Anexo A. Matriz de 58 casos

La matriz reproduce la clasificación U/M/T calculada caso por caso y la clasificación E/R/G posterior a la validación primaria; además, superpone H. La columna de actividad resume en una oración la conducta ofensiva, vulnerabilidad o modalidad observada. NC y H? conservan la incertidumbre deliberadamente: no se equiparan a una categoría conocida.

ID	Caso	Actividad / vulnerabilidad / ataque	U/M/T | E/R/G	H
C01	RedVDS	BEC/fraude con phishing personalizado, identidades sintéticas, face swap y clonación de voz/video.	U1/M1/T1 | E0/R3/A	H1
C02	VoidLink	Desarrollo acelerado con agentes de IA de un framework avanzado de malware Linux.	U2/M1/T4 | E1/R1/A	H?
C03	UNC1069	Ingeniería social para robo de criptomonedas con investigación, tooling y contenido sintético/deepfake.	U1/M1/T1 | E0/R2/B	H1
C04	SmartLoader	Ataque de supply chain sobre MCP/IA; no se demostró uso material de IA por el atacante en la campaña específica.	NC/NC/NC | —/R2/A	H?
C05	PromptSpy	Malware Android usa Gemini en runtime para interpretar UI, generar gestos y verificar persistencia.	U4/M2/T2 | E3/R1/A	H?
C06	FortiGate / CyberStrikeAI	Compromiso masivo mediante credenciales débiles, ausencia de MFA y administración expuesta; IA escala TTP conocidas.	U4/M3/T3 | E2/R3/A	H2
C07	Arkanix Stealer	Infostealer experimental atribuido a desarrollo asistido por LLM, con fuente primaria todavía insuficiente.	U2/M1/T2 | E1/R1/D	H?
C08	Moltbook	Prompt injection y contenido malicioso preparado para ser consumido por otros agentes.	U5/M4/T5 | E5/R2/A	H0
C09	Bob-p2p	Skill malicioso induce a agentes a almacenar claves, comprar tokens y desviar pagos.	U5/M4/T5 | E5/R2/A	H0
C10	AWS - 8 minutos	Reconocimiento, código y decisiones tácticas en tiempo real asistidas por IA durante intrusión cloud.	U4/M3/T3 | E3/R3/B	H?
C11	Google Coin	Chatbot fraudulento conversa, responde objeciones y persuade a la víctima para efectuar un pago cripto.	U1/M2/T1 | E3/R2/A	H1
C12	AI bot contra CI/CD	Bot autónomo prueba técnicas, abre PR maliciosos, explota GitHub Actions, logra RCE y exfiltra un token.	U4/M3/T4 | E3/R3/A	H0
C13	Slopoly	Malware probablemente generado/modificado	U2/M1/T3 | E1/R3/B	H?

ID	Caso	Actividad / vulnerabilidad / ataque	U/M/T | E/R/G	H
		con IA incorporado a una operación real de ransomware.		
C14	México / Monterrey	IA participa en intrusión IT, busca credenciales y dirige password spraying; identifica por iniciativa propia un activo SCADA/IIoT.	U4/M3/T6 | E3/R3/A	H?
C15	Fake OpenClaw / Bing AI	Manipulación del entorno de búsqueda para que un asistente recomiende un repositorio malicioso que entrega malware.	U5/M4/T5 | E5/R3/B	H1
C16	PhantomRaven	Campaña npm con dependencias invisibles. La fuente localizada no documenta un papel material de IA ni la modalidad de slopsquatting atribuida en el borrador.	NC/NC/NC | E5/R2/B	H?
C17	TroyDen Lure Factory	Repositorios GitHub con señuelos generados mediante LLM para distribuir infostealers.	U1/M1/T1 | E0/R2/C	H1
C18	DeepLoad	Distribución de stealer mediante ClickFix, con código/ofuscación con fuertes indicios de generación por IA.	U2/M1/T2 | E1/R3/B	H1
C19	Bubble	Phishing alojado en infraestructura legítima de Bubble; no se probó uso material de IA por el atacante.	NC/NC/NC | —/R2/D	H1
C20	prt-scan	Campaña masiva de supply-chain attack contra GitHub mediante PR y explotación automatizada/asistida por IA.	U5/M3/T4 | E2/R2/B	H?
C21	Pushpaganda	Generación industrial de contenido fraudulento, SEO abusivo y scareware para monetización criminal.	U1/M1/T1 | E0/R2/B	H1
C22	ATHR	Vishing automatizado mediante agentes de voz que realizan llamadas y obtienen OTP/credenciales.	U1/M2/T1 | E3/R2/B	H1
C23	NGate	Aplicación móvil troyanizada usada en fraude financiero, con artefactos que sugieren asistencia GenAI en el desarrollo.	U2/M1/T2 | E1/R2/B	H1
C24	Prompt injection web	Instrucciones ocultas para inducir a agentes de terceros a exfiltrar información o ejecutar acciones destructivas.	U5/M4/T5 | E5/R2/A	H0
C25	PromptMink	Compromiso de proyectos Web3 mediante dependencias/paquetes maliciosos incorporados con asistencia de un agente de coding.	U5/M4/T4 | E1/R2/A	H0
C26	Bluekit	PhaaS que integra generación automática de phishing y clonación de voz para industrializar campañas de engaño.	U1/M1/T1 | E0/R2/A	H1
C27	Hugging Face / ClawHub	Skills maliciosos explotan confianza/autoridad de agentes para distribuir infostealers y ejecutar fraude.	U5/M4/T5 | E5/R2/A	H0

