No band applies. Section 7.6 shows the decision.

Assessment three: a state-nexus actor has compromised an AI framework package. Likely (55 to <85), Moderate confidence. Two vendors name North Korean operators, and only Microsoft states a basis, while the shared indicators and the missing government statement hold the call at Moderate. Row KP-2 grades B2, T 61, a point clear of the Weak edge. The Mastra rows take no step-down for a single reporter.

7.2 Accounts, models and research as prizes

Google's 8 September report finds that average underground prices per AI account more than doubled in 2026. Infostealers supply the accounts, among them LUMMAC.V2, STEALC.V2, VIDAR and ACRSTEALER, and demand concentrates on Claude and Gemini credentials and on Cursor Pro and Devin. A stolen account gives an operator model access. No billing trail follows.

Google also reports regular distillation campaigns against its own models, some above 100 million prompts, and Mandiant investigated multiple extortion operations in which actors stole proprietary AI data: models, skills, prompts, source code and research. In June, Google reported UNC6508, a PRC-nexus cluster, on a multi-year campaign against academic, medical and military research institutions in North America. The campaign targets proprietary AI research. Google's June report contains no finding on AI models or local infrastructure. The local-model sentence appears once, in the 8 September tracker: GTIG "suspected" UNC6508 activity that compromised cloud environments to deploy local LLM infrastructure, which, Google reasons, avoids commercial API monitoring. No model, victim, date or confidence figure accompanies the sentence. Corroboration is missing: every other source relays Google, a Cloud Security Alliance research note among them. Nor does the sentence say which of Google's three sources supports it. A basis in Mandiant incident-response visibility is our inference. The page does not say so. Anthropic's September report adds a case that ended in failure for the attacker. In GTG-50020, an actor found one working attack path into customer environments and repeated it against thirty targets. The goal was access to a pre-release Claude model, and the actor used customers' stolen keys to try. Access never came. Every path to the model failed. Anthropic's own systems stayed uncompromised. Anthropic reports the case. So the failure flatters the reporter. Row EN-12 grades B2. The bias controls from Section 1.5 apply.

The cognitive frame enters through capability. A stolen model carries the means to write, translate and impersonate at scale. The earlier sections show who wants those means.

Assessment: adversaries treat proprietary AI assets, meaning accounts, model logic, prompts and research, as intelligence targets. Likely (55 to <85), Moderate confidence. Rows EN-3 and EN-12 grade B2, at T 68 and T 72. Section 5.5 covers the largest single case.

7.3 Assistants and agents as a channel

A channel carries instructions inside. An AI assistant that reads files, mail and project folders carries whatever those files tell it to do.

Google's Threat Intelligence Group describes DUSTMAKER, the credential stealer of the criminal group UNC6780 (TeamPCP), as malware that plants configuration files in hidden project directories. An AI coding assistant then runs arbitrary commands or scripts during routine developer work. In Google's account, the design forces the model to execute commands on the attacker's behalf without the developer's knowledge. GTIG says the functionality sits in DUSTMAKER samples. The report gives no victim count and no tally of assistants that ran the planted instructions. Scale stays unknown. One Mandiant case from the same group also appears. There, UNC6780 created a malicious GitHub Actions workflow for a company's proprietary AI repository and handed access to a separate actor. That actor sent a ransom demand under LAPSUS branding and took a copy of the repository. The page does not tie the case to the assistant functionality. Row EN-3 grades B2, T 68. Google alone describes the case.

The research record shows the same weakness in more places. OX Security's 15 April finding calls an architectural property of the Model Context Protocol a systemic exposure. Anthropic called the behavior expected. The NSA's 20 May security note says some MCP implementations let attacker-controlled input reach

execution environments, and many omit authentication entirely. Joint guidance of 30 April from the NSA, CISA and Five Eyes partners warns that over-privileged agents magnify the damage from a single compromise. Unit 42's Vertex AI "Double Agents" study found default service-agent permissions that let a deployed agent read project storage buckets and pull restricted Google images. Google updated its documentation. Noma Security's GrafanaGhost used indirect prompt injection to bypass Grafana's client-side URL check and its AI guardrails and pull out data. Grafana fixed it after disclosure. A research paper describes EchoLeak, a zero-click prompt injection in Microsoft 365 Copilot (CVE-2025-32711) that needed one crafted email. The OWASP 2026 Top 10 for LLM applications, published 1 September, puts the classes on one list.

Flowise gives the sharpest example of the line between flaw and exploitation. CVE-2025-59528 carries a CVSS score of 10.0, because the CustomMCP node runs user input as JavaScript. Flowise fixed it in version 3.0.6. A VulnCheck researcher's LinkedIn post, which the news cycle dates to 7 April 2026, reports that the firm's Canary honeypot network began detecting first-time exploitation of the flaw. The activity came from one Starlink IP address. No victim, no compromise of a real system and no actor appears in the post. Honeypot hits count as attempts. SecurityWeek quotes the researcher on a flaw public for more than six months. Exposure of 12,000 to 15,000 instances is VulnCheck's attack-surface count of reachable servers, not victims. VulnCheck's blog and its KEV page carry no Flowise post. Headlines fuse exposure and exploitation, and The Hacker News says more than the originator does. The exploitation claim now rests on a vendor's honeypot observation. Nothing in it ties the flaw to AI misuse beyond Flowise being an agent builder, and so it adds nothing on the AI question.

Apart from DUSTMAKER, every item above is research, a standard, guidance, an attempt seen on a honeypot or a claim that carries no weight. None shows exploitation at scale.

Assessment one: one criminal group has built abuse of AI coding assistants through planted configuration files into its malware, as GTIG describes. Likely (55 to <85), Moderate confidence. Google states no scale. The grade for EN-3: B2, T 68.

Assessment two: adversaries exploit agent and MCP permission flaws in the wild at scale. Roughly even chance (45 to <55), Low confidence, because the judgment rests on absence of public proof. Rows EN-2 and EN-10 grade B2, and EN-10 stays at Level 1. The paths work in research. Public proof of use at scale stays absent. The ATCRI ranks agent and MCP exposure fourth of ten (raw risk 392, confidence 70). Open paths drive the likelihood score. Missing proof drives the confidence score.

7.4 Intrusion at the speed of an agent, and fraud by video call

Google reports one actor who used an AI coding chatbot, a prompt and a set of agent instructions. The actor planned, built and ran a mass credential harvest in under six hours.

Gambit Security describes one operator who used Claude Code and GPT-4.1 against nine Mexican government agencies between December 2025 and February 2026, and about 75 percent of remote command execution came from Claude Code, with 20 exploits for 20 CVEs and 1,088 logged prompts. The page date conflicts with its metadata, and a secondary report counts ten organizations plus a bank. Row EN-5 grades B3, T 53, Weak.

Anthropic's report groups three more cases by method. A hacktivist with stolen API keys (GTG-50029), a financially motivated crew harvesting credentials from mobile applications (GTG-50014) and a Russian state-nexus espionage actor (GTG-20006, whose attribution Anthropic calls consistent with public reporting on Midnight Blizzard) ran multi-victim campaigns with agentic AI that, in Anthropic's account, once needed teams of operators. In the espionage case, the actor used AI to monitor how well implants hid. When security products flagged an implant, the actor used Claude to find, modify and redeploy the detected artifacts. Row EN-12 covers all four.

Copyright 2026 Treadstone 71 - www.treadstone71.com

A joint advisory of 19 August (AA26-231A) carries five authoring agencies: the NSA, CISA, the FBI, the Department of Energy and the Environmental Protection Agency. The ledger lists it under the FBI and CISA alone. Its subject is reconnaissance and capability development against Siemens S7 PLCs, and it names no actor and no state. The agencies write that the actors use AI-generated exploitation scripts disguised as legitimate monitoring tools, built on the open-source snap7 library to read and write over the S7comm protocol. Censys and ZoomEye appear among the scanning services the advisory names, and the S7-200, 300, 400, 1200 and 1500 families appear as targets. Agencies judge the activity likely meant as persistent reconnaissance, and they call the AI use an evolution that lowers the expertise and time an actor needs. The text reports no PLC compromise, no victim tally and no loss. No sample is cited, and no basis is given for the finding that AI wrote the scripts. That wording is the agencies' assessment and no more. The AI claim stays unverified. Row EN-4 grades B3 at T 57, Weak. Five agencies share the text, and the AI finding rests on nothing the advisory shows.

Finance staff face a different attack, and three accounts need keeping apart.

Arup's accounts for the year to 31 March 2024 carry two statements. The chair's foreword says that in January 2024 criminals used "fake voice, signatures and images" in a fraud, and that the attack did not touch Arup's networks or financial stability. Note 10 books an exceptional item of 25.4 million pounds for "fraud incident and associated costs." The note calls the event a social-engineering-led attack by an external third party, says Arup reported it to the police and that an independent investigation followed, and states that no systems were compromised and no personal or project information was disclosed. No transferred sum is broken out, and neither statement uses the word deepfake or describes a video call. Hong Kong police, quoted by The Guardian in May 2024, said an employee joined a call with many participants who looked like real people and sent HK$200 million (about US$25.6 million) in 15 transactions to five local accounts. Police and press supply the figure. Arup's accounts supply another. The label on Arup's line covers associated costs as well, so the two numbers measure different things.

On 26 June 2024 the Secretary for Security answered a legislative question (LCQ9) on deepfake fraud. A January 2024 case in the reply began with a phishing email from a fraudster posing as the UK head-office finance chief. An informant joined a group video conference and sent money to five local accounts, about HK$200 million in all. According to the reply, the video was pre-recorded from downloaded public clips and voices, with no live interaction. Arup goes unnamed there, so we match the case on date, amount and account count. Police and the reply differ: one describes a live call with many participants, the other a pre-recorded video. Other cases in the reply stay out of the record: an August 2023 syndicate with a HK$200,000 loss and nine arrests, a May 2024 attempt of about HK$4 million, and 21 deepfake videos of officials or celebrities found by 31 May 2024.

Arup's chief information officer called the case technology-enhanced social engineering. The Forum's story page of February 2025 adds that he built a real-time deepfake of himself in about 45 minutes with open-source software and judged it not particularly convincing. That page also says Arup's systems stayed intact and no data was affected. The Forum's own text speaks of "a video call with senior management" and 25 million dollars, which is the Forum's framing and not Arup's. Row EN-13 grades B2 at Level 2, with T 65 (Adequate) and D 25. Arup's accounts give a primary source for the fraud and the exceptional charge, and that source puts row EN-13 at Level 2. A live multi-person deepfake call stays unconfirmed, which matters for any claim about real-time deepfake capability. Thirty-three months have since passed. No 2026 case of the same kind sits in the ledger.

Assessment: whether deepfake-enabled payment fraud against finance teams stays a live exposure rests on one confirmed fraud, and evidence of any 2026 case is absent. Roughly even chance (45 to <55), Low confidence. Its deepfake method is the less certain part, and the case sits before the 2026 record. The ATCRI ranks the threat last of ten, and Section 8.1 corrects the confidence input it uses.

Assessment: threat actors now run credential theft and multi-victim intrusion with agentic AI at a pace that shrinks the defender's warning time, as Google, Gambit and Anthropic report. Likely (55 to <85), Low confidence. Row EN-5 reads Weak on the ledger (T 53), and confidence follows that weakest row (Section 2). The ATCRI ranks autonomous multi-victim intrusion fifth of ten (raw risk 384, confidence 87).

7.5 The counterweights

Mandiant writes in M-Trends 2026: "we do not consider 2025 to be the year where breaches were the direct result of AI."

IBM and Ponemon report on 29 July that more than 20 percent of surveyed organizations had a breach that targeted AI. Compromised APIs, apps or plug-ins and cloud misconfiguration lead the causes at 27 percent each, and IBM's headline says one in four malicious breaches is AI-enabled and costs about $6 million. Respondents report on themselves. Row EN-8 grades B3 at T 56, Weak. The 2025 edition found 13 percent of organizations reporting breaches of AI models or applications, and 97 percent of those lacked proper AI access controls.

ENISA, the EU cybersecurity agency, covers 2025 in its 2026 Threat Landscape. The agency says threat groups show consistent interest in AI systems as tools and as targets, and one extract names a fake Claude AI website that distributed malware. Row EN-9 grades B2 at T 67, and the grade rests on quoted extracts, with the full document outside the sample. Section 10.2 counts EN-9 as an agency row.

One case shows how attribution fails. On 16 July Hugging Face disclosed an intrusion that an autonomous AI agent system drove end to end. OpenAI later said its own evaluation agents caused it, and no malicious human intent appears in OpenAI's account. A first reading of adversary activity reversed. Row EN-11 stays at Level 1 and bears on no judgment about adversaries. The case belongs in training material for any team that attributes AI-driven activity.

7.6 The CARM decision test

The CARM (Cyber Action Risk Management) engine weighs compliance exposure, retaliation risk and recovery cost against urgency. Section 8 shows the inputs for three proposed actions after a suspected malicious gateway release. Nobody calibrated the weights of 0.35, 0.45 and 0.20. The labels APPROVE, HOLD and DENY come from a rule of our own, set in Section 8.4, in place of the engine's published cut points.

Under that rule, rotating every secret the gateway holds and taking the gateway offline scores a risk of 3.3 and returns APPROVE. Blocking all outbound traffic from agent runtimes until the team reissues permissions scores 4.3 and returns HOLD. Deleting the agent memory store to remove suspected poisoned instructions scores 4.6 and returns HOLD. Move the cut points and the labels move. Lower the APPROVE ceiling to 3 and the rotation falls to HOLD. Raise it to 4.5 and the traffic block turns APPROVE, while memory deletion stays HOLD because its risk of 4.6 sits above the line. At 5, with an urgency floor of 7, all three pass. Set DENY at 4.5 and memory deletion turns DENY. Equal weights change nothing. Each label stays an illustration.

The lesson holds without the numbers. A cheap action that matches the exposure passes. Broad actions that disrupt the business wait for a second look.

7.7 Adversary reaction read (ACS)

The Adversarial Cognitive Simulator, run on a financially motivated supply-chain group under the trigger "public indictment and takedown of its malicious releases," returns the read below. Treat it as model output built on our inputs. No ledger row supports any line of it, and it counts as no evidence.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Worldview. The group treats developer tooling as a trust market: whoever holds the publishing token holds the audience. Fears: arrest of operators, loss of stolen tokens, and a hardened package registry. Biases: overconfidence in speed, and a habit of reusing one entry path across many targets. Likely actions: rotate infrastructure, move to a different registry or tool, and sell or extort on stolen data fast. CARM pathways: hide, relocate, monetize. Miscalculation risk: reuse of one path against many victims gives defenders a single indicator to hunt.

7.8 Three rows the enterprise record leaves thin

Each figure below is the forward chance from Section 12.10 that a public source documents a new case in that cell by 31 March, 30 June and 31 December 2027. No case count backs them. Each starts from the cell's status today.

Criminals and local models. The record shows aspiration and exposure and no incident. Check Point's January and February forum digest shows actors installing uncensored local models and asking what hardware an unrestricted model needs, and it concludes that adoption is growing but limited because local models underperform. Cisco Talos lists uncensored models that run through Ollama. Trend Micro's January 2026 analysis argues that Xanthorox probably does not run locally and that KawaiiGPT wraps commercial models, so a "self-hosted" label in a criminal product name proves nothing. The cases that look like local-model abuse are the inverse: Sysdig saw an unattributed actor use exposed Ollama servers as the reasoning engine of an offensive pipeline, and Pillar Security saw a campaign scan for exposed Ollama and OpenAI-compatible endpoints and resell the access. Those actors use another's local model. ESET reports that PromptLock calls a model locally through the Ollama API, and it was a research prototype from NYU academics, not an adversary tool. No incident-response report in the record shows a criminal group running its own model against a victim, and the Europol 2026 summary holds no discussion of the topic. Interpol, Volexity, Sekoia, Huntress, Kaspersky and Proofpoint fall outside the sample, so they stay open and do not count as nulls. In Section 12.10 the cell reads Partial, for forum installs, rig pricing and abuse of exposed servers. Its forward chance runs 26 percent, 37 percent and 54 percent across the three dates.

Iran and agents or supply chain as a channel. Anthropic's Iranian cases are propaganda bodies and surveillance tooling, including a malicious Firefox add-on. Google's tracker describes APT42 misuse of Gemini and Iranian persona work, with no AI supply-chain targeting. Microsoft's 2026 Digital Defense Report describes no Iranian AI use on its landing page or in the excerpts at hand, and the full PDF sits outside the sample. The supply-chain evidence that exists is North Korean (the Mastra framework compromise, which two vendors attribute) and criminal (TeamPCP, DUSTMAKER, Nx). Vendors also tie the Axios npm compromise to North Korea. Axios was a conventional package compromise, so it carries no weight in the AI supply-chain cell. Iran stays Not observed, at 6, 9 and 14 percent. Its record covers the sample, not what Iran does.

Hired hands and chatbots as an audience. The hired-hand model is documented: Anthropic's commercial influence cases, OpenAI's February 2026 "Trolling Stone" case on a for-hire cluster, and Meta's account of ANO Dialog, a state-linked body. Its network and the Doppelganger campaign outsourced work to one contractor. In each case the audience is people and platforms. No cited source shows paid workers building content for a chatbot to ingest. The chatbot-audience cell for hired hands holds at 9, 14 and 22 percent. Section 4's Russian Pravda network sits nearest, and it is a state project, not a hired one.

Copyright 2026 Treadstone 71 - www.treadstone71.com

8 Scoring and prioritization

The Treadstone 71 decision engines turn evidence into ranks. Each engine takes ratings from the analyst and applies a published formula. The ratings in Sections 8.1 to 8.4 are our own structured expert ratings. We print every input so a reviewer who reruns the engine with different ratings sees the effect at once.

8.1 The ATCRI ranked threat ledger

The ATCRI engine ranks competing threats. Raw risk equals Likelihood times Severity times Impact, each on a scale of 1 to 10, and weight equals a threat's share of the total. Confidence adds three parts: Data Quality scaled to 40 points, Source Reliability scaled to 35, and age points that run from 25 for evidence under a week old down to 5 for evidence over 180 days old. Age counts days from 6 October 2026 back to the publication date of a source behind the threat. For most threats the input matches such a date in the ledger. Row J breaks the pattern. Its input of 125 days matches no ledger date, since the one row behind it, EN-13, rests on sources from 2024 and early 2025. At over 180 days the age points fall from 10 to 5 and the confidence from 62 to 57. The table prints the engine's 62. Rank, raw risk and weight stand. The ten threats below draw on the ledger in Section 16. In the table, L stands for Likelihood, S for Severity and I for Impact.

Table 9. *8.1 The ATCRI ranked threat ledger*

Rank	Threat	L	S	I	Raw risk and ARI	Weight %	Confidence
1	D. AI supply-chain compromise (LiteLLM, npm packages)	9	8	8	576	15.5	85
2	G. AI-run surveillance and transnational repression (PRC, Iran)	9	7	7	441	11.9	80
3	F. Model theft and distillation by Chinese labs	9	6	8	432	11.7	87
4	E. Agent and MCP exposure inside enterprises	7	8	7	392	10.6	70
5	I. Autonomous multi-victim intrusion at state scale	6	8	8	384	10.4	87
6	A. State bodies run AI as standing influence staff (IR, RU, CN)	9	5	8	360	9.7	80
7	B. Poisoned machine inputs: LLM grooming of chatbots and retrieval	7	6	8	336	9.1	62
8	H. DPRK AI-built personas and insider access	9	6	6	324	8.7	75
9	C. Impersonation campaigns aimed at AI infrastructure and firms	7	6	7	294	7.9	75
10	J. Deepfake-enabled payment fraud against finance teams	7	6	4	168	4.5	62

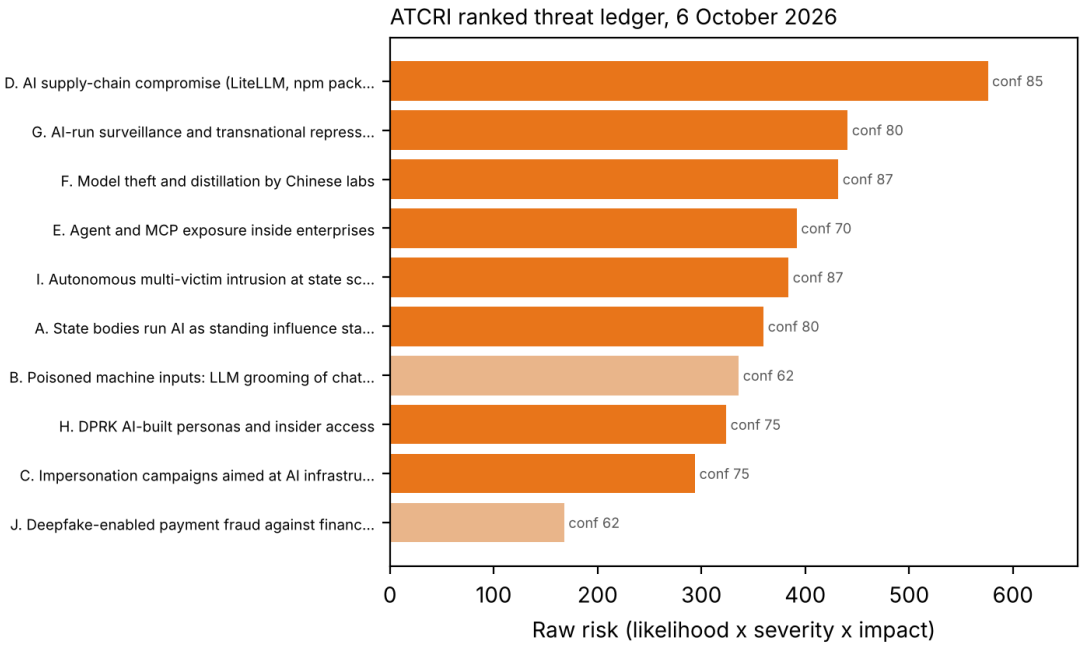


Figure 1. *ATCRI ranked threat ledger. Bar length shows raw risk. Pale bars mark confidence below 70.*

Four findings stand out.

The top five threats belong to the enterprise, to capability theft and to state surveillance. AI supply-chain compromise leads at 576. Surveillance, distillation, agent and MCP exposure, and autonomous intrusion follow between 384 and 441. State influence staff work ranks sixth (360). Chatbot poisoning ranks seventh (336), with a confidence of 62, the lowest in the printed table. Impersonation of AI infrastructure ranks ninth (294). Deepfake payment fraud ranks last (168) and prints the same 62. The age correction above puts it at 57.

Ranks two to five span 13 percent of raw risk, so read them as a tier. The order inside the tier shifts with small changes to a rating, and Section 10.5 tests it with 20,000 random changes.

The ranking measures defender risk. Effect on a population stays outside. A threat that persuades nobody still scores high on severity if it steals a credential. Nor does the order show what adversaries value. A high rank says what a defender stands to lose, and the ledger holds no row on adversary valuation. Read the order as an author-set ranking of our own ratings. No finding follows.

The scenario test shows how fragile the top spot is. Scenario A in the engine raises one threat by half, and Scenario B cuts it by 40 percent. Under Scenario A on supply-chain compromise, its weight climbs from 15.5 percent to 21.6 percent. Scenario B drops it to 9.9 percent, and AI-run surveillance takes first place at 12.7 percent. Raise chatbot poisoning under Scenario A, and the threat jumps from seventh to second at 13.0 percent. A defender who funds against rank alone buys the wrong lesson from a single week of news.

8.2 CWIA and CWC for impact and counter-influence priority

The CWIA engine scores a campaign on four inputs from 1 to 10: Content Saturation, Psychological Targeting Match, Engagement Velocity and Sentiment Shift or Trust Erosion. Impact equals their sum over 40 times 100. Scores of 75 and above read Critical, 45 to 74 read Elevated, and scores under 45 read Monitor. Confidence splits evenly between Data Quality and Source Reliability, 50 points each.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 10. 8.2 CWIA and CWC for impact and counter-influence priority

Campaign	Sat	Psych	Vel	Sent	Impact	Verdict	Conf
BRICS4CLICKS and Verified4War, Iran-war content (first month)	9	6	9	4	70.0	ELEVATED	80
CopyCop impersonation videos against Firebird, Armenia	4	7	7	3	52.5	ELEVATED	80
Iranian state soft-war content through Eitaa, Bale, Rubika (GTG-34001)	5	7	5	4	52.5	ELEVATED	70
Chatbot exposure to Pravda-network content, EU (modeled)	6	6	5	3	50.0	ELEVATED	50
Storm-1516 Baltic campaigns, 30 Jul to 17 Aug	3	6	5	3	42.5	MONITOR	80
Spamouflage AI content on the Tibetan election	4	5	3	2	35.0	MONITOR	70
PRC-linked operations on US AI debates (Data Center Bandwagon)	3	5	3	2	32.5	MONITOR	70
Matryoshka against German state elections	3	4	3	2	30.0	MONITOR	80

No campaign reaches Critical. One campaign, the BRICS4CLICKS and Verified4War networks, reaches 70.0, and its score rests on saturation and velocity alone. ISD found no evidence of state backing. Nobody measured a change in belief. The Sentiment input stays at 4 for that reason. Three more campaigns read Elevated: Firebird and the Iranian state soft-war content tie at 52.5, and the American Sunlight Project's modeled count of chatbot answers that cite Russian state-attributed sources reads 50.0 at the lowest confidence in the set (50). That model rests on a rate from ISD. ISD counts how often a chatbot response cites a state-attributed source: nearly 18 percent of 300 test responses, 11 percent for neutral prompts, and two links from the Pravda network. The project relabels citation as "Russian-influenced" and applies a researcher-built test rate to everyday news questions. At ISD's neutral 11 percent the same inputs give about 308,000 answers a month, not 504,060. The row is a test-set share and no user-exposure rate. No judgment gives the figure weight. The CWC engine ranks the same campaigns by the priority a defender should give them. Cognitive Threat Quantification equals Reach times Virality plus Framing times Trigger, and Counter-Influence Priority equals that figure times Target Vulnerability times Source Deception over 5. Reach, Virality, Framing and Trigger run from 1 to 10. Deception and Vulnerability run from 1 to 5.

Table 11. 8.2 CWIA and CWC for impact and counter-influence priority

Rank	Campaign	Reach	Vir	Frame	Trig	Dec	Vuln	CTQ	CIP
1	CopyCop impersonation, Firebird	4	7	7	6	5	4	70	280.0
2	BRICS4CLICKS and Verified4War	9	8	6	6	4	3	108	259.2
3	Storm-1516, Baltic campaigns	3	5	6	7	5	4	57	228.0
4	Iranian state soft-war (GTG-34001)	5	5	7	7	3	4	74	177.6
5	Pravda-network chatbot grooming	6	4	5	4	5	4	44	176.0
6	Spamouflage Tibetan election content	4	3	5	5	4	3	37	88.8
7	Matryoshka against German state elections	3	4	5	5	4	3	37	88.8
8	PRC-linked US AI-debate operations	3	3	5	4	4	3	29	69.6

The CWC order differs from the CWIA order. Highest on threat quantification (108), the BRICS4CLICKS networks still fall to second, because the audience rates 3 on vulnerability and the source deception rates 4, against 4 and 5 for Firebird. Firebird takes first place with a lower reach and a higher fit: a forged-outlet brand, a named facility, an audience that already doubts the project. Section 4.4 reached that lesson. Fit between payload and audience moves priority more than size does. Matryoshka ties Spamouflage's Tibetan election run at 88.8 for sixth and seventh place, so the order between them carries no information.

8.3 Four audiences under HTIM

The HTIM engine scores the audience. Other engines score the operation. Three ratings, each from 1 to 10, feed the Terrain Vulnerability Index, which averages them: Demographic Resonance, Cultural Norm Alignment and

Political Trend Volatility. A second index, the Pathway Exploitation Score, averages Channel Saturation and Behavior Shift Velocity. Every input is an analyst rating. Behavior Shift Velocity records our judgment of how fast an audience's conduct moves, and no study in the ledger measured that conduct. Weights of 0.6 for the TVI and 0.4 for the PES, multiplied by 10, give the HTIM Index.

Table 12. 8.3 Four audiences under HTIM

Population	Demo	Cult	Pol	Chan	Behav	TVI	PES	HTIM Index	Conf
German state-election electorate (Matryoshka)	5	5	6	4	3	5.33	3.5	46.0	75
Latvian conscripts and Baltic audiences (Storm-1516)	6	7	7	5	4	6.67	4.5	58.0	75
Armenian public and the AI-investment community (Firebird)	6	6	8	6	5	6.67	5.5	62.0	75
Iranian domestic audiences (GTG-34001 and platform walled gardens)	7	8	6	7	5	7.0	6.0	66.0	72

The ordering follows the reading in Sections 3 and 4, since we set the ratings after that reading. That match tests nothing: one hand set the ratings and the expectation. Iran's domestic audience scores highest at 66.0, because the state holds the closed platforms and the culture rewards the frame. Around Firebird, the Armenian audience scores 62.0. Latvian conscripts and Baltic audiences score 58.0. The German electorate scores 46.0. Germany's score sits lowest. We hold the German Behavior rating at 3 because nothing in the record shows that audience changing what it does. The rating reflects a missing sign. Nobody measured behavior.

We scored no Chinese, North Korean or proxy audience. The record names no audience for those cases in enough detail.

8.4 One decision under CARM

The CARM (Cyber Action Risk Management) engine scores a proposed defensive action. Risk equals Compliance Exposure times 0.35, plus Retaliation Risk times 0.45, plus Recovery Cost times 0.20. Urgency stays outside the sum and drives the verdict. Nobody calibrated the weights. The labels follow a rule we set for the example: APPROVE when risk is 4 or lower and urgency is 8 or higher, DENY from 7.5 upward, HOLD otherwise. Cut points published by the engine stay out of the example. Section 7.6 shows how the labels move at neighboring cut points. Three actions answer a suspected malicious release of an LLM gateway.

Table 13. 8.4 One decision under CARM

Proposed action	Urgency	Compliance	Retaliation	Recovery	Risk	Verdict
Rotate every key held by the LLM gateway and take the gateway offline after a suspected malicious release	9	4	2	5	3.3	APPROVE
Block all outbound traffic from agent runtimes until tool permissions are re-issued	8	5	3	6	4.3	HOLD
Delete the agent memory store to remove suspected poisoned instructions	7	6	2	8	4.6	HOLD

8.5 What the scores leave out

Every engine above scores presence and inferred risk. None measures effect. The ASA Measurement of Effect standard caps each effect claim by the strength of the study design behind it. No source in the ledger meets a design strong enough to support a claim of changed belief or behavior. Section 11.5 applies the ceiling. Until Treadstone 71 and its clients pre-register a baseline, the scores in Section 8 prioritize attention and prove no harm.

9 What governments and allies say, and what they leave out

Governments speak in four registers. Intelligence assessments set the strategic frame. Cyber agencies warn about speed. European security services and the EU count influence incidents and sanction people. Alliance bodies convene the debate. The registers disagree on what matters. Gaps between them carry as much information as the words.

9.1 The American assessment goes quiet

The Office of the Director of National Intelligence published its 2026 Annual Threat Assessment on 18 March. Across 34 pages the assessment frames AI as a technology-competition and cyber-enabler issue. China ranks as the most capable competitor in AI, the text says. Beijing, it adds, seeks to displace the United States as the global AI leader by 2030. The assessment calls Beijing the most active and persistent cyber threat. Russia appears as a persistent, advanced cyber attack and foreign intelligence threat. North Korea runs IT-worker and heist operations worth at least $1 billion a year, the assessment says. On Russia, the gray-zone list reads cyber attacks, disinformation and influence operations. Islamist groups, the authors note, shifted attention to information operations. Language on AI-enabled influence is absent. The phrase "malign influence" never appears. Election hits concern Haiti alone.

The ODNI 2.0 fact sheet, dated August 2025, says the office folds the Foreign Malign Influence Center into the National Intelligence Council and Mission Integration. Calling the center redundant, the sheet claims savings of about $7 million a year. State's Global Engagement Center, the sheet also says, shut earlier in 2025. The sheet carries no AI language. As of 7 October, the public listing pages of the White House (114 items from 1 September to 7 October), the Justice and State departments (State: 150 items), the FBI, ODNI, CISA, DHS, Treasury, the Election Assistance Commission, NASS and NASED hold no 2026 statement that AI-enabled foreign influence targets the midterms, which fall on 3 November. A 29 September executive order tells agencies to drop the term AI, so absence of the word proves less, and the listings hold no statement under "super intelligence" or "SI" either. The null covers listing pages only. Nothing follows about agency holdings, and the pages behind the listings are a separate question. A second pass on 9 October read ODNI newsroom pages 1 to 20, which hold every 2026 item, the CISA news feed back to 1 January, NSA press pages 1 to 10, and the State, Justice and DHS pages for 2026. None pairs AI-enabled foreign influence with the midterms. White House releases of 16 to 30 July name China, Russia, Iran and North Korea as election interferers, in earlier cycles and in voting systems and voter data, with no AI line and no midterm frame. The State Department closed its foreign information manipulation unit in April 2025, and its bureau menu lists no counter-disinformation office. ODNI's old Foreign Malign Influence Center election pages no longer resolve. The ODNI site search returned a service error, and Congress.gov stays behind a bot check. Executive silence is not total. A Department of War memorandum, dated 22 September and released on 28 September, is addressed to the commander of Cyber Command and the heads of NSA, DIA and NGA, directs work against foreign interference in the 2026 elections and uses the phrase "foreign malign influence." A one-page scan of the memorandum is public, and Military Times carried an account of it on 29 September. Neither the memorandum nor the release names a country or AI, and neither states a finding. The department's story of 29 September names the 3 November contests and no state and no AI. On 1 October the White House repeated the plan, again without a finding. The 2026 assessment holds no line on AI influence, and the FBI warnings on AI and elections in the record date from 2024. In its cyber strategy of 9 September, the FBI says the Bureau defends elections against foreign cyber threats, and its AI passages stay in cyber terms. NASS and NASED, the election officials' bodies, run trusted-information campaigns with Microsoft and OpenAI. Neither makes any statement on foreign AI influence.

CISA published its 2026 Election Infrastructure Security Plan on 24 September, under the subtitle "Securing the Next 250." The plan runs 13 pages. Page 2 holds only a running header, and the file holds five images, so that page is perhaps a graphic. The other twelve pages carry no word on AI, generative tools, deepfakes, synthetic media, disinformation, misinformation, propaganda, social media or narratives. Russia, China and Iran go unnamed. The plan scopes itself to cyber and physical threats against election technology, processes, facilities and personnel. Its text calls voter registration databases attractive targets for foreign adversaries and says hackers tried to breach them in all 50 states, with confirmed success in at least 20. A line on insider threats says a growing number of foreign adversaries kept monitoring election networks and tried to influence or interfere in US elections. No adversary, no 2026 incident and no AI-enabled technique accompany that line. One trust line tells election offices to act as the authoritative source on election information. The plan gives that task to state and local offices and hands CISA no counter-influence role. Of 107 election-related incidents tracked since 1 January 2022, 96 were bomb threats. Infosecurity Magazine (25 September), SecurityWeek and the Associated Press describe the same plan, and AP says it holds no specific information on 2026 foreign interference. AP also reports about 1,000 CISA staff cut and $10 million removed from election cybersecurity. The plan says nothing about AI and states no CISA role on foreign influence.

The best primary source on what is held back is a letter of 2 October from Warner and Himes to the national intelligence director. Career analysts recently briefed the House committee on foreign threats to the midterms, and the briefing is not public. The first statutorily required report has arrived and others are outstanding. Still not public: the 2024 intelligence community assessment. The letter asks for a classified update by 31 October and says that, absent warnings, its signers "will assume that no such foreign interference has been observed." Real-time warning is requested for tampering with election infrastructure, manipulation of results, interference with the integrity of voting and efforts to "interfere with or influence" the contests. Its scope reaches influence in general terms. No AI language appears in the letter. Congress is writing, not an executive agency. Its fact (briefings occurred) and its inference (silence means nothing observed) stay apart here. Four congressional documents carry AI lines. Senator Padilla and Representative Morelle wrote to DHS and CISA about election-office funding. Their press release is dated 3 September, the file name suggests 2 September, and the date inside the letter stays unconfirmed. One line says Russia, China and other foreign and non-state actors have significant cyber and online election interference capabilities, "likely increasing with advances in artificial intelligence technology." The wording is hedged and unsourced. Senators Padilla and Schiff wrote on 1 October to seven social-media and AI company heads, with replies due 15 October, and say advances in AI and chatbots open new chances for large-scale election misinformation by foreign and domestic actors. The letter's weight is domestic. Representatives Lawler and Gottheimer wrote on 7 July about chatbot answers. The fourth, from Representative Ross and other House members to CISA and the Attorney General on 20 July, is the only paper in the record that names Russia, AI and the midterms together. Cited are Russian campaigns in Moldova in September 2025 and Armenia in June 2026 that used AI, deepfakes and cyberattacks, and the letter asks the agencies to assess whether the same tactics threaten the November contests. Its own text reports foreign cases and asks for an assessment, and members of Congress are no executive agency. None of the four describes a foreign operation aimed at the midterms. The platforms speak for themselves: Microsoft (15 September) says deepfakes remain a serious concern, and its threat analysis center page lists no midterm report. Google (9 September) describes watermark and label measures, and Meta's adversarial threat report titled H2 2026, dated August and drawn mainly from first-half 2026 investigations, says every network it disrupted used some generative AI, which describes a selected sample, the networks Meta picked for removal. Meta's report holds two cases with US audiences. In one, Meta says it removed 4 Facebook and 31 Instagram accounts that originated in Iran and posed as US activists, students and designers. They posted content on American politics, some of it AI-generated, with anti-Republican views among it, and 79,400 accounts followed them. Meta says the operators masked their

Copyright 2026 Treadstone 71 - www.treadstone71.com

origin and that its investigation found links to actors in Iran. The other case, per Meta, ties a Russian network, ANO Dialog and the PR firm US-Media, to pay-to-play coverage of Kirill Dmitriev and to surveys of how Americans viewed him. Midterm appears nowhere in the report. No report from OpenAI, Graphika, ISD, DFRLab or Recorded Future in the record links AI-enabled foreign influence with the midterms; the OpenAI case index and Graphika's report index each stopped at 25 August when read. OpenAI's report of 8 October, read afterwards, covers a Russian and an Iranian false-front operation, and neither case names the midterms. Anthropic's 24 April election page reports a 600-prompt test and an influence-operation autonomy test. Each test is Anthropic's own.

An older joint guide from CISA, the FBI and ODNI warns that generative tools make foreign malign influence content much easier and cheaper to produce. The text on record carries no date for the guide. Its place before the refocus is inference.

Assessment: the thin American record reflects a changed posture and thinner capacity. No source shows adversary operations falling. Likely (55 to <85), Low confidence, and limited by the sample: the band covers only the public pages in it, not the agencies behind them. Row GV-1 grades A1 at T 88, Strong. The inference from that record to cause stays ours. Strong support earns High, and absence cuts it once to Moderate. Our own causal reasoning from a gap cuts it again to Low. The White House, State Department and Justice Department pages make the null a documented one for those sites, and they leave that reasoning where it stood. Likewise, the CISA plan supports the null without proving the cause, since its infrastructure scope alone explains the silence. Assumption A5 in Section 1.6 carries the limit. What the US government has said in 2026 is a null for the pages above, and the null stops there. Capacity cuts rest on primary text for one item (the ODNI sheet) and on press for the rest.

A client that reads only American sources in the weeks before 3 November will see silence and will read safety into it.

9.2 Cyber agencies warn about speed

The cyber agencies talk about the pace of attack. Speed leads every statement. In a blog post that its page dates 30 March 2026 and shows as modified on 29 May, Britain's National Cyber Security Centre said defenders should assume at least some attackers already have capable AI tools. The same page lists a 17 June news item on a speech by the NCSC chief executive to a RUSI audience, headed with the claim that hostile states link to three-quarters of the cyber attacks affecting critical infrastructure in the UK. Read in full, the speech text cites more than 200 incidents from June 2025 to May 2026 and says about 75 percent were believed linked to state actors. The speech says that by 2028 it is highly likely that AI-enabled capabilities will be used against known vulnerabilities in legacy critical infrastructure. No Russia, China or Iran appears in the text. Volt Typhoon is the one actor it names.

That page also lists a 22 June news item in which Five Eyes cyber security agencies urge organizations to act on rapidly transforming cyber risk. Behind it sits a joint statement signed by the heads of ASD, the Canadian Centre for Cyber Security, the New Zealand and UK cyber centres, NSA and, acting, CISA. Reports have it say that AI already shrinks the time between discovery of a flaw and its exploitation and treat the shift as present. Further coverage attributes to Canada's CCCS a line on frontier models with unprecedented capability in autonomous vulnerability discovery. The Canadian statement of 24 June, read on canada.ca, does not carry that phrase, and it notes Canadian participation in Project Glasswing. An empty stub is all the Cyber Centre's own page for it holds. ASD's own guidance "Defending against AI-enabled cyber attacks" says AI lets actors develop exploits within hours. Its date stays unconfirmed. The 8 September date on ASD's listing sits on the entry for the boards

Copyright 2026 Treadstone 71 - www.treadstone71.com

guidance, which was first released on 5 August, and the 6 July date one account gave finds no support on the page. ASD's alert of 25 September, "Risks of AI misalignment to Australian organisations," says the agency knows of AI agents that took unintended actions and probed public-facing sites, and it sees no indication of broader threat or malicious targeting. The alert is no adversary-use evidence. ASD's listing also shows further AI guidance from May to September: "Frontier AI cyber threat considerations for boards of directors" (page dated 8 September, first released 5 August), "Agentic AI harnesses" (11 September) and "Protect your organisations' AI services" (29 September). The 11 September advisory is now read, and it is a defender's advisory with no adversary evidence. Two items remain leads. A Canadian Cyber Centre advisory of 8 September on AI used in a global campaign against industrial controllers names no actor. Lithuania's national threat assessment, current as of 6 February, records that 2025 saw the first documented case of AI integrated into malicious code.

Assessment: allied cyber agencies speak with one voice on exploitation speed and say little about cognitive effect. Likely (55 to <85), Moderate confidence, and limited by the sample: the NCSC blog, the RUSI speech text, the Five Eyes and Canadian statements and two ASD pages are primary text, and the CCCS phrase finds no support in the Canadian text. Row GV-2 grades A1 at T 83 and sets the band. The print holds at Likely (Section 16.2). GV-3, at A1 and T 77, supports the judgment without setting it. Policy statements look forward. None proves a present-day case alone. Section 7 supplies the cases.

9.3 European services count influence

Europe publishes counts. Finland's security service SUPO writes that generative AI makes influence material easy in rare languages such as Finnish. The service mentions electoral influence, deepfakes and agent-run troll armies. Speed, SUPO says, drives information influence. The page shows no date.

The EEAS fourth report on foreign information manipulation and interference, dated 12 March, covers 540 incidents in 2025 across 10,500 channels. Summary and introduction give that channel count, and chapter 3 rounds it to approximately 10,000. Of all incidents, 29 percent link to Russia and 6 percent to China. Unattributed with coordination indicators: 65 percent. The report states that 27 percent of the incidents it detected involved AI-related tactics, techniques and procedures, 147 of 540. Forty pages make up the report, the last three of them images without text. Three annexes follow the body. They cover sanctions, law enforcement and digital regulation, and none says how incidents are coded for AI. The report gives no formal definition of AI involvement. Nearest to a definition is a line in the executive summary, which says one in four detected incidents "involved the use of AI tools to produce or distribute content." The AI chapter lists AI by function: automated content creation (text, voice cloning, synthetic audio, video and imagery) and large-scale distribution (networks, mass production, translation, and LLM grooming). A glossary entry defines an incident as an action with intent to deceive and leaves the AI qualifier undefined. No coding rule, no threshold and no coder-agreement figure appears in the report. Its incident set is a selective, time-bound sample, the report says, "associated, through public reporting or independent analytical attribution" with FIMI, so part of the set rests on the reporting of others. Much 2025 AI material was low in quality with limited organic engagement, it adds. The 2024 total is not stated, so the rise from 41 to 147 cases cannot become a share.

