Source dossier

The vote was an argmax — CLOSEDQUORUM polls four model providers, and the AI Act hands the Union an addressee rather than an off switch

Edition 2026-09-25 · narrative: agent security meets the GPAI chapter · compiled 2026-09-25

Claims index

Claim	Statement	Status	Sources
CLAIM-2026-09-25-01	Cisco Talos published its CLOSEDQUORUM analysis on Tuesday 22 September 2026, authored by Ryan Fetterman, describing it as to their knowledge the first publicly documented Windows implant to apply a model-vote model to tactical command and control.	verified	ciscotalos2026closed
CLAIM-2026-09-25-02	CLOSEDQUORUM queries DeepSeek, Qwen, Mistral and Google Gemini in sequence and acts on the plurality of their verdicts.	verified	ciscotalos2026closed
CLAIM-2026-09-25-03	The tally uses a strict less-than comparison so the first-encountered maximum wins; the preference order is DeepSeek, then Qwen, then Mistral, then Gemini, which gives DeepSeek the deciding vote in any tie.	verified	ciscotalos2026closed
CLAIM-2026-09-25-04	The implant is a 16.4 MB 64-bit Windows executable compiled in Go, re-executing at randomised 5 to 15 minute intervals and exfiltrating AES-256-GCM encrypted data via Discord webhooks.	verified	ciscotalos2026closed
CLAIM-2026-09-25-05	The public distribution build contains placeholder API keys and a dummy webhook, so Talos observed no complete end-to-end execution; development builds show build-time injection of provider credentials, and Talos has no confirmation of in-the-wild deployment.	verified	ciscotalos2026closed
CLAIM-2026-09-25-06	Talos states that the most useful detection strategy is still to focus on behavioral characteristics rather than domain blocking, and names as a signal similar requests potentially sent to several model providers within a short interval.	verified	ciscotalos2026closed
CLAIM-2026-09-25-07	CAIRN is MIT-licensed Python released by Cisco Talos on 22 September 2026, analysing binaries entirely from VirusTotal metadata with 26 YARA rules across three tiers, and its documentation states that T1 and T2 hits without genuine AI integration are common.	verified	ciscotalos2026introducin, github2026cognitive

Claim	Statement	Status	Sources
CLAIM-2026-09-25-08	AI Act Article 54(1) requires GPAI providers established in third countries to appoint, by written mandate, an authorised representative established in the Union before placing the model on the Union market, and Article 54(3) limits that representative's tasks to verifying documentation, keeping it available for ten years, providing information on reasoned request and cooperating with the AI Office.	verified	euartificia2026authorise
CLAIM-2026-09-25-09	AI Act Article 55(1)(c) requires providers of general-purpose AI models designated as having systemic risk to keep track of, document and report without undue delay to the AI Office relevant information about serious incidents.	verified	euartificia2026obligatio
CLAIM-2026-09-25-10	AI Act Article 52(6) requires the Commission to ensure that a list of general-purpose AI models with systemic risk is published and kept up to date.	verified	euartificia2026procedure

Only claims marked **verified** may carry a post. **single-source** may appear as hedged context; **disputed** must not be published at all.

Closed Quorum: inside the first reported autonomous AI C2 implant

PRIMARY

Publisher	Cisco Talos
Published	2026-09-22
Retrieved	2026-09-25
Cite key	ciscotalos2026closed
URL	https://blog.talosintelligence.com/the-closed-quorum-inside-the-first-reported-autonomous-ai-c2-implant/

Published Tuesday, 22 September 2026, 06:00, by Ryan Fetterman (Cisco Talos).

The claim. "CLOSEDQUORUM is, to our knowledge, the first publicly documented Windows implant to apply this model to tactical command and control (C2)."

The vote. The implant queries four commercial LLM providers - DeepSeek, Qwen, Mistral and Google Gemini - in sequence, tallies their independent verdicts and acts on the plurality. Talos: "four models are queried in sequence, their independent verdicts tallied, and the binary acts, based on their judgment."

The tie-break. "In any tie, an order of preference kicks in: DeepSeek first, then Qwen, then Mistral, then Gemini." The implementation uses a strict "<" comparison, so the first-encountered maximum wins; DeepSeek therefore carries the deciding vote in any tie. The ordering is an artefact of the comparison, not a designed quorum rule.

The artefact. A 16.4 MB 64-bit Windows executable compiled in Go. Six SHA256 hashes across development builds spanning seven days; first analysis date 17 June 2026. Re-execution at randomised 5-15 minute intervals.

Capabilities. LSASS credential dumping; browser passwords (Chrome, Edge, Firefox); cryptocurrency wallets (MetaMask, Exodus, Ethereum); process injection and persistence. Stolen data is AES-256-GCM encrypted and exfiltrated via Discord webhooks.

Credentials. "The public distribution build, however, contains placeholder API keys and a dummy webhook, so we did not observe a complete end-to-end execution." All LLM API credentials initialise to `dummy_api_key`. Development builds demonstrate build-time injection of provider credentials, indicating per-operator compilation.

Deployment. "While we do not have confirmation of in-the-wild deployment, artifacts from the binary were used to connect the developer to postings on criminal forums related to carding, dating back to 2025."

Detection guidance. "The most useful detection strategy is still to focus on behavioral characteristics, rather than domain blocking." Named signals: AI-provider API traffic

originating from an unexpected Windows executable; similar requests potentially sent to several model providers within a short interval; Discord webhook communication from the same process or host; repeated execution at randomized 5-15-minute intervals. Talos also published a YARA rule.