ID	Caso	Actividad / vulnerabilidad / ataque	U/M/T | E/R/G	H
C28	Zero-day 2FA	Desarrollo o weaponization asistido por IA de un exploit para evadir autenticación de dos factores.	U2/M1/T3 | E1/R2/A	H0
C29	Hextrike / Strix	Frameworks agénticos automatizan reconocimiento y explotación; probable nexo estatal, a separar del cibercrimen estricto.	U4/M3/T3 | E5/R2/A	H0
C30	Cryptojacking	Poisoning de recomendaciones para que asistentes conduzcan usuarios a recursos que terminan instalando minería.	U5/M4/T5 | E5/R2/B	H1
C31	Marimo	Agente LLM utilizado en post-explotación para interpretar el entorno, encadenar acciones y facilitar exfiltración.	U4/M3/T3 | E3/R3/A	H0
C32	Toolkit ransomware / EDR	IA empleada para desarrollar, probar e iterar ransomware y evasión de EDR en laboratorio ofensivo.	U2/M3/T3 | E1/R2/A	H?
C33	Meta AI Support	Account takeover explotando una falla de autorización/verificación en recuperación automatizada de cuentas (confused deputy).	U5/M4/T5 | E5/R3/B	H0
C34	Outsider Enterprise	PhaaS/smishing que utiliza Gemini para generar mensajes, señuelos y páginas de phishing.	U1/M1/T1 | E0/R3/A	H1
C35	Crypto-clipper	Tutoriales fraudulentos con narración/contenido sintético usados como señuelo para instalar malware que sustituye direcciones cripto.	U1/M1/T1 | E0/R2/A	H1
C36	Phantom Squatting	Dominios que los LLM tienden a inventar son registrados y weaponizados para phishing/distribución.	U5/M4/T5 | E5/R2/A	H1
C37	JadePuffer / EncForge	Agente LLM ejecuta, observa resultados, corrige errores, reintenta y adapta una cadena ransomware multietapa.	U4/M3/T5 | E4/R3/A	H0
C38	Forg365	PhaaS combina señuelos IA con phishing AiTM, device-code phishing, robo de tokens y post-compromiso.	U1/M1/T1 | E0/R2/B	H1
C39	PowerShell / AD	Script con indicios de generación por IA usado en una intrusión real para enumerar Active Directory.	U2/M1/T3 | E1/R3/B	H0
C40	TuxBot v3	Nueva versión de botnet con evidencia clara de desarrollo asistido por LLM, todavía de baja madurez operacional.	U2/M1/T2 | E1/R1/A	H?
C41	Gemini CLI	Agente usado para migración C2, troubleshooting, coding y despliegue dentro de una botnet real.	U4/M2/T3 | E3/R3/A	H0
C42	FakeGit / AgentBaiting	Repositorios, skills y servidores MCP falsos diseñados para que agentes los descubran,	U5/M4/T5 | E5/R2/A	H0