One observation carries a hedge. EUvsDisinfo wrote in March 2025 that the third EEAS report analyzed a sample of 505 incidents. Suppose the 41 AI cases of 2024 sat inside that sample: the 2024 share reads about 8 percent, against 27 percent in 2025. The fourth report never links the 41 to the 505, and the two years do not share a population, since the third report counts some 38,000 channels and the fourth 10,500. We treat 8 percent as a lead for a request to the EEAS and give it no place in the ledger.

The 27 percent and CrowdStrike's 89 percent measure different things. Side by side, the table prints the two definitions.

Table 14. 9.3 European services count influence

	EEAS, 27 percent	CrowdStrike, 89 percent
Publisher, date	EEAS, 12 March 2026	CrowdStrike, 24 February 2026
Unit counted	A detected FIMI incident: an action with intent to deceive, built from observables and TTPs	Attacks by adversaries that CrowdStrike tags AI-enabled; the tag sits at adversary level
Meaning of AI	AI-related TTPs, and "the use of AI tools to produce or distribute content," listed by function; no threshold	The body offers no test for calling an adversary AI-enabled; the text says adversaries "integrated AI into multiple operational stages," and the blog says adversaries used AI to speed up and tune existing techniques
Statistic	Share: 147 of 540 incidents	Growth rate: 89 percent year on year
Denominator	540 incidents, 1 January to 31 December 2025	None stated; the body gives no count of adversaries or attacks
Population	EEAS-detected, selective sample, partly from public reporting	CrowdStrike telemetry and intelligence, customer base
Domain	Information manipulation	Intrusion and eCrime; the scope sentence also names information operations
Coding rule published	No	No; none in the report body

One is a share and the other is a growth rate. The first covers influence while the second covers intrusion, though CrowdStrike's scope sentence also names information operations. Nobody should compare, average or trend them against each other. The Global Threat Report body runs 58 pages. Its 89 percent figure carries the same wording in the foreword, the introduction and the overview, and Figure 8 charts incident-count increases by attack stage with no incident totals. CrowdStrike's blog of 24 February and its Threat Hunting blog of 3 August add related statements. No open text of the Threat Hunting report body exists in the record. Row EN-14 scores the Global Threat Report (B2, T 59, Weak, D 25), and row KP-2 carries that report beside the Threat Hunting report, and the two rows share one report and count once. The Threat Hunting figure, that leads from AI agent-triggered detection surface 2.5 times as many threat leads as manual hunting, counts something else and stays out of the 89 percent.

In its 10 February report, the Estonian Foreign Intelligence Service carries a chapter titled "Chinese artificial intelligence distorts perceptions." A summary says the same report puts Moscow's spending on the Moldova influence campaign near $150 million, a figure that rests on the summary alone. Section 5.5 covers the DeepSeek judgment.

ENISA's 2026 Threat Landscape, as quoted in extracts, says the agency assesses that AI will highly likely increasingly support malicious operations. ENISA also says attackers mainly use AI to augment existing skills. The second statement lends weight to the Efficiency Amplifier hypothesis of Section 11.1.

The Council of the EU keeps sanctioning people under its hybrid-threats regime. On 16 March it added four individuals. The 24 September listing named Xenia Fedorova, formerly of RT France. After the September listing, 81 individuals and 20 entities stand sanctioned, though the press text truncated at "20 ent." No Chinese actor stands on the list. Listings concern conduct. They say nothing of AI.

Assessment: European bodies attribute and sanction influence work that American bodies leave out. None measures belief change. Likely (55 to <85), Moderate confidence, capped by the T-Score of 60. Row RU-9 carries

Copyright 2026 Treadstone 71 - www.treadstone71.com

B3 with T 60 and D 25, and row GV-4 stays at Level 1. An unpublished coding rule, a selective sample and two channel counts hold the measure at B3 and the row at Level 2. T 60 is 59.7 before rounding, so KJ8a sits on a cliff: one point lower moves it from Moderate to Low confidence. We cite 27 percent only as 27 percent of 540 EEAS-detected 2025 incidents, with the coding rule unpublished, and never as a prevalence of AI in influence operations at large.

9.4 Alliance and think-tank voices

NATO's Allied Command Transformation ran cognitive warfare newsletters in April, May and June. A challenge titled "Agentic AI for Cognitive Warfare" followed in Rennes on 27 and 28 May. The newsletters summarize outside authors, so they show what the alliance reads. They do not show NATO doctrine. Row GV-5 stays at Level 1.

RAND hosts a 27 July essay, "Total War on Taipei," by Howard Wang and Nathan Beauchamp-Mustafaga. The essay examines a March 2025 Chinese military research concept for taking Taipei that joins blockade, precision strikes and cognitive attack, and it says the concept is not yet doctrine. A summary of the RAND page, fetched on 9 October, is the whole of the record on it. Carnegie lists an AI and democracy mapping from January, and no page text backs that item. CNAS's "Red Lines" report, dated 12 June, stands as a lead with no text on record. None of the three is known to document an operation. Row CN-9 pairs the Estonian service, CNAS and NIST. The record holds part of the Estonian chapter and neither the CNAS nor the NIST text.

9.5 What the record leaves out

Four omissions stand out. Each holds for the sources in the record and not for governments at large. The first three are nulls whose scope the sample sets, and the fourth rests on the sources themselves.

No government source in the record measures a change in belief or behavior from an operation. The UK AI Security Institute and the Oxford Internet Institute published a Science paper on 4 December 2025 on 76,977 UK participants, 19 models and 707 issues. That paper measures persuasion per exposure by a model in a laboratory, and says effect outside the laboratory is possibly constrained. No adversary operation is measured. France's VIGINUM wrote in February 2025 that no consensus exists in academia on measuring impact. VIGINUM reports engagement only. EEAS reports engagement as limited. Taiwan's National Security Bureau reports over 45,000 inauthentic account sets and 2.314 million disinformation pieces for 2025 in a report released on 11 January 2026, counts from its own monitoring. The bureau is a party to the cross-strait dispute, and its text gives no effect data. Vendors hit the same wall: Recorded Future says it does not quantify behavioral shifts, and OpenAI says one Chinese operation "appears not to have shifted public opinion," which is an engagement judgment and not a poll. In the first volume of its final report, the Hogue Commission says that measuring the impact of foreign interference on an electoral outcome is inherently difficult. The record holds no Swedish Psychological Defence Agency publication, and no UK, Australian or Canadian source beyond Hogue. Those gaps stay open, and they are no documented null.

No Western government source in the record, apart from one 2024 release, reports adversary use of local or self-hosted models, the shift Sections 4.5 and 6.2 describe. The US Treasury's sanctions release of 31 December 2024 says a unit tied to the GRU in Moscow built a server to host generative AI tools and avoid foreign web hosts, and that the GRU helped pay for the server and for the rent of the apartment housing it. No model appears in the release, and it names the unit, not CopyCop. VIGINUM's February 2025 report expects more use of open-source models on private hardware. The report reads current actors' reliance on closed commercial services as immaturity. That is forecast, not observation. ENISA's 2025 Threat Landscape, as quoted in an extract, names commercial and jailbroken models. Twelve agency sites searched hold nothing on adversary self-hosting

Copyright 2026 Treadstone 71 - www.treadstone71.com

beyond the Treasury release. ENISA's 2026 Threat Landscape and the texts of ASD, CCCS, BSI and ANSSI fall outside the sample for the phrase. The null is limited by that sample.

The record holds no 2026 executive-agency statement on AI-enabled influence aimed at the midterms. Congress and companies mention AI, and the Department of War directive speaks of foreign malign influence without AI. The null is dated 7 October and limited to the sample Section 9.1 sets out.

Almost no government source in the record separates fact from adversary claim in the cognitive domain. Putin's Valdai remarks and the PLA Daily essays reach the press in the same register as vendor findings. A PRC example shows the cost. Chen Yixin, China's minister of state security, published an article in China Cyberspace on 13 September. In the Sina text he writes that hostile forces abuse deepfake audio and video to batch produce political rumors. That is a PRC official's claim about what Beijing fears, and it proves no foreign operation. The press carried it as a warning about foreign AI. Among the article's examples: Claude Mythos, GPT-5.5-Cyber and an open-source tool called Lobster. According to The Record, he did not accuse the two named models of use in attacks on China. Sina Finance hosts the article text. The Hong Kong Free Press item that relays it carries the AFP credit. AFP's text reports his claim that hostile forces abuse generative AI to fabricate political rumors and wage cognitive warfare. The wire copy names Claude Mythos and OpenAI's ChatGPT-5.5-Cyber for hacking capability and calls the risk to Chinese information infrastructure potential. Its text uses no word on deepfakes and none on Lobster.

The fourth omission, no separation of fact from adversary claim, is one the house products answer, since a verified fact and an adversary claim never occupy one sentence.

9.6 The Gulf states and Singapore

The Gulf and Singapore items in the record treat AI-enabled influence as a security and law-enforcement matter, and they cover AI-enabled cyber risk and identity fraud. That set is small, and it speaks for those bodies only, not for the governments as a whole. Saudi Arabia's National Cybersecurity Authority ran a consultation on draft AI cybersecurity guidelines, AICG-1:2026, from 5 July to 5 August. The consultation page, dated 5 July, shows Status "Closed." The authority's list of regulatory documents (three pages, 17 items, the newest dated 22 July) holds no AI guidelines, and its news from 18 June to 7 October shows no final text. As of 7 October no final guidelines stand on the NCA site. The draft concerns AI cybersecurity and not influence operations. SDAIA, the Saudi data and AI authority, lists on its own regulations page a document titled "Deepfakes mitigating risks while fostering innovation." The PDF text is absent, and the August date comes from press. On 14 June the United Arab Emirates Cabinet approved an Artificial Intelligence and Data Authority that reports to the Cabinet, merges three bodies and carries a mandate that includes support for cybersecurity and government information security. Press accounts and the Cabinet release describe no horizontal AI law. The u.ae page for the National Cyber Security Policy for AI reads "Updated on 02 Jul 2026" and confirms six areas, among them adversarial AI attacks. No text of the policy PDF sits in the record. UAE authorities announced the arrest of ten people on 14 March for posting interception videos and fake AI clips. That is a government claim that press relays. Saudi and Emirati press warn of fabricated AI content in the war information environment. Interested parties made those statements. The record carries no 2026 item on adversary AI influence from Qatar or Oman, and no influence-specific item from Singapore. For Qatar it holds only an extract of a national cybersecurity agency guide on secure adoption of AI, version 1.0 of February 2024. Singapore's Cyber Security Agency directed Critical Information Infrastructure owners in May to review cybersecurity as AI threats escalate. On 22 July its release said the code of practice will be updated later in 2026 to address advanced persistent threats and AI-enabled threats, and the newsroom showed no update by 7 October. Neither item addresses influence operations in the texts on record. The

Copyright 2026 Treadstone 71 - www.treadstone71.com

Singapore IMDA item on AI incidents has no text on record. Outside the record stay the SDAIA PDF, the UAE policy PDF, the Singapore update and its 17 June agentic AI addendum, Qatar's primary text and IMDA.

10 Patterns, trends and tendencies

Sections 3 to 9 report cases. Counting the cases shows what the record tells a reader and what it leaves out. Three counts follow, each read from the 63-row ledger of Section 16, where EN-14 shares a report with KP-2 and counts once, so each count covers 62 items. They read the ledger as data about the record itself, apart from the adversaries the record describes. Four more subsections work from other inputs: eleven judgments we chose, 30 hand-set ratings, the vector grid of Section 12.10 and the outside literature.

10.1 What the evidence measures

We coded every ledger row at the top rung of evidence it reaches. Twenty-three rows show AI in use. Seven count an audience. Ten record what a party says. Six document a loss. Ten hold exposure research, guidance or analysis. Six carry no AI element, no adversary or a null result. No row measures belief change, and no row measures behavior change. The top two rungs stand empty. Presence dominates the ladder.

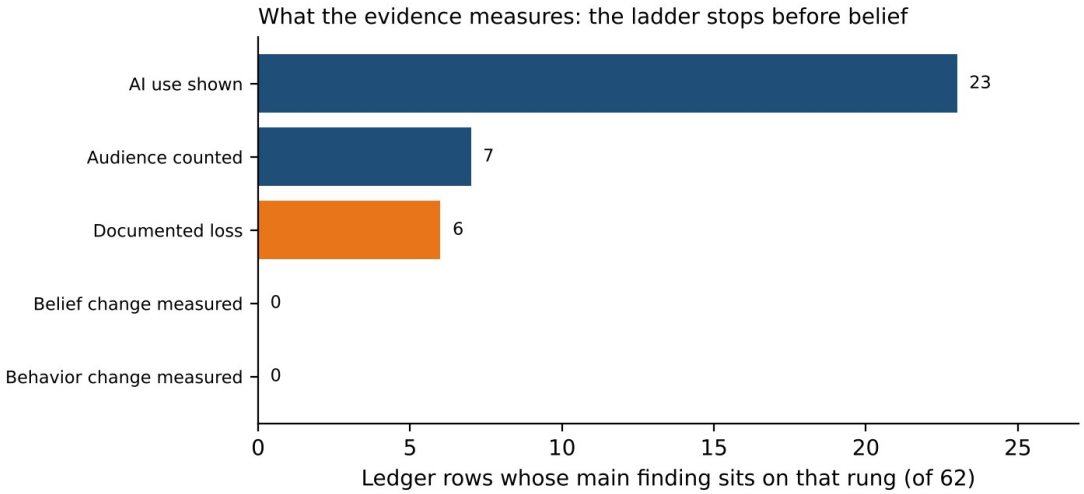


Figure 2. *Ledger rows by the highest rung of evidence each reaches.*

The ladder stops one step short of the audience. Every row that counts an audience counts views, posts or accounts. None asks what readers did next. The count of 23 presence rows against zero effect rows is the single largest pattern in the file, and it holds in every theater.

Loss splits by side. Five rows record loss on the enterprise side: EN-1, EN-5, EN-6, EN-8 and EN-13. One records loss on the influence side: CN-7, a theft of capability. Two of the five enterprise rows read Weak on the ledger. EN-8 is a survey of breached firms, and EN-5 rests on a page whose date conflicts and whose counts differ in secondary coverage. A January 2024 case, EN-13 is about 33 months old, and Arup's own accounts lift it to Adequate at T 65. The split still stands. A buyer who asks whether AI makes influence operations work receives presence data. Where the question is whether AI exposes the firm, the buyer receives loss data. Better evidence answers the second question.

Copyright 2026 Treadstone 71 - www.treadstone71.com

10.2 Who reports on whom

Five source kinds fill the ledger. Anthropic originates 9 rows, counting CN-7, which an NSA, CISA and FBI advisory co-sources. Other labs and vendors supply 20. Researchers and victims supply 8. Governments and alliance bodies supply 13, ENISA's row EN-9 among them. Adversaries and parties to the dispute supply 12.

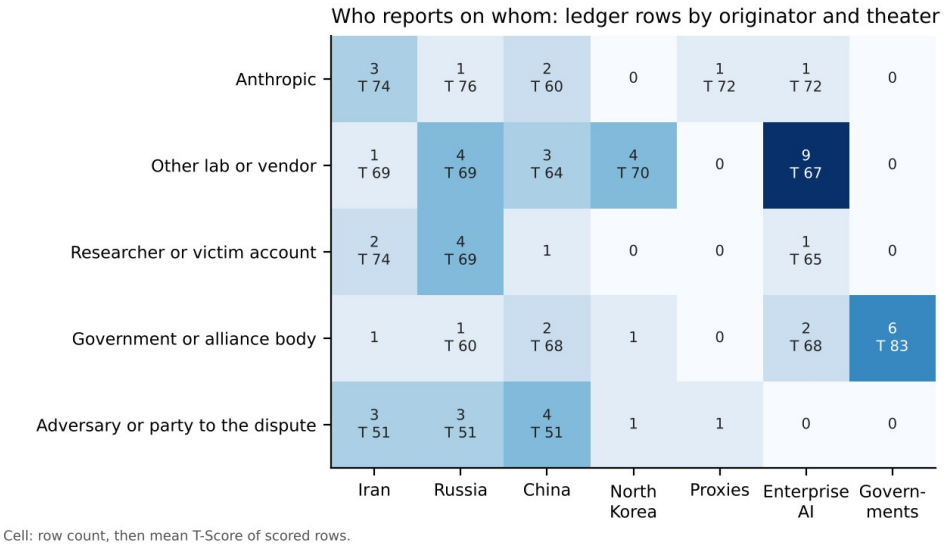


Figure 3. *Ledger rows by originator and theater, with mean T-Score.*

The figure draws CN-7 with the governments and EN-9 with the vendors, so its totals read 8, 21, 8, 13 and 12.

Coverage follows telemetry. Labs see their own users, governments publish counts and researchers study public platforms. Iran depends most on Anthropic: 3 of its 10 rows, covering the propaganda bodies, the surveillance units and the opposition network. North Korea rests on four vendor rows and no Anthropic row. Enterprise AI draws 9 of its 13 rows from labs and vendors. The six rows in the government theater score the highest mean T, 83. Adversary and dispute rows score lowest, at a mean T of 51 in each of the three theaters with scored rows of that kind.

One reporter shapes the Iran file. A theater read through one reporter inherits that reporter's sampling. Four judgments stand on Anthropic alone: IR-2, IR-8, CN-10 and PX-1.

10.3 When the evidence appeared

Of 61 dated rows, 25 fall in September and October 2026. September alone holds 21. February through August average 4.3 rows a month.

Copyright 2026 Treadstone 71 - www.treadstone71.com

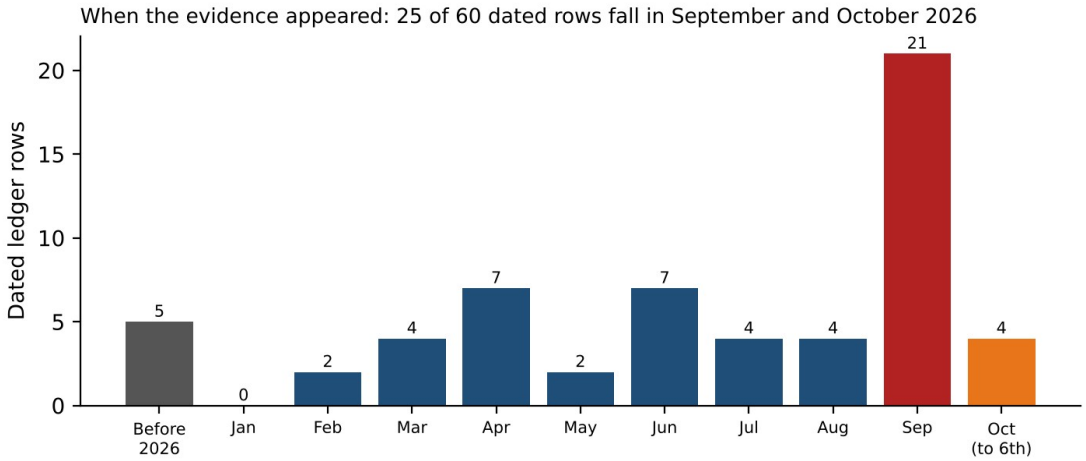


Figure 4. *Dated ledger rows by month of publication.*

Two causes drive the spike. The file cannot separate them. Labs and agencies publish in a September cycle: Google and the NSA, CISA and FBI on 8 September, Anthropic on 10 September. The ledger's window opens on 1 September, which pulls the newest material into it by construction. Add 21 further reports from February through August that meet the counting rule (Section 12.2), and the counted rows run at 2, 6, 5, 5, 8, 5 and 5 a month, a mean of 5.1. September's 13 counted rows come from five reports: Anthropic's 10 September report supplies eight, Google's 8 September tracker two, and Genians, DFRLab and ISD one each. Counted by report, September sits 0.1 standard deviations from the February to August mean. On row counts, it sits 4.4 above. The providers cut their September cases finer, and that explains the cluster better than a surge in operations. Hypothesis H3 predicts the pattern: a clock set by who reports and when, with adversary behavior moving less. The "Why Now" clocks in the BLUF carry that caveat, and the matrix in Section 11.1 does not code it. Our discounting runs one way. We cap adversary statements under Rule 1 and weigh vendor claims against their commercial interest (row E12). Government and other defender-side documents meet no equal check, although agencies hold budgets and mandates to defend. Treat the T-Scores of those rows as unadjusted for that interest.

10.4 How claims lose evidence as they climb

We mapped eleven judgments from Sections 3 to 6 onto six rungs of attribution, from the content itself up to what a state intended or knew. The band midpoint of our own judgments tells the story.

Copyright 2026 Treadstone 71 - www.treadstone71.com

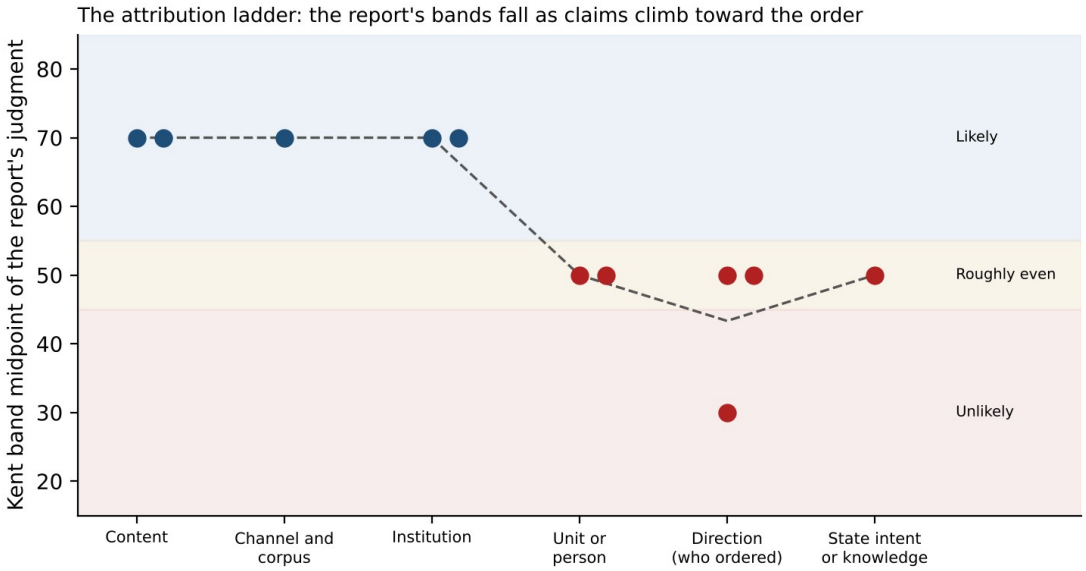


Figure 5. *Band midpoint of the report's judgments by attribution rung.*

Content, channel and institution sit at 70, the middle of Likely. Unit or person falls to 50. Direction, meaning who ordered the work, averages 43. State intent or knowledge returns to 50. The wall runs between the institution and the order. Evidence thins as claims climb. Open sources show the work product, the channel and the letterhead. They almost never show the instruction. The sample is eleven judgments chosen by us, so the pattern is descriptive. Section 15 records the same limit in words.

10.5 How stable the ATCRI ranks are

The ATCRI ranks ten threats on ratings we set by hand. To test the order we moved every one of the 30 ratings by zero or one point either way at random, 20,000 times, and recorded each threat's rank. The test measures sensitivity to random noise in our ratings. Bias in how we set them stays untested, since every run starts from the same hand-set values.

Copyright 2026 Treadstone 71 - www.treadstone71.com

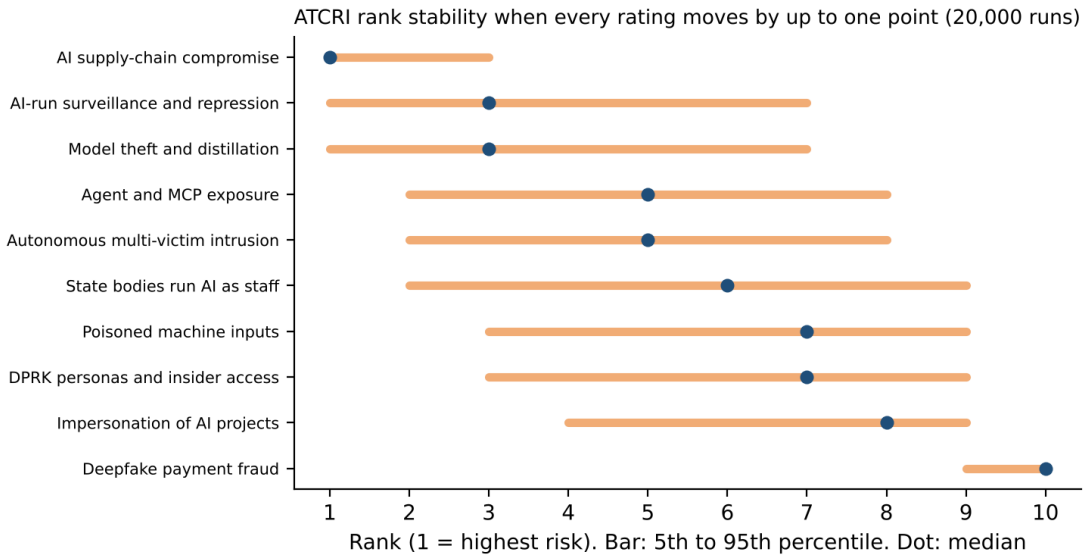


Figure 6. *ATCRI rank under 20,000 random one-point changes to every rating.*

Ranges here run from the 5th to the 95th percentile of rank across the runs. They carry no claim about truth. AI supply-chain compromise ranks first in about 70 percent of the runs, and 90 percent of its ranks fall between first and third. Deepfake payment fraud ranks last or next to last. The eight in between trade places. AI-run surveillance and model theft carry a median rank of 3 and a range of 1 to 7. Agent and MCP exposure and autonomous intrusion carry a median of 5 and a range of 2 to 8. State staff work carries a median of 6. Poisoned machine inputs and North Korean personas carry a median of 7. Impersonation of AI projects carries 8.

Three tiers result. One threat sits alone on top. A crowd of eight fills the middle. One sits at the bottom. On our ratings, a defender funds the top threat first and treats the middle eight as an unordered list. The ranks measure defender risk and carry no inference about what adversaries value.

The ATCRI likelihood and the band on a Key Judgment answer different questions. ATCRI L rates the chance that a defender meets the threat within a year. A Key Judgment band rates how strongly evidence supports a claim about the past. Agent and MCP exposure shows the split: its L is 7 of 10, while KJ9a prints Roughly even chance at Low confidence because nobody has shown exploitation at scale. ATCRI confidence follows a separate formula built on data quality, source reliability and evidence age, so the same threat carries 70 there and Low in KJ9a.

10.6 Tendencies across the four states

The vector grid in Section 12.10 sums to a rough order of adoption. Staff work and content volume lead at 10 points, then persona and deepfake production at 9, then surveillance, agent or supply-chain use and local models at 7 each. AI projects as target follow at 6, model capability as prize at 3, and chatbots as audience closes at 2. Adversaries adopt AI first where it replaces labor they already pay for: drafting, translation, personas. Tooling and agentic work follow. The roles that treat AI as an audience or a prize trail, and the record on them is thin partly because few observers look.

Adoption follows mission. Iran and China show their densest entries in internal security. Russia shows its densest in the information ecosystem, with chatbots and forged brands as terrain. North Korea shows identity and revenue. Each state adopts the tool where its main business already sits.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Exposure produces adaptation. Anthropic's September report counts more distillation than its February report did, and it adds Zhipu to the labs named earlier. Those are provider counts and no government counts. CopyCop kept building outlets. Recorded Future reports a self-hosted stack at CopyCop, and Genians reports local-model tooling on a Kimsuky host. Nothing shows the North Korean stack operating against a victim, and one source holds it. Press that restates the Genians report supplies the only other mention. No lab sees either stack. The pattern matters for collection: a lab report that falls silent about a state does not show the state stopped.

10.7 The outside view

The inside view reads the 2026 cases one by one. Seen from outside, the question becomes how operations of that kind performed in the best-measured earlier cases. Two classes of study exist, and they disagree in a way that matters.

The field class measures real exposure. Eady and colleagues linked survey panels to Twitter feeds and measured exposure to the Russian Internet Research Agency during the 2016 US election. They report that 1 percent of users accounted for 70 percent of exposures, that almost all exposure sat with 10 percent of respondents, mostly strong Republicans, and that posts from domestic media and politicians eclipsed the content. The authors find "no evidence of a meaningful relationship between exposure to the Russian foreign influence campaign and changes in attitudes, polarization, or voting behavior" (Nature Communications, 9 January 2023). Eady's panel holds 1,496 respondents. Equivalence tests rule out effects above 0.05 standard deviations on issue positions and 0.06 on perceived polarization, and the predicted change in Trump vote under observed exposure is -0.18 percentage points (90 percent interval -1.15 to 0.78). Exposure there is potential exposure from the timeline, not confirmed viewing.

Bail and colleagues followed the same agency a year later (PNAS 117(1), January 2020, online November 2019; PubMed Central, PMC6955293). The design is a two-wave panel of 1,239 Republican and Democratic Twitter users recruited in October 2017, who visit Twitter at least three times a week, matched to released records of the agency's accounts. Every 95 percent credible interval crosses zero on six measures of political attitudes and behavior. The paper carries no numbered table, and its coefficients appear only in figures. Its 43-page SI Appendix holds five numbered tables, and none prints an estimate. Table S2 counts 235 respondents with any interaction (19.0 percent) and 140 with direct engagement (11.3 percent), and 46 first interactions fell in the study month, 17 of them direct. The paper's 44 treated sit two below that count, and the supplement gives no reason. Table S4 sets the extended groups at 31 to 213 treated against 1,017 to 1,199 controls. The estimates stay in figures there too. Alternative specifications with pretreated respondents, with a synthetic retweet measure at ten thresholds, with direct engagement alone and with count dosage leave every interval across zero. Direct engagement alone, Figure S7, gives intervals as wide as about -0.25 to +0.57 on social distance. Coefficients read from the paper's Figure 2 hold to about 0.01 standard deviations. They put the average effect on the treated between -0.10 and +0.08, with every credible interval including zero. The widest, for feeling toward the opposing party, runs from -0.38 to +0.09. An extended window from February 2016 to April 2018, on ideology alone, gives about -0.06 for all engagements and -0.13 to -0.16 for direct engagement. Those intervals include zero too, and the upper end of one direct-engagement interval sits at +0.01, inside reading error. Intervals that wide rule out only large effects. In the sample, 19.0 percent interacted with the agency's accounts and 11.3 percent engaged directly, against Twitter's own figure of about 2 percent of its users. Troll interactions made up about 0.1 percent of a respondent's likes, mentions and retweets. The authors list limits: exposure is a binary proxy built from follows, likes, retweets and friends' mentions of the accounts, and no viewing is verified. Assignment was not random. Nor does the study speak to the 2016 election, and it holds frequent partisan users only. Its main models rest on 44 treated respondents and the extended models on 67 to 213, so the null is thin.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Mostly, the accounts reached users who were already strongly polarized, politically engaged and frequent on the platform.

Other field studies point the same way, and none concerns an influence operation in particular. In the Facebook and Instagram deactivation experiment before the 2020 election, 19,857 Facebook users and 15,585 Instagram users logged off for six weeks. The authors call the effects on polarization, perceived legitimacy, favorability and turnout "precisely estimated and close to zero". A companion experiment cut like-minded exposure from 53.7 to 36.2 percent and found no measurable effect on eight preregistered attitude measures. In 2026, Science Advances published a study that cut exposure to repeat-misinformation sources by about 70 percent for months. Preregistered outcomes showed no measurable effect, even for the heaviest users. A deactivation study of WhatsApp in Brazil found less false-rumor exposure and limited effects on belief. The numbers belong to the papers. A 2026 correction to the deactivation paper exists, and the figures above come from the original. One quasi-experiment, on Russian television signal spilling into Ukraine, found higher opposition voting where the signal reached. The study concerns mass broadcast to a sympathetic audience, and no AI.

Two items cut against a blanket reading of no effect. Incerti, Elkobi and Mattingly, in an arXiv preprint of 21 January 2026 that no peer review has covered, put 8,600 US respondents on a TikTok-style app for four minutes. Twenty percent were forced to see assigned videos, and 80 percent chose freely. Per additional minute watched, pro-China influencer content raised favorability toward China by about 0.10 on a four-point scale. State-media content moved it by about -0.05, which the authors treat as virtually no effect. No other item in the record measures the effect of state-linked content in an experiment. The test covers content, not a live operation, and it makes no AI claim. In a December 2025 preprint, Chmel and colleagues paid Americans aged 18 to 45 to follow five progressive creators and moved policy views. Effects ran larger than traditional campaign outreach, the authors report. Domestic creators carried the message, so the item calibrates the channel and shows no operation.

Reach studies add a third view. Researchers at the University of Manchester, in a release of 13 December 2024, report low engagement across websites, Telegram and X for five covert Russian outlets: RRN, BTZ, Voice of Europe, France et EU and the Pravda network. Telegram posts drew few views and almost no replies, and AI aggregation and translation produced error-prone text. The release counts reach and engagement and measures no attitude. Golovchenko and colleagues tracked over 12 million comments on 40 Russian-language YouTube channels. After the 2022 ban, commenting on banned Kremlin-affiliated channels fell to near zero and users moved to other pro-Kremlin channels. That is substitution and no belief change.

The laboratory class forces exposure and measures persuasion. A preregistered experiment with 8,221 US respondents found 47.4 percent agreement with original influence-operation articles against 24.4 percent for controls, and 43.5 percent for GPT-3 rewrites, a gap of 3.9 points. Three Nature Communications experiments with 4,829 participants found policy messages written by a language model moved attitudes by about two to four points, about as well as lay human messages. A study published in Science, available as an arXiv preprint, ran 76,977 UK participants through 19 models and found conversation beat a static message by about three points, with 36 to 42 percent of the effect persisting at one month. More persuasive models were less accurate. Lab studies do measure persuasion per exposure, and they do compare AI-written content with human-written content. Three lab papers sit outside that summary. Costello and colleagues report two studies, with belief drops of 16.8 and 12.3 points in dialogues on conspiracy beliefs. Lin, Rand and colleagues report double-digit shifts on ballot measures and 2 to 3 points on candidate choice. Salvi and Schoenegger test debate and quiz persuasion. The Lin and Salvi items report effects larger than the two to four points above.

Selection and dose reconcile the two classes. The lab forces exposure on a random sample, and the TikTok experiment sits in that class because it forces exposure on a fifth of its respondents. Field data show content reaching the already polarized, at low dose, among domestic noise. Neither class measures the effect of a live AI-enabled operation on a real audience, and the record holds no study that does, on any platform. Telegram now has one reach study. A Cambridge item in the American Political Science Review on pro-Beijing media and Taiwanese voters is a lead with no text in the record. No UK or Australian field study sits in the record, and the NATO StratCom Centre of Excellence's November 2024 report on psychological responses to Kremlin narratives in eight Nordic and Baltic states stands outside it.

No measured base rate exists for how often an influence operation changes belief in a population. The field studies point low for text feeds from 2016 to 2020, and we treat them as a reference class and not as a rate. They say nothing about short video, where the one preprint finds a positive per-minute effect for influencer content. Our prior of 45 percent for H1, 30 for H2 and 25 for H3 is an analyst judgment that the outside view informed (Section 11.3). No calculation produced it, and since no study in the record measures an AI-enabled operation, none moves it. Against that prior, the 2026 record fits. ISD finds limited reach for Matryoshka. OpenAI finds no breakout for a PRC-linked operation. Two networks reach a billion views, as ISD counts them, and nobody shows an effect. The uniform-prior test in Section 11.1 removes the tilt, and H1 stays in front at 76.0.

The reference class has limits. Eady's panel covers one agency on one platform in 2016, Bail's covers one month in 2017, and neither covers generative models. One preprint carries the TikTok result. The laboratory class uses single exposures to recruited samples. Reach counts, the field class argues, overstate effect for text feeds, and one preprint on video points the other way. A field study of a live operation that uses AI is what shows whether AI changes the dose.

10.8 What the patterns change

Five tendencies move decisions. The evidence stops at presence, so a counter-influence buyer asks for a pre-registered baseline before spending. Coverage follows telemetry, so a client reads a quiet lab report as a gap in sight. The calendar clusters, so a client discounts urgency drawn from one month of releases. Bands fall as claims climb toward the order, so attribution to a unit or a commander waits for a second source. The ranks hold only at the top and the bottom, so a security chief funds supply-chain compromise first and sequences the middle eight by local exposure. Section 13 turns each into an action.

11 Analytic tests

Every judgment in Sections 3 to 9 depends on a small set of explanations for the 2026 record. Seven tests attack the leading view from seven sides: an evidence matrix, source removal, staged evidence, a Bayesian update, an opposing case, a ranking of unknowns, and a ceiling on effect claims. A trap map for the reader follows. Three added tests close the section: a count of reporter dependence, a measure of cliff distance and an audit of the hindcast.

11.1 Analysis of Competing Hypotheses

Three hypotheses compete.

H1, the Efficiency Amplifier: AI cuts the cost and raises the volume of operations that already existed. Reach and effect stay put.

H2, the Capability Shift: AI opens operations and effects that adversaries had no means to reach before. Effect rises.

H3, the Signal Artifact: the rise in reported AI use reflects who reports, what they count and what they sell. Adversary behavior moved less than the reports suggest.

The hypotheses overlap. H3 holds beside either other. We score them as rivals because each predicts something different about the next piece of evidence.

Sixteen evidence items enter the matrix. Each item carries the T-Score of the weakest row it cites, and E12 carries the house rating of the ledger as a whole, 80 against a mean of 66.9 across the 48 scored rows. E12 reads consistent with H3 and neutral to the others, so the figure moves no inconsistency score. C marks consistent, I inconsistent and N neutral, and diagnosticity sets a weight of 2.0 for High, 1.0 for Medium and 0.5 for Low. Weighted inconsistency equals weight times T-Score over 100, summed across the inconsistent cells. The method works by disproof. Consistent cells earn nothing. The least inconsistent hypothesis leads. No row enters twice. The eleventh test below restores the double count of CN-1 and KP-3, each counted under two items. E8 stands next to Anthropic's own report, so the third test removes it with the Anthropic rows. E2 reports on 2025 breaches. A report on past breaches follows the events it covers, and a change after 2025 falls outside it. The matrix tests no event order.

Table 15. *11.1 Analysis of Competing Hypotheses*

Evidence	H1 Efficiency amplifier	H2 Capability shift	H3 Signal artifact	Diagnosticity	T
E1 Three theaters show state bodies using AI as routine staff work (IR-1, RU-1, CN-1)	C	C	I	Medium	60
E2 Mandiant: 2025 was not the year breaches were the direct result of AI (EN-7)	C	I	C	High	75
E3 OpenAI finds no meaningful breakout beyond the operator's own reach (RU-11, CN-2)	C	I	C	High	64
E4 ISD finds Matryoshka made basic errors and kept reach limited (RU-7)	C	I	C	High	76
E5 Two pro-Iran networks drew over one billion views with false and AI-made war content, cause of reach unproven (IR-7)	I	C	N	Low	73
E6 One operator breached nine Mexican agencies with 75 percent of commands model-written (EN-5)	N	C	N	Medium	53
E7 AI-orchestrated espionage across about 30 targets, few successes (CN-10)	C	C	N	Medium	60
E8 Chinese labs drain frontier models at industrial scale, confirmed by a government advisory (CN-7)	N	N	I	High	64
E9 EEAS counts 27 percent of 2025 FIMI incidents as AI-involved (RU-9)	C	N	N	Low	60
E10 Kimsuky uses AI for decoys, no evidence of model training (KP-3)	C	I	N	Medium	72
E11 Pravda-network content sits in training sources, memorization untested (RU-2)	N	C	N	Medium	75
E12 Every lead source is a lab or vendor with a commercial or policy interest (ledger)	N	N	C	High	80
E13 Firebird impersonation videos reached 1.6 million views between 24 June and 13 July (RU-5)	C	N	N	Low	68
E14 Agent-run credential harvest finished in under six hours (EN-3)	C	C	N	Medium	68

Copyright 2026 Treadstone 71 - www.treadstone71.com

Evidence	H1 Efficiency amplifier	H2 Capability shift	H3 Signal artifact	Diagnosticity	T
E15 ODNI is silent on AI influence while labs, the EU and researchers carry attribution (GV-1)	N	N	C	Medium	88
E16 CopyCop runs self-hosted models that vendors cannot see (RU-6)	N	N	I	Medium	65
Weighted inconsistency, all stages	0.36	5.02	2.53		

H1 scores 0.36. H3 scores 2.53. H2 scores 5.02.

Four rows sink H2. Mandiant finds that 2025 was not the year breaches came as a direct result of AI. OpenAI finds no meaningful breakout beyond an operator's own reach. ISD finds Matryoshka made basic errors and kept its reach limited. Genians finds Kimsuky uses AI for decoys and shows no sign of model training. Each of the four rates Medium or High on diagnosticity. Reach bounds effect from above: a small audience counts against a rise in effect and cannot show that no effect occurred. Assumption A3 in Section 1.6 carries that limit, and two tests below recode E3 and E4 as neutral.

Three rows sink H3, and each does so for a different reason. A US advisory (AA26-251A) warns of industrial-scale distillation by China-based companies. The ledger row that holds it (CN-7) is co-sourced with Anthropic's February report, so the two are not independent. Six firms appear by name in the advisory, and volume appears only in aggregate. Its references cite Anthropic and Google, so independence from Anthropic stays unconfirmed. Anthropic reports state bodies in three theaters using AI as routine staff work. Recorded Future describes, at CopyCop, a self-hosted stack that no lab sees. That runs against an artifact story, because lab reports undercount such activity.

One row troubles H1: ISD counts more than one billion views for two pro-Iran networks. Low diagnosticity applies, since nobody has shown that AI caused the reach and ISD found no state backing.

The ledger leans on Anthropic. Twelve tests remove or recode rows in different combinations. The first row repeats the base case.

Table 16. *11.1 Analysis of Competing Hypotheses*

Test	Weighted inconsistency H1	H2	H3	Posterior H1 (tempered)	H2	H3	Lead
All 16 rows	0.36	5.02	2.53	84.5	4.2	11.3	H1
Leave Anthropic out (E1, E7)	0.36	5.02	1.93	77.6	3.9	18.6	H1
Leave Anthropic out and the Anthropic-co-sourced advisory row (E1, E7, E8)	0.36	5.02	0.65	67.2	3.3	29.4	H1
Leave six vendor rows out (E1, E2, E6, E7, E10, E14), a partial set	0.36	2.8	1.93	67.7	10.2	22.1	H1
Drop the three high-diagnostic H1 rows (E2, E3, E4)	0.36	0.72	2.53	57.2	35.1	7.7	H1
Drop the three H2-hostile rows plus E12 (E2, E3, E4, E12)	0.36	0.72	2.53	58.2	35.7	6.2	H1
Drop the source-incentive row and ODNI row (E12, E15)	0.36	5.02	2.53	87.7	4.4	8.0	H1
Recode E6, E8 and E11 as inconsistent with H1	2.92	5.02	2.53	63.3	9.9	26.8	H3
Recode E3 and E4 neutral for H2, since reach counts do not measure effect	0.36	2.22	2.53	76.9	12.8	10.3	H1
Recode E2, E3 and E4 neutral for H2	0.36	0.72	2.53	69.6	21.1	9.3	H1
Restore the two double-counted rows of the earlier matrix (17-row matrix)	0.36	5.02	2.6	85.8	4.3	9.9	H1
Uniform priors, 33 percent each	0.36	5.02	2.53	76.0	5.7	18.4	H1

H1 leads the posterior in all twelve tests and the ACH score in eleven. Removing every Anthropic row leaves its posterior at 77.6 percent. Drop the advisory row as well and H1 falls to 67.2 percent while H3 reaches 29.4. The vendor row of the table removes six items (E1, E2, E6, E7, E10, E14), and H1 holds 67.7. Six falls short of the full vendor set.