Introducing CAIRN: frontier tracking for AI-integrated malware

PRIMARY

Publisher	Cisco Talos
Published	2026-09-22
Retrieved	2026-09-25
Cite key	ciscotalos2026introduci
URL	https://blog.talosintelligence.com/introducing-cairn-frontier-tracking-for-ai-integrated-malware/

Published Tuesday, 22 September 2026, 06:00, by Ryan Fetterman (Cisco Talos) - the same morning as the CLOSEDQUORUM analysis.

CAIRN (Cognitive Artifact Intelligence Research Network) is a research framework for discovering and analysing malware that operationalises or exploits AI systems. It identifies AI-integrated malware through embedded artefacts: prompts, API endpoints, jailbreak strings.

Metadata-first. Binaries are analysed "entirely from metadata - no binary downloads or execution required," combining rule-based detection, semantic clustering and relationship graphing over extracted strings, sandbox behaviour and antivirus detection labels.

Tracked fields. LLM provider endpoints; Python AI framework imports; AI analysis evasion strings; local model inference indicators; agentic tooling syntax.

Stated limitation. "T1/T2 hits without genuine AI integration are common" - false positives arise from PyInstaller bundles and certain framework structures.

CLOSEDQUORUM is named as the first confirmed family, "the first reported autonomous AI C2 implant." Repository: https://github.com/Cisco-Talos/Cognitive-Artifact-Intelligence-Research-Network

Cognitive Artifact Intelligence Research Network (CAIRN)

PRIMARY

Publisher	GitHub
Published	2026-09-22
Retrieved	2026-09-25
Cite key	github2026cognitive
URL	https://github.com/Cisco-Talos/Cognitive-Artifact-Intelligence-Research-Network

Repository confirmed live. Description: a research toolkit for identifying, attributing and tracking AI-related artefacts embedded in malware, using VirusTotal metadata exclusively.

Licence. MIT (Copyright (c) 2026 Cisco Systems, Inc.). **Language.** Python 3.11+.

Tiered YARA ontology - 26 rules across three tiers.

- T1, Primitive Artifacts (9 rules): individual cognitive signals such as API endpoints and hardcoded keys; "high recall, FPs expected".
- T2, Behavioral Context (8 rules): two or more co-occurring primitives in operationally meaningful combinations.
- T3, Operational Families (9 rules): family-level attribution anchored to confirmed seed hashes.

What it does not do. It "does not run YARA rules on raw binary content", and operates without binary downloads, file uploads to VirusTotal, URL submissions or active scans, scheduled automatic pulls, or touching malware directly.

Authorised representatives of providers of general-purpose AI models

LEGAL

Publisher	EU Artificial Intelligence Act
Published	n.d.
Retrieved	2026-09-25
Cite key	euartificia2026authorise
URL	https://artificialintelligenceact.eu/article/54/

Article 54(1). "Prior to placing a general-purpose AI model on the Union market, providers established in third countries shall, by written mandate, appoint an authorised representative ... established in the Union."

Article 54(3) - the representative's tasks.

1. Verify that the technical documentation required by Article 53 has been drawn up and that all obligations under Articles 53 and, where applicable, 55 have been fulfilled by the provider.
2. Keep a copy of the technical documentation at the disposal of the AI Office and national competent authorities for 10 years after the model has been placed on the market.
3. Provide the AI Office, upon a reasoned request, with all information and documentation necessary to demonstrate compliance.
4. Cooperate with the AI Office and competent authorities on any action taken in relation to the model.

Reading for this edition. The task list contains no duty to detect, police or terminate misuse of the model by a paying customer. The Union gains an addressee for documentation and cooperation, not an enforcement lever over an individual API account.

Obligations for providers of general-purpose AI models with systemic risk

LEGAL

Publisher	EU Artificial Intelligence Act
Published	n.d.
Retrieved	2026-09-25
Cite key	euartificia2026obligatio
URL	https://artificialintelligenceact.eu/article/55/

Binds providers of general-purpose AI models **designated as having systemic risk** only.

Article 55(1)(a). "perform model evaluation in accordance with standardised protocols and tools reflecting the state of the art, including conducting and documenting adversarial testing".

Article 55(1)(b). assess and mitigate possible systemic risks at Union level, including their sources, that may stem from the development, the placing on the market, or the use of general-purpose AI models with systemic risk.

Article 55(1)(c). "keep track of, document, and report, without undue delay, to the AI Office and, as appropriate, to national competent authorities, relevant information about serious incidents" and possible corrective measures.

Article 55(1)(d). "ensure an adequate level of cybersecurity protection for the general-purpose AI model with systemic risks and the physical infrastructure of the model".

Reading for this edition. The serious-incident duty is a reporting duty owed upward to the AI Office by a designated provider. It is not a duty owed to a deployer, and it does not attach to non-designated models.

Procedure for classification and the published list of systemic-risk models

LEGAL

Publisher	EU Artificial Intelligence Act
Published	n.d.
Retrieved	2026-09-25
Cite key	euartificia2026procedure
URL	https://artificialintelligenceact.eu/article/52/

Article 52(1). A provider whose general-purpose AI model meets the condition in Article 51(1)(a) must notify the Commission "without delay and in any event within two weeks".

Article 52(6). "The Commission shall ensure that a list of general-purpose AI models with systemic risk is published and shall keep that list up to date, without prejudice to the need to observe and protect intellectual property rights and confidential business information or trade secrets in accordance with Union and national law."

Reading for this edition. Whether an Article 55 duty attaches to any particular model is answerable by consulting a published list rather than by assumption. This edition deliberately does not assert the designation status of DeepSeek, Qwen, Mistral or Gemini; it points the reader at the list.