		confíen y consuman contenido malicioso.		
C43	Dolphin X	AI Profiler para puntuar, clasificar y priorizar víctimas automáticamente.	U3/M2/T1 | E2/R2/C	H0
C44	Hermes / Tailandia	Agente ejecuta reconocimiento y post-explotación desatendida dentro de un entorno ya comprometido.	U4/M3/T3 | E3/R3/A	H0
C45	DeepSeek vía Hermes	Agente enumera más de 460 objetivos, busca exploits y lanza intentos de explotación sin intervención humana continua.	U4/M3/T3 | E4/R2/A	H0
C46	Prompt injection cripto	Contenido web induce a agentes a validar plataformas fraudulentas o efectuar pagos cripto.	U5/M4/T5 | E5/R2/A	H0
C47	Ollama / LiteLLM	LLMjacking de endpoints de inferencia expuestos para reutilizar cómputo/modelos ajenos como backend ofensivo.	U5/M5/T5 | E5/R2/A	H2
C48	AWS - 72 horas	Flujos asistidos/agénticos aceleran reconocimiento, tooling, comandos y adaptación en una extorsión cloud.	U4/M3/T3 | E3/R3/B	H?
C49	Text salting	Más de un millón de correos usan texto oculto/alterado para degradar filtros, incluidos clasificadores basados en IA.	U5/M4/T5 | E5/R3/A	H?
C50	Trim / AI Pentest Checker	Plataforma ofensiva transforma instrucciones de alto nivel en reconocimiento, scans, validaciones y reportes automatizados.	U4/M3/T3 | E2/R2/A	H0
C51	Sandworm_Mode	Worm de supply chain inyecta instrucciones MCP/prompt, roba claves LLM y usa coding agents/CI como ejecución y propagación.	U5/M4/T5 | E5/R2/A	H0
C52	DeepSeek proxyjacking	Agente administra propagación, reintentos y failover de una campaña que afectó a más de mil hosts.	U4/M3/T3 | E4/R3/A	H0
C53	Poipet	Scams con identidades falsas, traducción, conversaciones y documentos generados con ChatGPT.	U1/M1/T1 | E0/R3/A	H1
C54	ProKYC	Documentos y video/liveness deepfake para superar KYC y crear cuentas fraudulentas; fuente primaria base anterior a 2026.	U1/M1/T1 | E0/R3/A	H0
C55	Siemens S7	Generación con IA de scripts ofensivos contra PLC para acceso, reconocimiento, credenciales y DoS.	U2/M1/T6 | E1/R2/A	H0
C56	RedC2 4.0	LLM integrado al C2 convierte instrucciones en lenguaje natural en secuencias de comandos para el beacon.	U4/M2/T3 | E3/R2/A	H0
C57	UAT-10147	PentestGPT/DeepAudit y artefactos IA para reconocimiento, explotación, payloads, validación, troubleshooting y persistencia.	U4/M3/T3 | E3/R3/A	H0
C58	AnonyMousKIT	Agentes de voz representan personas y	U1/M2/T1 | E3/R3/A	H1

		conversan con la víctima para obtener passcodes.		

Anexo B. Tablas de datos utilizadas en los gráficos

Mes	**Casos deduplicados**	Cobertura 2+ medios
Ene	2	2
Feb	9	3
Mar	8	2
Abr	7	2
May	5	2
Jun	5	3
Jul	15	7
Ago*	7	1

U	Casos
U1 Engaño/fraude	13
U2 Ingeniería ofensiva	10
U3 Recon/targeting	1
U4 Ejecución/orquestación	16
U5 Distribución/mediación	15
NC No clasificable	3

M	Casos
M1 Asistente generativo	19
M2 IA embebida runtime	7
M3 Orquestación agéntica	15
M4 Manipulación/confused deputy	13

Mes	**Casos deduplicados**	Cobertura 2+ medios
M5 Abuso infraestructura IA	1	
NC No clasificable	3	
T	Casos	
T1 Humano/identidad/finanzas	14	
T2 Endpoint/mobile/IoT	5	
T3 IT/cloud/red	16	
T4 Desarrollo/supply chain	4	
T5 Ecosistema IA/agentes	14	
T6 OT/ICS	2	
NC No clasificable	3	
R	Casos	
R1 Desarrollo/prototipo	4	
R2 Operación observada	32	
R3 Impacto confirmado	22	
E	Casos	
E0 Amplificación generativa	10	
E1 Ingeniería acelerada	11	
E2 Copiloto operacional	4	
E3 Ejecución agéntica	13	
E4 Operador adaptativo	3	
E5 Ataque mediado por IA	15	
Sin E defendible	2	

Mes	**Casos deduplicados**	Cobertura 2+ medios

G	Casos
A Directa/casi directa	39
B Inferencia técnica fuerte	15
C Indirecta corroborada	2
D Insuficiente/media-only	2

H	Casos
H0 Sin dependencia humana	25
H1 Acción activa de víctima	18
H2 Error/omisión previa	2
H? Indeterminado	13

E	**H0**	**H1**	**H2**	**H?**	**Total**
E0	1	9	0	0	10
E1	4	2	0	5	11
E2	2	0	1	1	4
E3	6	3	0	4	13
E4	3	0	0	0	3
E5	9	3	1	2	15
Sin E	0	1	0	1	2

Anexo C. Fuentes y trazabilidad

C.1 Arquitectura de fuentes

Las cuatro fuentes periodísticas núcleo se utilizan como detectores y mecanismos de corroboración. Las fuentes técnicas/institucionales se utilizan para validar el papel material de la IA, la realidad de implementación y la fuerza de evidencia. Esta tabla ofrece directorios oficiales para revisión independiente.