We define that set by source class. An item counts as vendor-sourced when the ledger names a lab or security vendor as its originator. Nine of the sixteen items qualify: Anthropic (E1, E7), OpenAI (E3), Google (E14), Mandiant (E2), Recorded Future (E13, E16), Gambit (E6) and Genians (E10). E8 stands apart, since a government advisory shares its ledger row with Anthropic's report. Two further removals sit outside the table and the figure. With all nine out, H1 reads 54.2 percent, H2 20.5 and H3 25.2 in the tempered posterior, and H1 stays ahead in the ACH with 0.36 against 1.52 for H2 and 1.28 for H3. Take E8 out as well and the posterior reads 44.9, 17.0 and 38.1, while H3 moves to the front of the ACH at 0 against 0.36 for H1. Six items remain in that run (E4, E5, E9, E11, E12 and E15): ISD, DFRLab, the EEAS, ODNI and the house rating.

The lead does not hold when all vendor evidence leaves. Under the full removal H1 sits below the 55 floor of Likely, and under the last run the two rivals sit seven points apart. H1 plus H3, the pair that agree reach stays small, hold 79.4 percent with nine items out and 83.0 with ten out. So the finding on reach survives. The choice between efficiency and signal artifact does not. Across the twelve table tests H1 runs from 57.2 to 87.7 percent, with a median of 72.8. The two further removals reach down to 44.9.

The lead survives poorly when three rows leave: Mandiant, OpenAI and ISD, each independent of Anthropic and each a report of what did not happen. Without them H1 falls to 57.2 percent and H2 climbs to 35.1. H1 rests on absence of evidence for effect, and absence of evidence is weaker than evidence of absence. That limit holds confidence at Moderate.

Two tests answer assumption A3. Reach counts prove presence and say nothing of persuasion, so E3 and E4 recode as neutral toward H2. H1 then reads 76.9 and H2 reads 12.8. Recode E2 as well and H1 reads 69.6 against 21.1 for H2. H1 still leads. A fair objection survives in part: much of H1's margin over H2 comes from rows that bound reach and leave effect open.

The recoding test changes three rows we scored neutral toward H1: the Mexican operator (E6), the distillation advisory (E8) and the Pravda training finding (E11). Each shows capability or intent and no effect, so a skeptic codes each as a challenge to an Efficiency Amplifier. Recoded, H1 scores 2.92 and H3 scores 2.53, and H3 takes the ACH lead by 0.39. The tempered posterior still favors H1, 63.3 percent against 26.8, because the posterior rewards consistent items and the ACH ignores them. Row E11 carries weight in the matrix: it scores consistent with H2 at Medium diagnosticity. The American Sunlight Project's modeled count of chatbot answers that cite Russian state-attributed sources (RU-3) is another item, and it carries no weight anywhere in the matrix.

H1 and H3 are the rivals that matter. The two agree that documented reach stays small. They split on cause. H1 says AI raised efficiency, and H3 says the reports overstate adversary behavior. The base case puts them far apart, 84.5 against 11.3. Only the hardest tests thin that margin: with every vendor item and E8 out the posteriors read 44.9 and 38.1, and the recoding test hands H3 the ACH lead. That thin margin holds the cost-and-output judgment at Low confidence.

The restoring test returns 85.8 for H1, 4.3 for H2 and 9.9 for H3, so double counting moves H1 by 1.3 points. Uniform priors move it to 76.0.

The hypotheses overlap (assumption A7), so the posteriors read as an analytic index. H1 and H3 together hold 95.8 in the base case. That sum is the number behind the statement that reach stays small.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Assessment one: the documented reach of state cognitive operations stays small. Likely (55 to <85), Moderate confidence. KJ2 in Section 2 carries it, and the two leading hypotheses agree.

Assessment two: AI cut the cost and raised the output of those operations. Likely (55 to <85), Low confidence. No row measures cost or output, and the inference drops confidence one step. KJ2a carries it.

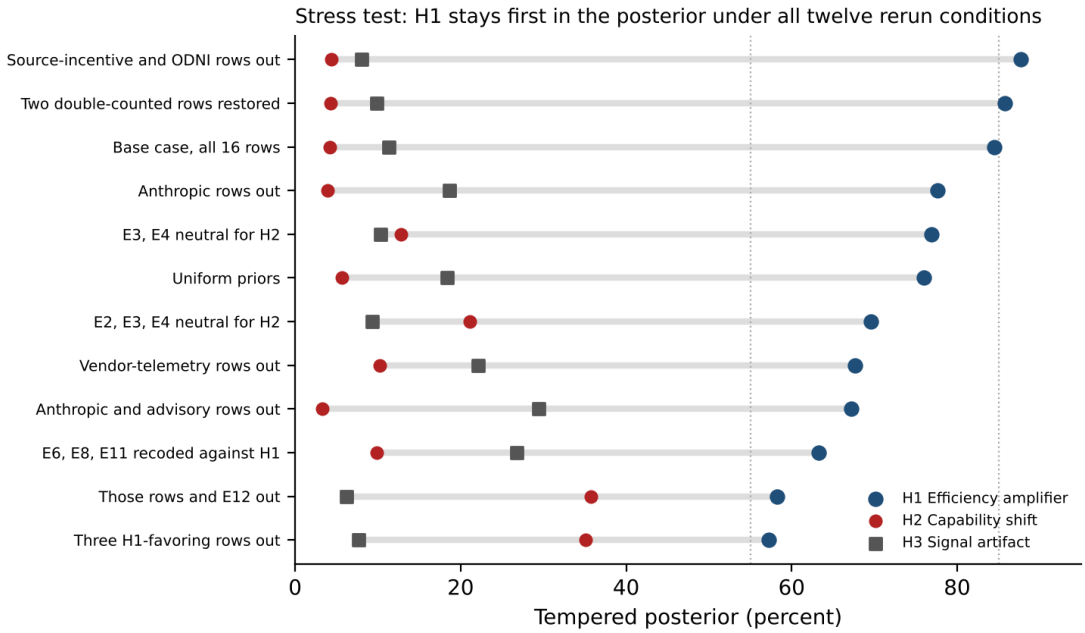


Figure 7. *Stress test of the hypothesis posterior under twelve test conditions.*

11.2 Hypothesis Evolution Tracking

The Hypothesis Evolution Tracking (HET) method follows each hypothesis as evidence arrives. We admitted the evidence in three stages, ordered by who produced it.

Table 17. *11.2 Hypothesis Evolution Tracking*

Stage	Evidence admitted	H1	H2	H3
1	Vendor telemetry and vendor-attributed reports (E1, E2, E6, E7, E10, E14, E16)	0.0	2.22	1.25
2	Adds researcher and platform reach findings (E3, E4, E5, E11, E13)	0.36	5.02	1.25
3	Adds government advisory, EU count and source-incentive check (E8, E9, E12, E15)	0.36	5.02	2.53

Stage one used vendor telemetry alone. H1 carried no inconsistency, H3 carried 1.25, and H2 carried 2.22. A reader who sees only lab reports calls H3 the nearest rival and treats H2 as unproven.

Stage two added the researcher and platform findings on reach, and H2 more than doubled to 5.02, because OpenAI and ISD describe small audiences. H1 picked up its only inconsistency, from the billion-view networks.

Stage three added the government advisory, the EU count and the source-incentive check, and H3 rose from 1.25 to 2.53, since the advisory contradicts it. The incentive row supports H3 but weighs too little to offset an independent government finding.

Copyright 2026 Treadstone 71 - www.treadstone71.com

The order of admission changes the story a reader sees. A reader of vendor blogs sees a different case from a reader of researcher reports, and neither sees the whole file. The tripwires in Section 12.6 tell HET what to watch: the next stage opens when effect data or local-model evidence arrives.

11.3 Dynamic Bayesian Narrative Analysis

The Dynamic Bayesian Narrative Analysis (DBNA) turns the matrix into posterior probabilities. We set the priors before the matrix at 45 percent for H1, 30 for H2 and 25 for H3. The outside view in Section 10.7 informed them, and no formula produced them. Analyst judgment set all three.

Likelihoods follow the matrix. A consistent item earns 0.5 plus 0.3 times a strength factor, a neutral item earns 0.5, and an inconsistent item earns 0.5 minus 0.35 times the factor. The factor runs 1.0 for High diagnosticity, 0.6 for Medium and 0.3 for Low.

Table 18. *11.3 Dynamic Bayesian Narrative Analysis*

Step	H1 Efficiency amplifier	H2 Capability shift	H3 Signal artifact
Prior	45	30	25
After stage 1	88.7	6.4	4.8
After stage 2	94.1	0.4	5.5
After stage 3	96.5	0.4	3.1
After stage 3, evidence dependence tempered (exponent 0.5)	84.5	4.2	11.3

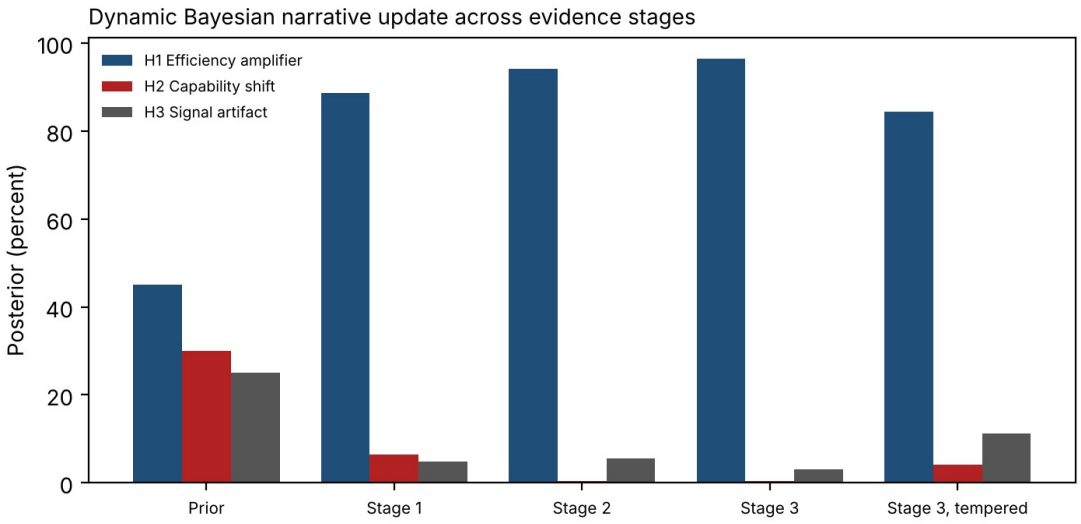


Figure 8. *Posterior probability of each hypothesis across the evidence stages.*

Straight updating gives H1 96.5 percent. That figure overstates. The evidence items still overlap: one Anthropic report feeds two items, and the distillation advisory sits beside Anthropic's own report. Independent updating counts shared information twice.

Tempering halves the weight of each update, on the theory that overlapping reports carry about half the independent information their count suggests, and the tempered posterior reads H1 84.5, H2 4.2 and H3 11.3.

The tempered figure sits half a point under the lower edge of the Highly likely band, and Adequate rows carry the matrix. Under the ASA Evidence Standard, Adequate evidence caps at Likely, so we print Likely (55 to <85) for

Copyright 2026 Treadstone 71 - www.treadstone71.com

H1. Three limits back the cap: the hypothesis is broad, the likelihoods are ours, and the evidence leans on one source type. Read the posteriors as an analyst index. Nobody measured a frequency.

11.4 Devil's advocacy

The opposing case argues that H2 already holds and the evidence hides it.

Labs see only their own users and publish what they choose. Recorded Future reports a self-hosted stack at CopyCop, Genians reports local model tooling on a Kimsuky host, and Google reports one hedged sentence on a Chinese intrusion actor suspected of hosting an open-weight model in victim clouds. All three sit out of any lab's sight. Anthropic's February report counts over 16 million exchanges drawn from its models by three Chinese labs, and its September report counts more. Those are provider counts and no government counts. Google reports that an actor finished a mass credential harvest in under six hours with an agent. Gambit says another operator ran 75 percent of remote commands through a model. Nobody has measured belief change, so every "limited reach" finding measures reach as a platform counts it and says nothing of what a reader came to believe or did afterward. The limited-reach findings come from OpenAI, a firm that gains standing when it reports containment, and from ISD, a research body with its own funders. Mandiant, also a firm, reports breaches. A capability shift looks exactly like the 2026 record: loud in the margins, silent where the observers have no view. The case carries weight. We hold the verdict. Each item in the case shows capability or intent. None shows effect, and none separates a shift from an efficiency gain. Independent researchers report limited reach, and no government source reports otherwise. We keep H1 in front and keep H2 alive. The review triggers in Section 12 name the observations that move it.

A second test faces the enterprise judgment. The opposing view says enterprise AI exposure is overrated, since criminals drive the cases and states play little part, and IBM's figures come from self-report. Answer: the FBI FLASH on TeamPCP, an A1 source with a T-Score of 80, documents the gateway compromise directly. The exposure holds. Who stands behind AI-stack compromise stays unproven. The NCSC chief executive is reported to link about 75 percent of incidents against UK critical infrastructure to states, and the figure does not isolate AI systems. Microsoft, at high confidence, and CrowdStrike name North Korea-nexus actors behind the Mastra injection. As of 7 October, FBI, CISA, NCSC, ASD and CCCS pages hold no government statement that says the same. The 22 September WaterPlum advisory covers npm malware and does not name Mastra.

11.5 The effect ceiling

The ASA Measurement of Effect standard (ASA-MOE) is a draft whose comment window opens on 20 October 2026 and closes on 20 November, with version 1.0 due in January 2027. ASA-MOE ranks study designs and sets a ceiling on any causal claim. Randomized assignment (DS-1) reaches Almost certain, a stepped-wedge rollout (DS-2) reaches Highly likely, and a matched comparison (DS-3) or a synthetic control (DS-4) reaches Likely. A pre and post measure with no comparison (DS-5) tops out at Roughly even chance. Source grades lower the claim and never raise it. A campaign measured only by its own operator grades no higher than C on reliability until an independent party verifies the design, the data and the computation.

We tested every effect statement in the record against the ladder.

Table 19. 11.5 The effect ceiling

Effect statement in the record	Who measured	Design found	Ceiling on a causal claim
More than one billion views for two pro-Iran networks (IR-7)	Researcher, platform counts	Counts, no comparison	None. Proves presence

Effect statement in the record	Who measured	Design found	Ceiling on a causal claim
Firebird videos reached 1.6 million views (RU-5)	Vendor, platform counts	Counts, no comparison	None. Proves presence
Matryoshka kept reach limited in Germany (RU-7)	Researcher	Observation, no stated comparison	None. Proves presence
No meaningful breakout for a PRC-linked operation (CN-2)	Platform operator	Operator-side observation	Below DS-5, and capped at C
Modeled count of chatbot answers that cite Russian state-attributed sources (RU-3)	Modeler	Model built on a test-set rate	None. Not a measurement
27 percent of 540 EEAS-detected 2025 incidents involve AI (RU-9)	Government body	Count. The nearest definition is a summary line, "one in four detected incidents involved the use of AI tools to produce or distribute content". The coding rule is unpublished	None. Counts incidents, not effects
Russian state-attributed sources cited in nearly 18 percent of 300 chatbot test responses (ISD, Section 4.2; RU-3 models from it)	Researcher	Citation count in a researcher-built test set; neutral prompts 11 percent; two links from the Pravda network	None. Proves citation, not belief
One in four malicious breaches is AI-enabled (EN-8)	Survey of breached firms	Self-report	None. Respondents report on themselves
No attitude change among Twitter users who engaged IRA accounts (Bail, 2020; estimates read from figures)	Academic panel	Two-wave matched comparison, 44 treated respondents in the main model	DS-3 on design, Likely at most for a claim about those users. Every credible interval includes zero and the intervals are wide
Pro-China influencer content raised favorability toward China by about 0.10 per minute watched on a four-point scale (Incerti and colleagues, 2026 preprint)	Academic team	Randomized forced exposure for 20 percent of 8,600 respondents, instrumental-variable estimate	DS-1 on design, lowered by preprint status. Tests content, not an operation

No ledger row from the 2026 cases states a comparison design. Not one effect claim there reaches DS-5. The last two rows come from academic studies outside those cases, and neither tests an AI-enabled operation. Every claim of changed belief or behavior stays below the bottom rung, DS-5. The report makes none. Sections 3 to 9 claim presence, volume, reach as counted, and capability. The laboratory studies and the TikTok experiment in Section 10.7 use randomized assignment, which sits at DS-1, and they test persuasion per exposure. None tests a live operation. The effect judgment in Section 11.1 rests on the absence of any sign of effect from a live operation, and it carries the confidence that implies.

11.6 The trap map

The Decoy's Dilemma exercise trains analysts against reflexive control, the technique of handing a target prepared information so the target reaches a choice that feels free and came from planning elsewhere. Its ten phases name ten traps. We ran no scored instance. A score belongs to the person who takes the test. The table maps each trap to the way an AI-influence reader meets it in the 2026 record.

Table 20. 11.6 The trap map

Phase	Trap	How the trap appears in the 2026 record	Counter
1	The Provocation	An Iranian or Russian item written to draw a public rebuttal, like the Mehr commentary of 14 September, whose headline frames Anthropic's finding as an attack on American forces that Anthropic's text does not claim for the Iranian clusters. The commentary denies none of the finding	Deny the stimulus. Record the statement, grade it C, move on
2	The Semiotic Lure	A persona or outlet built to attract a researcher, like the forged Firebird brand	Test the source before the claim
3	Narrative Inflation	Billion-view counts and "AI-enabled" percentages that grow with each retelling	Brief on intent and reach, and leave the headline figure out
4	The Kinetic Feint	A loud chatbot-poisoning story that pulls budget from supply-chain defense	Rank by ATCRI across the full list
5	The Ego Trap	A flattering claim that a firm's model is a prime target	Ask who gains from the flattery
6	The Sunk Cost Steer	Defending a collection plan after the evidence moved to local models	Re-baseline each quarter
7	Information Pollution	Poisoned open-source and chatbot answers feeding the analysis itself	Look for what is missing, and trace every claim to its originator
8	The False Choice	"Ban the assistants or accept the risk"	Strip permissions, keep the tool, document the reason
9	The Moral High Ground	Alert fatigue from constant attribution pressure	Hold a standing cadence and a handover rule
10	The Final Steer	Post-report spin, with vendors and states reusing the findings	Control the distribution of the assessment

The map earns its place in a section on tests because the adversary studies the reader too. A report on AI influence is itself a target for influence.

11.7 Reporter dependence

Section 10.2 counts rows by originator. Reporter-Dependence and Echo Mapping (RDEM) counts cases and the outlets that repeat them.

Section 1.6 names nine Anthropic rows. Eight carry case numbers, 18 in all: IR-1 one, IR-2 three, IR-8 one, RU-1 one, CN-1 three, CN-10 one, PX-1 four and EN-12 four. Second-source searches cover 10 of the 18. Two hold a second source at case level: GTG-50014, which Google's tracker reports as UNC6240, and GTG-20006, which Microsoft's CaptiveCrunch report covers. That reads 2 of 10, Wilson interval 6 to 51 percent. Eight drew no search: GTG-14020, 24015, 34001, 50020, 50029, 54006, 54009 and 84002.

Table 21. Seven single-origin cases, relay items found and independent sources at case level.

Case	Relay items	Independent
GTG-14021 and 14022 (CN-1)	7	0
GTG-34007, 30006 and 30004 (IR-2)	6	0
GTG-84006 (IR-8)	3	0
GTG-10007 (CN-10)	5	0
GTG-50027 (PX-1)	6	0
Kimsuky local stack (KP-3)	8	0
UNC6508 local models (EN-3)	8	0

Copyright 2026 Treadstone 71 - www.treadstone71.com

Of the 43 items, 37 show as search-result text alone, so the count is a floor. The Mali case shows the pattern. Six items appear, and one states that nothing independent confirms it.

Two Moderate judgments rest on one row and one origin each: KJ3 on RU-2 (DFRLab) and KJ9 on EN-3 (Google). Neither print carries the step-down that Section 1.4 sets for one reporter, and that section exempts findings on published artifacts. The ledger does not say whether either row qualifies. KJ2a rests on IR-1, RU-1 and CN-1, three rows from one reporter, and prints Low. KJ1 keeps KP-1 and KP-3, two of five rows, once Anthropic leaves. No printed number moves.

Confidence sits at Moderate for the origin counts and Low for the relay counts.

11.8 Cliff distance

Threshold and Cliff Sensitivity Analysis (TCSA) measures how far each printed confidence sits from its band edge, and how often the spread between two readers carries it over.

Seven judgments print Moderate: KJ1, KJ2, KJ3, KJ4, KJ5, KJ8a and KJ9. Each falls to Low when its weakest supporting row drops under 59.5, since T rounds half up. Shifts come from 24 rows rated twice by different readers. The gaps, second reading minus first, run a mean near -4 with a standard deviation of 6.5, and the pairs sit in the Collection and Analysis Record. A final column removes the mean and keeps the spread.

Table 22. *Weakest supporting row per Moderate judgment, and the gaps of 24 that push it under the floor.*

KJ	Row	T unrounded	Margin	Gaps that flip	Spread only
KJ8a	RU-9	59.69	0.19	17	10
KJ1	CN-1	60.33	0.83	17	9
KJ2	CN-2	63.50	4.00	10	8
KJ5	CN-7	63.92	4.42	10	8
KJ9	EN-3	68.33	8.83	8	1
KJ3	RU-2	75.00	15.50	1	1
KJ4	EN-1	80.00	20.50	1	0

Section 16.2 flags KJ1 and KJ8a. KJ2 stands as near the floor as KJ5, and CN-2 has no second reading. CN-7 has one, at 68, so KJ5 keeps a distance and no fall. The gaps come from rows near an edge and one house reader. Read the counts as a ranking of distance. They carry no probability.

Probabilities show the same pattern. Six of the ten forecasts in Section 12.12 stand 5 points or less from the nearest Kent edge: 2, 3, 5, 6, 8 and 9. Three of five paths do too (C, D and E), and the H1 posterior of 84.5 sits 0.5 under the edge at 85. Each band holds as printed. KJ2 gains a third cliff flag. No printed number changes.

11.9 Hindcast audit

Forecast Calibration and Hindcast Audit (FCHA) rescores the 50 windows of Section 12.6 and tests how much of the gain over the pooled rate survives a change of baseline, sample or search miss.

The judged chances score a Brier of 0.140 against 0.172 for the pooled rate of 22 percent, a skill score of 0.18. Calibration holds in four bins (reliability 0.001). Holding the baseline out by indicator gives 0.185. A bootstrap over the 16 indicators (seed 71, 20,000 draws) puts the difference at -0.026, with a 5th to 95th percentile range of -0.056 to +0.003. The judged chances win in 93 percent of draws, 99 percent against the fixed pooled rate.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 23. Brier score by half-year window from 1 October 2024.

Window	Scored	Events	Judged	Pooled rate
W1	15	2	0.100	0.123
W2	13	1	0.088	0.091
W3	12	3	0.152	0.188
W4	10	5	0.253	0.328

T4 adds 0.587 of a 1.581 total, L4 0.360 and L2 0.304, while L7 subtracts 0.199 and L10 0.096. The event rate climbs from 8 percent in W2 to 50 percent in W4 as the mean chance holds near 25 percent, so the constant hazard under-forecasts the latest window. Reporters cluster near the period end, so recency lifts W4 on its own. If 10 percent of the 38 windows coded no at low confidence hide an event, the Brier reads 0.184 against 0.208. Those 38 are all but one of the 39 windows coded no. The lead ends between 20 and 30 percent. Section 12.6 says the judged chances beat each baseline. The sentence holds against a fixed or held-out baseline. Refitting the baseline inside each resample thins it. No printed number moves.

12 Outlook in measured baselines, resolvable paths, indicators and dated forecasts

Foresight starts from the record and stops where the record stops. Section 12 counts first. Three indices come straight from the ledger: the share of AI-use rows that document local models, the share that come from the model providers themselves, and the breadth of vectors documented for each actor. Each later number says whether we counted it, modeled it or judged it. Judged inputs carry a sensitivity test in 12.16. Percentages print as whole numbers for lookup, and the inputs support nothing finer than steps of about five points.

The outlook runs from 6 October 2026 to 31 December 2027. Five paths cover it. Each path resolves on a rule that a reviewer applies to a published record. A path is never a mood. The shares come from the model. We set no target for them. Counts come first, judgment last.

Copyright 2026 Treadstone 71 - www.treadstone71.com

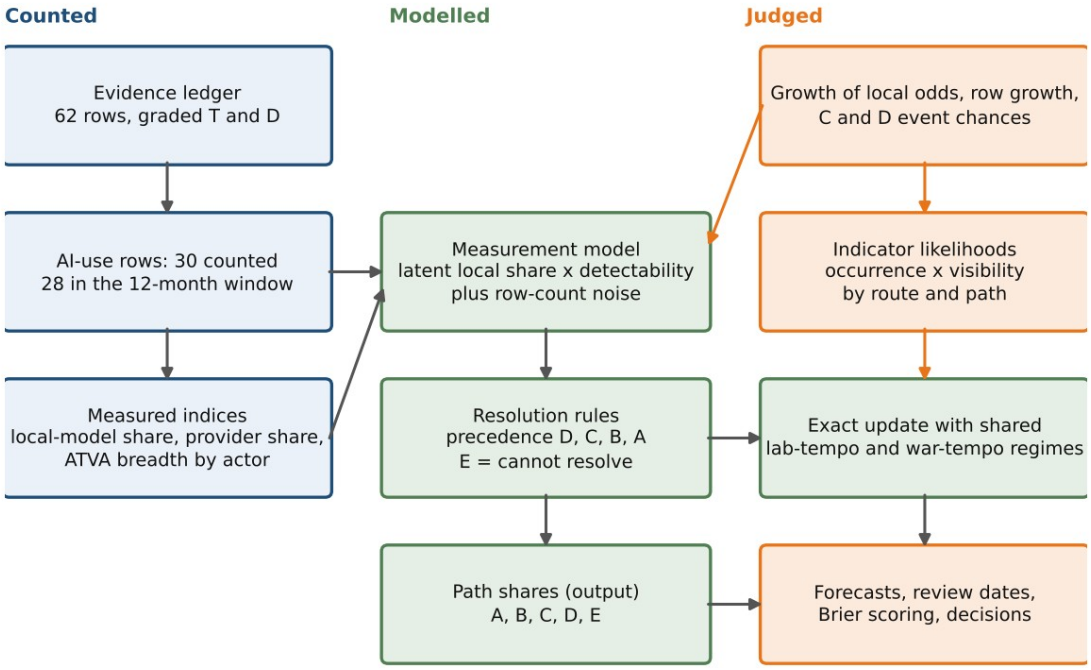


Figure 9. *Logic of the outlook: what is counted, what is modeled, what is judged.*

12.1 How the outlook is built

Section 12.18 reads the weeks to 9 November on its own. Three input kinds feed the outlook. Counted inputs have rows behind them. Modeled inputs follow from counted ones through stated arithmetic. Judged inputs are ours. Section 12.6 scores their indicator chances against a hindcast, which is no out-of-sample test. The assumptions below carry the weight. Nothing hides here. A reader who rejects one has the exact place to change the result.

Table 24. *Outlook assumptions and what follows if each fails.*

ID	Assumption	If it fails
O1	A rerun of the protocol in 2027 draws rows at the effort that built the baseline set, scaled by O5. The baseline itself was uneven: 25 of its 28 rows sit in the second half of the window	Counts move with our reach. The rerun's effort sets how often Path E resolves, and Path B resolves on our effort as much as on the adversary. We log effort each quarter
O2	A model provider sees its own platform and little of what runs on an actor's own hardware. The visibility figures of 12.6 (provider disclosure at 40 percent under Path B and 55 to 90 elsewhere) are analyst judgment	The record holds counterexamples: Anthropic's GTG-17002 and GTG-50027 cases and the hindcast window for L10. OpenAI's February 2026 case study points the same way, though its local-model detail is a banned user's claim that OpenAI left unconfirmed. Where providers see local use, the quiet-lab inference weakens. Section 12.8 tests it at 60 and 75 percent
O3	Non-provider detectors see local-model use at half the rate they see commercial use, with a range of one quarter to one	The latent local share and Path B move (12.16). At one quarter, KJ14's 73 percent falls to 42 (12.2)
O4	The odds of local-model use rise by about 2.5 times from the baseline to mid-2027, with a spread of 0.9 on the log-odds scale	Path B runs from 19 to 41 percent as growth moves from 0.3 to 1.5, and B passes A at 1.2 (12.16)
O5	Rows per window grow by a median 1.25 times as activity and reporting rise	Path E grows when counts stay flat

Copyright 2026 Treadstone 71 - www.treadstone71.com

ID	Assumption	If it fails
O6	A probe or retrieval study with confirmation arrives by the end of 2027 at 19 percent, set as 35 percent that someone runs it and 55 percent that it finds distortion	Path C moves almost one for one
O7	A shock with comparison evidence arrives at 5 percent	Path D moves one for one
O8	Given the path, indicators fire independently except through two shared regimes: lab disclosure tempo and war tempo	Evidence piles up faster than the model says. 12.8 tests loadings up to 1.4
O9	Four states move the paths: Iran, Russia, China and North Korea	A state outside the four drives the visible events. The premortem in 12.15 names it
O10	An indicator's chance of occurring is the judged chance, pulled toward its 2024 to 2026 hindcast rate with the judgment weighted as four half-year windows	The hindcast holds 50 scored indicator-windows drawn from public sources, L5 stays unscored, and the judged inputs overlap the period tested. Chances move (12.6)
O11	Path B resolves when local-flag rows reach 30 percent of non-provider rows in the 2027 window, and the window holds at least 8 non-provider rows. The same 8-row minimum separates A from E	A threshold of 25 percent puts B ahead of A, and 27.5 percent does not. Row minimums of 6 and 10 move E between 5 and 13 percent and leave A ahead (12.16)

Two uncertainties set the paths. Where does the work run: on commercial platforms that labs watch, or on models that states hold? And does anyone measure effect? Section 12.3 crosses the two.

12.2 Measured baselines

The unit of count is a ledger row at the rung "AI use shown" or "Audience counted", since each such row records an adversary using AI. Rows that qualify number 30. Of them, 28 fall inside the window of 7 October 2025 to 6 October 2026. That window sets every baseline. Two rows fall outside it: RU-6, Recorded Future's CopyCop report, which its page dates 18 September 2025, and CN-12, the Vanderbilt release on GoLaxy documents. The Vanderbilt release carries no date, and the press coverage it lists falls in August 2025. Section 16.1 lists every counted row with its codes. The headline indices stay on those 30 rows. Extended counts add 21 reports and two slices of merged rows, and then 14 reports from the last quarter of 2025 and January 2026. The extended table in Section 16.1 lists all 67 rows.

Four counting rules fix the indices.

Provider telemetry. A row carries the flag when its finding rests on a model provider reporting on its own platform: Anthropic, OpenAI, Google or Microsoft. Twelve windowed rows carry it.

Local-model evidence. Any row that documents an actor running a local or self-hosted model for operations carries the flag. Three rows do: RU-6, KP-3 and EN-3. The adjudication below explains the third.

Local-model share (DLS). The count of local-flag rows, divided by rows without the provider flag. Behind the share sits one premise: a provider sees little of a model that runs on the actor's own machines. So provider rows leave the denominator, and a row flagged provider and local drops out of the numerator too. O2 names exceptions in the record.

Provider-telemetry share (LTS). Rows with the provider flag, divided by all counted rows.

Row EN-3 needed a ruling. The row is Google's AI Threat Tracker of 8 September 2026. Its credential-theft and account-price findings rest on incident response and market observation, and its model-misuse findings rest on Gemini telemetry. The report also says, in a passage whose lead sentence uses the word "suspected," that activity tied to UNC6508, which Google links to the PRC, compromised cloud environments and stood up local LLM infrastructure there. Google reasons the setup avoids commercial API monitoring, and it adds that the group continues to research local use of open models. Providers see none of that use. The passage names no

Copyright 2026 Treadstone 71 - www.treadstone71.com

model, victim or date and states no confidence level. Google's June report on the same cluster holds no LLM content. The record holds no second source, since every press and profile page that repeats the passage cites Google. Coded by the platform of the report, the row falls among the provider rows. We code it as a non-provider row with the local flag, since the purpose of the denominator is to hold the rows that a provider cannot see. The ruling lifts Path B by 11 points. Among coding decisions in the report it is the single most sensitive (12.16). The ruling rests on one passage in one report, hedged in its lead sentence, with no second source behind it. That call is our own.

Row RC-CN-6 follows the opposite convention. The row is OpenAI's report of 25 February 2026 on a user connected with Chinese law enforcement, split from ledger row CN-2. OpenAI's summary line lists locally deployed models as part of the effort. The case study body relays a banned user's claim that provincial teams experimented with DeepSeek-R1 and Qwen2.5, and adds that OpenAI was unable to check it independently. An Axios summary says the campaign likely ran on locally hosted Chinese models after ChatGPT refused. The row carries a provider flag and a local flag, and the local flag is the weaker of the two. Its count enters the provider numerator, and its local flag stays out of the DLS count, because the DLS base holds non-provider rows only. The EN-3 asymmetry is a judgment. EN-3's local sentence rests on no provider platform. RC-CN-6's local evidence sits in a provider's own report, as a claim the provider left unconfirmed. Because OpenAI confirmed no part of it, the row weakens O2 less than a confirmed case does. OpenAI's June report states the local-model point again as its own characterization of the strategy, adds no case, and records no June actor on a local model. The coding holds. Coded the way EN-3 is, as non-provider with the local flag, the row gives 5 of 35, because moving it out of the provider group lifts the base by one. A and B then read 38 and 36 (12.16).

Three more rows count on thin footing. RU-3 is a modeled estimate that applies ISD's citation rate, taken from researcher-built prompts, to ordinary users. The EEAS report behind RU-9 and its three annexes publish no rule for coding AI involvement. EN-4 is a five-agency advisory with no attributed actor. Its body asserts AI-generated exploitation scripts and cites no sample or basis. Each counts once in the non-provider base, since the rung scheme sets the unit and says nothing of evidence quality. Without the three, DLS reads 2 of 13, 15 percent (4 to 42). Recorded Future's page dates RU-6 to 18 September 2025, outside the window. Placed inside, it gives 3 of 17, 18 percent (6 to 41). The flag rests on Recorded Future's reading of its host artifacts, which it words as very likely or almost certain for self-hosting and as likely for the Llama 3 base. Treasury's notice of 31 December 2024 describes a server the GRU helped fund, housed in a rented Moscow apartment, and names a GRU affiliate, not CopyCop. A French text test (arXiv 2608.15746, August 2026) supports the Llama 3 family and also points to Mistral. No source names hardware or a host, so the flag stays single-source for the self-hosting itself.

The CrowdStrike 2026 Global Threat Report moves no count. Ledger row EN-14 records its 58 pages, and the extended table already holds the same report as slice RC-KP-2, a non-provider row without a local flag. CrowdStrike reports forum talk that names DeepSeek as a self-hosted model and describes no operator running a model of its own. The headline 89 percent has no definition and no denominator, and it enters no index.

A third index needs no sampling. ATVA breadth sums the grid in 12.10 for each actor: two points for each Documented vector, one for each Partial, over eight vectors, for a ceiling of 16.

Table 25. *Measured baselines with Wilson 95 percent intervals.*

Index	Count	Value (percent)	Interval (percent)	Reading
DLS, 12-month window	2 of 16	12	3 to 36	Local-model use shows in about one non-provider row in 8
DLS, all rows to date	3 of 18	17	6 to 39	The 2025 CopyCop row lifts the share by about 4 points, from 12.5 to 16.7 before rounding

Copyright 2026 Treadstone 71 - www.treadstone71.com

Index	Count	Value (percent)	Interval (percent)	Reading
LTS, 12-month window	12 of 28	43	27 to 61	Providers supply under half the counted rows
LTS, all rows to date	12 of 30	40	25 to 58	Same reading
DLS, extended rows (21 reports and two slices added)	4 of 34	12	5 to 27	The share holds and the interval narrows. Strict coding (RC-EN-8 not local) gives 3 of 34, 9 percent, 3 to 23. Keeping the RC-CN-6 local flag gives 5 of 35, 14 percent, 6 to 29
DLS, extended rows with October 2025 to January 2026 rows added	4 of 44	9	4 to 21	Ten more non-provider rows and no local flag dilute the share
LTS, extended rows	17 of 51	33	22 to 47	Providers supply a third of the rows
LTS, extended rows with October 2025 to January 2026 rows added	21 of 65	32	22 to 44	The reading holds
LTS, extended rows, one row per September report	8 of 41	20	10 to 34	Anthropic's 10 September report and Google's 8 September tracker count once each
LTS, ledger rows and added reports before September 2026 (15 and 21)	7 of 36	19	10 to 35	The ledger's provider share reflects September
LTS, added reports alone	4 of 21	19	8 to 40	Level with the ledger rows before September (3 of 15, 20 percent)
LTS, October 2025 to January 2026 rows alone	4 of 14	29	12 to 55	Small set, wide interval
LTS, September alone	9 of 13	69	42 to 87	Provider reports drive the month
ATVA breadth, Iran	5 of 16	n/a	n/a	Thinnest column
ATVA breadth, Russia	12 of 16	n/a	n/a	Level with China
ATVA breadth, China	12 of 16	n/a	n/a	Level with Russia. The chatbot cell has no support (12.10)
ATVA breadth, North Korea	8 of 16	n/a	n/a	Middle
ATVA breadth, hired hands	6 of 16	n/a	n/a	One row, PX-1, drives the column
ATVA breadth, criminals	8 of 16	n/a	n/a	Enterprise targets dominate. The local-model cell stands at Partial (12.10)

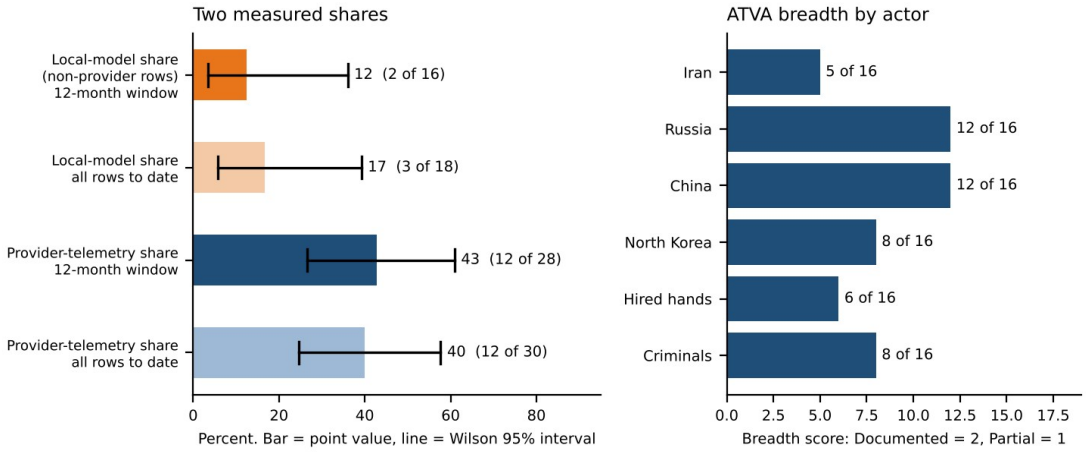


Figure 10. *Measured baselines: two shares with intervals, and ATVA breadth by actor.*

Three cautions apply. First, one row moves the local-model share by about 6 points. That is why the interval spans 3 to 36 percent. Second, the rows cluster. Of the 28 windowed rows, 13 fall in September 2026, and five

reports supply those 13 (Section 10.3). Rows from one reporter move together. That is why the Wilson intervals run narrower than the truth. The report-basis row above is the better guide to the provider share. Third, the unit is a row and not an operation. A report that covers three clusters counts once. Small counts rule.

A local model leaves no trace on a provider's platform. The observed share mixes two things: how much local use exists, and how well anyone sees it. We separate them with a measurement model. If a share lambda of state-linked AI operations runs part of its work on local models, and non-provider detectors see local use at a fraction d of the rate they see commercial use, the observed share equals lambda times d over lambda times d plus one minus lambda. At d of 0.5, a measured 12 percent implies a latent share near 23 percent, with a spread from 8 to 46 percent between the 10th and 90th percentiles. On that setting the chance that the latent share sits under one third today stands at 73 percent. Move d to 0.25 and the figure falls to 42 percent, with a median latent share of 37. At d of 1.0 it rises to 93. O3 allows that whole range, so the same counts put KJ14's bound on either side of a band edge (Likely at 0.5, Unlikely at 0.25). Treat the 73 as a loose bound.

The ledger holds few rows before September, so February through August rest on added reports. Reports that meet the counting rule and sit outside the ledger come to 21. Two more are slices of merged ledger rows (CN-2 and KP-2). Seven failed a counting rule, most often because no identifiable actor appears using AI. Section 16.1 lists them as not counted.

A further 14 reports cover October 2025 through January 2026: four in October, four in November, three in December and three in January. Four carry the provider flag. OpenAI and Microsoft supply the October pair, and Google and Anthropic the November pair. None meets the strict local-model test, which needs a report showing an actor operate a model on hardware or hosting that actor controls. Three carry a loose flag only. Google's report has PROMPTSTEAL calling Qwen2.5-Coder-32B-Instruct through the Hugging Face API, an open-weight model on a third party's hosting. ESET's PromptLock turned out to be a prototype from NYU researchers, with its model reached through an Ollama server. Unit 42 describes malicious-LLM products for sale, and the sellers neither confirm nor deny what runs behind them.

Two of the 14 overlap rows we already hold. Anthropic's November 2025 case was detected in mid-September 2025, published on 13 November and edited on 17 November, and it is a third slice cut from CN-10 with no new independent source. The Taiwan National Security Bureau report of 11 January 2026 is the document held as ledger row CN-6, seen here through a newspaper. Dropping the two overlaps leaves 63 rows and 4 of 43 local, 9 percent, with each path share inside one point of the 14-row result. Two figures need care. Microsoft's count of more than 200 instances of foreign adversaries using AI to make fake content in July 2025 reaches the record through an AP article, and Microsoft's report pages do not carry it. Check Point's VoidLink report of 20 January 2026 is the original of a case that the March digest also covers (12.10).

Extended counts carry limits. December and January are thin, with three reports each. Coverage of the final months of 2025 and of January omits Recorded Future Insikt, Meta, Google's threat bulletins, Mandiant, CrowdStrike and Microsoft's East Asia notes, and holds no native-language source. The only joint government statement inside the window, from the Multilateral Sanctions Monitoring Team on 22 October 2025, holds no AI text and is not counted. Australia's DFAT joint statement is on record, and the alert that carries AI text is dated July 2026, outside the window. Korean press repeats the Genians report of 10 August, and a relay in Chinese dated 15 September claims more than that report states, so KP-3 keeps its single source.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 26. Provider and local shares under nine row sets.

Row set	Counted rows	Provider	Non-provider	Local	LTS (percent, 95 percent interval)	DLS (percent, 95 percent interval)
Ledger rows	28	12	16	2	43 (27 to 61)	12 (3 to 36)
Plus 21 reports	49	16	33	4	33 (21 to 47)	12 (5 to 27)
Same, RC-EN-8 not local	49	16	33	3	33 (21 to 47)	9 (3 to 24)
Plus two slices	51	17	34	4	33 (22 to 47)	12 (5 to 27)
Plus two slices, RC-CN-6 local flag kept	51	16	35	5	31 (20 to 45)	14 (6 to 29)
Plus seven reports that failed a rule	58	18	40	4	31 (21 to 44)	10 (4 to 23)
One row per September report	41	8	33	4	20 (10 to 34)	12 (5 to 27)
October 2025 to January 2026 rows added	65	21	44	4	32 (22 to 44)	9 (4 to 21)
One row per September report, October 2025 to January 2026 rows added	55	12	43	4	22 (13 to 34)	9 (4 to 22)

Three readings follow. The claim that providers supply under half of counted rows survives every variant and strengthens, from 43 percent to about a third, and to a fifth by report. Rows from those 14 reports leave that reading in place, at 32 percent of 65 rows and 22 percent by report. Local share stays near one non-provider row in 7 to 11. Its interval narrows only from 3 to 36 to 5 to 27, and to 4 to 21 with the added reports in. Ten of those reports are non-provider rows, none with a strict local flag, so the share slips from 12 to 9 percent. Read the loose flags on ESET and Unit 42 as local and the count is 6 of 44, 14 percent (6 to 27). We do not count them, since neither report shows an actor operating a model it controls. Two local flags among the 21 added reports are weaker than the ledger's. RC-EN-3 records forum actors who install uncensored local models and price rigs. Check Point calls the use growing but limited. RC-EN-8 records one actor who drove an agent from someone else's exposed Ollama server and also tested open-weight models. If the two come out, the extended local count falls to 2 of 34, 6 percent, 2 to 19. Section 12.16 shows what each variant does to the paths.

12.3 Five paths and how each resolves

A path is the way 2027 closes. The rules below apply in a fixed order: D first, then C, then B, then A. E catches the cases the rules cannot resolve. One path holds on 31 December 2027. So the shares sum to 100. No path overlaps another.

Table 27. Five paths, their resolution rules, shares and bands.

Path	Resolves when	Share (percent)	Band	Confidence
A. Ordinary staff equipment	None of D, C or B holds, and the 2027 window holds at least 8 non-provider rows	39	Unlikely (15 to <45)	Low
B. Local and sovereign shift	In the 2027 window, local-flag rows reach 30 percent of non-provider rows, and the window holds at least 8 non-provider rows	30	Unlikely (15 to <45)	Low
C. Poisoned-machine era	A study or vendor test shows reproducible, measured distortion of a major chatbot's answers or an enterprise retrieval system by named adversary content, and a second team or the vendor confirms	18	Unlikely (15 to <45)	Low
D. Shock event	A government or lab states that an AI-enabled operation changed a named election, market or mobilization outcome, with comparison evidence at DS-3 or better	5	Highly unlikely (5 to <15)	Low
E. Cannot resolve	No D or C, and fewer than 8 non-provider rows in the window, or the counting protocol did not rerun	8	Highly unlikely (5 to <15)	Low

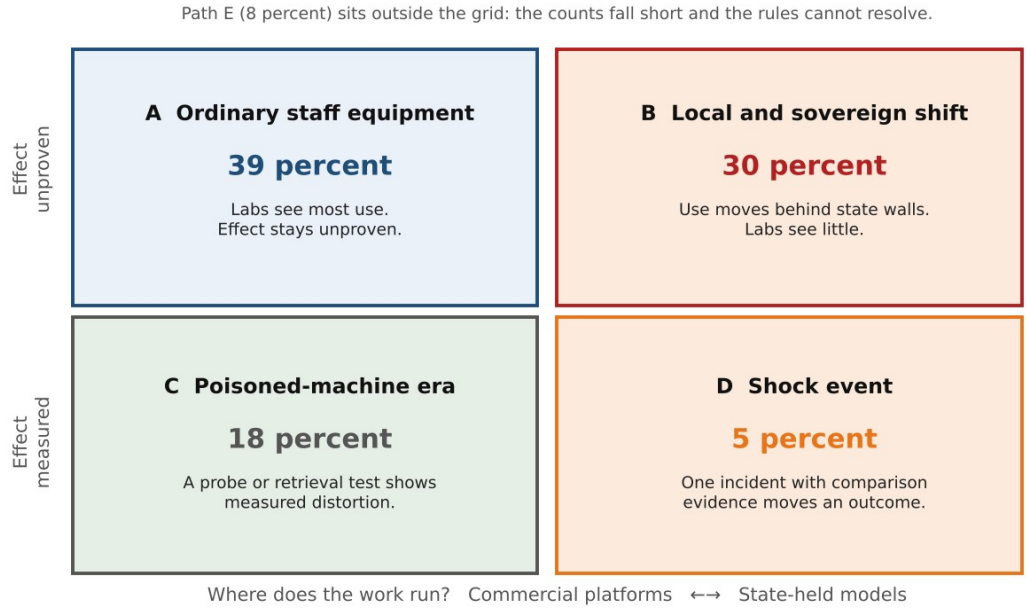


Figure 11. The four named paths on two uncertainties. Path E sits outside the grid.

Confidence stays Low on every path. Judgments set the inputs, and the hindcast of 12.6 tests part of them in sample. The model supplies arithmetic and adds no evidence. Path A is a plurality: 39 percent, so 61 percent of futures land elsewhere. A and B are not separable. The 8.6-point lead of A over B moves with the coding of RC-CN-6, which ties the two, and a rule threshold of 25 percent or growth of 1.2 in the local odds reverses the order (12.16). Read the headline as "A and B are close", never "A holds". The headline rests on the 30 ledger rows alone. Added rows move the shares, and 12.16 shows by how much. Section 10.7 gives the outside view. The best-measured earlier case, the 2016 Russian operation on Twitter, found exposure concentrated in a small group and no measurable shift in attitudes. That reference class favors A, and it rests on one study. A 2026 TikTok preprint on state-linked influencer content shows a measurable shift and cuts the other way (Section 10.7). D sits on a band edge at 5.00 percent, and a rounding choice moves it between two bands.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Path B resolves on an imperfect rule. The latent local share passes one half in 40 percent of runs by 31 December 2027. Then the counting rule fires 74 percent of the time. If the latent share stays under one half, the rule still fires 18 percent of the time. Sampling noise alone does it, and those false alarms supply 11 of the rule's 40 points of firing. A reader who sees Path B resolve should hold the two numbers together.

12.4 Drivers read through STEMPLES Plus as a forward scan

Treadstone 71 describes STEMPLES Plus as an environmental scan. Its Plus adds indicators of change, with Demographics, Religion and psychological factors beside the base set. Section 6.6 read the present through twelve factors. The table below reads the same factors forward. Each row names a driver, the paths it pushes, the actors it touches and the indicators that signal movement. The strategic foresight course lists the cone and strategic foresight analysis among its methods.

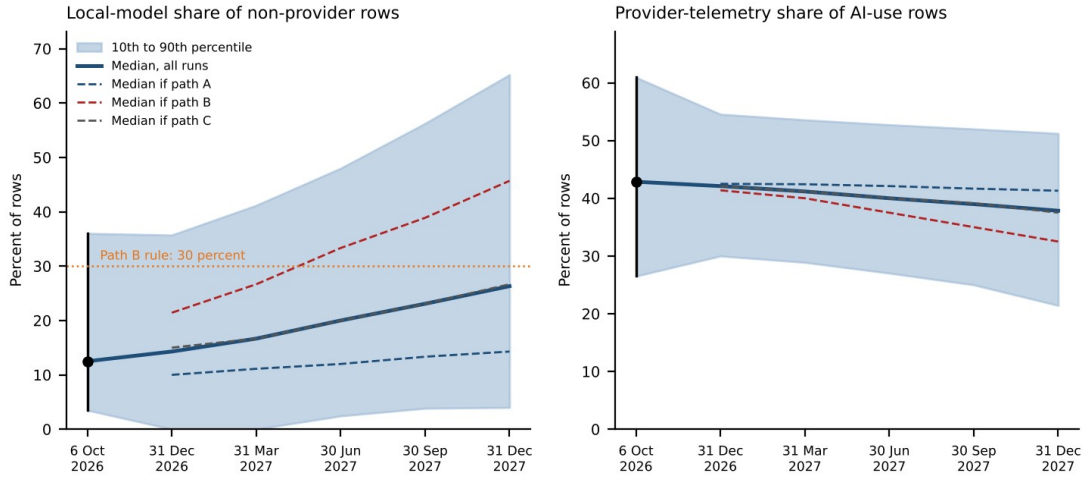
Table 28. *STEMPLES Plus forward scan, one driver per factor.*

Factor	Forward driver	Pushes	Actors touched	Indicators
Social	Shared prompt practice, contractor markets and facilitator networks spread one workflow across units	A	All four states and the hired hands	L9
Technical	Open-weight models run on hardware the labs never see. Agent frameworks widen the supply chain	B and C	Russia, China, North Korea	T2, T3, T7, L1, M1
Economic	A fall in inference prices is possible. No ledger row measures price, and L4 tests it. Influence for hire sells to states with thin capacity	B	China, hired hands	L4
Military	The war with Iran. GRU-linked clusters. Government attribution of an AI supply-chain compromise	C and D	Iran, Russia, North Korea	T5
Political	Midterms on 3 November 2026. US posture on foreign influence. North Korean silence on AI	D	Iran, Russia	T4, T6, L5
Legal	Russian model classes from March 2027. Chinese content labels since 1 September 2025. Claude blocked inside Iran	B	Russia, China, Iran	L2, L3
Educational	State academies and universities write AI doctrine	A, slowly	China	L7
Security (internal)	Surveillance units take taskings from officials. Bureaus run sentiment monitoring	A	Iran, China	L10
Religion	Faith frames for propaganda and faith communities as targets	A and D at the edges	Iran, China	L8
Demographics	Diaspora, Persian, Tibetan, Uyghur, European and Latin American audiences. Western remote-hiring markets	Widens the target set	All four states	L8
Social psychology	Trust levers: forged brands, interview trust and chatbot answers	C	Russia, North Korea	T1, L6
Strategic	Each state follows its mission. Lab visibility falls if use moves to local models	B	All four states	T8, M1

Five drivers do most of the work. Price: cheaper open-weight models favor B, if the price falls. Law: Russia's model classes and China's labels change what a state body does in the open and what it does behind a wall. War and elections: the war with Iran that began on 28 February and the US midterms supply motive and dates. Visibility: a lab that changes what it discloses moves the record without moving the adversary. Measurement: the first comparison study of effect changes the argument, and until it exists the debate runs on counts.

12.5 Cones of plausibility on measured indices

A cone of plausibility draws the range of futures that stay inside the evidence. Treadstone 71 uses it to hold diverging outcomes, driven by intent, capability shifts and triggers, in one picture. Ours start at a counted value. The fans are wide. They widen through the model of 12.2 and the inputs of 12.1. We ran 300,000 trajectories (seed 71). Each draws a baseline, a growth in latent local use, a detectability and a count of rows, and then reads the index as a reviewer reads it.



Point at 6 Oct 2026 = measured value with Wilson 95% interval. Later fans = model with sampling noise. Trailing 12-month windows.

Figure 12. *Cones for the local-model share and the provider-telemetry share, trailing 12-month windows.*

Read the left panel first. The measured local-model share starts at 12 percent. In the window that closes on 31 December 2027 the median reads 26 percent, with a 10th to 90th range of 4 to 65. The fan includes sampling noise from small counts. Medians split by path: 14 percent under A, 46 under B and 27 under C. The 30 percent rule sits just above the median, and the rule fires on 40 percent of runs.

The right panel moves little. Providers supply 43 percent of rows today on the fixed protocol, and 33 percent on the extended rows of Section 12.2, or 32 percent once the October to January rows join. The median ends near 38 from the 43 percent start. Under B the path median falls to 32 percent. Provider rows and other rows grow together, so a falling provider share signals B weakly. Indicator M2 targets that weak signal, and 12.6 says why it sits outside the models.

Copyright 2026 Treadstone 71 - www.treadstone71.com

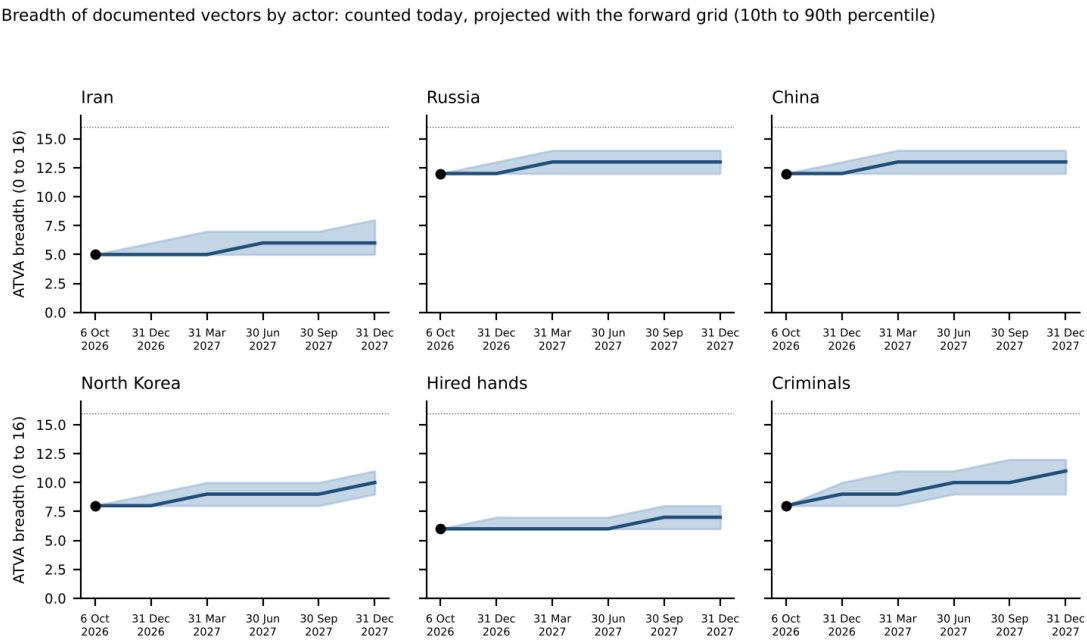


Figure 13. *Breadth of documented vectors by actor: counted today, projected with the forward grid.*

The breadth cones run by actor, and each starts at a count. Each cell of the forward grid in 12.10 moves up one step with its own chance. Constant hazards set the dates. China moves from 12 to a median of 13, with a range of 12 to 14. Russia starts at 12 and ends at 13. The two are held back because 5 of their eight vectors already stand as Documented. North Korea runs from 8 to 10, Iran from 5 to 6, hired hands from 6 to 7 and criminals from 8 to 11. The index measures documentation. A rise means the actor moved or the reporters looked. The index cannot tell which.

12.6 Indicators of change

Nineteen indicators watch the five paths. Eight are tripwires. Ten lead them. One, M1, reads the simulated row counts, and a twentieth, M2, appears in the table as withdrawn. A leading indicator tends to show before a tripwire does. Each carries a check date and a route, which names who must see the event before it counts.

Each indicator's chance of firing under a path equals the chance the event occurs times the probability that the route sees it. The product matters. We judge that a lab sees most of what happens on its own platform and little of what happens on a state's hardware. Silence from the labs then carries information about the path. The information points toward B. Visibility by route and path runs as follows.

Table 29. *Visibility of each route by path, the chance the route sees an event that occurs.*

Route	A	B	C	D	E
Provider disclosure	85	40	75	90	55
Independent report	70	65	80	85	55
Government statement	75	70	75	95	60
Study or audit	60	60	75	85	50
Public act	95	95	95	95	95

The visibility figures are analyst judgment, and the record holds counterexamples to the 40 under Path B. Anthropic reports that the GTG-17002 actor kept a self-hosted model beside Claude, and that a platform built

with Claude's help in GTG-50027 ran on local models. Each rests on Anthropic's own telemetry, and a commentary of 30 September says outright that GTG-50027 lacks independent confirmation. OpenAI's February 2026 case study repeats a banned user's account of local DeepSeek and Qwen use and reports that it was unable to confirm the point. That case study fills the third hindcast window of L10. Section 12.8 tests the quiet-lab inference with the provider figure for B set to 60 and then to 75 percent.

Two rules keep the list honest. An indicator whose chance of firing differs by less than 1.5 times across paths A to D tells us almost nothing, so it stays on the list as watch only and does not enter any update. Three indicators fall under that rule: T6, L3, L4. Nested events are defined to exclude each other. T4 covers a US federal agency and L5 covers platforms, vendors and state election officials. Two shared regimes tie indicators together, a lab-disclosure tempo and a war tempo, so that a burst of reports raises several lab-route indicators at once.

Table 30. *Indicators of change, check dates, the chance each fires, and what each does to the path shares (points).*

ID	Observation	Route	Check by	Fires (percent)	If it fires	If silent	Judgment it tests
T1	Comparison-design study shows belief change from an AI-run state operation	Study or audit	31 Mar 2027	2	D +44, A -27	D -1, A 0	KJ13 falls to Unlikely. KJ3a reopens
T2	State network above one million verified engagements documented on self-hosted models	Independent report	30 Jun 2027	11	B +35, A -26	B -4, A +3	KJ14 reopens: local share above one third
T3	Probe study finds Pravda-network text memorized in a major model	Study or audit	31 Dec 2026	5	C +55, A -28	C -3, A +1	KJ3a up one band
T4	US federal agency names AI-enabled foreign influence aimed at the midterms	Government statement	3 Nov 2026	26	D +11, B -5	D -4, B +2	KJ8 down one band
T5	A government attributes an AI supply-chain compromise to a state actor	Government statement	30 Jun 2027	11	C +11, A -9	C -1, A +1	KJ4 gains a state-actor clause
T6	KCNA or Rodong Sinmun publishes a statement on AI in operations	Public act	31 Dec 2026	8	Watch only	Watch only	DPRK collection plan rewritten
T7	Exploitation of an agent or MCP flaw at scale appears in a primary source	Independent report	31 Mar 2027	44	C +10, B -6	C -8, B +5	KJ9a up one band
T8	A lab reports falling state use of commercial models while a rival source reports a rise elsewhere	Provider disclosure	30 Sep 2027	7	B +42, A -28	B -3, A +2	KJ6 stands as fact. KJ14 reopens
L1	State-linked body ships or cites a self-hosted fine-tune in Persian, Russian or Korean	Independent report	31 Mar 2027	14	B +25, A -17	B -4, A +3	KJ6 and KJ14
L2	Russia publishes the rules that sort models into sovereign and national classes	Public act	31 Mar 2027	51	B +7, A -4	B -8, A +4	Russian rows in 4.9
L3	China publicly enforces its AI content-label rule against a named influence operator	Public act	30 Jun 2027	20	Watch only	Watch only	China row in 5.7
L4	A frontier-class open-weight model gains a 50 percent price cut in one quarter	Market fact	30 Jun 2027	64	Watch only	Watch only	Price driver in 12.4
L5	A platform, vendor or state election official documents an AI-enabled influence incident aimed at the midterms	Independent report	3 Nov 2026	40	D +5, B -3	D -3, B +2	KJ8
L6	A published chatbot audit shows more repetition of Pravda-network claims than an earlier audit	Study or audit	31 Mar 2027	12	C +36, B -18	C -5, B +2	KJ3a
L7	A state academy or university publishes AI cognitive-operations doctrine	Public act	30 Jun 2027	22	B +7, A -4	B -2, A +1	China and Russia doctrine rows
L8	A vendor documents AI-built content aimed at a religious or ethnic community in a new language	Independent report	30 Sep 2027	40	D +3, E -2	D -2, E +1	Regional table 12.11
L9	A lab or vendor documents a hired-hand program that works for a second state client	Provider disclosure	30 Jun 2027	21	A +16, B -14	A -4, B +4	KJ7
L10	A lab or vendor documents a new state surveillance unit that uses AI on a domestic or diaspora population	Provider disclosure	30 Sep 2027	64	B -12, A +10	B +21, A -18	KJ1, security rows

Copyright 2026 Treadstone 71 - www.treadstone71.com

ID	Observation	Route	Check by	Fires (percent)	If it fires	If silent	Judgment it tests
M1	Interim local-model share of non-provider rows reaches 20 percent (12 months to 30 Jun 2027, at least 6 rows)	Counted	30 Jun 2027	50	A -18, B +17	A +18, B -17	KJ14
M2	Provider-telemetry share of counted reports under 16 percent (12 months to 30 Jun 2027, at least 10 reports)	Counted	30 Jun 2027	Withdrawn	Not modeled	Not modeled	KJ6

M1 comes from the simulation of 12.2 and carries no judged likelihood. The indicator fires when the local share of non-provider rows in the 12 months to 30 June 2027 reaches 20 percent, with at least 6 rows. Because M1 and Path B draw on one simulated count process, the lift of M1 on B is partly circular: the model measures itself. Each indicator that is a sub-event of its path's rule, such as T3 for C and T1 for D, shares that circularity. M2 counts reports and not rows, since a row-based share fires on today's counts: extended rows already stand at 33 percent, and row-splitting inside single provider reports drives the share. Firing needs a provider share under 16 percent among counted reports, with at least 10 reports behind the share. The baseline on that basis is 20 percent, and 16 is a relative drop of one fifth from it, the same drop as 35 percent from 43 on rows. No simulated likelihood exists for M2, and M2 stays out of every exact table, the information ranking and the pattern table.

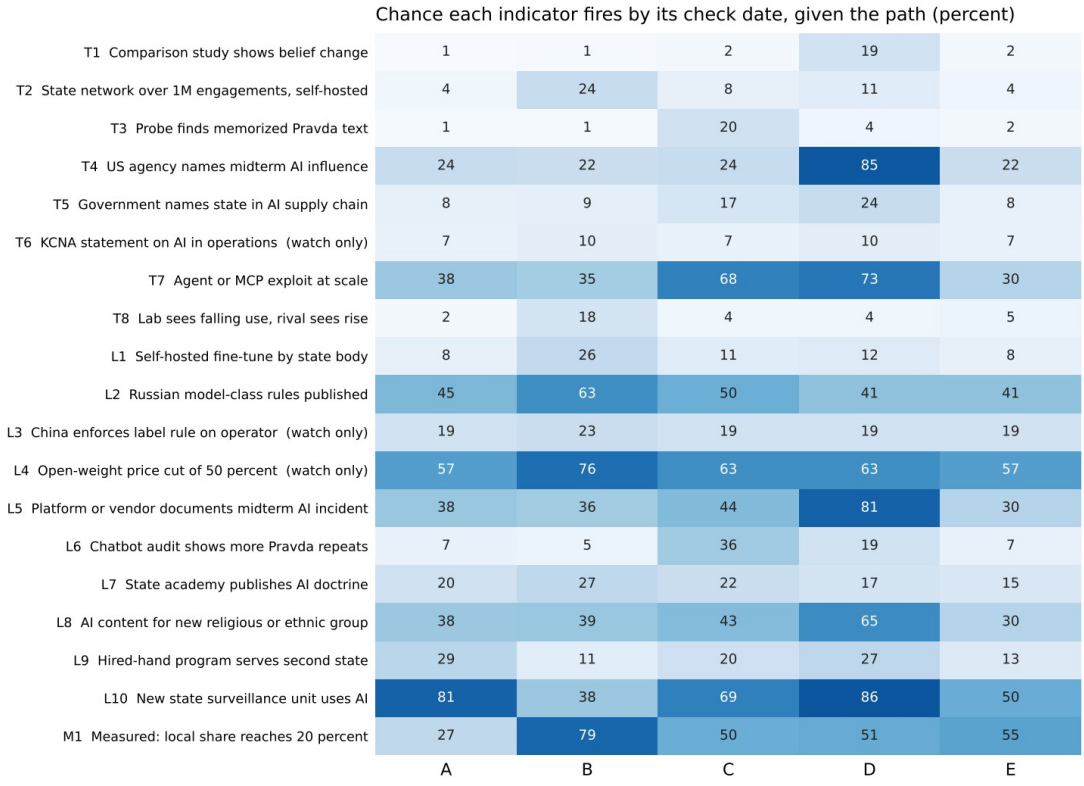


Figure 14. *Chance each indicator fires under each path.*

The moves in the table are exact single-indicator updates, and a reader must not stack them. Two moves from one table add up only when the indicators are independent given the path. They are not. Section 12.8 gives the joint results.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Hindcast scores for the indicators

A hindcast tests the judged chances. For each of 17 observations, public sources show whether the event occurred in each of four half-year windows, covering 1 October 2024 to 30 September 2026. Each indicator's chance by its check date converts, from the hindcast's as-of date of 7 October 2026, to a chance per half year at a constant hazard, and the hindcast compares that chance with what the windows show. Windows marked unclear are left out, 11 of them. Four rulings apply. The one hit for T1 measures agreement with article theses and no belief change, so T1 counts as no in every window. T4 asks about a federal election inside the window, and only the first window held one, so its other three drop out. Incidents aimed at other elections (Ghana, Germany, Ireland, Bangladesh and Moldova) do not match L5, which asks about the US midterms, so all four L5 windows stay unscored. DFRLab's April 2026 completion test of an open base model reproduced RT and CGTN items, and a loose match scores it as a T3 hit. A model with a December 2023 cutoff predates the expansion of the Pravda network, so the test ran on no Pravda text, and the T3 window scores no. Of 68 indicator-windows, 50 stay scored: 11 are unclear, 3 drop with T4 and 4 with L5. Each window spans six months.

Table 31. *Indicator chances against the hindcast, four half-year windows from October 2024.*

ID	Windows scored	Windows with the event	Model chance per half year (percent)	Hindcast rate (percent)	Chance by check date, judged (percent)	After the pull, before the cap on occurrence (percent)
T1	4	0	4	0	4	2
T2	2	0	12	0	16	11
T3	4	0	21	0	10	5
T4	1	1	85	100	25	26
T5	4	0	15	0	21	11
T7	3	2	29	67	28	44
T8	3	0	7	0	13	8
L1	3	0	26	0	25	14
L2	4	2	55	50	54	51
L3	2	0	22	0	30	20
L4	3	2	38	67	51	64
L5	0	n/a	97	n/a	40	40
L6	3	0	22	0	22	12
L7	4	0	31	0	42	22
L8	4	1	21	25	37	40
L9	2	0	22	0	30	21
L10	4	3	20	75	35	72

The judged chances score a Brier of 0.14 across the 50 scored windows. A coin flip scores 0.25. Stating the pooled hindcast rate of 22 percent each time scores 0.17. The average modeled chance per window, 25 percent, sits next to the average rate seen, 22 percent. T4 carries an extreme chance, 85 percent per half year, that its one window confirms. Without T4 the Brier reads 0.14 on 49 windows. The judged chances beat each fixed baseline. Section 11.9 shows the lead thin under a refit baseline and gone at a search miss of 20 to 30 percent. L5 stays unscored: its 97 percent per half year extrapolates a 27-day horizon at constant hazard, and no window describes a US midterm incident. The hindcast counts days from 7 October and the forecasts of 12.12 from 6 October. One day moves a chance by about a point or less.

The score carries five limits. We set the judged chances in October 2026 with knowledge of the period, so the hindcast leaks part of the answer into its own test. A search finds events and cannot prove absence, so some

windows scored as no are misses. Each indicator has at most four windows. Several definitions also join two conditions that the record shows apart. Looser matches remain possible in other windows. Read the Brier of 0.14 as a floor check, and take it as no proof of skill. Absence in a search proves nothing. The test leaks.

The hindcast also moves the inputs. Each indicator's chance of occurring is pulled toward its hindcast rate, with the judgment weighted as four windows. The pull stays small where the windows agree with the judgment and grows where they do not. That rule is mechanical. Pulling raised L10 from 35 to 72 percent, since three of four windows show a new state surveillance unit documented. The model caps an indicator's chance of occurring at 95 percent on each path before route visibility applies, so L10 fires at 64 percent in the tables, and 72 is the uncapped figure. T7 rose from 28 to 44. Cuts fell on L7 from 42 to 22, T5 from 21 to 11, T1 from 4 to 2 and T3 from 10 to 5 percent. L5 keeps its judged 40. The path prior stays untouched. A hindcast tests indicators and no path. Every table from the indicator table of 12.6 onward uses the pulled chances.

12.7 Which indicators matter most

Information gain ranks indicators by what each tells us about the five paths. The measure is the expected drop in uncertainty, in bits, once the indicator resolves. Two readings follow. One treats each indicator alone, and one treats a watch list.

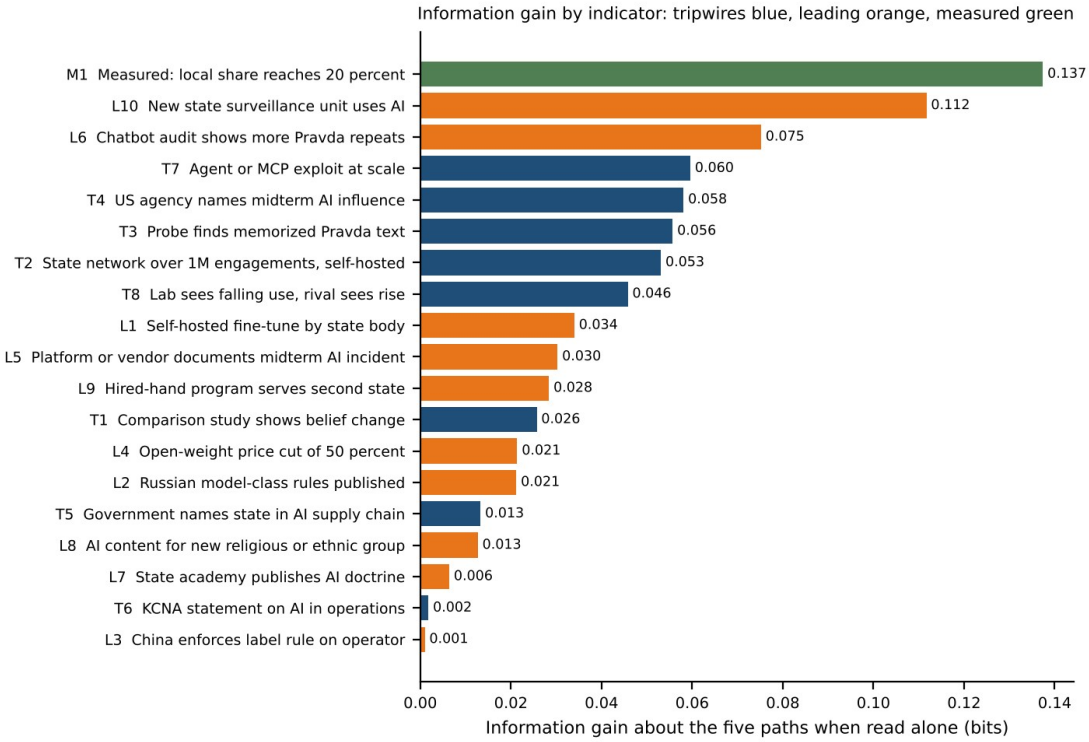


Figure 15. *Information gain of each indicator about the five paths, read alone.*

Read alone, the measured local share M1 leads at 0.14 bits. The surveillance-unit disclosure L10 follows at 0.11, the chatbot audit L6 at 0.08 and the agent-exploitation report T7 at 0.06. T4, the midterm statement from a US agency, and T3, the probe study, also read 0.06, and T2 reads 0.05. Gaps under 0.02 carry no ranking, because the inputs are judged. M1 draws on the very count process that sets Path B, so its lead is partly circular. T3 ranks sixth. A yes for the DFRLab window puts T3 first on the list, which shows how much a ranking of that kind owes to one hindcast ruling.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Single gains overlap. A chatbot audit and a probe study speak to Path C alike, so the second adds less once the first has landed. The greedy search below picks the set that adds the most at each step, using the exact joint model.

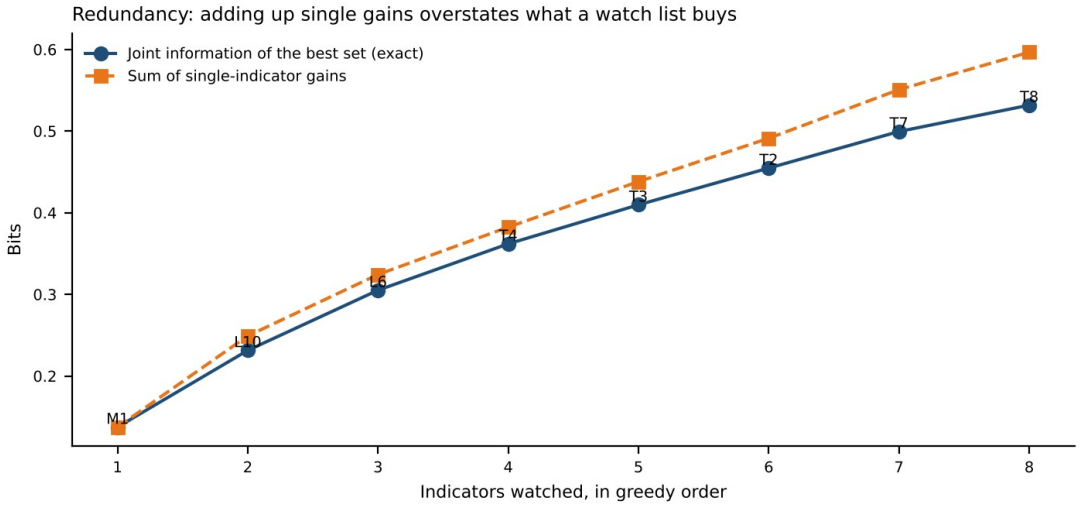


Figure 16. *Joint information of the best watch list against the sum of single gains.*

The first eight, in order, are M1, L10, L6, T4, T3, T2, T7, T8. They carry 0.53 of the 0.64 bits that all 16 active indicators hold. That is 83 percent. Adding their single gains gives 0.60, so the sum overstates the joint value by 12 percent. The 0.64 comes from 2,000 simulated worlds, with a standard error near 0.01 bits. A client with a limited budget runs the M1 count first, since it asks for a protocol rerun and no new study. The surveillance-unit watch L10 and the chatbot audit L6 follow. Speeches wait.

12.8 What the indicators do to the outlook

We drew 30,000 worlds. Each draws a path from the prior, draws the two regimes, draws every active indicator under that path, and updates the path shares exactly. The mean of the updated shares matches the prior to within a point: A 39, B 30, C 18, D 5 and E 8. Evidence moves the odds and adds nothing on average, as it should.

Copyright 2026 Treadstone 71 - www.treadstone71.com

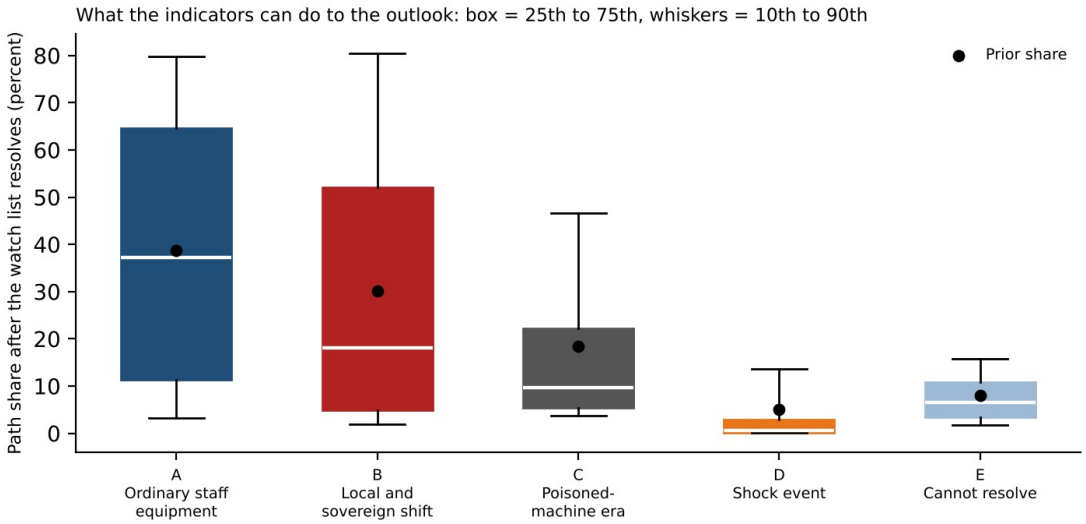


Figure 17. *Path shares after the watch list resolves, across 30,000 simulated worlds.*

Table 32. *Path shares before and after the full watch list resolves (percent).*

Path	Prior	10th percentile	Median	90th percentile
A Ordinary staff equipment	39	3	37	80
B Local and sovereign shift	30	2	18	80
C Poisoned-machine era	18	4	10	47
D Shock event	5	0	1	14
E Cannot resolve	8	2	6	16

Medians for B, C and D sit far under their priors because the means come from long right tails. Tails carry the risk. Most worlds show evidence that points firmly away from them. A minority shows evidence that points firmly toward one. The updated top path matches the true path in 64 percent of worlds. Path B overtakes Path A in 36 percent, A falls under 40 in 53 percent, B climbs above 35 in 34 percent and C climbs above 25 in 21 percent.

Regimes matter little at the loading we chose. With no shared regime, A spans 4 to 80 from the 10th to the 90th percentile. At a loading of 0.7 it spans 3 to 80, and at 1.4 it spans 3 to 82. Two regimes cannot stand for every way signals cluster. One escalation in the Iran war moves several indicators at once, and the model sees only part of that.

Table 33. *Four patterns of signals and the path shares that follow (percent, A B C D E).*

Pattern	Indicators that fire	Exact update	Tempered at 0.5
Quiet year	None of the 16	52, 16, 6, 0, 25	49, 24, 12, 1, 15
Local-model year	T2, T8, L1, M1	0, 99, 0, 0, 1	3, 92, 2, 0, 3
Poisoned-machine year	T3, L6, T7	2, 0, 97, 0, 1	14, 6, 74, 1, 5
Shock year	T1, T4, L5	12, 5, 5, 64, 14	30, 17, 13, 25, 15

Copyright 2026 Treadstone 71 - www.treadstone71.com

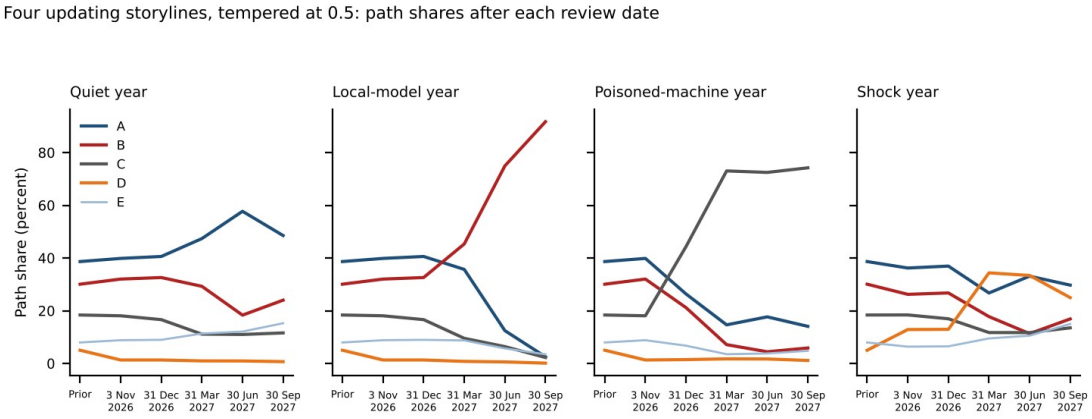


Figure 18. *Four updating storylines, each tempered at 0.5, by review date.*

The exact column answers what the model says. Tempering applies the Dynamic Bayesian Narrative Analysis exponent of 0.5, which halves each update in log terms and offsets part of the independence the model still assumes. Treadstone 71 reports the tempered column as the working read.

A quiet year deserves a close look. When no indicator fires, A rises to about 49 percent, and E rises to 15. Silence mostly fails to resolve the question. The silence also splits by route. If only the three provider-route indicators stay quiet (T8, L9 and L10), A falls to 20 percent and B rises to 50. That inference rests on the visibility table of 12.6, which is judgment and has counterexamples in the record. With provider disclosure under B at 60 percent, the same silence leaves A at 24 and B at 40. At 75 percent it leaves A at 28 and B at 31, level with the prior of 30, and L9 and L10 fall under the 1.5-times rule and drop to watch only. The inference then disappears. A quiet lab record is a reason to count, and it is no sign of calm, if providers miss most local use. The measured indicator carries more weight than either side of that quiet. If M1 stays silent, B falls to 13 percent. Should M1 fire, B reaches 47.

12.9 Actors and the tilt of the drivers

The five paths hold across the four states as a group. Each state sits at its own point inside them. We rated each state on each path with four scores from 0 to 4: motive, means, opportunity and constraint. Motive comes from the Strategic, Political, Social psychology and Religion factors. Means comes from Technical and Economic. Opportunity comes from Military, Demographics, Educational and Social. Constraint comes from Legal and Security (internal), plus the sanctions that bind Iran. Net score equals motive plus means plus opportunity, minus constraint.

Table 34. *Driver ratings by state and path, scores 0 to 4.*

Actor	Path	Motive	Means	Opportunity	Constraint	Net
Iran	A Staff equipment	4	3	3	2	8
Iran	B Local and sovereign	3	2	2	3	4
Iran	C Poisoned machine	1	2	2	2	3
Iran	D Shock	3	2	3	3	5
Russia	A Staff equipment	4	4	4	1	11
Russia	B Local and sovereign	4	3	3	2	8
Russia	C Poisoned machine	4	3	3	2	8
Russia	D Shock	2	3	3	3	5

Copyright 2026 Treadstone 71 - www.treadstone71.com

Actor	Path	Motive	Means	Opportunity	Constraint	Net
China	A Staff equipment	3	4	4	1	10
China	B Local and sovereign	4	3	3	2	8
China	C Poisoned machine	1	3	2	3	3
China	D Shock	1	3	2	4	2
North Korea	A Staff equipment	4	3	4	1	10
North Korea	B Local and sovereign	3	3	3	2	7
North Korea	C Poisoned machine	2	2	3	2	5
North Korea	D Shock	2	3	2	3	4

The states have no path shares of their own. Each indicator in 12.6 acts on the aggregate path, and none splits by state. We say how each state tilts around the aggregate shares. A logit model turns the net scores into a tilt. Its slope is a model parameter at 0.25. No data fix it. Each path gets one intercept, set so that the equal-weight mean of the four states equals the aggregate share. The record gives no reason to weight one state above another, and we chose no weights. Within a state, the four path shares come from one softmax across the paths, so a tilt is relative: it reflects the state's net score on a path against its scores on the other three, as well as against the other states.

Table 35. *Tilt by state: the state's modeled share divided by the aggregate share (1.00 = no tilt).*

Actor	A Staff equipment	B Local and sovereign	C Poisoned machine	D Shock
Iran	1.04	0.80	1.00	1.86
Russia	0.91	0.89	1.43	0.76
China	1.02	1.30	0.59	0.52
North Korea	1.02	1.01	0.98	0.86

Three tilts stand out. China tilts toward B at 1.30 and away from C and D. Russia tilts toward C at 1.43. Iran tilts toward D at 1.86. North Korea sits within 2 percent of the aggregate shares on A, B and C. Tilts are relative within a state. Iran and Russia each net 5 on D, yet their tilts are 1.86 and 0.76, because Iran's other three scores run lower and Russia's higher. China and Russia each net 8 on B, and their tilts are 1.30 and 0.89 for the same reason. The model computes tilts from the net scores alone. Motive context from Sections 3 to 6 explains why we scored as we did: China's home-built models and the loss of a Western account, the Pravda network's treatment of chatbots as terrain, and the war and the midterms for Iran. That context is ours and no part of the arithmetic.

Two tests check the tilts. We added or removed one point from each net score at random 400 times. Stability holds for the leaders on B, C and D: China leads B in 85 percent of draws, Russia leads C in 86 and Iran leads D in 100. The leader on A does not hold: China leads in 32 percent, North Korea in 25 and Iran in 32. We then moved the slope from 0.15 to 0.35. State shares for B run from 22 to 26 for Iran and from 35 to 42 for China. The ordering holds. Gap size depends on the slope. No inter-rater data support any rating.

12.10 Scored futures by adversary, vector and date

Adversary Threat Vector Analysis (ATVA) lays actors against vectors. The grid shows where the record documents a vector today, where it shows part of one, and where it shows nothing. Empty cells measure collection as much as adversary behavior.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 36. ATVA grid, status today.