Tipo	Fuente	Directorio oficial	Función en el estudio
Medio especializado	The Hacker News	https://thehackernews.com/	Detección y seguimiento periodístico
Medio especializado	BleepingComputer	https://www.bleepingcomputer.com/	Detección y seguimiento periodístico
Medio especializado	SecurityWeek	https://www.securityweek.com/	Detección y seguimiento periodístico
Medio especializado	Dark Reading	https://www.darkreading.com/	Detección y seguimiento periodístico
Fuente técnica primaria	Microsoft Security / Threat Intelligence	https://www.microsoft.com/en-us/security/blog/	Telemetría, campañas, legal disruption, fraude y técnicas
Fuente técnica primaria	Google Threat Intelligence Group / Mandiant	https://cloud.google.com/blog/topics/threat-intelligence	Atribución, campañas, uso de Gemini y explotación
Fuente técnica primaria	AWS Security / Threat Intelligence	https://aws.amazon.com/blogs/security/	Incidentes cloud y escalado operacional
Fuente técnica primaria	Palo Alto Networks Unit 42	https://www.paloaltonetworks.com/unit42	Telemetría, ataques agénticos, supply chain y control contra hype
Fuente técnica primaria	Cisco Talos	https://blog.talosintelligence.com/	Intrusiones, C2, exploits y workflows agentic
Fuente técnica primaria	Sysdig Threat Research	https://www.sysdig.com/threat-research	Forense cloud y agentes dentro del loop operacional
Fuente técnica primaria	Trend Micro Research / TrendAI	https://www.trendmicro.com/en_us/research.html	Logs/sesiones, malware y operaciones asistidas por IA
Fuente técnica primaria	ESET Research / WeLiveSecurity	https://www.welivesecurity.com/en/	Malware, mobile y separación PoC/telemetría/producción
Fuente técnica primaria	Check Point Research	https://research.checkpoint.com/	Malware, agentes, OPSEC y desarrollo spec-driven
Proveedor / plataforma	OpenAI	https://openai.com/	Reportes sobre abuso de modelos y disrupciones
Proveedor / plataforma	Anthropic	https://www.anthropic.com/research	Investigación y reportes sobre uso/seguridad de modelos
Agencia pública	CISA	https://www.cisa.gov/	Alertas y confirmación institucional de amenazas
Agencia pública	FBI	https://www.fbi.gov/	Acciones legales, alertas y evidencia institucional
Agencia pública	NSA Cybersecurity	https://www.nsa.gov/Cybersecurity/	Guías y alertas sobre amenazas a infraestructura

C.2 Referencia primaria utilizada por caso

El manifiesto siguiente vincula cada caso con una URL completa y registra la consulta realizada y validada el 1 de septiembre de 2026. El estado distingue validación, reserva y límites de alcance; una reserva no se convierte automáticamente en una categoría U/M/T.

Caso asociado	Fuente	URL completa	Consulta	Estado de validación
C01 — RedVDS	Microsoft	https://www.microsoft.com/en-us/security/blog/2026/01/14/inside-redvds-how-a-single-virtual-desktop-provider-fueled-worldwide-cybercriminal-operations/	01/09/2026	Validada
C02 — VoidLink	Check Point Research	https://research.checkpoint.com/2026/voidlink-early-ai-generated-malware-framework/	01/09/2026	Validada