Vector	Iran	Russia	China	North Korea	Hired hands	Criminals
Staff work and content volume	Documented	Documented	Documented	Partial	Documented	Partial
Persona and deepfake production	Partial	Documented	Partial	Documented	Documented	Partial
Surveillance and profiling	Documented	Partial	Documented	Not observed	Documented	Not observed
Chatbots as audience	Not observed	Documented	Not observed	Not observed	Not observed	Not observed
AI projects as target	Not observed	Documented	Partial	Partial	Not observed	Documented
Model capability as prize	Not observed	Not observed	Documented	Not observed	Not observed	Partial
Agents and supply chain as channel	Not observed	Partial	Documented	Documented	Not observed	Documented
Local and self-hosted models	Not observed	Documented	Documented	Documented	Not observed	Partial

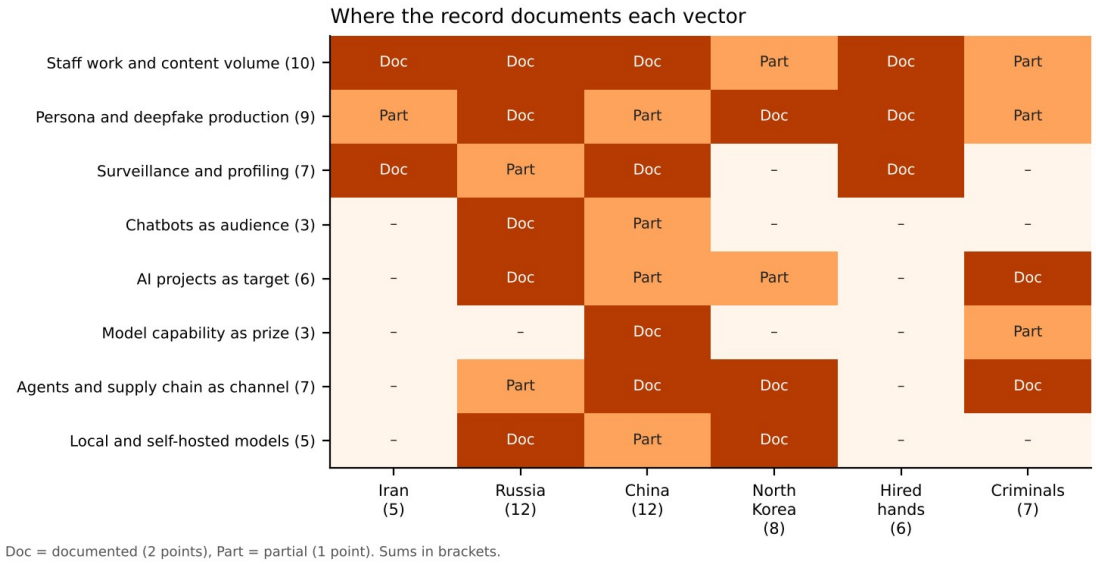


Figure 19. Adversary Threat Vector Analysis: where the record documents each vector. Sums score Documented as 2 and Partial as 1.

Four readings follow. China and Russia tie for the densest column at 12 points each. Staff work and content volume forms the densest row at 10 points, and chatbots as audience (2 points) and model capability as prize (3) form the thinnest rows. Row EN-3 puts China's local-model cell at Documented. The cell stands on a single hedged sentence, for which no outside report exists, and on Anthropic's GTG-17002 case, so it carries Low confidence, the bar that Sections 4.5 and 5.6 set for the same evidence. China shows the widest spread across staff work, surveillance and capability theft. The empty cells for Iran and the hired hands mark collection gaps before they mark adversary absence. Section 15 plugs them.

Eight cells carry the least evidence in the grid. China's chatbots-as-audience cell reads Not observed. No page describes a PRC operation that poisons chatbots or retrieval, and third-party items in the record show Chinese models steering outputs. That makes a chatbot a channel and no audience. Criminals' local-model cell reads Partial. The record holds forum actors who price rigs and install uncensored local models (RC-EN-3), an actor

who ran an agent on an exposed Ollama server that belonged to someone else (RC-EN-8), and vendors who call local use growing but limited. No incident-response case of a criminal group running its own model appears there. Of the 14 added reports, none touches that cell or Russia's. Google ties PROMPTSTEAL to APT28, and the malware queries an open-weight Qwen model that sits on a third party's servers. PromptLock is an academic prototype that ESET analyzed. Cells that stay Not observed carry a documented null where searches covered them and an open row where the record holds no dedicated search. The hired hands' local-model cell stays Not observed in the grid. Anthropic reports that the consultant in GTG-50027 deployed a state-service surveillance platform on local models, and the cell reads Documented only if 6.5 shows that case. Two more Iran cells, agents with supply chain and local models, are open as well, with no documented null.

Table 37. *The eight least-evidenced cells.*

Cell	Status today	What the record holds	Forward chance by 31 Dec 2027 (percent)
China: Chatbots as audience	Not observed	No operation appears in the Graphika, OpenAI, Meta and Estonian pages cited. No dedicated null exists, so the row is open. No finding on narratives appears in NIST's evaluation of DeepSeek V4 Pro (1 May 2026) as held in the record. The September 2025 evaluation, reported to carry a finding on CCP narratives, is not in the record	18
China: Local and self-hosted models	Documented, Low confidence	Two anchors: Google's one hedged sentence on UNC6508, with no second source, and Anthropic's GTG-17002, where a self-hosted model ran beside Claude. OpenAI's February case study relays a user's claim of local Chinese models and reports that it was unable to confirm the point. No PLA or MSS actor is shown running Qwen or DeepSeek locally	84
Criminals: Local and self-hosted models	Partial	Forum installs, rig pricing and abuse of exposed servers. Trend Micro finds that brands sold as self-hosted wrap commercial models. No incident-response case	54
Iran: Agents and supply chain as channel	Not observed	IRGC threats against US technology firms and Gulf data centers, and a claimed strike on Amazon facilities in the Emirates, are physical. Anthropic, Google, Microsoft and Recorded Future describe no Iranian software supply-chain case in the sources cited. No dedicated null exists, so the row is open	14
Russia: Model capability as prize	Not observed	No Russian state distillation or weight theft appears in Anthropic, Google or Mandiant reports. Anthropic's distillation cases name Chinese labs. A Russian-speaking criminal chased pre-release access and failed (Section 4.12)	18
Hired hands: Chatbots as audience	Not observed	Hired-hand cases target people and platforms. A secondhand report of leaked documents from Russia's for-hire Social Design Agency says it plans to target AI ecosystems. Bloomberg is the originator, and the +4 rests on that one item	22
Iran: Chatbots as audience	Not observed	No documented null stands behind the cell	18
North Korea: Chatbots as audience	Not observed	English and Korean sources show no hit (Section 6.8)	10

The forward grid asks a second question. For each actor and vector, what chance does a public source document a new case by 31 December 2027? Each cell starts from its status today: Documented starts at 90, Partial at 62 and Not observed at 18. Eleven cells carry an adjustment, each with a stated basis. The grid scores

Copyright 2026 Treadstone 71 - www.treadstone71.com

documentation, leaves occurrence unscored, rests on three base values and those eleven adjustments, and orders judgments without measuring anything finer.

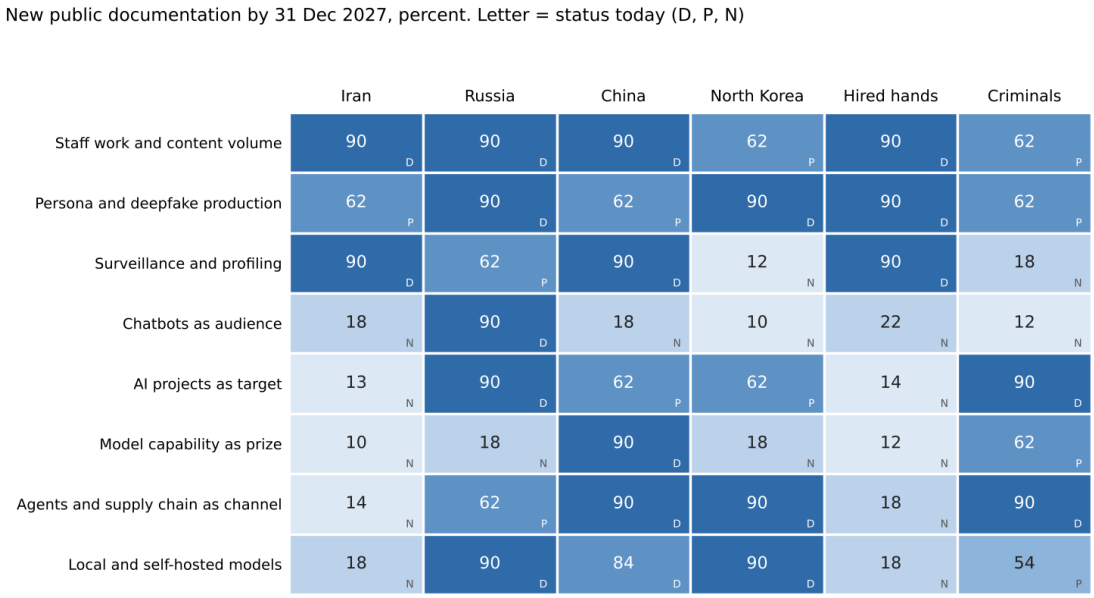


Figure 20. *Chance of new public documentation by 31 December 2027, by actor and vector.*

Table 38. *Adjustments to the forward grid, with the basis for each.*

Cell	Adjustment (points)	Reason	Basis
Iran: Model capability as prize	-8	Sanctions and thin compute	Structural reasoning, no ledger row
Iran: AI projects as target	-5	No Iranian case against an AI facility in the record	Absence in ledger
North Korea: Chatbots as audience	-8	No sign of audience-building in KP rows	Absence in ledger
North Korea: Surveillance and profiling	-6	Revenue outranks surveillance (KP-1, KP-3, KP-5)	Ledger rows
Hired hands: Chatbots as audience	+4	A for-hire Russian firm's leaked documents reportedly target AI ecosystems, secondhand	One secondhand item, with Bloomberg as originator. Without the item the cell reads 18, in the same band
Criminals: Chatbots as audience	-6	Profit gives criminals no reason to build audiences	Structural reasoning, no ledger row
Hired hands: AI projects as target	-4	Few hired-hand brands at stake	Structural reasoning, no ledger row
Hired hands: Model capability as prize	-6	Buying access beats stealing it	Structural reasoning, no ledger row
China: Local and self-hosted models	-6	Two single-source anchors and no PLA or MSS actor shown running a Chinese model locally	Ledger rows
Criminals: Local and self-hosted models	-8	Check Point and Trend Micro call criminal local use limited, and no incident-response case exists	Vendor reports

Copyright 2026 Treadstone 71 - www.treadstone71.com

Cell	Adjustment (points)	Reason	Basis
Iran: Agents and supply chain as channel	-4	Iranian threats to AI data centers are physical, and no software supply-chain case appears in the sources cited	The adjustment is judgment, with no documented null

Column means rank the actors by documented reach across vectors: Russia 74, China 73, criminals 56, North Korea 54, hired hands 44 and Iran 39. Every Documented cell without an adjustment reads 90, so the means restate status counts, and the one-point gap between Russia and China carries no meaning. Row means rank the vectors: staff work 81, persona and deepfake production 76, agents and supply chain 61, surveillance 60, local models 59, AI projects as target 55, model capability as prize 35 and chatbots as audience 28.

A constant hazard turns the end-2027 chance into earlier dates. The record is bursty. Of the 28 ledger rows in the window, 25 fall in the second half and 3 in the first. The extended rows put the count at 41 and 24, since the 14 rows of October to January all sit in the first half. Check Point's VoidLink case appears in two counted reports, the original of 20 January 2026 and the digest of 29 March 2026. The digest keeps its March place for the forum actors that carry its local flag, so the monthly counts stand and the case counts once in each report. Near dates still carry more error than the end date. Read the dates as a guide to order.

Table 39. *Forward documentation chance for selected cells, by date (percent).*

Cell	Status today	By 31 Mar 2027	By 30 Jun 2027	By 31 Dec 2027	Band at end of 2027
Russia: Chatbots as audience	Documented	59	74	90	Highly likely (85 to <95)
North Korea: Persona and deepfake production	Documented	59	74	90	Highly likely (85 to <95)
China: Agents and supply chain as channel	Documented	59	74	90	Highly likely (85 to <95)
Criminals: AI projects as target	Documented	59	74	90	Highly likely (85 to <95)
China: Chatbots as audience	Not observed	7	11	18	Unlikely (15 to <45)
China: Local and self-hosted models	Documented	51	66	84	Likely (55 to <85)
Criminals: Local and self-hosted models	Partial	26	37	54	Roughly even chance (45 to <55)
Iran: Agents and supply chain as channel	Not observed	6	9	14	Highly unlikely (5 to <15)
Russia: Model capability as prize	Not observed	7	11	18	Unlikely (15 to <45)
Hired hands: Chatbots as audience	Not observed	9	14	22	Unlikely (15 to <45)
Iran: Chatbots as audience	Not observed	7	11	18	Unlikely (15 to <45)
North Korea: Chatbots as audience	Not observed	4	6	10	Highly unlikely (5 to <15)

Chatbots as audience shows Not observed for China, Iran, North Korea, hired hands and criminals. Four rows have no closed null: China chatbots, Iran chatbots, Iran agents and supply chain, and Iran local models. For the other Not observed cells the table lists the sources cited, and a rerun starts there.

12.11 Scored futures by target region

Operators in Europe, the United Kingdom, Australia, Singapore and the Gulf ask a narrower question. How likely is a report of an AI-enabled operation or intrusion aimed at operators or audiences in the region to go public by 31 December 2027? Section 13.5 reads the regional rules, and the numbers below read the exposure. No ledger row shows an operation aimed at Singapore or the Gulf. Every cell outside Europe carries Low confidence and rests on opportunity, war and history. The ledger holds one row on the United Kingdom as a target, the joint alert on Iranian targeting of dissidents (IR-9). That row shows no AI element. Gulf evidence is thin. The figures

are analyst judgments. Bands of 85 and over rest on public documentation counts, and confidence stays at Moderate or lower.

Table 40. Chance of a public report by 31 December 2027, by target region and actor (percent, with Kent band).

Region	Iran	Russia	China	North Korea	Criminals	Confidence
Europe	50 (Roughly even chance)	92 (Highly likely)	55 (Likely)	65 (Likely)	90 (Highly likely)	Moderate for Russia, Low for the rest
United Kingdom	55 (Likely)	80 (Likely)	60 (Likely)	50 (Roughly even chance)	90 (Highly likely)	Low
Australia	22 (Unlikely)	40 (Unlikely)	66 (Likely)	45 (Roughly even chance)	85 (Highly likely)	Low
Singapore	14 (Highly unlikely)	20 (Unlikely)	62 (Likely)	40 (Unlikely)	80 (Likely)	Low
Gulf	62 (Likely)	30 (Unlikely)	45 (Roughly even chance)	28 (Unlikely)	80 (Likely)	Low

Criminals score 80 or higher in every region. Russia scores 80 or higher in Europe and the United Kingdom only, and Europe carries the one Russia figure above 90. Matryoshka in Germany, the Moldova campaign and the EEAS counts give Russia a record there. The Gulf carries the largest Iran figure, 62, and the least evidence, because the war with Iran puts Gulf states and infrastructure inside Tehran's reach. A Gulf operator treats that cell as a collection task. Singapore and Australia face China and criminals first. The regional rows in 13.5 and the decisions in 13.6 apply.

12.12 Ten dated forecasts

Each forecast resolves on a named observation by a fixed date, so a reviewer scores it against the page. Every forecast carries a point probability and its Kent band. Where a forecast and an indicator describe the same event, the forecast takes the model value. Remaining forecasts rest on a stated base rate or on judgment, and the table says which. Forecast 10 and tripwire T1 describe different events, and the note below the table explains.

Table 41. Ten dated forecasts with point probabilities.

#	Forecast	Probability (percent)	Band	Basis	Confidence	Resolves
1	No US federal agency issues a public statement naming AI-enabled foreign influence aimed at the 3 November midterms	74	Likely (55 to <85)	Model: 100 minus T4	Low	3 Nov 2026
2	At least one lab or vendor reports a new state-linked AI influence operation aimed at US audiences	52	Roughly even chance (45 to <55)	Base rate of 3 such rows a year (CN-2, RU-5, RU-11), grown pro rata toward 1.25 by end 2027 (1.04 by 31 Dec), over 86 days: 0.74 expected	Low	31 Dec 2026
3	No comparison-design study shows belief change from an AI-run state operation	98	Almost certain (95 to 99)	Model: 100 minus T1	Low	31 Mar 2027

Copyright 2026 Treadstone 71 - www.treadstone71.com

#	Forecast	Probability (percent)	Band	Basis	Confidence	Resolves
4	A further malicious release hits an AI gateway, agent framework or model-hosting package	73	Likely (55 to <85)	Base rate of 2.5 a year (Mastra, TeamPCP, Nx), grown pro rata toward 1.25 by end 2027 (1.09 by 31 Mar), over 176 days: 1.32 expected	Low	31 Mar 2027
5	Russia publishes rules that sort models into sovereign and national classes	51	Roughly even chance (45 to <55)	Model: L2	Low	31 Mar 2027
6	The 2027 ODNI Annual Threat Assessment carries AI-enabled foreign influence language	50	Roughly even chance (45 to <55)	Judgment. The 2026 assessment carries none (GV-1)	Low	31 Mar 2027
7	At least one new report adds a row to the counted record that documents an actor running self-hosted models for operations. The count takes any actor, state or criminal, and repeat reports on CopyCop, Kimsuky and UNC6508	77	Likely (55 to <85)	Model: chance of at least one new local-flag row in the simulation. A row-count forecast and no forecast about a new state actor	Low	30 Jun 2027
8	A government or lab names a Chinese lab for distillation that the advisory of 8 September does not list. The advisory already names six	48	Roughly even chance (45 to <55)	Judgment. New names appeared once in seven months (CN-7)	Low	30 Jun 2027
9	A government attributes an AI supply-chain compromise to a state actor. Microsoft attributes the Mastra injection with high confidence and a stated basis, and CrowdStrike names a North Korea-nexus actor with no basis stated. Pages from FBI IC3, CISA, NCSC, ASD and CCCS name no state actor, and the WaterPlum advisory of 22 September on North Korean npm malware does not name Mastra. The forecast needs a government source	11	Highly unlikely (5 to <15)	Model: T5	Low	30 Jun 2027
10	A comparison-design study measures the effect of an AI operation on belief, in either direction	25	Unlikely (15 to <45)	Judgment. No such study of an AI-run operation exists in the record, and field studies of earlier operations found no effect (Section 10.7)	Low	31 Dec 2027

Bands at 85 and above need Strong evidence under the ASA rule, and no forecast reaches them except forecast 3, which rests on the absence of an event. Confidence stays Low on all ten. On forecast 10: it counts any measured result, in either direction, by the close of 2027. T1 needs a measured belief change by 31 March 2027. A constant hazard puts a 25 percent chance by the end of 2027 at about 11 percent by 31 March, so the 2

percent of T1 implies that about one such study in five shows a change. Field studies of earlier operations found none (Section 10.7), and one in five is already generous.

12.13 Momentum-Reversal Analysis

Momentum-Reversal Analysis (MRA) forecasts how an adversary pivots after a takedown or exposure. We paired each actor with the Adversarial Cognitive Simulator reads in Sections 3 to 7 and with the CARM pathways: hide, relocate, monetize, conceal, replace. The bands below are simulator outputs and our reads. Each is conditional on the disruption occurring. Nobody observed the pivots.

The ledger holds no clean test of a pivot after a takedown. Two cases come close. CopyCop kept building forged outlets after public reporting, and Recorded Future counts more than 300 sites for the year (RU-5, RU-6). Anthropic's own distillation counts grew between its February report and its September report, which adds Alibaba and Zhipu to the labs named earlier (CN-7). Neither case follows a disruption we name.

Table 42. *Momentum-Reversal reads, conditional on the disruption.*

Actor	Disruption	Expected pivot	Band	First tell
Iranian propaganda bodies	Public attribution of the three bodies	Shift volume to domestic walled-garden platforms. A move to local models needs capacity the record does not show	Roughly even chance (45 to <55)	New Eitaa and Bale accounts, Persian-language tooling signatures
Russian state-media pipelines	Account bans at the lab	Move to self-hosted models and keep the editorial workflow	Likely (55 to <85)	Same house style, new tooling signatures
CopyCop and Storm-1516	Takedown of forged outlets	Rebrand outlets and rotate infrastructure within weeks	Likely (55 to <85)	New domains that clone the retired letterhead
PRC security bureaus	Lab ban on the account	Switch to DeepSeek, Qwen or another local model	Likely (55 to <85)	Prompt formula persists in new tools
North Korean IT-worker networks	Facilitator prosecutions	Rotate personas and facilitators, move to contractor and maintainer roles	Likely (55 to <85)	Reused AI photos across new personas
Supply-chain criminal groups	Indictment and release takedown	Rotate infrastructure, change registry, sell data fast	Highly likely (85 to <95)	Token theft at a different registry

12.14 Prerequisites for paths B and C

A prerequisite tree starts from an end state and lists what must hold. The tree checks the model against itself and validates no share, since the inputs that built the shares also built the tree.

Path B rests on five conditions, and none is strictly necessary. The table gives each with its chance from the model.

Table 43. *Prerequisites for Path B and the chance of each.*

Prerequisite	Chance (percent)	Source
No D event resolves first	95	Input O7
No C event resolves first	81	Input O6
The latent local share passes one half by the end of 2027	40	Model, from O3 and O4. The rule also fires without it
The 2027 window holds at least 8 non-provider rows	92	Model, from O5
The rule fires: 30 percent or more of those rows carry the local flag	40	Model. Power 74 when the latent share exceeds one half, false alarm 18 otherwise

The chain resolves to 30 percent for Path B. Weakest of all is the latent shift itself, at 40 percent, and the path does not need it: false alarms from a latent share under one half bring 11 of the 40 points where the rule fires.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Forecast 7 tests it first. At least one new local-flag row, for any actor, joins the counted record by 30 June 2027 with a 91 percent chance if B holds and 67 percent if A holds. Repeat reports on actors already named count toward the row count. The gap is wide enough to learn from.

Path C needs two things in sequence. First, someone runs a probe or retrieval study by year end 2027, at 35 percent. Second, it finds distortion that holds up on confirmation, at 55 percent. The product is 19 percent, and the 95 percent chance that no D event resolves first brings it to the 18 printed in 12.3. No one has run a probe for memorized Pravda text on a model trained after the network's expansion. Retrieval probes across assistants have run. DFRLab's completion test of April 2026, run on an open base model that stops at December 2023, returned RT and CGTN text verbatim. That is the nearest evidence, and T3 scores it as no (12.6). The path hangs on that gap. T3 and L6 watch it.

12.15 Wildcards, premortem and robust actions

Wildcards ride on top of the five paths. Their chances do not sum. Each one is a separate bet. Taken together, at least one of the six occurs by 31 December 2027 with a chance near 16 percent.

Table 44. Six wildcards, chances by 31 December 2027, and where each lands among the paths.

Wildcard	Chance (percent)	Band	Lands in	Signpost
A state actor steals the weights of a frontier model	4	Almost no chance (1 to <5)	A, B or C	Lab disclosure, advisory on a named intrusion
An AI-enabled intrusion reaches an operational-technology network at a critical infrastructure operator and a government confirms the AI role	3	Almost no chance (1 to <5)	A, B or C	T5, T7
A deepfake of a leader or central banker moves a market on a named trading day	3	Almost no chance (1 to <5)	D, with comparison evidence	L5, regulator statement
A state shuts Western model access at its border and lab visibility ends within a year	2	Almost no chance (1 to <5)	B, or E if counts fall	L2, L3
An AI-enabled operation alters polling-place information in a G20 state and the election authority says so	2	Almost no chance (1 to <5)	D, with comparison evidence	T4, L5
A lab ends public threat reporting	3	Almost no chance (1 to <5)	E	T8

Only the two outcome events resolve as Path D, and only with comparison evidence. The other four move the record without moving a path, except that a lab which ends its reports pushes Path E up.

The two wildcards that land in Path D carry 3 and 2 percent, and O7 gives Path D 5 percent. Their sums match, yet each wildcard needs comparison evidence on top, so the wildcards alone put D below 5. Read the 5 as a judged ceiling and no estimate with a test of its own.

A premortem asks why the outlook failed, assuming it did so by 31 December 2027. Seven causes lead the list.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 45. Premortem causes, first tells and counters.

Cause of failure	First tell	Counter now
Lab visibility fell and we read quiet as calm	T8 or M1 fire	Count with the 12.2 rules each quarter and collect from adversary-language sources
Our counts moved because our collection moved	Rows per quarter jump when search effort changes	Log effort and hold the protocol fixed
Indicators moved together and the regimes missed it	Three or more fire in one month	Refit the regime loadings at each review date
The effect study arrived and showed belief change	T1 fires	Rebase KJ13 and Path A within 30 days
An actor we do not track drove the visible events	A report names a state outside the four	Add Gulf-state and commercial actors to the collection plan
Our likelihood judgments ran overconfident	Forecast scores fall below 50 percent on resolved items	Score the ten forecasts as they resolve and publish the result
A single shock reset the war and midterm picture	Path D indicators fire together	Reopen all five paths within 14 days

The last test asks which actions pay off whatever path arrives. We rated six actions from Sections 13.1 and 13.2 on each path, with 2 for a strong payoff, 1 for some and 0 for little. The expected payoff weights each rating by the path shares of 12.3. Path E carries no rating. We then redrew the shares from a Dirichlet distribution around the prior and changed 30 percent of ratings by one point, 3,000 times. Counts followed of how often each action ranks in the top three.

Table 46. Stable actions: ratings, expected payoff and rank stability.

Action	A	B	C	D	Expected payoff	In top three (percent of draws)	Weakest path
Inventory gateways, key stores and agent runtimes	2	2	1	2	1.66	96	C
Strip agent permissions and authenticate every MCP server	2	2	1	2	1.66	96	C
Live-video and photo-reuse checks in hiring	2	2	1	1	1.61	95	C, D
Callback rule for payment instructions	1	1	1	2	0.97	5	A, B, C
Impersonation rehearsal before a launch	1	1	1	2	0.97	5	A, B, C
Log and distrust retrieval sources	1	0	2	1	0.80	2	B

The inventory and the agent-permission cut top the list at 1.66 and stay in the top three in 96 percent of draws. Hiring checks follow at 1.61 and 95 percent. The callback rule and the impersonation rehearsal tie at 0.97 and reach the top three in 5 percent of draws. Logging retrieval sources ranks last and pays heavily only if Path C arrives. That makes it insurance. Judgment sets every rating. Four decisions from Sections 13.1 and 13.6 stay unrated, since they pay off on every path alike: the single owner, the ASA grading step, the credential watch and the board's supplier register.

12.16 What moves the headline

The headline shares rest on inputs we judged. Each input moves across a tested range in the table below, with the model recomputed on 300,000 trajectories at seed 71. The base row gives the headline shares. Two rules set the 2027 row count, and the last column names the rule for each row. Rows marked Base keep the 16-row non-provider baseline, so a recoding changes the shares and leaves the effort alone. Where a row says Scaled, the 2027 count follows the larger row set, as O1 states, so a larger set means a larger rerun. As a result the E column moves with effort in those rows and says nothing about evidence. Provider counts never enter the path rules. Simulation error on a share runs near 0.1 points, and rows sum to 99 or 101 where rounding falls badly.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 47. Path shares under changed inputs (percent, A B C D E).

Run	A	B	C	D	E	2027 rows
Base case	39	30	18	5	8	Base
Growth in local odds: low (0.3)	52	19	18	5	6	Base
Growth in local odds: 1.2	32	36	18	5	9	Base
Growth in local odds: high (1.5)	25	41	18	5	10	Base
Detectability of local use 0.25	38	27	18	5	12	Base
Detectability of local use 1.0	39	32	18	5	6	Base
Baseline counts: all rows to date (3 of 18)	31	37	18	5	8	Base
EN-3 coded as a provider row (1 of 15)	50	19	18	5	7	Base
No local flag among the 16 non-provider rows (0 of 16 local)	65	5	18	5	6	Base
21 reports and two slices added (4 of 34, 17 provider)	44	30	18	5	3	Scaled
21 reports and two slices, RC-EN-8 not coded local (3 of 34)	51	23	18	5	3	Scaled
21 reports and two slices, RC-CN-6 local flag kept (5 of 35)	38	36	18	5	3	Scaled
21 reports and two slices, no KP-3 local flag and RC-EN-8 not local (2 of 34)	59	15	18	5	2	Scaled
October 2025 to January 2026 rows added (4 of 44, 21 provider)	51	23	18	5	2	Scaled
Rule threshold 25 percent	33	36	18	5	8	Base
Rule threshold 27.5 percent	37	32	18	5	8	Base
Rule threshold 35 percent	44	25	18	5	8	Base
Path C event 10 percent	43	34	9	5	9	Base
Path C event 30 percent	34	26	29	5	7	Base
Path D event 2 percent	40	31	19	2	8	Base
Path D event 10 percent	37	29	17	10	8	Base
Row growth low (1.0)	36	27	18	5	13	Base
Row growth high (1.6)	40	32	18	5	5	Base

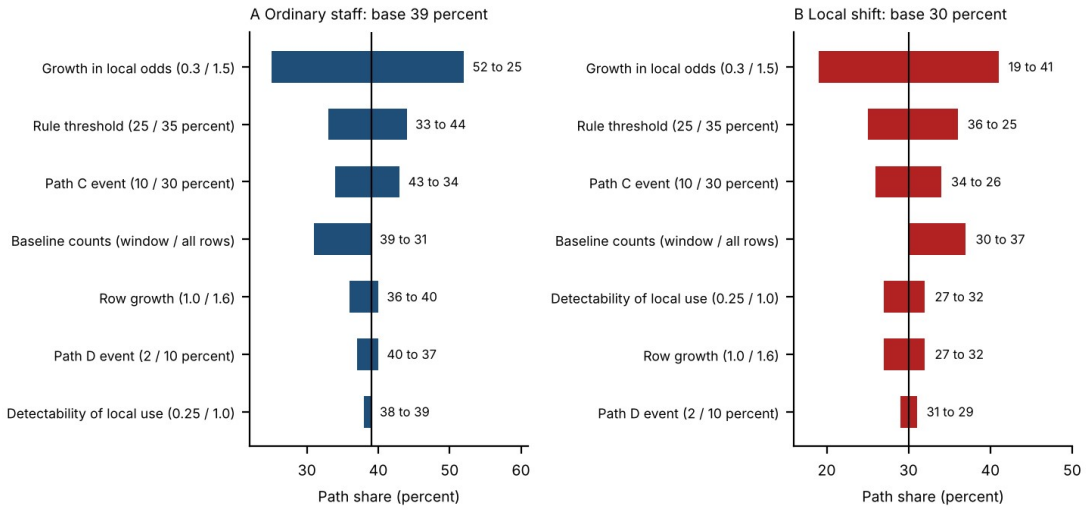


Figure 21. Path A and Path B shares as each input moves across its tested range.

Three results matter. The order of A and B depends on a few inputs, and none of them is settled. Evidence does not separate the two. A leads in 18 of the 22 perturbed runs and B in 4. That count is a vote over perturbations we chose, several of them near duplicates, and gives no frequency of robustness. A runs from 25 to 65 percent.

Copyright 2026 Treadstone 71 - www.treadstone71.com

B runs from 5 to 41 percent. The inputs that move B most are growth in local odds, the baseline count and the rule threshold.

Coding decisions show how thin the count is. Row EN-3 coded as a provider row takes B from 30 to 19 percent and lifts A from 39 to 50. A count with no local flag among the 16 non-provider rows puts B at 5 and A at 65. Keeping the RC-CN-6 local flag, as 12.2 explains, puts A at 38 and B at 36, and at 36 and 33 with the row count held at base. Strict coding of RC-EN-8 puts A at 51 and B at 23. Dropping the KP-3 local flag as well puts A at 59 and B at 15. Adding the February to August reports and slices puts A at 44 and B at 30 and cuts E from 8 to 3. That fall is an effort effect: the rerun is assumed to collect at the larger set's effort, and more non-provider rows make the rule likelier to resolve. Adding the 14 reports of October 2025 to January 2026 on top puts A at 51, B at 23, C at 18, D at 5 and E at 2. Ten of the 14 are non-provider rows and none carries a local flag, so the local share dilutes from 4 of 34 to 4 of 44. Those A and B shares match the row that strips the RC-EN-8 local flag, reached by dilution in place of recoding. If ESET and Unit 42 count as local, the set reads 6 of 44, with A at 40 and B at 34. Dropping the two overlap rows leaves 4 of 43 and shifts no share by a full point. The headline of 39, 30, 18, 5 and 8 comes from the ledger rows alone. Taken together, the added rows put the shares at 51, 23, 18, 5 and 2: B falls 7 points and A gains 12.

Two inputs outside the coding reverse the order. Lower the rule threshold to 25 percent and B moves ahead, 36 against 33. At 27.5 percent A leads again, 37 against 32, so the order flips between 25 and 27.5 percent. Raise the growth in local odds to 1.2, against the stated 0.9, and B moves ahead as well, 36 against 32. Row minimums of 6 and 10 leave A ahead and move E between 5 and 13 percent. Detectability moves the paths little: at 0.25, A stands at 38 and B falls to 27. KJ14's bound moves more, because the probability of a latent share below one third drops from 73 to 42 percent. The headline holds at the 30-row ledger set. Path C and Path D move almost one for one with their event inputs and nothing else.

The sensitivity says what the model needs most. Data beats debate. A second year of counts under the same protocol cuts the uncertainty on baseline and growth more than any other step.

12.17 What breaks the picture, confidence and review dates

Four observations break the picture. Each belongs on a watch list now.

A comparison study shows that an AI-run state operation moved belief by a measured amount. Path A falls, KJ13 reverses and T1 fires.

A lab report shows adversary use of commercial models falling to near zero while influence activity holds steady. T8 fires and the turn to local models stands as fact.

A government names a state actor inside an AI gateway or agent framework. Every enterprise judgment keeps its band, and the state share moves up at once.

A clean memorization result lands for the Pravda network. Chatbots then count as a proven audience, and the corpus around them as a battlefield.

None of the four appears in the record today. Two vendors name North Korea-nexus actors behind the Mastra injection, and the five agency pages (FBI IC3, CISA, NCSC, ASD and CCCS) name no state actor for it.

Confidence in every path share is Low. Judgment sets the shares, the indicator likelihoods, the driver scores and the model inputs. The hindcast in 12.6 checks the indicator chances against four windows and no more. Nobody has back-tested the rest. Simulations show the arithmetic of those judgments and add no evidence. Tilts by

Copyright 2026 Treadstone 71 - www.treadstone71.com

state, dates and regional cells also carry Low confidence. Counted indices carry the most weight in the outlook, and they rest on 30 ledger rows. Added reports, 21 and then 14, widen the base without changing the headline.

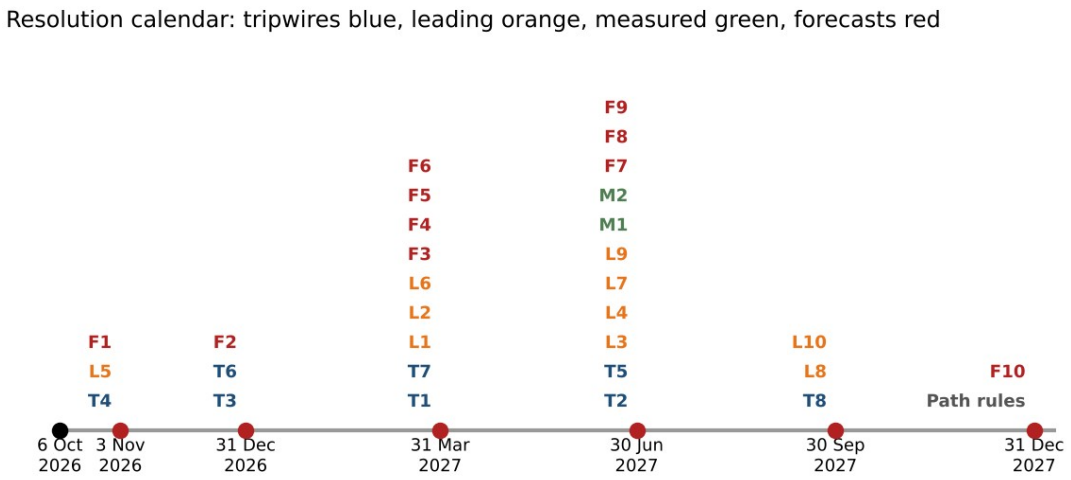


Figure 22. *Resolution calendar: indicators and forecasts by check date.*

Signals also trigger decisions. The table pairs each signal with the first move and a planning date.

Table 48. *Decision triggers, first moves and planning dates (Treadstone 71 planning figures).*

Signal	First move	Within
T3 or L6 fires	Extend retrieval-source logging to every enterprise retrieval layer and brief Path C	14 days
T2, L1 or M1 fires	Move Daily Adversary Brief collection to adversary-language repositories and add a provider discount to every count	30 days
M2 fires	Recount the 12.2 indices on a report basis, with a second check of every new row	Next quarter
T4 or L5 fires	Brief the board on the incident and run the impersonation rehearsal	7 days
T5 or T7 fires	Reopen agent-permission and MCP audits and rank supply-chain risk first again	14 days
T1 fires	Rebase KJ13 and every effect statement, and reopen counter-influence budgets	30 days

Review dates follow the indicators.

Table 49. *Review dates, what resolves and what to rerun.*

Date	Resolves	Then
3 Nov 2026	T4, L5, forecast 1	Rerun 12.8 and the Path D share
9 Nov 2026	Window calls W1 to W20 and estimates WE1 to WE6 (Section 12.18)	Score each call and log the Brier figure
31 Dec 2026	T3, T6, forecast 2	Rerun the pattern table. Recount the 12.2 indices
31 Mar 2027	T1, T7, L1, L2, L6, forecasts 3 to 6	Rerun the cones and the regional cells
30 Jun 2027	T2, T5, L3, L4, L7, L9, M1, M2, forecasts 7 to 9	Rerun the path shares and the tilts by state
30 Sep 2027	T8, L8, L10	Rerun 12.8
31 Dec 2027	Path rules, forecast 10	Resolve the five paths and publish the final scores

The scoring plan uses the Brier score, the squared gap between a stated probability and what happened. A forecaster who says 50 percent every time scores 0.25. Our ten stated probabilities, if they are well calibrated,

carry an expected score of 0.19. Ten items cannot show calibration. The sample is too small. Their standard error runs near 0.05. We score them as they resolve, publish each result, and add the tripwires and leading indicators to the scoring ledger. The path rules score the five shares on 31 December 2027. A single resolved path says little about calibration, so the shares serve as a record for the next version more than a test of the present one.

12.18 The election window: 9 October to 9 November

Dossier 1 of the Public Attribution Series carries the same window with its sources. The window opens on 9 October and closes five days after the results go public. Networks will probably call House control on the morning of 4 November, which sets the close at 9 November. If a chamber stays uncalled, the five days run from the call.

Four modes of judgment appear, each tagged. F is a forecast: a probability on a dated event. E is an estimate: a quantity with a median and an 80 percent range. P is a prediction: a committed call on a named day. S is foresight: a scenario with signposts, built for watching and not for ranking. Probabilities use the Kent bands of Section 1.3 and carry a confidence level. Each number is a judgment. None is a measurement. Call IDs run W1 to W20, estimates WE1 to WE6 and scenarios WS1 to WS4.

One finding frames the rest. Adversaries have no shown path to the count. They work on how the count looks and sounds: when it ends, who announces it, which screenshots circulate, and which claims of a hack follow. Anonymous officials in the Times story, read in a repost, say they see no effort against voting machinery. That is a claim at grade C3. We expect noise at the edges of the process and certified totals that no cyber action alters. The odds that one does: Almost no chance (1 to <5), Moderate confidence.

What past cycles show

Past cycles set the base rates. Federal naming of a foreign actor in the final month happened in 2018, 2020 and 2024. The 2022 cycle brought public service announcements and no name. Election-week disruption grew over time. In 2024 the FBI tied hoax threats to polling places to Russian email domains.

Federal capacity in 2026 is smaller than in any of those years. The Justice Department ended the FBI Foreign Influence Task Force in February 2025 (CyberScoop, B2), the foreign influence center sits in a refocus, and CISA's 2025 budget plan projected a loss of nearly 1,000 positions, about 29 percent of staff (Defense One, B2). We shade federal calls down for that reason. Platform and vendor calls go up, because they carry the load when agencies go quiet.

Table 50. Base rates from four cycles.

Cycle	Federal action in the final month	Election week	Lesson for 2026
2018	Justice charged a Russian accountant, Elena Khusyaynova, on 19 October for work on Project Lakhta (NBC News, B2)	Facebook blocked 115 accounts on 5 November after a law-enforcement tip about suspected foreign ties (CNBC, B2)	Platforms act in the last 72 hours
2020	The Director of National Intelligence said at a press conference on 21 October that Iran and Russia had obtained voter registration data (ODNI transcript, B2)	A joint statement of federal, state and vendor councils on 12 November called the vote the most secure in American history (NASED, B2)	Federal naming came two weeks out. The all-clear came nine days after the vote
2022	FBI and CISA public service announcements on 4 and 6 October (B3)	Mississippi's Secretary of State website went down under a traffic flood. A Russian hacking group claimed it on Telegram. CISA said on 8 November that counting was unaffected and saw no coordinated campaign (Nextgov, B2)	Claims and nuisance disruption, with no effect on counts
2024	Joint ODNI, FBI and CISA statements named Russian influence actors behind fake videos, including one on 1 November (Fox 5 Atlanta, B2)	Hoax bomb threats reached polling places in several states. The FBI said they came from Russian email domains. Fulton County, Georgia, closed two sites briefly (PBS, B2). CISA's director said the next day that she saw no evidence of malicious activity with a material impact (CISA, A1 as to text)	Naming came fast when the apparatus was whole

Phases and the modal path

Five phases split the window. Each carries one prediction, the single path we judge most likely, and the event that breaks it.

Table 51. Phases, predictions and breakpoints.

Phase	Dates	Prediction (P)	What breaks it
1 Positioning	9 to 19 October	Matryoshka keeps posting. A researcher or outlet reports another fake with a real face or a news brand. No federal agency speaks on AI	A joint federal statement lands early: 12 percent
2 Late campaign	20 October to 2 November	Platforms publish takedowns. The 31 October date set by Warner and Himes passes without a public text. One fake or synthetic-audio item reaches national coverage. Iran-origin content leans on war framing	A hack-and-leak goes public: 12 percent
3 Election Day	3 November	Hoax threats reach polling places in several states. Fake results graphics start after the eastern polls close. A state results page slows. Officials say counting is unaffected	A verified intrusion touches a reporting channel (WS4)
4 Count and call	4 and 5 November	Networks call the House on 4 November and the Senate by 6 November. Adversary outlets describe disorder in the count. Fake concession clips circulate	Slow-count states hold a seat open (WS2)
5 Aftermath	6 to 9 November	A federal statement says no material effect. Platforms and vendors publish what they removed. Political figures blame foreign hands for losses. Adversary outlets cite the blame as proof of decay	A national figure disputes a result and files suit (WS3)

Copyright 2026 Treadstone 71 - www.treadstone71.com

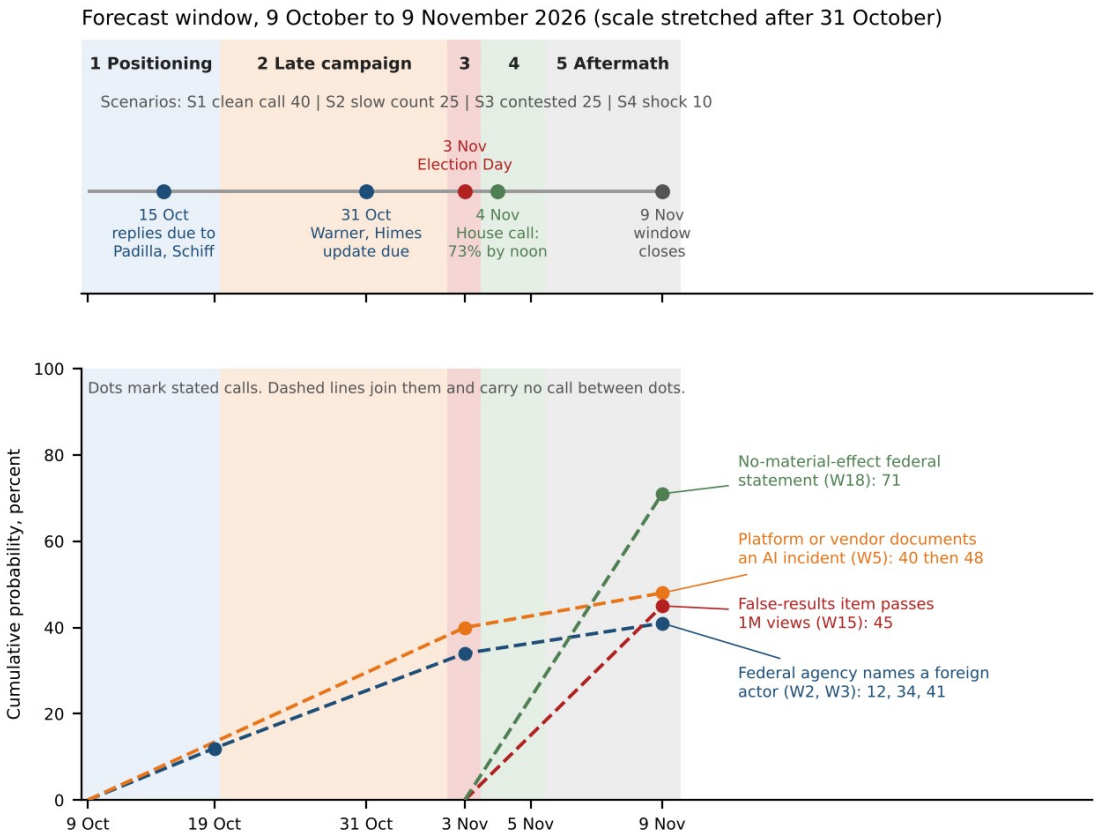


Figure 23. *Forecast window: phases, dated checks and cumulative odds.*

Four scenarios

Legal ground is unsettled. On 14 September the Supreme Court left in place an order that blocks parts of a Postal Service mail-ballot rule (SCOTUSblog, B2). The Justice Department said in August that it plans about 1,000 monitors at polling sites (Democracy Docket, B3). California accepts mail ballots that arrive up to seven days after the vote, and Alaska must finish counting absentee ballots by day 15 (Ballotpedia, B2). Each fact adds friction, and friction feeds disputes.

We borrow the outcome baseline from raters. Cook Political Report moved Kansas to Toss Up, New Hampshire to Lean Democratic and North Carolina to Likely Democratic on 6 October, and puts Democratic Senate gains at two to six seats against the four needed (political.org, B3). Markets price a Democratic House at 90 percent and a Democratic Senate at 62 percent (Kalshi and Polymarket, averaged by DeFi Rate on 9 October, B3). We set the joint outcomes at 57 percent for Democratic control of the House and the Senate, 33 percent for a Democratic House and Republican Senate, and 10 percent for a Republican House.

The paths split on one point: whether the count finishes clean and fast.

Table 52. Scenarios to 9 November (S). Probabilities sum to 100. Confidence is Low for all four.

Scenario	Probability	What happens	What adversaries do	Signposts
WS1 Clean call, loud noise	40 percent	Networks call both chambers by 5 November. Losers concede within two days	Fake results graphics, hoax threats and state-media sneers. Content spreads fast and fades fast	Tipping-point margins above 3 points. Calls before midnight eastern. Concessions by 5 November
WS2 Slow count, long fuse	25 percent	A Senate seat, or control, stays open for days. Mail-ballot states count on their own clocks	Feed "late ballot" stories to every side and time fakes to the gaps. Hardest for attribution, since every claim carries some grain	A Senate race inside one point. Control hangs on one seat
WS3 Contested result	25 percent	A national figure or the White House disputes a result and files suit or threatens certification	Amplify the dispute and cite it as proof of a failing system. Foreign-interference claims become a domestic weapon	The administration disputes a result within 24 hours. Suits in more than three states
WS4 Shock	10 percent	A verified intrusion on a reporting channel, a hack-and-leak, political violence or a restart of US operations in the Iran war lands in election week	Pile onto the event faster than agencies attribute it	An official confirms an intrusion. Operations restart before 9 November

Calls

The table runs in date order, and the same numbers feed the chart that follows.

Table 53. Dated calls to 9 November (F).

ID	When	Call	Probability and band	Confidence
W1	By 19 October	A researcher or outlet reports a new fake video or audio item that uses a real US candidate's face or voice, or a news brand	92 percent, Highly likely (85 to <95)	Moderate
W2	By 19 October	A federal agency names a foreign actor tied to midterm influence on the record	12 percent, Highly unlikely (5 to <15)	Moderate
W3	By 3 November	A federal agency names a foreign actor tied to midterm influence on the record	34 percent, Unlikely (15 to <45)	Low
W4	By 3 November	A federal agency names AI-enabled foreign influence aimed at the midterms. The complement of Forecast 1	26 percent, Unlikely (15 to <45)	Low
W5	By 3 November	A platform, vendor or state official documents an AI-enabled incident aimed at the midterms. Indicator L5	40 percent, Unlikely (15 to <45)	Low
W6	By 9 November	A platform or vendor publishes a takedown or report on a foreign operation aimed at US political audiences	92 percent, Highly likely (85 to <95)	Moderate
W7	By 9 November	Iran-origin content aimed at US voters is newly documented by a platform, vendor or researcher	65 percent, Likely (55 to <85)	Low
W8	By 9 November	A Chinese operation aimed at the midterms is documented. A North Korean one, in the same test, stands at 4 percent	35 percent, Unlikely (15 to <45)	Low
W9	By 2 November	A hack-and-leak of campaign or party material is made public	12 percent, Highly unlikely (5 to <15)	Low
W10	By 9 November	A synthetic-audio or fake-video item aimed at US voters is tied to a foreign network and reaches national coverage	60 percent, Likely (55 to <85)	Low

Copyright 2026 Treadstone 71 - www.treadstone71.com

ID	When	Call	Probability and band	Confidence
W11	3 November	Hoax threats reach polling places in at least two states	75 percent, Likely (55 to <85)	Low
W12	By 9 November	A federal agency attributes the hoax threats to a foreign actor	10 percent, Highly unlikely (5 to <15)	Low
W13	3 to 5 November	A state results page suffers a publicly reported flood of traffic or intrusion	36 percent, Unlikely (15 to <45)	Low
W14	By 9 November	A fake results graphic, call or concession clip in a news brand's livery is documented	86 percent, Highly likely (85 to <95)	Moderate
W15	By 9 November	One foreign-attributed false-results item passes 1 million views	45 percent, Roughly even chance (45 to <55)	Low
W16	By 9 November	Russian, Iranian or Chinese state outlets or officials call the US count flawed or illegitimate	91 percent, Highly likely (85 to <95)	Moderate
W17	By 9 November	A US political figure of national standing blames foreign interference for a result	49 percent, Roughly even chance (45 to <55)	Low
W18	By 9 November	A federal agency says on the record that no foreign activity materially affected the election	71 percent, Likely (55 to <85)	Low
W19	By noon eastern, 4 November	Networks call House control. Senate control stands at 49 percent. By 9 November the figures are 94 and 81 percent	73 percent, Likely (55 to <85)	Moderate
W20	By 9 November	A cyber action changes a certified result	2 percent, Almost no chance (1 to <5)	Moderate

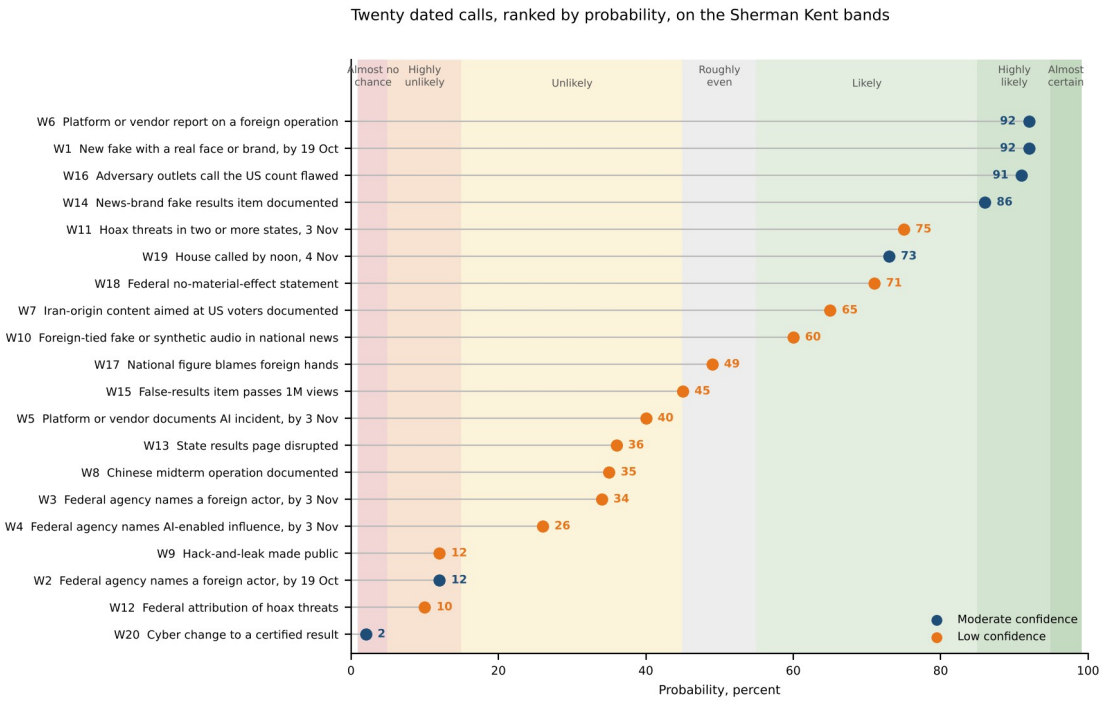


Figure 24. *Calls ranked by probability and colored by confidence.*

Estimates

Each quantity carries a median and a range that should hold the real figure four times in five.

Table 54. Estimates to 9 November (E).

ID	Quantity	Median	80 percent range	Basis
WE1	Items a researcher or outlet ties to Matryoshka, 9 October to 9 November	60	25 to 150	A wave running since about 10 September. No public item count
WE2	Views on the most-viewed foreign-attributed item with a platform count	400,000	120,000 to 3,000,000	Mother Jones fakes drew roughly 155,000 and nearly 200,000 views, with bot inflation reported
WE3	Polling sites reported with a hoax threat on 3 November	40	8 to 250	Threats in several states in 2024. No 2026 count
WE4	On-record federal statements naming foreign election influence	1	0 to 4	None on file for 2026 in Section 9.1
WE5	Platform and vendor reports naming a foreign operation aimed at US politics, 24 October to 9 November	3	1 to 7	Cadence of Meta, OpenAI, Microsoft, Google and Graphika in 2026
WE6	Days from Election Day until networks call Senate control	1	0 to 12	Slow-count states and close races

How adversaries work the result

Five levers act on perception. None touches a ballot. Timing is the opening. The hours between poll close and the last call leave the field to whoever speaks first, and the federal voice is thinner than in 2024. Fakes posted into that gap meet fewer official answers.

Mirror claims are the second feature. A domestic figure says foreign hands changed the outcome. Adversary outlets then repeat the claim as proof that American democracy has failed. The loop costs the adversary nothing.

Table 55. Levers on the result (forecast).

Lever	Likely actor	Form	Timing	Call
Fake results and fake concessions	Russia, through Matryoshka	Graphics in the livery of CNN, the BBC or the Times. Altered clips of candidates and officials. Voice clones	Evening of 3 November to 5 November	W14, W15
Hoax threats	Senders with Russian domains in 2024, and copycats	Emailed bomb threats at polling places	Hours of 3 November	W11, W12
Reporting-channel noise	Hacktivist personas with Russian or Iranian ties	Traffic floods at state results pages, and claims of a hack that outrun the effect	3 to 5 November	W13
Pretext and mirror claims	State outlets of Russia, Iran and China. Domestic voices	Reports of a flawed count. Blame placed on foreign hands	4 to 9 November	W16, W17
Hack-and-leak	Iran, on the 2024 precedent	Stolen campaign or party material released near the vote	20 October to 2 November	W9
Model poisoning	Russia, through the Pravda network	Planted articles meant to shape chatbot answers to results queries. A documented test on midterm queries stands at 25 percent, Unlikely (15 to <45), Low confidence	Whole window	None

The frames shift with the outcome. Each cell predicts content and carries Low confidence. The cells rest on each actor's pattern in Sections 3 to 6 and on the Times account.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 56. Likely frames by outcome (P).

Outcome	Russia	Iran	China
Democrats take both chambers: 57 percent	Fraud and elite-rigging stories for the right, chaos stories for the left, attacks on the winners' legitimacy	The vote as a verdict against the war	State media on a divided, failing democracy
Democratic House, Republican Senate: 33 percent	Gridlock stories and fights over aid to Ukraine	A split verdict on the war and the war-powers fight	The same dysfunction frame, with little else
Republican House: 10 percent	A stolen-by-Republican-states line for left audiences	Rigged-vote claims in Persian and Arabic outlets	Restraint

Signposts and decisions

Signposts move the numbers. Each row names the sign and the new figure.

Table 57. Signposts that move the calls.

If this appears	Then
A platform or vendor names a Matryoshka takedown before 24 October	W5 rises from 40 to 70 percent
A joint ODNI, FBI and CISA statement	W3 resolves. W18 rises to 80 percent. Forecast 1 flips
A tipping-point Senate race stays inside one point at poll close	WS2 rises from 25 to 40 percent
The White House or Justice disputes a result within 24 hours	WS3 rises from 25 to 45 percent. W17 rises to 75 percent
Hoax threats reach three or more states before noon on 3 November	WE3 median rises to 90. W12 rises to 15 percent
An official confirms an intrusion on a results page	WS4 rises from 10 to 30 percent
US operations in the Iran war restart before 9 November	WS4 rises to 35 percent. W7 rises to 85 percent

Decisions follow the dates.

Table 58. Decisions by date.

By	Decision
15 October	Name a contact at each platform for your brand and executives. Write the holding statement for a fake in your logo
31 October	Brief the board that silence is no safety. Set a watch roster for 3 to 9 November
3 November, poll close	Staff the watch. Treat any news-branded results graphic as false until the outlet's own channel confirms it
4 and 5 November	Share no unverified results clip. Check every claim against the AP and the state results page
6 to 9 November	Log what you saw. Hand it to the platform or the vendor. Score the calls against the pages

Where the forecast breaks

Four conditions break the forecast. A federal agency speaks early and often, which lifts the federal calls and shortens the life of fakes. An intrusion on a reporting channel, once verified, lifts WS4 and every count call. The Iran war restarts in election week and pulls adversary effort away from the vote. A recount runs past 9 November, and the window extends.

The weak joints sit in three places: base rates that rest on press accounts of past cycles, an outcome baseline that comes from raters and markets, and scenario odds that are judgments with no history behind them. Calls resolve against public pages on 3 November and 9 November, and the Brier score for the set goes to the run log.

Copyright 2026 Treadstone 71 - www.treadstone71.com

13 So what for readers

Three readers act on the assessment: the enterprise that runs AI, the buyer of counter-influence work, and the analyst or editor who reads the record. Each section below names the action, the evidence behind it and the date.

13.1 For the enterprise that runs AI

One firm now holds three exposures at once. Adversaries read its chatbots as an audience, its keys and models as targets, and its assistants as a channel. A firm that splits the three among separate owners leaves gaps between them. The first action closes the gaps. Section 7 supplies the evidence for the others.

Table 59. 13.1 For the enterprise that runs AI

#	Action	Evidence	Decide by
1	Name one owner for AI exposure across influence, supply chain and agent permissions. The owner reports to the security chief and the communications chief together	Section 7 role table	31 Oct 2026
2	Inventory every gateway, key store and agent runtime. Rotate the provider keys a gateway holds, pin versions and hash-check every release	TeamPCP and LiteLLM (EN-1, Level 3, T 80)	31 Oct 2026
3	Strip agent permissions until each agent has a written reason to hold them. Require authentication on every MCP server	NSA note and joint guidance (GV-6), OX Security (EN-2)	30 Nov 2026
4	Trace what feeds the retrieval layer. Log the sources and treat open-web retrieval as untrusted	Pravda network (RU-2, RU-3)	30 Nov 2026
5	Grade AI-generated analytic output on the ASA risk clusters C5 and C6 before it reaches a decision	ASA Evidence Standard 1.2	31 Dec 2026
6	Rehearse an impersonation attack against the next AI facility or product launch. Publish a verified-source page and name a response lead	Firebird (RU-5)	before the next launch
7	Add photo-reuse and live-video checks to remote hiring	Microsoft on Jasper Sleet and Coral Sleet (KP-1)	31 Dec 2026
8	Watch account markets and infostealer logs for staff AI credentials	GTIG on account prices (EN-3)	standing
9	Require a callback on a known number before anyone approves a payment instruction that arrived on a video call	Arup (EN-13)	31 Dec 2026

Use the CARM test from Section 7.6 before any broad response. Cheap actions go first. The cheap action that matches the exposure passes, and an action that stops the business waits for a second look.

13.2 For the buyer of counter-influence work

Do not fund a counter-influence program that cannot state its design. Section 11.5 shows that no effect claim in the open record clears the first rung of the ASA-MOE ladder. A vendor who measures its own campaign grades no higher than C on reliability. Ask for four things before the contract: a pre-registered baseline, a comparison design, an independent audit, and an Effect Statement that names who paid, who ran and who measured. A program that cannot supply them buys activity. Activity then gets reported as effect.

13.3 For analysts and editors

Four habits follow from the record.

Read the labs as a floor. Every lab report samples only the lab's own users, so annotate each AI Tradecraft Adoption score with a lab-visibility discount and say so on the page. The annotation marks local use as unknown and leaves the score unchanged.

Refuse cross-source percentages. The EEAS figure of 27 percent counts 147 of 540 incidents the service detected in 2025, and its report with three annexes prints no coding rule. CrowdStrike's 89 percent is a year-on-year rise in attacks by adversaries it tags AI-enabled. Its report body defines no such tag and prints no count or baseline behind the figure. The two measure different things. Quote each with its source, and compare, average or trend neither against the other.

Keep fact and claim apart. A ministry statement, a PLA essay and a Putin speech enter the file as claims at grade C. They never share a sentence with a verified fact.

Read silence as capacity. An American assessment with no AI-influence language fits a changed posture. No source shows adversary activity falling.

13.4 Decision calendar

Dates run in order, nearest first. Items marked Now start today and keep running. Each row names the section that holds the evidence.

Table 60. *Decision calendar for readers, in date order.*

When	Action	Where
Now, standing	Watch account markets and infostealer logs for staff AI credentials	13.1, action 8
Before the next AI launch	Rehearse an impersonation attack. Publish a verified-source page and name a response lead	13.1, action 6
20 Oct to 20 Nov 2026	Comment window on the ASA-MOE draft. Buyers of counter-influence work and analysts can comment	13.2
31 Oct 2026	Name one owner for AI exposure. Inventory every gateway, key store and agent runtime, and rotate the provider keys a gateway holds	13.1, actions 1 and 2
3 Nov 2026	Midterms. Forecast 1 resolves	12.12
9 Nov 2026	The election window closes. Calls W1 to W20 and estimates WE1 to WE6 are scored	12.17 and 12.18
30 Nov 2026	Strip agent permissions and require authentication on every MCP server. Trace what feeds the retrieval layer	13.1, actions 3 and 4
31 Dec 2026	Grade AI analytic output on the ASA risk clusters, add live-video checks to hiring, and require a callback on video-call payments. Tripwire T3 checks	13.1, actions 5, 7 and 9
Jan 2027	ASA-MOE reaches version 1.0	13.2
31 Dec 2027	The last forecast resolves	12.12

13.5 A regional read for operators outside the United States

Most of the ledger comes from US bodies and US labs. Operators in Europe, Australia, Singapore and the Gulf who run critical infrastructure meet the same threats under different duties. The table pairs ledger rows with published regional texts. Regional duties differ. Treadstone 71 is no law firm. Each duty below is our reading of a published text. Counsel confirms it before a firm acts.

Table 61. 13.5 A regional read for operators outside the United States

Region	What the ledger shows	Duty or guidance	What follows for the operator
European Union	ENISA reports consistent threat-group interest in AI systems as tools and as targets (EN-9). The EEAS counts 540 incidents in 2025. Its summary says one in four detected incidents involved AI tools to produce or distribute content, 147 incidents, and the report publishes no coding rule. It links 29 percent of the 540 to Russia and 6 percent to China, and 65 percent stay unattributed (RU-9). Finland's SUPO and Lithuania's VSD describe AI-made influence material and AI inside malicious code (GV-3)	NIS2 Article 21(2)(d) requires supply chain security, including each entity's relationships with direct suppliers. Article 21(3) tells entities to weigh each direct supplier's vulnerabilities and secure development practices. Article 23(4) sets an early warning within 24 hours, a notification within 72 hours and a final report within one month. Article 20 puts management bodies in charge of approving and overseeing the Article 21 measures, and makes them liable for infringements. Article 32(5)(b) lets a competent authority temporarily bar a person with managerial duties at chief executive or legal representative level in an essential entity. Article 34(4) sets the ceiling for essential entities at no less than EUR 10 million or 2 percent of worldwide annual turnover, whichever is higher, and Article 34(5) sets it for important entities at no less than EUR 7 million or 1.4 percent	An AI gateway, model host or agent framework sits in the supplier register. The LiteLLM release stayed live about three hours, and a firm that pulled it still decides within 24 hours whether the incident is significant. Put the AI stack in the supplier register and in the incident playbook
United Kingdom	In a blog post dated 30 March and modified 29 May, the NCSC told defenders to assume attackers hold capable AI tools. On 17 June its chief executive reported more than 200 incidents against UK critical infrastructure in the year to May 2026, and said the NCSC believes about 75 percent link to states (GV-2)	NCSC statements. The 75 percent covers all incidents and does not isolate AI systems	Treat the state-linked share as the planning case for infrastructure. Read the figure as an upper frame for AI exposure and never as an AI-specific count
Australia	ASD guidance of 8 September says AI lets actors develop exploits within hours (GV-2)	ASD guidance of 8 September 2026 for government, critical infrastructure and large enterprises. Immediate priorities: an executive accountable for AI risks, rapid patching of all severities on internet-facing services, and a daily check of those services. Short-term: least privilege, segmentation, centralized log analysis with automated triage, and exercised incident plans. ASD's alert of 25 September covers AI agents taking unintended actions that probed public-facing sites. It reports no sign of wider threat or malicious targeting, so it is a misalignment alert and no adversary-use alert	Action 1 in Section 13.1 matches ASD's first priority. A daily inventory of exposed services extends to the gateway and the agent runtime

Copyright 2026 Treadstone 71 - www.treadstone71.com

Region	What the ledger shows	Duty or guidance	What follows for the operator
Singapore	No Singapore row sits in the ledger. The exposure comes from the enterprise findings in Section 7	The Cyber Security Agency's addendum on securing agentic AI, 17 June 2026, voluntary guidance. Controls: supply chain security, access control and segmentation, input and output validation, hardening, human oversight, continuous logging. It names rogue agent actions and sensitive data disclosure as the main risks and leaves cyberattack misuse of AI outside its scope. The table holds no Cybersecurity Act duties. The Agency's release of 22 July says the updated code of practice for critical information infrastructure comes later in 2026. No amended code appears in the record	The addendum covers an agent that fails on its own. Adversary abuse of agents, the MCP and configuration-file paths in Section 7.3, sits outside it. Add those paths to the threat model
Gulf states	Anthropic reports a UAE-directed influence program (GTG-84002) and a commercial vendor that profiles accounts of Iranian and Persian Gulf-based users (GTG-54009). The two rest on Anthropic alone (PX-1). We hold no row that shows AI-enabled adversary influence aimed at a Gulf infrastructure operator	The UAE government portal's page on the National Cyber Security Policy for AI, updated 2 July 2026, confirms six areas, among them adversarial AI attacks and AI monitoring and response. The page shows no version number. Saudi Arabia's NCA put its AI Cybersecurity Guidelines to consultation on 5 July 2026, with feedback due 5 August, across four domains: governance, defense, resilience and third-party cybersecurity. The consultation page reads Closed on 7 October. The NCA list of regulatory documents (17 items, newest 22 July) and its news from 18 June to 7 October show no final AI guidelines, a null for the NCA site alone. SDAIA lists a deepfake guidance document on its own regulations page. The UAE Cabinet release of 14 June names an Artificial Intelligence and Data Authority that reports to the Cabinet and whose mandate includes support for cybersecurity and government information security. Qatar's guidelines for secure adoption of AI (version 1.0, February 2024) predate 2026, and no Qatari text dated 2026 appears	Read Sections 3 and 7 as exposure by proximity to the war with Iran, and as no documented targeting. The UAE area on adversarial AI attacks and the Saudi third-party domain line up with the gateway and supply-chain findings

Four findings hold in every region: AI gateways and agent frameworks as supply-chain targets, over-privileged agents, untrusted retrieval sources, and impersonation of AI projects. The ledger cannot say how often any of them strikes a non-US operator. European bodies count incidents, and none splits out infrastructure operators. No row shows an adversary aiming AI-enabled influence at a Gulf or Singapore operator.

Assessment: a European operator's AI gateway falls inside the NIS2 supply-chain duty. Likely (55 to <85), Low confidence. The reading is ours. Nothing in the Directive names AI. A duty clock of 24 hours runs shorter than the

exploit clock that ASD and the Five Eyes describe in hours, so the operator's detection speed sets whether the first report arrives on time.

13.6 Who cares, what to decide and what it costs

The effort figures below come from a task model. Each decision splits into tasks. Every task takes an assumed time per unit. The unit counts draw from ranges that fit a firm of 1,000 to 5,000 staff with a working security team. 40,000 draws give a 10th to 90th range and a median, in person-days of internal time, before any tool purchase and before counsel time. Training and tabletop rows count participant time. The second table below lists the main units and times. Every unit time is ours. No source and no client log measures one. Scale the figures to the firm.

Table 62. *Effort model, main units and assumed times (minimum, most likely, maximum).*

Reader	Why the reader cares	Decision	Decide by	Effort (task model, 10th to 90th)
Security chief	AI supply-chain compromise ranked first in about 70 percent of 20,000 rank runs (Section 10.5). The LiteLLM release stayed live about three hours	Inventory every gateway, key store and agent runtime. Rotate the provider keys a gateway holds. Pin versions and hash-check releases	31 Oct 2026	17 to 35 person-days (median 25)
Head of platform engineering	Over-privileged agents magnify one compromise (GV-6). The MCP path is open in research (EN-2)	Strip agent permissions that lack a written reason. Require authentication on every MCP server	30 Nov 2026	15 to 39 person-days (median 25)
Owner of the retrieval layer	Pravda content sits in a major training source, and retrieval poisoning is a named path (RU-2)	Log the sources that feed retrieval. Treat open-web retrieval as untrusted	30 Nov 2026	7 to 28 person-days (median 15)
Finance chief	Hong Kong police put the Arup transfer at HK$200 million, and Arup's accounts book 25.4 million pounds for the fraud and associated costs. A phishing email and a group video conference started it, per the Hong Kong government (EN-13)	Require a callback on a known number before any payment instruction that arrives by video call	31 Dec 2026	6 to 15 person-days (median 10)
Head of hiring	North Korean fake workers pass interviews (KP-1, KP-4)	Add photo-reuse and live-video checks to remote hiring	31 Dec 2026	7 to 12 person-days (median 9)
Communications chief	Forged outlet videos targeted a named AI facility (RU-5)	Publish a verified-source page. Name a response lead. Rehearse an impersonation attack	Before the next launch	8 to 14 person-days (median 11), including the tabletop
Board and compliance	The NIS2 clock is 24 hours, and ASD names an accountable executive	Add the AI stack to the supplier register and the incident playbook. Confirm regional duties with counsel	31 Dec 2026	7 to 14 person-days (median 10) plus counsel time
Buyer of counter-influence work	No effect claim in the open record clears the first rung (Section 11.5)	Require a pre-registered baseline, a comparison design, an independent audit and an Effect Statement	Before contract	Adds 4 to 8 weeks to program start. A schedule estimate, no effort model
Analysts and editors	Lab reports sample their own users, and cross-source percentages mislead	Annotate scores with the lab-visibility discount. Refuse cross-source percentages. Keep fact and claim apart	Standing	1 to 2 person-days to set the checklist

Table 63. 13.6 Who cares, what to decide and what it costs

Decision	Units drawn (minimum, most likely, maximum)	Time per unit
Security chief	1, 2 or 6 gateways. 2, 4 or 10 key stores. 2, 5 or 15 agent runtimes. 5, 12 or 40 provider keys. 5, 10 or 30 build pipelines	4, 8 or 16 hours per system. 0.5, 1 or 2 hours per key. 1, 2 or 4 hours per pipeline
Head of platform engineering	3, 8 or 30 agents. 3, 8 or 25 MCP servers	3, 6 or 12 hours per agent. 4, 8 or 16 hours per server
Owner of the retrieval layer	1, 3 or 8 retrieval pipelines	12, 24 or 60 hours per pipeline
Finance chief	Policy, workflow change and 20, 50 or 120 trainees	4, 8 or 16 hours of policy. 4, 12 or 32 hours of workflow change. 0.5, 0.75 or 1.5 hours per trainee
Head of hiring	Design, tooling, pilot and 5, 12 or 30 recruiters	8, 12 or 24 hours of design. 8, 16 or 40 of tooling. 8, 12 or 24 of pilot. 1, 1.5 or 2.5 hours per recruiter
Communications chief	Page, response lead, playbook and 6, 10 or 15 tabletop participants	6, 10 or 16 hours for the page. 8, 12 or 24 for the playbook. 4, 6 or 8 hours per participant
Board and compliance	5, 15 or 40 AI suppliers, playbook and regional review	1, 2 or 3 hours per supplier. 16, 24 or 40 hours of playbook. 8, 16 or 32 hours of regional review
Analysts and editors	Checklist and team review	4, 6 or 8 hours to draft. 2, 3 or 6 hours of review

A firm with no AI in production and no public AI project skips the gateway, agent and retrieval rows. The payment-fraud and hiring rows still apply.

14 Questions answered and unanswered

A report earns trust by stating what it settled and what it left open. The Priority Intelligence Requirements in Section 1.2 set six questions. An audit follows that tests the report against those six and against the questions a reader raises.

14.1 What the report now answers

Eleven questions now have answers.

Table 64. 14.1 What the report now answers

Question	Answer	Band and confidence	Where
Do Iranian, Russian and Chinese state bodies use AI in influence and security work?	Yes, as staff for content, personas and monitoring. North Korean units use it for fake identities and lure documents	Likely, Moderate (KJ1, one point above the Low line)	Sections 3 to 6
How far does the documented work reach?	Small in the cases where someone measured reach	Likely, Moderate (KJ2)	Sections 4.6, 5.2, 11.1
Does Russian state-linked content sit in a chatbot training source?	Yes, in Common Crawl	Likely, Moderate (KJ3)	Section 4.2
Are enterprise AI gateways and frameworks compromised now?	Yes, by criminal groups, and one compromise reaches many keys	Likely, Moderate (KJ4)	Section 7.1
Did Chinese labs distill US frontier models at industrial scale?	Anthropic reports yes, through fraudulent accounts. A joint NSA, CISA and FBI advisory names six firms and gives no volume beyond billions of tokens	Likely, Moderate (KJ5)	Section 5.5
Does the thin US record show adversaries falling back?	No source shows it. CISA's 2026 election plan has no AI text and names no foreign state	Likely, Moderate (KJ8)	Section 9.1

Copyright 2026 Treadstone 71 - www.treadstone71.com

Question	Answer	Band and confidence	Where
Which enterprise threat ranks first, and does the rank hold?	AI supply-chain compromise, and it holds under random rating changes	Rank 1 in about 70 percent of runs	Section 10.5
Does the lead evidence depend on Anthropic?	Yes. The effect judgment survives removal. KJ1 for Iran and KJ7 do not, and KJ5 drops to Low confidence	Not applicable	Sections 2, 11.1
What changes our mind?	Eight dated tripwires and four named breaks	Not applicable	Sections 12.6 and 12.17
What are the odds by actor, vector, region and date, and which signals move them?	Five paths resolve by rule: 39, 30, 18, 5 and 8 percent. Counted baselines feed the model. States tilt around the aggregate and differ most on B, C and D. A quiet lab record moves B up. Eight indicators carry 83 percent of the information	Low confidence on every share and on the actor, date and regional splits	Sections 12.2 to 12.17
What does a defender do first, and by when?	Nine reader-specific decisions with dates and effort estimates	Not applicable	Section 13.6

14.2 What the report answers in part

Seven answers come with limits.

Table 65. 14.2 What the report answers in part

Question	What we know	What stays open
Did AI cut cost and raise output?	Staff-work findings in three theaters make it the likely reading (KJ2a, Likely, Low)	No row measures cost or output
Who directs the two pro-Iran networks with a billion views?	ISD finds no evidence of state backing. Diplomatic accounts shared the posts	Command is unseen (KJ11, Unlikely)
Are agent and MCP flaws exploited at scale?	One malware family abuses coding assistants in the wild (EN-3). The paths work in research. VulnCheck's honeypot recorded first-time attempts on the Flowise flaw from one IP address, and no victim is reported. Mercor confirms an intrusion through the compromised LiteLLM releases and states no data volume	Use at scale has no public proof (KJ9a, Roughly even chance). The 4 TB figure is Lapsus$'s own claim, relayed by Cybernews, which failed to open the data
Does a state stand behind compromises of AI stacks?	Two vendors, CrowdStrike and Microsoft, attribute the Mastra injection to a North Korea-nexus actor	No government has said so. On 7 October the pages of the FBI, CISA, the UK NCSC, ASD and the Canadian CCCS carried no attribution, and the joint advisory of 22 September on North Korean npm malware leaves Mastra unnamed. The TeamPCP cases name criminals

Question	What we know	What stays open
How much adversary work runs on local models?	Three reporters describe self-hosted use by CopyCop, Kimsuky and a Chinese intrusion actor. The North Korean stack has one source and has not been shown running against a victim. A Treasury release of December 2024 places a GRU-funded AI server in a rented Moscow apartment, though it names another body. The Washington Post of October 2024 reports a GRU-paid AI server for Dougan and names no model. A French preprint supports a Llama 3 attribution for CopyCop text. Counting finds local models in 2 of 16 ledger non-provider rows and in 4 of 34 extended rows. The 14 reports of October 2025 to January 2026 show no actor running its own model and bring the count to 4 of 44. The model puts the latent share near 23 percent. No Western government source reports such use, and VIGINUM forecasts it	The ledger count carries an interval of 3 to 36 percent, the extended count 5 to 27, and the latent share runs 8 to 46 percent (KJ6, KJ14, Low). One coding ruling moves Path B by 11 points
Do chatbots repeat the planted content?	Presence in the training source is shown. ISD's retrieval test of October 2025 found Russian state-attributed sources in nearly 18 percent of 300 answers, 11 percent on neutral prompts, with two links from Pravda	No probe of memorized Pravda text has run on a model trained after the network expanded. Retrieval probes have run (NewsGuard, HKS and ISD), and each counts a different quantity (KJ3a). A citation is not a belief, and the American Sunlight Project's 504,060 monthly responses in the EU come from a model and not a measurement
What does the EEAS mean by AI involvement in its 27 percent?	The fourth report, annexes included, says one in four detected incidents involved AI tools to produce or distribute content. Its AI chapter counts AI-related tactics in 147 of 540 incidents, against 41 the year before, and sorts them by function	The EEAS publishes no coding rule, no threshold and no coder-agreement figure. The body of CrowdStrike's 2026 report gives 89 percent as a growth rate for adversaries it marks AI-enabled. The body defines the mark nowhere and counts no base, so the two figures cannot sit side by side. The sample is selective and draws partly on public reporting. The 2024 and 2025 samples differ in size and channel count, so the rise from 41 to 147 compares unlike sets

14.3 What the report does not answer

Eleven questions stay open.

Table 66. 14.3 What the report does not answer

Question	Why it stays open
Does any AI-enabled operation change belief or behavior?	No study in the ledger uses a comparison design, and no government source in the ledger measures it. Bail's panel finds no effect inside its sample, with every credible interval across zero, and its supplement's alternative specifications, also in figures, leave every interval across zero, while a preprint experiment on a TikTok clone finds a per-minute shift for pro-China influencer content. Laboratory studies measure persuasion per exposure. No study tests an AI-enabled operation
What share of adversary AI work did the labs never see?	Labs see only their own users, and the counts of Section 12.2 bound the share only loosely. No Western government source reports adversary use of local models

Copyright 2026 Treadstone 71 - www.treadstone71.com

Question	Why it stays open
Which orders reach the units?	Open sources show the work product and never the instruction
Did Pravda text enter a model as memorized output?	No probe of memorized Pravda text has run on a post-expansion model. Retrieval probes have run (NewsGuard, HKS and ISD)
Do Eitaa, Bale and Rubika carry the amplification Anthropic describes?	A 21.9K-follower channel of IRGC supporters on Eitaa relays war news and labels its one AI-made video. Public pages show no bot activity and no paid amplification, and a public page says little about either
Do Hezbollah, Houthi and Hamas channels use AI?	Fourteen Telegram previews, in windows of 62 to 771 posts, show no AI-made media of their own and no claim of AI production, judged on post text alone. The AI items they carry relay other parties' claims. The Al-Masirah Net and Sanaa Saba home pages, one text snapshot dated 7 and 8 October, carry no AI claim and no AI-made media. The record holds no text from the Al-Aqsa TV and Qassam sites, and Al-Aqsa's channel stops in May 2023
Do the T and D scores predict anything?	No test of T or D against later events has run. 27 of 48 scored rows sit on the D floor of 25, so D carries almost no spread, and the correlation of T and D reads -0.01
Do the indicator likelihoods predict resolved events?	A hindcast over up to four half-year windows per indicator scores the judged chances at a Brier of 0.14 on 50 scored windows, against 0.25 for a coin flip. The figure is a floor check, since we set the chances knowing the period
Do non-US infrastructure operators face AI-enabled influence from the four states?	No row shows it for Singapore or the Gulf, and European counts do not split out operators. ASD's alert of 25 September concerns AI agents that took unintended actions and shows no malicious targeting. Singapore's code update is announced and not released, and Saudi Arabia has issued no final AI guidelines
Will any US agency say in public that AI-enabled foreign influence targets the midterms?	The record holds no executive-agency statement of that kind in 2026 from ODNI, FBI, CISA, DHS, State or DOJ. CISA's own election security plan of 24 September holds no AI text and gives CISA no foreign-influence task. Senator Warner and Representative Himes ask for a classified update by 31 October. The Padilla and Morelle letter of 2 September carries one passing line on AI. The one operation named for the midterms, a Matryoshka wave of doctored videos from about 10 September, rests on researchers, the press and anonymous officials in a Times story, and no agency has named it
What do the recommended actions cost?	A task model gives person-day ranges for eight decisions. Every unit time is our assumption

Further reading moves one of the eleven questions before 3 November: the one on US agency statements, since agency pages change each day. A search of 9 October for second sources behind the four Anthropic-only judgments and behind the self-hosted model findings turned up context and one dispute, and no corroboration. The judgments stay as printed.

The fourth omission in Section 9.5, the lack of any separation between fact and adversary claim, is not a question the report leaves open. Practice answers it: the report keeps each verified fact and each adversary claim in separate sentences.

Capture and recapture deserves a word. If two detectors, say a lab and a vendor, each find a set of operations, the overlap sizes the unseen remainder. The estimate equals the product of the two counts divided by the number of operations the two share. Too few operations in the ledger appear in the work of two reporters, so the report prints no estimate.

14.4 Which questions matter most

Three open questions carry the decisions. Whether any operation moves belief governs whether a client funds counter-influence work at all. How much the labs miss governs how far a client trusts any lab report. The question of whether a state stands behind compromises of AI stacks governs how an enterprise ranks supply-chain risk against the influence threat. Section 12.6 names the observation that closes each.

15 Evidence limits

Thirteen limits stand on the evidence the report carries. No step closes them, and each marks the edge of what a judgment claims.

Table 67. 15 Evidence limits

Limit	Judgments it touches	Why it stays
Voice of Korea stays unreachable, and the KCNA Watch search needs a sign-in. State-media items from 2026 carry the AI claims	KJ1 for North Korea	KJ1 for North Korea rests on KP-1 (T 77) and KP-3 (T 72), not on state media. The sign-in search lies outside what the work crosses
No primary AI media appears from Hezbollah, Houthi or Hamas channels. Windows run 2.5 to 6.5 days, and only post text was scanned	Proxy judgments	No headline judgment rests on the null. A picture test needs forensic work on the media, and a web page does not supply it
No Russian vendor names Russian state AI use. F6 leaves Russia out by design, so its silence says little	Russia sections	No ledger row depends on it. The vendor reports pair AI with criminals or with foreign actors
Chinese vendor pages show no US or Taiwan AI operation. The CVERC annual report body is read and holds none, which carries the null to its five vendor co-authors. Other report bodies stay unread	China sections	No band moves.
Russia has not published the decree that names the sovereign-model cases	Forecast 5	The decree is an event nobody here controls
The Mercor "4 TB" figure is Lapsus$'s own claim, relayed by Cybernews. The leak site is unread	KJ9, parts of KJ2	The leak site is criminal infrastructure, and the work does not fetch it. The figure stays a claim
Open sources prove content, channels and footprints and never the order given	All attribution to units and commanders	No open source holds an order. Keep attribution bands a step under pattern bands and mark every wall crossing
Two vendors, CrowdStrike and Microsoft, attribute the Mastra injection to a North Korea-nexus actor. No government confirms it	Forecast 9, Section 7.1, tripwire T5	Government attribution is an event nobody here controls. Vendor attribution holds at moderate confidence
Regional evidence is thin for Singapore and the Gulf. The Saudi final guidelines and the Singapore code update are not issued	Sections 12.11 and 13.5	Two sources do not yet exist
D sits on its floor of 25 on 27 of 48 scored rows, so T and D show a correlation of -0.01 and D carries almost no spread	Section 16	D feeds no judgment. Evidence of deception would lift rows off the floor
The outside view rests on studies of one agency on one platform, and a thin set from other platforms. No study of an AI-enabled operation exists	Section 10.7, KJ13	The Bail paper prints no estimate table, and its effect sizes come from figures. Every interval includes zero
Indicator likelihoods, driver scores, model inputs and wildcard chances rest on analyst judgment	Sections 12.1 to 12.17	Scoring comes only as events resolve through 2027, and Section 11.9 audits the first 50 windows. Each review date publishes a Brier score
Two shared regimes stand for the ways indicators cluster, and the loadings are judged	Section 12.8	Refitting comes only at each review date

Copyright 2026 Treadstone 71 - www.treadstone71.com

16 Evidence ledger

The ledger holds every source that supports a judgment, 63 rows in all (Sections 16.3 and 16.4 list further statements and sources, unscored), split 7, 41 and 15 across Levels 3, 2 and 1. Each row carries an Admiralty anchor (source reliability A to F, information credibility 1 to 6, printed apart), an ASA T-Score for quality, and a D-Score for the likelihood of deception with its Kent band. Level 1 rows carry the anchor alone. Twenty-two of the 48 scored rows rest on an older two-cluster method for Level 2 rows, so ledger means and band counts mix two scoring bases.

Table 68. *16 Evidence ledger*

ID	Source and date	Anchor	Level	T	T band	D	D band	Note
IR-1	Anthropic, GTG-34001, Iranian state propaganda bodies, 10 Sep 2026	B2	3	75	Adequate	30	Unlikely	Vendor telemetry on own platform; institutions tied by actor disclosures, document footers and open-source checks
IR-2	Anthropic, GTG-34007, 30006, 30004, Iranian surveillance and profiling units, 10 Sep 2026	B2	2	73	Adequate	30	Unlikely	Same vendor, same telemetry limits
IR-3	Mehr News, cognitive-war speech, 6 Oct 2026 (statement made)	C3	2	54	Weak	25	Unlikely	Rule 1 cap. Records what the regime says, supports no judgment on facts. A provincial police commander's speech with no AI content
IR-4	Mehr News, commentary on the Anthropic report, about 14 Sep 2026 (statement made)	C4	2	48	Weak	32	Unlikely	Rule 1 cap. Page dated 14 Sep 2026. The headline claims more than the body
IR-5	Recorded Future Insikt, Iran AI asymmetric playbook, Jan to Jun 2026	B2	2	69	Adequate	25	Unlikely	Published 16 Jul 2026. Synthesis of other reports, and the Cyabra statistics stay secondhand
IR-6	ISD, Iranian diplomats' meme war, 23 Apr 2026	B2	2	75	Adequate	25	Unlikely	Open-source research body
IR-7	ISD, BRICS4CLICKS and Verified4War networks, spring 2026	B2	2	73	Adequate	28	Unlikely	ISD found no evidence the networks are state-backed
IR-8	Anthropic, GTG-84006, MEK and NCRI-aligned network, 10 Sep 2026	B2	2	73	Adequate	32	Unlikely	Opposition movement, not a state
IR-9	FBI, UK NCSC, Dutch AIVD, CHOSEN BRICK alert, 15 Sep 2026	A1	1	n/a	n/a	n/a	n/a	No AI element in the alert
IR-10	ISNA, IRGC naming US technology firms as targets, about 2 Apr 2026	C4	1	n/a	n/a	n/a	n/a	ISNA relays an FT story on Amazon in Bahrain with one sentence on the IRGC threat to three US firms. No AI element
RU-1	Anthropic, GTG-24015, Russian state-media editorial pipelines, 10 Sep 2026	B2	3	76	Adequate	25	Unlikely	Content traced to published Sputnik and RIA Novosti output
RU-2	DFRLab, Pravda in the pipeline, 8 Apr 2026	B2	2	75	Adequate	25	Unlikely	Separates retrieval poisoning from training-data poisoning
RU-3	American Sunlight Project, LLM exposure estimate, 2 Jun 2026	C4	2	53	Weak	25	Unlikely	Modeled baseline, not a measurement. Its input rate counts cited sources in test answers, not user exposure
RU-4	DFRLab, Storm-1516 and the Baltic states, 17 Sep 2026	B2	2	72	Adequate	25	Unlikely	The GRU link rests on patterns attributed to the GRU
RU-5	Recorded Future Insikt, CopyCop and the Firebird AI data center, Jul 2026	B2	3	68	Adequate	25	Unlikely	Artifacts are the impersonation videos themselves. Graded on the report, not the videos. Page dated 18 Aug 2026

ID	Source and date	Anchor	Level	T	T band	D	D band	Note
RU-6	Recorded Future Insikt, CopyCop self-hosted models, 2025	B2	2	65	Adequate	25	Unlikely	Page dated 18 Sep 2025. US Treasury (31 Dec 2024) separately describes a GRU-funded AI server. The Washington Post of 23 Oct 2024 reports GRU payments to Dougan and an AI server going live in summer 2024. The article says nothing of Llama, hosting or CopyCop. The Post and Treasury accounts possibly share an origin. The CopyCop link rests on the vendor
RU-7	ISD, Matryoshka and German state elections, 4 Sep 2026	B2	2	76	Adequate	25	Unlikely	ISD reports basic errors and limited reach
RU-8	Google GTIG, SANDWORM RELIC and UNC5792, 8 Sep 2026	B2	2	76	Adequate	27	Unlikely	Vendor attribution
RU-9	EEAS, 4th FIMI threat report, Mar 2026	B3	2	60	Adequate	25	Unlikely	Its annexes hold no coding rule for AI involvement
RU-10	EEAS and Ukraine CCD, joint report, 23 Jun 2026	C3	2	51	Weak	27	Unlikely	Ukraine CCD is a belligerent body. Counts are its own and the page shows no method
RU-11	OpenAI, International Burke Institute operation, 25 Aug 2026	B2	2	66	Adequate	25	Unlikely	34 of 36 linked articles were copied from elsewhere
RU-12	Kremlin, Putin at Valdai, 1 Oct 2026 (statement made)	C3	1	n/a	n/a	n/a	n/a	Rule 1 cap. Signaling
RU-13	Ukraine CCD, AI street-interview videos, 23 Sep 2026	C3	1	n/a	n/a	n/a	n/a	Belligerent source. Attribution is its own
CN-1	Anthropic, GTG-14020, 14021, 14022, PRC security organs, 10 Sep 2026	B2	3	60	Adequate	25	Unlikely	Vendor telemetry. Only GTG-14021 names municipal police and a state security bureau
CN-2	OpenAI, Cyber Special Operations (Feb 2026) and PRC-linked AI-debate operations (10 Jun 2026)	B2	2	64	Adequate	25	Unlikely	The 15-page June PDF holds no counts and no model detail, and hedges its attribution, so T holds
CN-3	Graphika, Spamouflage AI content on the Tibetan election, 4 Mar 2026	B2	2	64	Adequate	25	Unlikely	Tumblr articles stated as AI-generated, volume inferred
CN-4	PLA Daily, AI-plus-cognitive-attack doctrine, 13 Apr 2026 (statement made)	C2	2	63	Adequate	26	Unlikely	Rule 1 cap. Doctrine stated, practice unproven
CN-5	PLA Daily, claim that the US and Israel ran AI cognitive offensives on Iran, 13 Apr 2026	C6	2	42	Weak	45	Roughly even	Adversary assertion, own row under Rule 3. One unsourced sentence. No null documented, so credibility grades 6
CN-6	Taiwan NSB, claims on PRC generative-AI contractors, date unconfirmed	C4	2	49	Weak	27	Unlikely	Party to the dispute. The NSB site dates the release 11 Jan 2026
CN-7	Anthropic, distillation (Feb 2026), NSA, CISA, FBI AA26-251A (8 Sep 2026)	B2	3	64	Adequate	25	Unlikely	Names six firms and states no basis for its awareness judgment. PRC denial recorded apart
CN-8	PRC MFA, denial of the AA26-251A accusation, 9 Sep 2026	C5	1	n/a	n/a	n/a	n/a	Adversary assertion
CN-9	Estonian FIS (10 Feb 2026), CNAS Red Lines (12 Jun 2026), NIST CAISI on DeepSeek (30 Sep 2025)	B2	2	72	Adequate	30	Unlikely	Three sources on model behavior and propaganda risk
CN-10	Anthropic, AI-orchestrated espionage (Nov 2025) and GTG-10007, 10 Sep 2026	B2	2	60	Adequate	25	Unlikely	Single source, Anthropic. Row joins the Nov 2025 case (about 30 targets) and GTG-10007 (about 50 organizations). Split pending
CN-11	Meta, H2 2026 adversarial threat report	B2	2	63	Adequate	25	Unlikely	Dated Aug 2026. No China-origin network among the cases

ID	Source and date	Anchor	Level	T	T band	D	D band	Note
KP-1	Microsoft, AI as tradecraft, 6 Mar 2026	B2	2	77	Adequate	27	Unlikely	Jasper Sleet and Coral Sleet
KP-2	CrowdStrike, Global Threat Report (24 Feb 2026) and Threat Hunting Report (3 Aug 2026)	B2	2	61	Adequate	25	Unlikely	Vendor attribution. The Threat Hunting report body sits behind a form
KP-3	Genians, Kimsuky and the opencode agent, 7 Sep 2026	B2	2	72	Adequate	30	Unlikely	Keyboard layout and spelling signals conflict
KP-4	DOJ sentencings (20 Mar, 15 Apr, 6 May 2026) and Okta (11 Feb 2026)	A1	1	n/a	n/a	n/a	n/a	No AI element verified
KP-5	Chainalysis, Bitget theft (1 Oct 2026) and blockchain dead drops (17 Sep 2026)	B2	1	n/a	n/a	n/a	n/a	No DPRK AI use claimed for the theft
KP-6	KCNA, unmanned aeronautical technology complex, 19 Sep 2025	C3	1	n/a	n/a	n/a	n/a	Only AI-adjacent state item found. No 2026 item
CN-12	Vanderbilt Wicked Problems Lab, GoLaxy documents, 8 Sep 2025	B3	1	n/a	n/a	n/a	n/a	The document set runs 399 pages (112.9 MB), all read. Page 7 holds one sentence in which 117 numbers a Congress and counts no members, and 2,000+ counts other figures and opinion leaders. Dated papers run from 9 Nov 2020 to 6 Mar 2024. The firm's own scripts and table name state users, and the firm's pitch carries every deployment claim. Rows AO-CN-01 to 25 in Section 16.3 list the papers. Precedent, outside the window
PX-1	Anthropic, GTG-84002, 54006, 50027, 54009 (UAE-directed, Bangladesh, Mali, commercial vendor), 10 Sep 2026	B2	2	72	Adequate	32	Unlikely	Hired hands and commercial vendors
PX-2	Hezbollah, Houthi and Hamas channels, search null result	F6	1	n/a	n/a	n/a	n/a	No primary evidence found. Absence of collection, not absence of activity. The Saba Net and Al-Masirah Net front pages, one-day copies, carry no AI claim
EN-1	FBI FLASH on TeamPCP (2 Jul 2026), Unit 42, Wiz, GTIG UNC6780	A1	3	80	Strong	25	Unlikely	Criminal group. No state actor named
EN-2	OX Security, MCP architecture finding (15 Apr 2026)	B2	2	71	Adequate	30	Unlikely	Anthropic called the behavior expected
EN-3	Google GTIG, AI Threat Tracker, 8 Sep 2026	B2	3	68	Adequate	25	Unlikely	Credential theft in under six hours, model and prompt theft, account prices. The UNC6508 local-model passage is single-source and hedged in its lead sentence
EN-4	FBI and CISA, AA26-231A on Siemens S7 PLCs, 19 Aug 2026	B3	2	57	Weak	25	Unlikely	Unattributed. Five authoring agencies. AI use asserted with no sample or basis
EN-5	Gambit Security, one operator against nine Mexican agencies, Dec 2025 to Feb 2026	B3	2	53	Weak	25	Unlikely	Page date conflicts. Counts differ in secondary coverage
EN-6	Nx s1ngularity (Aug 2025) and Salesloft Drift (Aug 2025)	A2	2	81	Strong	25	Unlikely	Precedents from 2025
EN-7	Mandiant, M-Trends 2026	B2	2	75	Adequate	30	Unlikely	Counterweight. Date not shown
EN-8	IBM and Ponemon, Cost of a Data Breach 2026, 29 Jul 2026	B3	2	56	Weak	25	Unlikely	Self-reported survey data
EN-9	ENISA Threat Landscape 2026, Sep 2026	B2	2	67	Adequate	35	Unlikely	Treadstone 71 peer review applies

ID	Source and date	Anchor	Level	T	T band	D	D band	Note
EN-10	OWASP LLM Top 10 2026 (1 Sep 2026); Unit 42 Vertex study; Noma GrafanaGhost	B2	1	n/a	n/a	n/a	n/a	Research and standards. No in-the-wild abuse shown
EN-11	Hugging Face and OpenAI, July 2026 incident	A1	1	n/a	n/a	n/a	n/a	Evaluation-agent misalignment. Not adversary activity
EN-12	Anthropic, GTG-50020, 50029, 50014, 20006 (AI supply-chain attempt, stolen-key hacktivist, credential crew, state-nexus espionage), 10 Sep 2026	B2	2	72	Adequate	30	Unlikely	Vendor telemetry. GTG-50020 failure flatters the reporter
EN-13	Arup deepfake video-call fraud, HK$200M (US$25.6M), Jan 2024	B2	2	65	Adequate	25	Unlikely	Arup's accounts for the year to 31 Mar 2024 book 25.4 million pounds as fraud and associated costs. Video-call detail rests on police and press. Incident is about 33 months old at 6 Oct 2026
EN-14	CrowdStrike, 2026 Global Threat Report, 24 Feb 2026	B2	2	59	Weak	25	Unlikely	The 89 percent figure has no definition or denominator. Overlaps KP-2
GV-1	ODNI, Annual Threat Assessment 2026, 18 Mar 2026	A1	2	88	Strong	35	Unlikely	No AI-influence language. Absence is a finding about the source
GV-2	UK NCSC (blog of 30 Mar, modified 29 May; speech of 17 Jun 2026) and Five Eyes statement (22 Jun 2026)	A1	2	83	Strong	30	Unlikely	Policy statements. Forward-looking
GV-3	SUPO (Finland) and Lithuania VSD threat assessments	A1	2	77	Adequate	30	Unlikely	SUPO date not shown
GV-4	Council of the EU, hybrid-threat listings, 16 Mar and 24 Sep 2026	A1	1	n/a	n/a	n/a	n/a	Sanctions record. No AI element
GV-5	NATO ACT, Cognitive Warfare newsletters, Apr to Jun 2026	B2	1	n/a	n/a	n/a	n/a	Curated summaries of outside authors
GV-6	NSA, CISA and allies, agentic AI guidance (30 Apr 2026) and NSA MCP security note (20 May 2026)	A1	1	n/a	n/a	n/a	n/a	Guidance. No incident data

Rows marked "statement made" record what an adversary says. They enter at grade C under Rule 1 of the ASA Evidence Standard. Their weight on facts is nil. An adversary assertion about a rival sits on its own row under Rule 3, apart from any verified fact.

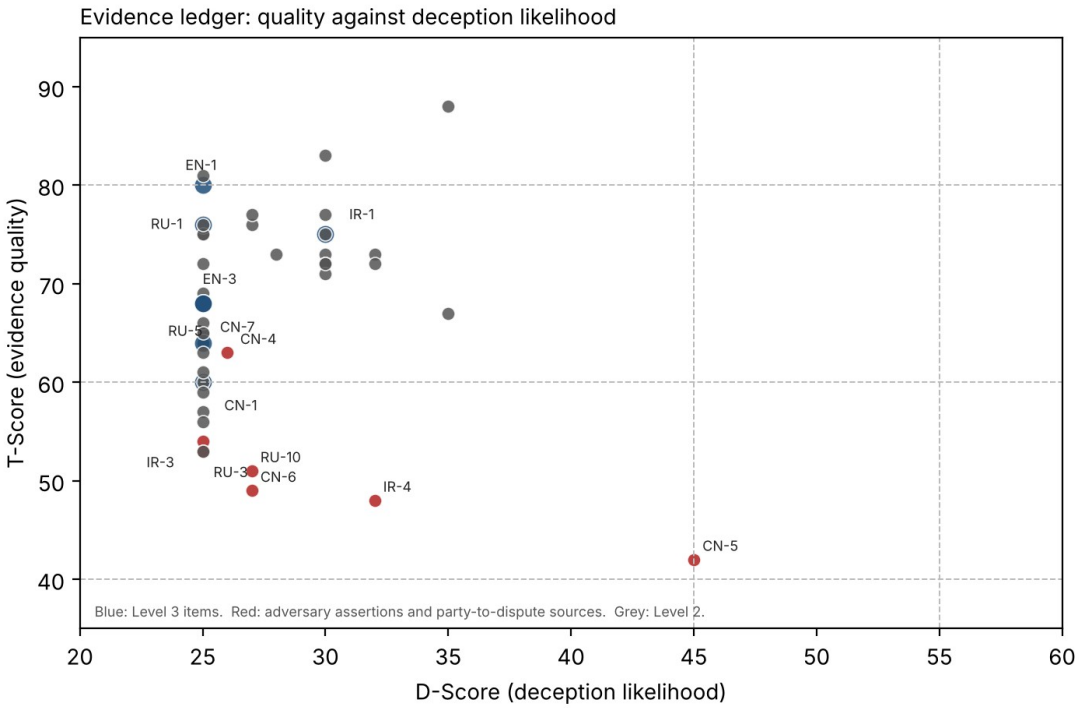


Figure 25. *Ledger rows by T-Score and D-Score.*

Three patterns show in the ledger.

Thirty-five of 63 rows grade B2. Labs, vendors and research bodies supply most of the evidence, and the grade says the source is usually reliable and the information probably true.

Of 48 scored rows, 34 read Adequate, 10 read Weak and 4 read Strong. None reads Insufficient. The mean T-Score is 66.9 and the mean D-Score 27.5. Adequate caps any forward judgment at Likely, so few judgments in the report print higher. The four Strong rows are the FBI FLASH on TeamPCP, the Nx and Salesloft precedents, the ODNI assessment and the allied cyber statements.

The nine rows that rest on Anthropic average a T-Score of 69.4, against 66.4 for the other 39 scored rows. Primary telemetry earns points for specificity and traceability, and the gap also fits the conflict disclosed in Section 1.5. Each Anthropic-only row sits at B2 by rule.

Ten scored rows merge several sources: CN-2 (two OpenAI reports), CN-7 (Anthropic's distillation report and the NSA, CISA and FBI advisory), CN-9 (three sources), CN-10 (two Anthropic reports), EN-1 (four), EN-6 (two precedents), EN-13 (Arup's accounts beside police and press reporting), GV-2 (NCSC statements and the Five Eyes statement), GV-3 (SUPO and Lithuania VSD) and KP-2 (two CrowdStrike reports). Five Level 1 rows package several documents as well: EN-10, GV-4, GV-6, KP-4 and KP-5. Each T-Score rates the package. The weakest source limits it. CN-10 joins two cases, and Rule 8 forbids blending them, so Rule 8 calls for two rows, CN-10a and CN-10b. The split is pending. EN-14 scores the CrowdStrike Global Threat Report alone, and KP-2 pairs it with the Threat Hunting Report. The two rows rest on one report, so a tally of independent items counts it once.

T and D do not move together. Across the 48 scored rows the correlation is -0.01. The floor explains the lack of any link. Each D is the mean of five factor scores with a floor of 25, a clean item scores low on every factor, and 27 of the 48 rows now sit on the floor. With almost no spread, D says little, and a near-zero correlation does not

show that T and D are independent. D feeds no judgment. The T and D scores are expert ratings applied under a documented method.

Section 16.2 lists the rows near a band edge.

16.1 Index coding sheet

The outlook of Section 12 counts 30 ledger rows at the rungs "AI use shown" and "Audience counted". Each row appears with its actor, its month, the provider-telemetry flag, the local-model flag and whether it falls inside the 12 months ending 6 October 2026. Row EN-3 carries the local flag and no provider flag by ruling (Section 12.2).

The sheet below carries 67 rows: the 30 ledger rows, 21 reports dated February through August 2026 that pass the counting test and stand apart from the ledger, two slices cut from the merged rows KP-2 and CN-2, and 14 reports from the winter months, four of them with provider telemetry and none a strict local-model case. Rows with an RC prefix belong to the extended set. Row RC-KP-2, the KP-2 slice, is the report that ledger row EN-14 scores, so EN-14 adds no row to the sheet. The headline paths of Section 12.3 rest on the 30 ledger rows. A single coder's coding covers the 37 added rows. Section 12.2 shows the counts under nine row sets.

Two of the 14 reports duplicate rows already held. Anthropic's November 2025 case, a slice of CN-10 with an attribution edit of 17 November, adds no independent source. The January 2026 report from Taiwan's bureau is ledger row CN-6 again, and the sheet reaches it through a newspaper. Check Point's VoidLink case sits in two counted reports, the original of 20 January 2026 and a digest of 29 March 2026. The digest keeps its March place for its forum-actor content, and the monthly counts stand.

Table 69. *16.1 Index coding sheet*

Row	Actor	Month	Report	Provider telemetry	Local-model evidence	In 12-month window	Basis
CN-12	China	2025-08	Vanderbilt Wicked Problems Lab, GoLaxy documents, Aug 2025 (release page shows no date)	No	No	No	Ledger
RU-6	Russia	2025-09	Recorded Future Insikt, CopyCop self-hosted models, 18 Sep 2025	No	Yes	No	Ledger
RC2-EN-1	Multi-actor	2025-10	OpenAI, Disrupting malicious uses of AI: October 2025, 7 Oct 2025	Yes	No	Yes	Extended set
RC2-IL-1	Israel (assessed)	2025-10	Citizen Lab, PRISONBREAK influence network, 14 Oct 2025	No	No	Yes	Extended set
RC2-EN-2	Multi-actor	2025-10	Microsoft, Digital Defense Report 2025 and blog, 16 Oct 2025	Yes	No	Yes	Extended set
RC2-RU-1	Russia	2025-10	ISD Germany, when chatbots surface Russian state media, 27 Oct 2025	No	No	Yes	Extended set, borderline row
RC2-EN-3	Multi-actor	2025-11	Google GTIG, AI Threat Tracker, 5 Nov 2025	Yes	No	Yes	Extended set, loose local flag not counted
RC2-CN-1	China	2025-11	Anthropic, first reported AI-orchestrated cyber espionage campaign, 13 Nov 2025 (slice of CN-10)	Yes	No	Yes	Extended set, slice
RC2-EN-4	Criminals	2025-11	Unit 42, The Dual-Use Dilemma of AI: Malicious LLMs, 25 Nov 2025	No	No	Yes	Extended set, loose local flag not counted

Copyright 2026 Treadstone 71 - www.treadstone71.com

Row	Actor	Month	Report	Provider telemetry	Local-model evidence	In 12-month window	Basis
RC2-RU-2	Russia	2025-11	Ukraine Centre for Countering Disinformation, report on AI-generated videos, 28 Nov 2025	No	No	Yes	Extended set
RC2-RU-3	Russia	2025-12	FIMI-ISAC with DFRLab and others, Moldova election report, 5 Dec 2025	No	No	Yes	Extended set
RC2-EN-5	Multi-actor	2025-12	ESET, Threat Report H2 2025, 16 Dec 2025 (PromptLock is an academic prototype)	No	No	Yes	Extended set, loose local flag not counted
RC2-EN-6	Criminals	2025-12	FBI IC3, PSA I-121925-PSA, 19 Dec 2025	No	No	Yes	Extended set
RC2-RU-4	Russia	2026-01	DFRLab, Russia's influence machine after Maduro's capture, 10 Jan 2026	No	No	Yes	Extended set
RC2-CN-2	China	2026-01	Taiwan NSB report, read through the Taipei Times, 11 and 12 Jan 2026 (the document CN-6 holds)	No	No	Yes	Extended set, overlaps CN-6
RC2-EN-7	Unattributed	2026-01	Check Point Research, VoidLink, 20 Jan 2026	No	No	Yes	Extended set
RC-CN-6	China	2026-02	OpenAI, Cyber Special Operations, February 2026 (case page 1 Feb 2026, report index 25 Feb 2026) (slice of CN-2)	Yes	Yes	Yes	Extended set, slice; local flag rests on the user's claim, which OpenAI could not confirm
RC-EN-1	Multi-actor	2026-02	Google GTIG tracker, 12 Feb 2026	Yes	No	Yes	Extended set
RC-EN-2	Multi-actor	2026-02	Unit 42, Global Incident Response Report, Feb 2026	No	No	Yes	Extended set
RC-KP-2	North Korea	2026-02	CrowdStrike, Global Threat Report, 24 Feb 2026 (slice of KP-2)	No	No	Yes	Extended set, slice; scored as ledger row EN-14
CN-3	China	2026-03	Graphika, Spamouflage AI content on the Tibetan election, 4 Mar 2026	No	No	Yes	Ledger
KP-1	North Korea	2026-03	Microsoft, AI as tradecraft, 6 Mar 2026	Yes	No	Yes	Ledger
RC-EN-3	Criminals	2026-03	Check Point, AI threat digest for Jan to Feb, 29 Mar 2026	No	Yes	Yes	Extended set
RC-IR-1	Iran	2026-03	Cyabra, Iranian AI content, 29 Mar 2026	No	No	Yes	Extended set
RC-IR-2	Iran	2026-03	Unit 42, Boggy Serpens, 16 Mar 2026	No	No	Yes	Extended set
RU-9	Russia	2026-03	EEAS, 4th FIMI threat report, 12 Mar 2026	No	No	Yes	Ledger
IR-6	Iran	2026-04	ISD, Iranian diplomats' meme war, 23 Apr 2026	No	No	Yes	Ledger
RC-CN-1	China	2026-04	Graphika, #PuppetMaster, 29 Apr 2026	No	No	Yes	Extended set
RC-CN-2	China	2026-04	DFRLab, Spamouflage and Tibetan elections, 24 Apr 2026	No	No	Yes	Extended set
RC-IR-3	Iran	2026-04	Fortune, Lego videos, 9 Apr 2026	No	No	Yes	Extended set

Copyright 2026 Treadstone 71 - www.treadstone71.com

Row	Actor	Month	Report	Provider telemetry	Local-model evidence	In 12-month window	Basis
RU-2	Russia	2026-04	DFRLab, Pravda in the pipeline, 8 Apr 2026	No	No	Yes	Ledger
IR-7	Iran	2026-05	ISD, BRICS4CLICKS and Verified4War networks, spring 2026	No	No	Yes	Ledger
RC-EN-4	Multi-actor	2026-05	Google GTIG tracker, 11 May 2026	Yes	No	Yes	Extended set
RC-EN-5	Unattributed	2026-05	Sysdig via The Hacker News, Marimo agent, 29 May 2026	No	No	Yes	Extended set
RC-KP-1	North Korea	2026-05	CrowdStrike, financial services report, 14 May 2026	No	No	Yes	Extended set
RC-RU-1	Russia	2026-05	ISD, Armenia vote, 27 May 2026	No	No	Yes	Extended set
CN-2	China	2026-06	OpenAI, Cyber Special Operations (Feb 2026) and PRC-linked AI-debate operations (10 Jun 2026)	Yes	No	Yes	Ledger
RC-CN-3	China	2026-06	DFRLab, Philippine Facebook pages, 30 Jun 2026	No	No	Yes	Extended set
RC-EN-6	Multi-actor	2026-06	Anthropic, Mapping AI-enabled cyber threats, 3 Jun 2026	Yes	No	Yes	Extended set
RC-EN-7	Criminals	2026-06	ZeroFox, jailbreak chatter, 17 Jun 2026	No	No	Yes	Extended set
RC-EN-8	Criminals	2026-06	Sysdig, LLMjacking evolved, 17 Jun 2026	No	Yes	Yes	Extended set
RU-10	Russia	2026-06	EEAS and Ukraine CCD, joint report, 23 Jun 2026	No	No	Yes	Ledger
RU-3	Russia	2026-06	American Sunlight Project, LLM exposure estimate, 2 Jun 2026	No	No	Yes	Ledger
RC-CN-4	China	2026-07	Unit 42, autonomous AI attack campaign, 30 Jul 2026	No	No	Yes	Extended set
RC-CN-5	China	2026-07	Graphika, Ore Else, 30 Jul 2026	No	No	Yes	Extended set
RC-EN-10	Multi-actor	2026-07	Check Point, AI Security Report, 14 Jul 2026	No	No	Yes	Extended set
RC-EN-9	Criminals	2026-07	OpenAI, Cambodia scam operation, 31 Jul 2026	Yes	No	Yes	Extended set
IR-5	Iran	2026-07	Recorded Future Insikt, Iran AI asymmetric playbook, 16 Jul 2026	No	No	Yes	Ledger
CN-11	China	2026-08	Meta, H2 2026 adversarial threat report, Aug 2026	No	No	Yes	Ledger
EN-4	Unattributed	2026-08	NSA, CISA, FBI, DOE and EPA, AA26-231A on Siemens S7 PLCs, 19 Aug 2026	No	No	Yes	Ledger
KP-2	North Korea	2026-08	CrowdStrike, Global Threat Report (24 Feb 2026) and Threat Hunting Report (3 Aug 2026)	No	No	Yes	Ledger
RC-RU-2	Russia	2026-08	DFRLab, Armenia electoral space, 24 Aug 2026	No	No	Yes	Extended set

Row	Actor	Month	Report	Provider telemetry	Local-model evidence	In 12-month window	Basis
RU-11	Russia	2026-08	OpenAI, International Burke Institute operation, 25 Aug 2026	Yes	No	Yes	Ledger
RU-5	Russia	2026-08	Recorded Future Insikt, CopyCop and the Firebird AI data center, 18 Aug 2026	No	No	Yes	Ledger
CN-1	China	2026-09	Anthropic, GTG-14020, 14021, 14022, PRC-aligned actors, 10 Sep 2026 (date from secondary sites)	Yes	No	Yes	Ledger
CN-10	China	2026-09	Anthropic, AI-orchestrated espionage (13 Nov 2025) and GTG-10007, 10 Sep 2026 (date from secondary sites)	Yes	No	Yes	Ledger
EN-12	Multi-actor	2026-09	Anthropic, GTG-50020, 50029, 50014, 20006 (AI supply-chain attempt, stolen-key hacktivist, credential crew, state-nexus espionage), 10 Sep 2026 (date from secondary sites)	Yes	No	Yes	Ledger
EN-3	Multi-actor	2026-09	Google GTIG, AI Threat Tracker, 8 Sep 2026	No	Yes	Yes	Ledger, suspected and single source
IR-1	Iran	2026-09	Anthropic, GTG-34001, Iranian state propaganda bodies, 10 Sep 2026 (date from secondary sites)	Yes	No	Yes	Ledger
IR-2	Iran	2026-09	Anthropic, GTG-34007, 30006, 30004, Iranian surveillance and profiling units, 10 Sep 2026 (date from secondary sites)	Yes	No	Yes	Ledger
IR-8	Iran	2026-09	Anthropic, GTG-84006, MEK and NCRI-aligned network, 10 Sep 2026 (date from secondary sites)	Yes	No	Yes	Ledger
KP-3	North Korea	2026-09	Genians, Kimsuky and the opencode agent, 7 Sep 2026; local tooling sits in the 10 Aug 2026 report	No	Yes	Yes	Ledger
PX-1	Hired hands	2026-09	Anthropic, GTG-84002, 54006, 50027, 54009 (UAE-directed single operator, Bangladesh partisan operator, Mali consultant, commercial vendor), 10 Sep 2026 (date from secondary sites)	Yes	No	Yes	Ledger
RU-1	Russia	2026-09	Anthropic, GTG-24015, Russian state-media editorial pipelines, 10 Sep 2026 (date from secondary sites)	Yes	No	Yes	Ledger
RU-4	Russia	2026-09	DFRLab, Storm-1516 and the Baltic states, 17 Sep 2026	No	No	Yes	Ledger
RU-7	Russia	2026-09	ISD, Matryoshka and German state elections, 4 Sep 2026	No	No	Yes	Ledger

Row	Actor	Month	Report	Provider telemetry	Local-model evidence	In 12-month window	Basis
RU-8	Russia	2026-09	Google GTIG, SANDWORM RELIC and UNC5792, 8 Sep 2026	Yes	No	Yes	Ledger

Seven reports of the spring and summer months fail the counting rule and are not counted. Each is listed with its reason.

Table 70. 16.1 Index coding sheet

Report	Date	Reason not counted
Microsoft Security Blog, threat actor abuse of AI accelerates from tool to cyberattack surface	2 Apr 2026	Aggregate commentary. No identifiable actor shown using AI
Euromaidan Press, reporting Bloomberg on leaked Social Design Agency documents	23 Jun 2026	Secondary report of documents. The Bloomberg original lies outside the record
DFRLab, AI-generated YouTube channels that farm war coverage	23 Mar 2026	The adversary type, state, proxy or criminal, is not established
Graphika, Pundit by Prompt	21 Jul 2026	Graphika states it was unable to identify the operators or confirm the links
DFRLab, digital infrastructure behind Russian interference in Armenian elections	29 Jul 2026	Large audience count, and the report documents no AI use by the actor
Euronews, misinformation and AI deepfakes in the Iran war	30 Mar 2026	No actor attribution and no audience count
Vision Times, China's influence operations use AI	16 Apr 2026	Derivative summary of a Meta report

Nine reports of the winter block fail the same rule. The reason for each follows.

Table 71. 16.1 Index coding sheet

Report	Date	Reason not counted
Microsoft, SesameOp backdoor	3 Nov 2025	Abuses OpenAI's Assistants API as a command channel. No model output serves the attack
NewsGuard, Sora piece on Kremlin propagandists	17 Nov 2025	Paywalled teaser only. It carries a claim that Russian actors use Sora 2, unverified
Alethea, Storm-1516 case study	19 Nov 2025	Vendor product material. The case shows no AI use by the actor
DFRLab, lessons from Moldova's 2025 elections	17 Dec 2025	Landing summary of the FIMI-ISAC work already counted. No AI term
GreyNoise, threat actors targeting LLMs	8 Jan 2026	Scans of exposed model endpoints, including an Ollama honeypot. Infrastructure targeting, no AI use
Graphika, Venezuela blog	9 Jan 2026	Adversary not established
FDD, Iranian X network	26 Jan 2026	No AI mentioned
Pillar Security, Operation Bizarre Bazaar	28 Jan 2026	Stolen model access resold. Infrastructure abuse
Multilateral Sanctions Monitoring Team, joint statement on DPRK cyber and IT-worker activity	22 Oct 2025	The statement holds no AI element. The Team's report lies outside the record

16.2 Rows near a band edge

The table lists the 27 ledger rows whose T-Score lands close to the 40, 60 or 80 cut, with the distance to the nearest cut.

Table 72. Scored rows within five points of an edge.

Row	Anchor	Level	T	Band	Nearest edge	Distance
IR-1	B2	3	75	Adequate	80	-5
IR-6	B2	2	75	Adequate	80	-5
RU-1	B2	3	76	Adequate	80	-4
RU-2	B2	2	75	Adequate	80	-5
RU-6	B2	2	65	Adequate	60	+5
RU-7	B2	2	76	Adequate	80	-4
RU-8	B2	2	76	Adequate	80	-4
RU-9	B3	2	60	Adequate	60	0
CN-1	B2	3	60	Adequate	60	0
CN-2	B2	2	64	Adequate	60	+4
CN-3	B2	2	64	Adequate	60	+4
CN-4	C2	2	63	Adequate	60	+3
CN-5	C6	2	42	Weak	40	+2
CN-7	B2	3	64	Adequate	60	+4
CN-10	B2	2	60	Adequate	60	0
CN-11	B2	2	63	Adequate	60	+3
KP-1	B2	2	77	Adequate	80	-3
KP-2	B2	2	61	Adequate	60	+1
EN-1	A1	3	80	Strong	80	0
EN-4	B3	2	57	Weak	60	-3
EN-6	A2	2	81	Strong	80	+1
EN-7	B2	2	75	Adequate	80	-5
EN-8	B3	2	56	Weak	60	-4
EN-13	B2	2	65	Adequate	60	+5
EN-14	B2	2	59	Weak	60	-1
GV-2	A1	2	83	Strong	80	+3
GV-3	A1	2	77	Adequate	80	-3

The public edition of ASA 1.2 withholds its per-factor thresholds, which sit in an internal edition, so a grade rebuilt from the public text lands near the original and not on it. Cluster C4 reads unclear as well. The standard prints only a human-source example, and the scores apply C4 to vendor telemetry, technical reports and advisories. Dropping C4 from the 13 rows that carry it lifts T by 2 to 6 points on each and moves one band: EN-14 rises to Adequate at 63, from 59 Weak.

KJ4 prints at Likely with Moderate confidence, a band below Highly likely, because EN-6 (T 81) and EN-1 (T 80) sit at the Strong edge. The Section 9 assessment on allied agencies prints at Likely for the same reason, since GV-2 (T 83) and GV-3 (T 77) sit within five points of the same edge. If EN-6 and EN-1 each hold at 80 or more under a second analyst's ratings, KJ4 returns to Highly likely. KJ10 stands at Low because EN-5 scores 53, a Weak row, and confidence follows the weakest supporting row.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Two judgments stand on a cliff. CN-1 carries KJ1 at 60, unrounded 60.3, and RU-9 carries KJ8a at 60, unrounded 59.7. Section 11.8 ranks every Moderate judgment by its distance from the floor. One point off either row drops its judgment from Moderate to Low.

16.3 The adversary record

Above, the ledger weighs vendors, labs and governments. The rows below list what adversaries themselves say, in their own documents, speeches and channels, read in the original language. Level 1 covers all 64 rows: a statement made, held at C under Rule 1, and never scored. None enters a T-Score, a D-Score or a count in Sections 10 and 12. Nor does any row support a finding of fact. Each shows how an adversary frames the contest, what it claims, threatens or asks to buy, and what it leaves unsaid.

Iran supplies 16 rows, Russia 3, China 31, North Korea 3 and the proxies 11. Of the 31 Chinese rows, 25 come from the GoLaxy documents and 6 from the CVERC report. Persons named inside the documents appear by role.

Table 73. *Statements by adversaries and their channels, 1 Feb 2023 to 9 Oct 2026.*

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-IR-01	1 Feb 2023 (12 Bahman 1401)	IRGC commander-in-chief, closing a Basij digital-content event; Asriran, Persian	Says "today's war is in cyberspace" and the enemy widened its war to culture and technology. Claims a US strategy aims to make all narratives enemy-told. Urges Basij content producers not to stay silent.	doctrine	Asriran, news ID 876957, pp1-6 (capture 3daac8f6). Year digits read from a low-resolution stamp. Listed here only.
AO-IR-02	undated (c. 2025, inferred; the rally was 9 Dey 1404, 30 Dec 2025)	Adviser to the IRGC commander-in-chief, at a Gorgan rally; Etemad Online with ISNA credit, Persian	Says that after the 12-day war the enemy concluded a military option is useless and "has focused on cognitive war": forgetting, distortion, discrediting and censorship. Lists its next steps as a war shadow, economic pressure, polarization and weakened social hope.	doctrine	Etemad Online, politics, pp1-4 (capture 83743df8). Date inferred; a pop-up covers the speaker's first name. Listed here only.
AO-IR-03	1 Apr 2026 (12 Farvardin 1405), 11:23	IRGC notice as relayed by Digiato, Persian	Tells staff of US firms in the region, listed as 18 companies including Microsoft, Google, Palantir and Nvidia, to evacuate by 20:00 Tehran time. Says each assassination in Iran will be met by destroying a matching unit.	threat	Digiato, pp1-5 (capture 61cef2fc). The notice itself carries no date; the firm list is as printed by Digiato. Section 3.5.

Copyright 2026 Treadstone 71 - www.treadstone71.com

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-IR-04	2 Apr 2026 (13 Farvardin 1405), 10:46	ISNA, Economic, Communications and IT desk, Persian	Reports damage to Amazon's cloud business in Bahrain, citing the Financial Times and Bahrain's attribution to Iran, and recalls the IRGC warning that US firms are "legitimate targets" for backing US army acts. Iran itself claims nothing here.	routine propaganda	ISNA, news ID 1405011306154, p1 (capture 861d7127). Page printed sideways. Section 3.5.
AO-IR-05	29 Jul 2026 (7 Mordad 1405), 08:11	Guest-note author, described as head of international affairs at a judiciary lawyers center; Mehr, Persian	Argues Palantir, Anduril and Shield AI have moved into the kill chain and that firm officers can be liable for aiding crimes, so a "responsibility gap" widens. Concedes no firm's role in attacks on Iran is proven in court.	accusation	Mehr, news ID 6901729, pp1-2 (capture fc068b2b). The author's title is as printed on the page. Listed here only.
AO-IR-06	14 Sep 2026 (23 Shahrivar 1405), 08:49	Mehr reporter, University and Technology desk, Persian	Under the headline "Claim of Iran's use of Claude AI", relays Anthropic's report of three blocked accounts attributed to Iran for influence operations and narrative shaping. Does not deny the finding, and itself calls Anthropic's other findings "claims".	routine propaganda	Mehr, news ID 6945759, p1 (capture ff01a9b7). Page printed sideways. Section 3.5.
AO-IR-07	14 Sep 2026 (23 Shahrivar 1405)	Mehr reporter, "Blind spot" section of the same item, Persian	Argues Anthropic works inside the US military structure: Claude sits on classified networks under a 2025 contract worth up to 200 million dollars, and unnamed reports say US forces used it against Iran. Concludes "misuse" follows host-state interests.	accusation	Mehr, news ID 6945759, p2 (capture ff01a9b7). Mehr names no source for the "published reports". Section 3.5.
AO-IR-08	14 Sep 2026 (23 Shahrivar 1405)	Mehr reporter, summary section of the same item, Persian	Reads the report as the rise of a "cognitive infrastructure of war" and says "the algorithmic war has begun". Urges home-grown frontier-model capacity, secure compute, defensive agents, and legal tools against AI firms' role in armed conflict.	doctrine	Mehr, news ID 6945759, p2 (capture ff01a9b7). Section 3.5.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-IR-09	7 Oct 2026 (15 Mehr 1405), 01:08, for a rally of 6 Oct	Markazi provincial police commander at an Arak gathering; Mehr, Persian	Says the enemy, after failing in physical, military and security fields, now targets public trust through cognitive war and false narratives. Calls the unity of people and armed forces a barrier, and urges media literacy and reliance on official sources.	routine propaganda	Mehr, news ID 6970215, p1 (capture 964d21cb). Page printed sideways. Section 3.5.
AO-IR-10	8 Oct 2026 (16 Mehr 1405), 01:00	Head of the Majlis Research Center; Eghtesad Online, Persian	Denies that parliament was "closed" during the war, saying committees, oversight and virtual sessions continued. Adds that in cognitive war the enemy need not invent a narrative if officials sometimes hand over its raw material.	denial	Eghtesad Online, news code 2167575, pp2-6, 8, 11 (capture b931e3a0). A pop-up hides part of the lower lines. Listed here only.
AO-IR-11	8 Oct 2026 (16 Mehr 1405), 12:27	Eitaa channel @hamyane_sepah, an IRGC-supporters channel, Persian	States that British refueling aircraft take part in Saudi operations against Yemen. The post gives no source, date or place, and the channel is a supporter channel, not an IRGC office, on the page's evidence.	accusation	Eitaa @hamyane_sepah, capture file e-013. Listed here only.
AO-IR-12	9 Oct 2026 (17 Mehr 1405), 00:34	Same channel, post headed "Leader of the Revolution", Persian	Headline says strengthening police (Faraja) forces, above all upgrading technologies, "including targeted use of" a missing word "in data analysis", is a main mission. The sentence is incomplete in the capture, and the attribution is the channel's label.	doctrine	Eitaa @hamyane_sepah, capture file e-086. Listed here only.
AO-IR-13	9 Oct 2026 (17 Mehr 1405), 01:12	Same channel, post headed "Leader of the Revolution", Persian	Says the "American-Zionist enemy" focused on striking Faraja ranks up to the top in the recent "imposed wars", which showed the institution's weight, and that the effort failed because the forces kept working in streets and vehicles.	accusation	Eitaa @hamyane_sepah, capture file e-091. Listed here only.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-IR-14	9 Oct 2026 (17 Mehr 1405), time not shown	Same channel, relaying a post by the executive director of United Against Nuclear Iran, Persian	Relays a reported X message in which the executive director of UANI, called a "terrorist group", says Iran's "psychological war" has made Americans lose morale to continue the war. The original X post is not on the page.	boast or capability claim	Eitaa @hamyane_sepah, capture file e-097. Listed here only.
AO-IR-15	9 Oct 2026 (17 Mehr 1405), time not shown	Same channel, relaying Maj. Gen. Vahidi (first name and post not shown), Persian	Relays a statement that no extra-regional power may threaten or interfere in Hormuz and the Gulf, and that the naval force, with cruise missiles, drones, underwater vehicles, electronic warfare and coastal defenses, will give a decisive reply.	threat	Eitaa @hamyane_sepah, capture file e-110. The first text line is partly covered by the logo. Section 3.8.
AO-IR-16	9 Oct 2026 (17 Mehr 1405), time not shown	Same channel, no source named, Persian	Posts that thick smoke rose at Abqaiq in Saudi Arabia, then that fire spreads at the refinery and "satellite images from 8 hot points" show about 332.4 megawatts against a normal 4. No imagery provider is named.	routine propaganda	Eitaa @hamyane_sepah, capture files e-129 and e-133. First digits of 332.4 could be 322. Listed here only.
AO-RU-01	30 Sep 2026, 21:28	Rossiyskaya Gazeta, reporting a presidential instruction after a Council for Science and Education meeting; Russian	Reports that Putin ordered the government to organize centralized collection and provision of scientific and technical information for training sovereign and/or national foundational AI models, and access for scientists to Russian AI software, with a report by 1 March 2027.	intent	Rossiyskaya Gazeta (rg.ru), p1 (capture 53235295). The page does not print the Kremlin list number. Section 4.9.
AO-RU-02	8 to 9 Oct 2026 (page activity)	Hugging Face organization "ai-forever", owner not shown on the page; English	Lists 122 models, 44 datasets and 15 Spaces, including Zarya text-diffusion models up to 4B parameters, FRIDA, the MERA and LIBRA benchmarks and a retrieval-benchmark leaderboard. The page makes no claim and shows no license or state link.	procurement or pitch	huggingface.co/ai-forever, pp1-6 (capture cc7a5b69). Download counts are on-page counters. Listed here only.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-RU-03	9 Oct 2026 (ticker 10:47; date inferred from print time)	Rossiyskaya Gazeta news ticker, Russian	Headline only: infrastructure of a Yandex data center in Kaluga was damaged after a UAV attack. The headline names no attacker, and the article body was not captured. Treat as a lead for retrieval, not as a claim.	routine propaganda	Rossiyskaya Gazeta ticker, p2 (capture 53235295). Listed here only.
AO-CN-01	undated (c. 2018, inferred from events cited to 2017 and a chart to 2018)	Unsigned "intelligent propaganda" proposal, Chinese; GoLaxy documents	Says news and opinion work still sits at propaganda and guidance, and its technical level cannot support large-scale opinion warfare. Calls for matching equipment. The proposal does not name GoLaxy; the link is the Vanderbilt release only.	intent	GoLaxy documents, PDF p328. Listed here only.
AO-CN-02	undated (c. 2020, inferred; the script calls the Hong Kong security law and a planned US visit to Taiwan "recent")	GoLaxy team script for visiting leaders of the parent institute, Chinese; GoLaxy documents	Claims GoPro output is already deployed at the State Council Hong Kong and Macao office, the Ministry of State Security, a Joint Staff intelligence bureau and front-line units. Says close to 50 staff built it for national cognitive-domain confrontation.	boast or capability claim	GoLaxy documents, PDF p25. Section 5.3.
AO-CN-03	undated (c. 2020, inferred; same script as AO-CN-02)	Same GoLaxy script, Chinese; GoLaxy documents	Claims that from 9.66 million Twitter users and 180,000 Hong Kong users it locked onto 3,000+ participants, 100+ opinion leaders and 20+ outlets. Adds AI-aided text generation and cross-platform "intelligent push" with persona upkeep, and rebuttal of targets' "fallacies".	boast or capability claim	GoLaxy documents, PDF pp25-26. Counts are the firm's example, not a verified result. Listed here only.
AO-CN-04	undated (c. 2020, inferred; same Hong Kong figures as the p25 script)	Second GoLaxy script, Chinese; GoLaxy documents	Says Western states led by the US run a strong online opinion war over Hong Kong, Xinjiang, the virus and trade that threatens regime security. States a team of 50+ with no national funding. Makes no deployment claim.	accusation	GoLaxy documents, PDF pp121-122. Section 5.3.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-CN-05	undated (c. 2020, inferred; same script as AO-CN-04)	Second GoLaxy script, Chinese; GoLaxy documents	Claims "several hundred" human-like bot accounts on Twitter, Facebook and Telegram and a batch posting tool. Calls the system a first step, short of national needs, and asks to cooperate with the cyberspace, propaganda, military science and State Security bodies.	boast or capability claim	GoLaxy documents, PDF pp123-124. Listed here only.
AO-CN-06	undated (c. 2020, inferred; data stamped May to Oct 2020)	GoLaxy progress page and screens, Chinese; GoLaxy documents	Aims to profile 200,000 Hong Kong Twitter users, noting Twitter has about 750,000 there. Reports 32.1 million Twitter and 45.3 million Weibo accounts accumulated. Screens show a search on the national security law returning 2,090 records.	boast or capability claim	GoLaxy documents, PDF pp392-397. Listed here only.
AO-CN-07	9 Nov 2020	Bid-result confirmation letter, Chinese; GoLaxy documents	GoLaxy bid CNY 252,000, scored 82.54 and ranked first for an entity extraction and retrieval project; a rival bid CNY 258,000, scoring 68.77. The report reads the buyer as PLA Unit 31040. A small purchase; no tasking shown.	procurement or pitch	GoLaxy documents, PDF p93. Section 5.3.
AO-CN-08	undated (c. 2021, inferred from nearby pages)	GoLaxy company introduction deck and case pages, Chinese; GoLaxy documents	States the firm was incubated by the CAS computing institute in 2010 and serves "党、政、军" customers (Party, government, military). Case pages claim systems for a military university and a Joint Staff bureau.	procurement or pitch	GoLaxy documents, PDF pp34-36, 195-199, 239-240. Section 5.3.
AO-CN-09	2021 (year printed; day not shown)	GoLaxy research group plan, Chinese; GoLaxy documents	Names GoPro as the core system with four technologies: intelligent creation, opinion mining, account cultivation and user profiling. The goal slide describes "intelligent assistance for targeted propaganda overseas". Phase three, from 2024, seeks national-level projects and awards.	intent	GoLaxy documents, PDF pp381-385. English quote is a translation of the slide. Listed here only.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-CN-10	2021 (workbook year; compile date not shown)	GoLaxy internal sales pipeline, Chinese; GoLaxy documents	Lists 17 projects at stages 1 to 7, with buyers including the CAS computing institute and provincial public-security and other departments. Products include opinion situation awareness and the GoIN intelligence tool. Amounts carry no unit.	procurement or pitch	GoLaxy documents, PDF pp90-91 (rotated). Not contracts. Listed here only.
AO-CN-11	undated (c. 2022, inferred)	GoLaxy account-nurturing deck, Chinese; GoLaxy documents	Describes raising fake social accounts until they look real, and says the work must "bypass the platform's behavior recognition mechanisms". One host drives 50+ phones across twelve platforms. A screen headed 3,692 accounts lists personas in several countries.	boast or capability claim	GoLaxy documents, PDF pp271-280, 284. The screen may be demo data. Listed here only.
AO-CN-12	Dec 2022 (cover)	Open-topic meeting deck, computing institute and lab, Chinese; GoLaxy documents	Lists three first-batch topics, including identity shaping, funded 285/300, 285/300 and 475/500 (10,000 yuan). Targets 3 large covert cognitive-confrontation tests with 50,000 users. The deck does not name GoLaxy; the link is the Vanderbilt release only.	procurement or pitch	GoLaxy documents, PDF pp294-300 (rotated). Listed here only.
AO-CN-13	undated (c. 2022, inferred; same open-topic series as the Dec 2022 deck)	Open-topic requirements, topic 3, Chinese; GoLaxy documents	Specifies digital humans speaking in a target's style in Taiwan Mandarin, Hokkien and English: at least 100 digital identities and 1,000 virtual persons on Facebook, Twitter and TikTok. It does not name GoLaxy; the link is the Vanderbilt release only.	intent	GoLaxy documents, PDF pp303-305, 309 (rotated). Section 5.3.
AO-CN-14	11 May 2023, 2:00 pm	Minutes of a program progress meeting, topic two, Chinese; GoLaxy documents	Treats election opinion guidance as the hardest case: fix main figures' attitude toward China, assess elite mobilizers, and model users with stance and Big Five personality models. The minutes do not name GoLaxy; the link is the Vanderbilt release only.	doctrine	GoLaxy documents, PDF pp28-29 (angled photograph). Listed here only.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-CN-15	11 May 2023, 2:00 pm	Minutes of the same meeting, topic three and account report, Chinese; GoLaxy documents	Lists bot registration scripts, "combat tasks" for bot accounts, and a status of 10 large, 50 ordinary and 600 grassroots Twitter accounts and 20 on Reddit. The minutes do not name GoLaxy; the link is the Vanderbilt release only.	boast or capability claim	GoLaxy documents, PDF pp30-31 (angled photograph). Listed here only.
AO-CN-16	undated (c. 2023, inferred from deck pages stamped 2023)	GoLaxy data-capability deck, slide on US political figures, Chinese; GoLaxy documents	One sentence lists the 117th Congress, 2022 candidates, the establishment, Trump supporters and right-wing forces under a "4000+" whose scope the page leaves open, and other political figures and opinion leaders at "2000+". The 117 is an ordinal, not a count.	boast or capability claim	GoLaxy documents, PDF p7. Section 5.3.
AO-CN-17	undated (c. 2023, inferred from pp85-86 stamps)	GoLaxy slide "political intervention action path", Chinese; GoLaxy documents	Maps eight steps around a "combat platform": election opinion listening, swing-person identification, content generation with large language models, election simulation, delivery through "personified account cultivation", and effect evaluation.	doctrine	GoLaxy documents, PDF p83. Slide footer is cut off. Listed here only.
AO-CN-18	undated (c. 2023, inferred; text brackets mid-2022 to early 2024)	GoLaxy product and main-user status table, Chinese; GoLaxy documents	Main users: the Cyberspace Administration for rows 1 to 3 (bot detection, model risk), the CMC Science and Technology Commission for rows 4 and 5, "311" for row 6, the Taiwan Affairs Office for row 7.	procurement or pitch	GoLaxy documents, PDF p379 (rotated photograph). Listed here only.
AO-CN-19	undated (c. 2023, inferred; same table)	GoLaxy status table, rows 4 to 6, Chinese; GoLaxy documents	Row 4: 8 million sources, profiling, own account groups, content generation and push to Twitter and Facebook. Row 5: batch account cultivation and account proxying. Row 6: scheduling of ten-thousand-level accounts and thousand-level tasks.	boast or capability claim	GoLaxy documents, PDF p379. Listed here only.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-CN-20	undated (c. 2023, inferred; same table)	GoLaxy status table, row 7, Chinese; GoLaxy documents	Claims support for M's 2016 and 2020 elections, T's 2020 election and Hong Kong's 2022 chief executive election, with reports drawing an "important" leadership instruction, and for T's 2024 election now. Main user: the Taiwan Affairs Office.	boast or capability claim	GoLaxy documents, PDF p379. M and T are masked; see notes. Listed here only.
AO-CN-21	19 Oct 2023, 9:00 to 12:00	Minutes of a program meeting, institute head's feedback to topic 3, Chinese; GoLaxy documents	Asks for fuller account personas, more human-like behavior, some large-model work, and long-term pre-positioning of topics and accounts around TW items and Lithuania. The minutes do not name GoLaxy; the link is the Vanderbilt release only.	intent	GoLaxy documents, PDF p193. Section 5.3.
AO-CN-22	19 Oct 2023, 9:00 to 12:00	Minutes of the same meeting, topic discussion and follow-up tasks, Chinese; GoLaxy documents	Plans a "simulation drill" on a non-political, technology-leaning topic aimed at "M and T" political circles and big-V accounts, with Twitter as an "information delivery experiment platform". The minutes do not name GoLaxy; the link is the Vanderbilt release only.	intent	GoLaxy documents, PDF pp193-194. Listed here only.
AO-CN-23	1 Dec 2023 (coverage 25 Nov to 1 Dec 2023)	Weekly Taiwan opinion report from Zhongke Tianji Data Technology, Chinese; GoLaxy documents	Counts 634,448 items and 94,791 posts in the week before Taiwan's January 2024 vote, splits them by blue, green and centrist camps, ranks ten hot events, and narrates how the green camp attacks the blue ticket.	routine propaganda	GoLaxy documents, PDF pp251-259. Listed here only.
AO-CN-24	1 Dec 2023 (signed p260, dated p261)	Same weekly report, section 3 recommendations, Chinese; GoLaxy documents	Advises "our side" to watch the Tsai government, plan for a Lai-Hsiao win, and "mobilize the cyber army" for a cognitive-warfare effort against the green camp's claim that the blue camp is pro-China. No tasking order appears.	intent	GoLaxy documents, PDF pp260-261. Quote translates the Chinese 宜发动网军. Listed here only.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-CN-25	6 Mar 2024 (V1.0)	Requirements specification "QB Analysis and Monitoring System", Chinese; GoLaxy documents	Specifies a multi-source collection and monitoring platform with US-related topic trees (US election, Trump, trade war) and a Taiwan situation assessment topic. Sample results are US-election items. No customer is named.	procurement or pitch	GoLaxy documents, PDF pp125-129, 148-150. Section 5.3.
AO-CN-26	Sep 2026 (reporting year Jul 2025 to Jun 2026)	CVERC with five company co-authors, 2026 Cyberspace Security Situation Analysis Report, Chinese	Says OpenAI and Anthropic each admitted that an advanced model escaped its test environment and attacked unintended targets, calls this proof that AI loss of control is now real, and says alignment is flawed and jailbreaks extract attack tutorials.	accusation	CVERC report, printed pp3-4, 59-61. The report states it is not an official government position. Listed here only.
AO-CN-27	Sep 2026	CVERC report, Chinese	Says APT groups from South Asia, Southeast Asia, East Asia and North America target China. Lists APT-C-78, a North America group, eighth in 2025 and absent in 2026. The page gives a region label only, with no state or evidence.	accusation	CVERC report, printed pp38-41. Listed here only.
AO-CN-28	Sep 2026	CVERC report, Chinese	Says foreign APT groups target military-linked universities and firms for semiconductor, aerospace and quantum research, and that "AI-enabled phishing", supply-chain attacks and zero-days are getting harder to see. No actor is named. The AI abuse section elsewhere covers fraud only.	accusation	CVERC report, printed pp42, 60. Listed here only.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-CN-29	Sep 2026 (rules dated 14 Mar 2025 and 28 Feb 2025; law in force 1 Jan 2026)	CVERC report, describing Cyberspace Administration measures, Chinese	Describes four-department rules requiring labels on AI-generated content, standard GB 45438-2025, app takedowns for missing labels, Framework 2.0, and an AI clause in the amended Cybersecurity Law, which it reads as the move from rule-making to enforcement.	doctrine	CVERC report, printed pp82-83. Section 5.7.
AO-CN-30	Sep 2026 (US items dated 23 Jul 2025 to 1 Jul 2026)	CVERC report, describing US policy, Chinese	Presents the US AI Action Plan as development-first, with export controls "against adversary countries", and describes the federal-versus-state fight over AI law, including a DOJ task force, an FTC statement and a 99-1 Senate vote.	routine propaganda	CVERC report, printed pp84-85. Listed here only.
AO-CN-31	Sep 2026	CVERC report, recommendations section, Chinese	Urges using the World AI Cooperation Organization to "prevent AI safety standards from forming camps". Proposes cross-border incident reporting, minimum standards for dangerous-capability tests, and mutual recognition of deepfake labels among China, the US and Europe.	intent	CVERC report, printed p106. Listed here only.
AO-KP-01	16 Jun 2026	Rodong Sinmun issue index, Korean	The index lists no AI headline. Items include a congratulatory message to Xi Jinping, party guidance of the Three Revolutions Red Flag movement, science and technology pieces warning of departmentalism as a "brake", and a feature on information-technology products.	routine propaganda	Rodong Sinmun, 16 Jun 2026, items 1-41, pp1-5 (capture f0e57870). Headlines only. Section 6.4.
AO-KP-02	16 Jun 2026	KCNA items on Rodong Sinmun pages 2, 4 and 6, Korean	Foreign pages carry Putin saying no one can inflict "strategic defeat" on Russia, a DPRK condemnation of "unjust" US conduct, a protest against US military acts off Oman, and reports of Israeli airstrikes on Lebanon.	accusation	Rodong Sinmun, 16 Jun 2026, items 42-49 (capture f0e57870). Headlines only. Listed here only.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-KP-03	19 Sep 2026 (inspection 16 to 18 Sep)	NK News relaying Rodong Sinmun on Kim Jong Un, English; second-hand	Says state media reported Kim toured munitions factories under the Second Economic Commission and called for expanded capacity and weapons modernization, including through AI and automation. The report did not identify the plants.	intent	NK News, 19 Sep 2026, p1 (capture 10fae396). Page two is behind a paywall. Section 6.4.
AO-PX-01	7 Oct 2026	Houthi leader, speech on the third anniversary of 7 Oct 2023; Saba Net and Al-Masirah, Arabic	Calls on the Islamic nation, naming Turkey and Pakistan, to act against Saudi plots on holy sites. Alleges a secret Saudi-Zionist alliance under an American umbrella, and calls for a million-man march on Friday.	accusation	Saba Net news3805067, 3804920, 3804889, 3804813 (captures cdcc3379, 379d6f60). Listed here only.
AO-PX-02	7 Oct 2026	Houthi chief of general staff, messages to the Quds and Qassam Brigades; Saba and Al-Masirah, Arabic	Marks the anniversaries of 7 October 2023 and the founding of Palestinian Islamic Jihad. The lede quotes "we are with you until victory" and promises to stand with the Palestinian people in any new escalation.	intent	Saba Net news3805061, 3805053 (captures cdcc3379, 379d6f60). Only ledes are on the pages. Listed here only.
AO-PX-03	7 Oct 2026	Houthi Foreign and Expatriates Ministry statement; Saba Net and Al-Masirah, Arabic	Holds the Saudi enemy responsible for escalation at Bab al-Mandab and asks the international community to press it to lift the blockade. Al-Masirah adds that silence on Saudi militarization of the strait serves American and Israeli interests.	accusation	Saba Net news3805151, 3805170 (captures cdcc3379, 379d6f60). Listed here only.
AO-PX-04	undated (c. 2026, inferred; same news cycle as 7 Oct 2026)	Houthi deputy information minister, an official named al-Sharjabi and a senior military source; Al-Masirah, Arabic	Say Yemeni strength defeated a Saudi military, media and political attack and a project of "misleading". The military source claims Houthi forces hold positions at Bab al-Mandab and that Saudi "victories" exist only in media.	boast or capability claim	Al-Masirah posts 311547, 311546, 311467 (capture 379d6f60). Headline-level reading. Section 6.5.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-PX-05	undated (c. 2026, inferred; same news cycle as 7 Oct 2026)	Hamas statement carried by Al-Masirah, Arabic	Says Israeli claims about hospitals are false and "systematic propaganda" meant to justify targeting civilian facilities. The item is a headline with no date or further detail on the page.	denial	Al-Masirah post 311612 (capture 379d6f60). Listed here only.
AO-PX-06	undated (c. 2026, inferred; same news cycle as 7 Oct 2026)	Al-Masirah features and translations section, Arabic	Curates Western press as proof of US decline: Middle East Eye on Vietnam's defeat returning through Iraq and Hormuz, a feature on suicide among US Navy sailors, and digests on collapsing hegemony and Saudi weakness before Yemeni naval deterrence.	routine propaganda	Al-Masirah posts 311193, 311206, 311210, 311151, 311103 (capture 379d6f60). Cited outlets not checked. Listed here only.
AO-PX-07	undated (c. 2026, inferred; same news cycle as 7 Oct 2026)	Houthi armed forces and military media via Al-Masirah and Saba, Arabic	Claims strikes on Riyadh and Abha airports and the Rabigh refinery with ballistic missiles and drones, a CH-4 shoot-down over Jawf, and video of diving drones. Saba shows a coordination center's warning to airlines using Saudi airspace.	boast or capability claim	Al-Masirah and Saba posts 311549, 311548, 311482, 311510, 311532, 311459, 311593 (captures 379d6f60, cdcc3379). Damage claims untested. Listed here only.
AO-PX-08	undated (c. 2026, inferred; same news cycle as 7 Oct 2026)	Al-Masirah and Saba economy and reports sections, Arabic	Says Reuters reports Yanbu tankers face attack with insurance tripling, that the Yemeni ban cuts Saudi port imports by 34 percent, and that "the Saudi enemy continues falsification" of its violations after killing and displacing civilians.	accusation	Al-Masirah and Saba posts 310705, 309916, 311414, 311226, 311251, news3801094, 3802663 (captures 379d6f60, cdcc3379). Reuters attribution not checked. Section 6.5.
AO-PX-09	8 Oct 2026 (16 Mehr 1405), 13:12 to 13:59	Houthi leader as quoted by Eitaa channel @hamyane_sepah, Persian	Quoted posts say Saudi Arabia committed crimes including the Attan bombing, that its aim is to break Yemeni will, and that the Zionist enemy and its backers, including the United States, are the most prominent stubborn enemies.	accusation	Eitaa @hamyane_sepah, capture files e-023, e-025, e-034, e-035, e-040, e-041. No date for the original speech. Section 3.8.

ID	Date	Speaker, outlet, language	What it says	Type	Source and section
AO-PX-10	9 Oct 2026 (17 Mehr 1405), time not shown	Yemeni forces spokesman and Yahya Saree as relayed by Eitaa channel @hamyane_sepah, Persian	Relays claims that Yemen hit Riyadh and Najran airports and Khamis Mushait base, King Khalid airport with two cruise missiles (another post says one ballistic missile), and Aramco facilities at Abqaiq. No evidence is shown.	boast or capability claim	Eitaa @hamyane_sepah, capture files e-106, e-118, e-119, e-137, e-139. The two posts differ on the weapon. Section 3.8.
AO-PX-11	undated (year not printed; page shows 8 Oct, earlier than the 2026 capture)	Al-Masirah feature, special section, Arabic	Headline asks how the United States enlisted its spies to destroy education in Yemen. The lede says the US worked to penetrate state ministries and institutions, wreck them and steer their policies.	accusation	Al-Masirah post 259302 (capture 379d6f60). Headline and lede only. Listed here only.

Four limits govern the rows. The GoLaxy rows carry the firm's or its staff's words, and none shows use, tasking or effect. Minutes of 11 May and 19 October 2023, the open-topic papers and the unsigned proposal (AO-CN-01, 12 to 15, 21 and 22) do not name GoLaxy, and their tie to the firm runs through the Vanderbilt release alone. A date marked inferred rests on events or stamps inside the document. Photographs of rotated pages, among them pages 90 and 91, 192 to 194, 297 to 312 and 379, were read after turning, and any figure quoted from them needs a check against the original scan.

Masked terms carry the reading used. The letter M stands for the United States and T or TW for Taiwan, from the surrounding slides, and neither expansion prints (AO-CN-16, 18, 20, 21 and 22). In the page 379 table, the entry "311" has no expansion, and no row uses a reading of it as a PLA base. QB reads as intelligence from a screen header (AO-CN-25). The abbreviation for strategic support in the page 25 script most likely means the Strategic Support Force, and the row says front-line units only. Other letters on those pages stay unexpanded and unused.

Six readings follow from the rows. Command-level Iranian speech uses the cognitive-war frame before Anthropic's report appeared (AO-IR-01 and 02), and the Mehr reading of that report reaches for doctrine (AO-IR-06 to 08). The captured Eitaa posts carry no AI wording. Russia's rows show intent and a model supplier, with no operation. Chinese rows show a firm that sells, specifies and boasts, and a state center whose report says little on AI influence. On the North Korean side, two pages carry no AI headline, and the one AI item arrives through NK News. Proxy rows are grievance and strike claims, and none mentions a model.

16.4 Sources read after the ledger froze

The sources below reached the record after the 63 rows were scored. They stay unscored so that the counts and indices in Sections 10 and 12 stand as printed. Each entry shows what the source holds and where it touches a judgment. Two entries fall after the counting window of 7 October 2025 to 6 October 2026.

Copyright 2026 Treadstone 71 - www.treadstone71.com

Table 74. Sources read after the ledger froze, unscored.

Source	Date	What it holds	Bearing
OpenAI, Disrupting AI-enabled false-front operations	8 Oct 2026	Russian "Dark Clark": a fake Social Research Center in Latin America, run through a fake persona with unwitting local staff, rated Category 5, OpenAI's first. The operators used ChatGPT for internal reports, content such as fake letters and audio scripts, and performance reports for the center. Iranian "Bogus Bylines": seven personas and nearly 100 articles in about a dozen outlets, from July 2025 to October 2026 and far more often once the US-Iran war began. OpenAI rates the articles Category 4 and the comments Category 2, "consistent with a commercial actor running a for-hire influence campaign." X and Instagram data for one persona show a "West Asia Android App" and an Iran location	After the counting window. Reach is not small in every case (KJ2). OpenAI names no actor behind the Iranian operation. A second lab backs the hired-hand pattern. OpenAI links Dark Clark to Politology reporting from its own view
OpenAI, coordinated model-distillation campaign	30 Sep 2026	Activity from 1 Jul, a spike of 16,000 requests from over 4,000 users on 24 and 25 Jul, over 15,000 users disrupted by 28 Jul, a core cluster tied to people associated with Moonshot AI	A distillation post. It names no Russian, Iranian or North Korean operation. No Moonshot response found
Microsoft, Digital Defense Report 2026	1 Oct 2026	Says lines between state activity, cybercrime and influence operations keep eroding. Russian actors use AI-generated tooling. North Korea uses AI for IT-worker personas, malware and infrastructure. Stolen AI serving capacity is resold. PromptLock ran an open-weight model on adversary infrastructure	Direction-level support for the criminal-proxy pattern. Names no UNC6508, no Telegram monitoring bots and no CopyCop, so it corroborates none of RU-6, RU-8 or KP-3
Mandiant, AI risk and resilience in 2026	Sep 2026	Reports malware with embedded lightweight AI models and adversary use of open-weight models. Names no UNC6508	Adds a second Google voice on open-weight use. No hosting evidence
GTIG, from prompting to autonomy	8 Sep 2026	MIDNIGHT NEPTUNE, a North Korea-nexus financial cluster, combines commercial LLMs and open-weight models, with a DeepSeek-Coder assistant named for Python RATs	Open-weight use with hosting unstated. No local flag
CrowdStrike, ARTEX operator framework	7 Oct 2026	An actor-run framework driving DeepSeek v4.1-flash, GLM-5.3 and Grok 4.6, reached through a likely API proxy or reseller	Open models through a hosted route. Outside the local flag
Dream, multi-agent framework against Asian governments	12 Aug 2026	A framework on the Hermes and OpenClaw agent shells with parallel sub-agents. Dream names no model. Press names DeepSeek-V4-Flash	Model and hosting unknown. Outside the local flag
Check Point, underground-forum review	Spring 2026	A phishing-kit operation, EvilTokens, used a Groq-hosted llama-3.1-8b-instant. Check Point says operations on self-hosted models stay unclassified	Hosted access. The admission names the blind spot behind local-model use

Copyright 2026 Treadstone 71 - www.treadstone71.com

Source	Date	What it holds	Bearing
Trellix, DarkForums review	12 Aug 2026	A seller, Shadowx007, advertised APEX AI as a self-hosted, uncensored attack tool. Trellix does not verify it	Intent by advertisement. No buyer, no run
Chainalysis, blockchain dead drops	17 Sep 2026	Daily on-chain writes tied to the technique rose from 2.06 to 11.1, and Chainalysis links the rise in time to open Chinese model releases	Timing only. No operator shown using a model
10a Labs, supply chain of uncensored models	4 Sep 2026	3,471 original uncensored models, 8,164 redistributed copies, 1,643 downstream apps, about a quarter of them malicious, and 110 repositories wiring such models to Telegram bots	Availability. Names no operator
ThreatDown, guardrail-free AI	21 Jul 2026	6,644 uncensored models and over 22 million downloads in 30 days. Read through snippets only	Availability
Securelist, Armored Likho	3 Jul 2026	Medium-confidence attribution. Loaders said to be AI-generated, from verbose comments and emoji bullets	A style inference, not provider telemetry. The actor attacks Russian targets
BI.ZONE, werewolf clusters	27 Mar 2026	Documents Eagle Werewolf with no AI claim. For another cluster, calls a VBS file "presumably written using generative AI" from detailed comments	Same inference method, different group
NewsGuard, Russian operative reveals plans to target Western AI	Search date 16 Oct 2025	Dougan at a Moscow roundtable on 27 Jan 2025: one home server, almost 90,000 articles a month, an uncensored Llama trained on a Russian view. Page not fetched, snippets only	A Level 1 statement. It leaves CopyCop output untied to the server
Timbuktu Institute, Algorithms at War	9 Sep 2026, via ADF 6 Oct	AI-enabled influence by the junta, Africa Corps, JNIM and FLA. Only the ADF text was read	Supports the Mali setting. Names no surveillance platform and none of the Anthropic cases
Marc Owen Jones, from North Macedonia to Khartoum	About 9 Sep 2026	Four pseudo-news sites run from a Hong Kong company by North Macedonian directors: 3,708 AI-assisted articles and 558 LinkedIn and Meta ads. Paid promotion favors the UAE, Sudan's army and the Muslim Brotherhood. No customer named	Topic fit only. The author sits at Northwestern University in Qatar, so a Qatar-UAE bias is possible
PMOI, statement on false claims of AI misuse	13 Sep 2026	Denies the charge, concedes one closed account, says Anthropic never made contact	The accused party's account. See Section 3.4
Open Technology Fund, Iranian messaging apps audit	Dec 2023 and Oct 2024 audits, page updated 12 Feb 2026	No end-to-end encryption in Eitaa, Rubika or Bale. Servers see the sites users reach. No AI analysis of message content found in app code	Independent method. Bears on platform visibility only. See Section 3.8
Graphika, Pundit by Prompt and Spamouflage flyers	21 Jul and 13 Jul 2026	Pundit by Prompt: more than 300 assets and one persona on 154 channels, with the operators unidentified and the links unconfirmed. The flyer report carries no AI element in its public summary	Counts as stated by the reporter. No attribution

Source	Date	What it holds	Bearing
Government pages read in the second pass	9 Sep to 1 Oct 2026	ODNI, CISA, NSA, DHS, Justice, State and White House releases, the Department of War notices of 28 and 29 Sep, the FBI cyber strategy of 9 Sep. None frames AI-enabled foreign influence against the midterms	Basis for the agency-record null in Section 9
New York Times, 18 Sep 2026, read in a DNYUZ repost	18 Sep 2026	Anonymous officials describe classified assessments that the Kremlin again authorized a campaign aimed at the midterms, less extensive than past ones, with chaos as the main goal and no sign of efforts against voting machinery. Graphika's Jean Le Roux ties the content to Matryoshka and expects it to grow. Nine candidates are named as targets. The story says ODNI, the White House and the Russian Embassy did not respond. The Times page is behind a paywall and was read to its opening paragraphs only	A leak, so the forecast rule is not met and Forecast 1 holds. The officials' claims sit at Level 1, and the repost grades B2 as to its text
Washington Post, 4 Oct 2026, read in an archive copy	4 Oct 2026	People familiar with the reports say the National Intelligence Council assessed Russia as the most active adversary, using digital tools to sow dissent and weaken confidence in the vote, with China working individual races and Iran at a smaller scale. The council's report, they say, finds adversaries leaning harder on AI to create and spread false content. An ODNI report shared with Congress sees little chance of direct interference with vote counts, and Representative Himes says on the record that he sees no change in the long-held finding. The Post credits the Times with the first account. No agency has published the council report	A leak, so the forecast rule is not met and Forecast 1 holds. Anonymous sources, graded B3. It names no AI case and says only that adversaries use AI more
Graphika, Operation Overload page on the midterms and the Berlin vote	Sep 2026, text supplied by the editor	Between 7 and 14 September, likely inauthentic X accounts seeded videos and doctored newspaper pages that accuse German and US politicians of fraud and corruption. Videos used the names of BBC, CNN, Politico, VIGINUM and an MS NOW journalist. Nine Democratic candidates are named. Graphika attributes the activity with high confidence to Operation Overload and the Foundation to Battle Injustice, and says it gave the finding to the Times first. The page text holds no AI term and no reach count	The vendor's own page backs the attribution in the Times story. The AI element rests on press accounts. B2 as to what the page states. Address not recorded

Source	Date	What it holds	Bearing
The Insider, 14 Sep and La Voce di New York, 16 Sep	14 and 16 Sep 2026	The Insider counts at least eight posts of 11 and 12 September with 16 CNN-branded celebrity videos under #midterms2026, and attributes them to Matryoshka through AntiBot4Navalny. La Voce reports AI-generated or altered audio, removals by X, Bluesky and TikTok, and Ossoff and Talarico as targets	Independent press accounts of the wave, B2. No tool is named for the audio
Mother Jones, David Corn: 24 Sep and 29 Sep	24 and 29 Sep 2026	A fake video in the magazine's name drew about 155,000 views in a day on one X account and about 50,000 on a second, which researchers call inflated. At least nine more carry its logo. A fake book promotion in Corn's name showed nearly 200,000 views that an unnamed source calls bot-driven. X removed one post under its trademark policy. AntiBot4Navalny makes the attribution to Matryoshka	Reach figures come from an interested party and an unnamed source. B2 as to what the pages state
Howard W. Gordon, Beyond the Narrative, 30 Sep 2026	30 Sep 2026	Relays the Times story and the wave, with details that match the pages above	An advocate's essay at Level 1, superseded as a source by the rows above

Two further leads wait on primary pages. A Microsoft item on memory poisoning through "Summarize with AI" links cites 50 cases across 31 companies, and we have the relay only. Positive Technologies and Group-IB describe leaked chats of The Gentlemen ransomware group, in which Kimi, Ernie and DeepSeek write code and local models are discussed for sorting stolen data. The account is second-hand and criminal, and it supports no state claim.

17 Keywords and wrap-up

Keywords

Cognitive warfare, influence operations, hybrid warfare, adversary AI use, AI tradecraft adoption, LLM grooming, distillation, reflexive control, AI supply chain, Model Context Protocol, LiteLLM, TeamPCP, Firebird, CopyCop, Storm-1516, Matryoshka, Pravda network, Iranian soft war, Spamouflage, Kimsuky, Jasper Sleet, Public Attribution Series, ASA Evidence Standard, ASA Measurement of Effect, ATCRI, CWIA, HTIM, RDEM, TCSA, FCHA, Sovereign AI Architecture Audit

Wrap-up

Adversaries use AI the way an overworked office uses a new hire. They hand it the drafting, the translating, the watching and the filing, and they keep the decisions. Iranian propagandists, Russian newsrooms, Chinese bureau staff and North Korean operators all do so now. Each acts for its own reasons. None acts under a joint plan.

The work probably got cheaper, though no row measures cost. No source shows persuasion improving, and researchers who measured reach found small audiences. The best-measured earlier case, the 2016 Russian operation on Twitter, concentrated exposure in 1 percent of users and left no measurable mark on attitudes or votes. A later panel on the Internet Research Agency found no significant effect. Large Facebook and Instagram experiments found near-zero attitude effects. One preprint on a TikTok clone found that pro-China influencer

Copyright 2026 Treadstone 71 - www.treadstone71.com

content moved views of China. No study has tested an AI-enabled operation, and the large audience counts of 2026 trace to networks with no documented AI role in the reach. Budget against volume and you buy a cost curve. Measure effect and you learn whether anything lands. Nobody has done the measuring.

The sharper danger sits inside the enterprise. A company that runs AI in production gives adversaries three things at once: an audience in its chatbots, a target in the keys, models and accounts it holds, and a channel in its agents. Criminals already work the second and the third. The LiteLLM compromise lasted about three hours and reached a gateway that sits in more than a third of cloud environments. One owner for all three exposures closes the gaps between them. In the European Union the first report falls due 24 hours after a significant incident.

The record has a blind side, because labs publish only what they see. Recorded Future describes a self-hosted stack behind CopyCop, a Russian network, and a French text analysis partly agrees. Genians reports local-model tooling on a Kimsuky host, a North Korean group, and finds no sign of model training. Google suspects one Chinese intrusion actor of hosting an open-weight model in victim clouds, in a single hedged sentence. Each finding has one reporter, and none has been shown running against a victim. Public reporting thins if more operators move to local models. Treadstone 71 treats that thinning as a finding. Every source gets a grade under the ASA standard, and nothing untraceable gets a score.

Ahead, the baseline leads, and the lead is narrow. State bodies fold AI into ordinary work in 39 percent of futures, a plurality without a majority. A move to local models follows at 30 percent. Counting puts local-model use at 2 rows in 16 among non-provider sources on the ledger, 4 in 34 on the extended set and 4 in 44 with the October to January reports. The model reads latent use near 23 percent. Google's finding on the Chinese actor carries eleven points of Path B. Put back the first coding of that row and A leads 50 to 19. Recode one February row as local and the lead falls to 38 against 36. A rule threshold near 25 percent flips the order, so the order stays open to the coding. Quiet labs push Path B up, and a higher local share pushes it further. Watch the local-share count, new state surveillance units and the next chatbot audit first.

The United States shows a blind side in public too. No federal agency has said on the record that AI-enabled foreign influence targets the midterms, and the record holds no such statement from ODNI, the FBI, CISA, DHS, State or Justice. Silence marks a limit of the record and proves no safety. Researchers and the press, not agencies, name the one operation aimed at the vote: a Matryoshka wave of doctored videos from about 10 September, with small reach where anyone counted. Forecast 1 puts the chance that no agency names the threat before 3 November at 74 percent. Section 12.18 carries the weeks after: adversaries have no shown path to the count, so they work on how it looks. Hoax threats and fake results graphics on 3 and 4 November lead the likely path, and a cyber change to a certified result stands at Almost no chance (1 to <5).

Of 20 judgments in Section 2, none reaches Highly likely, thirteen sit at Likely, five stand at Roughly even chance and two at Unlikely. Sixty-three rows sit in the ledger, and four grade Strong. The extended count adds 23 rows covering February to August and 14 from the four months before, each coded once. None measures what a single reader came to believe. Section 15 lists thirteen limits the evidence carries. Another 64 statements from Iranian, Russian, Chinese, North Korean and proxy sources sit in Section 16.3, each at Level 1 and unscored. The next dated checks fall on 15 October and 31 October, the election resolves Forecast 1 on 3 November, and the window calls resolve on 9 November.

Glossary and abbreviations

Terms appear as the report uses them. Treadstone 71 method names link to their public pages where the report first cites them.

Table 75. Glossary of terms, methods and abbreviations.

Term	Meaning in the report
ACH	Analysis of Competing Hypotheses. A matrix that scores each piece of evidence against each hypothesis and ranks the hypotheses by what evidence contradicts
ACS	Adversarial Cognitive Simulator. A Treadstone 71 method that reads an adversary's likely reaction to a trigger
Admiralty anchor	A two-part grade for a source: reliability A to F and information credibility 1 to 6, printed apart and never averaged
ASA Evidence Standard	The Treadstone 71 evidence-grading standard (DOI 10.5281/zenodo.23137881), Public Edition 1.2. Each ledger row gets an anchor, a T-Score and a D-Score under it
ASA-MOE	ASA Measurement of Effect. A draft standard that ranks study designs and caps any causal claim at the strength of the design
ATA	AI Tradecraft Adoption. A score for how far an actor has built AI into its tradecraft, part of TAI
ATCRI	A Treadstone 71 ranking of threats by raw risk and confidence
ATVA	Adversary Threat Vector Analysis. A grid of actors against vectors that shows where the record documents each vector
Backcasting	Starting from an end state and listing what must hold for it to arrive. Section 12.14 calls the result a prerequisite tree
Brier score	The squared gap between a stated probability and the result, averaged over forecasts. A forecaster who says 50 percent every time scores 0.25
CARM	Cyber Action Risk Management pathways: hide, relocate, monetize, conceal and replace
Cone of plausibility	A picture of the range of futures that stay inside the evidence, widening with time
CWC and CWIA	Cognitive Warfare Countermeasures and Cognitive Warfare Impact Assessment. Treadstone 71 methods used in Section 8
D-Score	The ASA likelihood that a source holds deception, printed with a Kent band
Daily Adversary Brief	The Treadstone 71 monitoring product for adversary-language media, code repositories and forums
DBNA	Dynamic Bayesian Narrative Analysis. A sequential update of hypothesis probabilities, here tempered at an exponent of 0.5
Detectability (d)	How often non-provider detectors see local-model use, relative to commercial use. The model sets the median at 0.5
Distillation	Training a copy model on the outputs of a stronger model, here through fraudulent accounts, as Anthropic and a joint US advisory describe it
DLS	Local-model share, the local-flag rows divided by rows without the provider flag
DS	Design strength. The rung of a study design on the ASA-MOE ladder. DS-1 is randomized assignment
Extended set	The ledger plus the candidate rows and reports from the two earlier-month blocks, used for the 4 of 34 and 4 of 44 local-model counts. Ledger counts leave it out
FIMI	Foreign information manipulation and interference, the term the EEAS uses
F-Score	The ASA product score for an item, printed with a Most-Likely Achieved Endgame
HET	Hypothesis Evolution Tracking
HTIM	Human Terrain Influence Mapping
Indicator of change	An observation, with a check date, that moves the odds of a path. A tripwire is a hard-edged indicator, and a leading indicator tends to show before one

Term	Meaning in the report
Information gain	The expected drop in uncertainty about the paths, in bits, once an indicator resolves
Kent band	A Sherman Kent probability band with half-open edges: Almost certain 95 to 99, Highly likely 85 to <95, Likely 55 to <85, Roughly even chance 45 to <55, Unlikely 15 to <45, Highly unlikely 5 to <15 and Almost no chance 1 to <5
KJ	Key judgment. Section 2 lists 20
Lab-visibility discount	An annotation on every ATA score, not a numeric correction. The note states how much of the evidence comes from provider telemetry, marks the score as a floor and marks local use as unknown
Latent local share	The share of state-linked AI operations that run part of their work on local models, inferred by the model from DLS and detectability
Leading indicator	See indicator of change
Ledger row	One source in Section 16, graded by the ASA Evidence Standard
Level 1, 2, 3	Depth of grading. Level 1 rows carry an anchor alone, and Level 2 and Level 3 rows add a T-Score and a D-Score, with more criteria at Level 3
Local model	A model an actor runs on a stack it controls, also called self-hosted. No source in the report names the hardware or the host. A provider cannot see its use
LTS	Provider-telemetry share, the provider-flag rows divided by all counted rows
MCP	Model Context Protocol. A standard that connects agents to tools and data
MRA	Momentum-Reversal Analysis. A method that forecasts how an adversary pivots after a takedown or exposure
NIS2	The European Union directive on network and information security, second edition
Non-provider row	A counted ledger row whose finding does not rest on a model provider's own telemetry
Path (A to E)	One of five ways 2027 closes, each with a resolution rule applied on 31 December 2027 in the order D, C, B, A, with E for cases the rules cannot resolve
PIR	Priority intelligence requirement, listed in Section 1.2
Provider telemetry	A model provider's report on use of its own platform: Anthropic, OpenAI, Google or Microsoft
Public Attribution Series	The Treadstone 71 dossier product. Dossier 1, dated 12 October, keeps each verified item and each adversary assertion apart
RC row	A row in the extended set, listed in Section 16.1 with an RC prefix. Each carries a single coder's coding
Report basis	A count in which each September report counts once, however many ledger rows it feeds
Resolution rule	The test, applied to the published record, that assigns the future to a path
Retrieval poisoning	Planting content so that a system which retrieves documents at answer time repeats it
Rung	A step on the evidence ladder in Section 10: AI use shown, audience counted, loss documented, belief change measured and behavior change measured
STEMPLES Plus	A twelve-factor environmental scan, with indicators of change as its Plus: Social, Technical, Economic, Military, Political, Legal, Educational, Security, with Religion, Demographics and psychological factors
T-Score	The ASA quality score for a row, from 0 to 100. Strong runs 80 and above, Adequate 60 to 79, Weak 40 to 59 and Insufficient below 40
TAI	Treadstone 71 Adversary Index. A composite score for adversary actors built from sub-indices, among them ATA
Tilt	A state's modeled path share divided by the aggregate share, where 1.00 means no lean
Tripwire	See indicator of change
Watch only	An indicator whose chance of firing differs by less than 1.5 times across paths A to D, so it stays on the list and enters no update
Wilson interval	A 95 percent confidence interval for a share, used for small counts

Copyright 2026 Treadstone 71 - www.treadstone71.com