Caso asociado	Fuente	URL completa	Consulta	Estado de validación
C03 — UNC1069	Google Threat Intelligence Group / Mandiant	https://cloud.google.com/blog/topics/threat-intelligence/unc1069-targets-cryptocurrency-ai-social-engineering	01/09/2026	Validada
C04 — SmartLoader	Straiker STAR Labs	https://www.straiker.ai/blog/smartloader-clones-oura-ring-mcp-to-deploy-supply-chain-attack	01/09/2026	Reserva: IA objetivo, no uso material
C05 — PromptSpy	ESET Research	https://www.welivesecurity.com/en/eset-research/promptspy-ushers-in-era-android-threats-using-genai/	01/09/2026	Validada
C06 — FortiGate / CyberStrikeAI	Amazon Threat Intelligence	https://aws.amazon.com/blogs/security/ai-augmented-threat-actor-accesses-fortigate-devices-at-scale/	01/09/2026	Validada
C07 — Arkanix Stealer	Fuente primaria pendiente	https://securelist.com/arkanix-stealer/119006/	01/09/2026	Reserva: la fuente técnica no prueba el uso de IA
C08 — Moltbook	Zenity Labs	https://labs.zenity.io/post/agent-to-agent-exploitation-in-the-wild-observed-attacks-on-moltbook-b929	01/09/2026	Validada
C09 — Bob-p2p	Straiker	https://www.straiker.ai/blog/built-on-clawhub-spread-on-moltbook-the-new-agent-to-agent-attack-chain	01/09/2026	Validada
C10 — AWS - 8 minutos	Sysdig Threat Research Team	https://www.sysdig.com/blog/ai-assisted-cloud-intrusion-achieves-admin-access-in-8-minutes	01/09/2026	Validada
C11 — Google Coin	Malwarebytes Threat Intelligence	https://www.malwarebytes.com/blog/ai/2026/02/scammers-use-fake-gemini-ai-chatbot-to-sell-fake-google-coin	01/09/2026	Validada
C12 — AI bot contra CI/CD	StepSecurity	https://www.stepsecurity.io/blog/hackerbot-claw-github-actions-exploitation	01/09/2026	Validada
C13 — Slopoly	IBM X-Force	https://www.ibm.com/think/x-force/slopoly-start-ai-enhanced-ransomware-attacks	01/09/2026	Validada
C14 — México / Monterrey	Dragos / Check Point Research	https://www.dragos.com/blog/ai-assisted-ics-attack-water-utility	01/09/2026	Validada
C15 — Fake OpenClaw / Bing AI	Huntress / Bing telemetry context	https://www.huntress.com/blog/openclaw-github-ghostsocks-infostealer	01/09/2026	Validada
C16 — PhantomRaven	Koi Security	https://www.koi.ai/blog/phantomraven-npm-malware-hidden-in-invisible-dependencies	01/09/2026	No validada: la fuente no sustenta slopsquatting/IA
C17 — TroyDen Lure Factory	AlienVault OTX / análisis derivado	https://otx.alienvault.com/browse/global/pulses?q=TroyDen	01/09/2026	Reserva: sin fuente primaria canónica localizada
C18 — DeepLoad	ReliaQuest	https://reliaquest.com/blog/threat-spotlight-deepload-malware-pairs-clickfix-delivery-with-ai-generated-evasion/	01/09/2026	Validada
C19 — Bubble	Fuente primaria no localizada	https://www.kaspersky.com/blog/bubble-no-code-phishing/55488/	01/09/2026	Reserva: uso material de IA no demostrado
C20 — prt-scan	Wiz	https://www.wiz.io/blog/six-accounts-one-actor-inside-the-prt-scan-supply-chain-campaign	01/09/2026	Validada
C21 — Pushpaganda	HUMAN Security	https://www.humansecurity.com/learn/resources/satori-threat-intelligence-alert-pushpaganda-manipulates-google-discovery-feeds-with-ai-generated-content-to-spread-malicious-notifications/	01/09/2026	Validada
C22 — ATHR	Abnormal Security	https://www.abnormal.ai/blog/athr-ai-voice-phishing-toad-attacks	01/09/2026	Validada
C23 — NGate	ESET Research	https://www.welivesecurity.com/en/eset-research/new-ngate-variant-hides-in-a-trojanized-nfc-payment-app/	01/09/2026	Validada
C24 — Prompt injection web	Google Security / Unit 42	https://unit42.paloaltonetworks.com/ai-agent-prompt-injection/	01/09/2026	Validada
C25 — PromptMink	ReversingLabs	https://www.reversinglabs.com/blog/claude-promptmink-malware-crypto	01/09/2026	Validada
C26 — Bluekit	Varonis Threat Labs	https://www.varonis.com/blog/bluekit	01/09/2026	Validada
C27 — Hugging Face / ClawHub	Palo Alto Networks Unit 42	https://unit42.paloaltonetworks.com/openclaw-ai-supply-	01/09/2026	Validada

Caso asociado	Fuente	URL completa	Consulta	Estado de validación
		chain-risk/		
C28 — Zero-day 2FA	Google Threat Intelligence Group	https://cloud.google.com/blog/topics/threat-intelligence/ai-vulnerability-exploitation-initial-access	01/09/2026	Validada
C29 — Hextrike / Strix	Google Threat Intelligence Group	https://cloud.google.com/blog/topics/threat-intelligence/ai-vulnerability-exploitation-initial-access	01/09/2026	Validada; fuera del núcleo criminal estricto
C30 — Cryptojacking	Microsoft Defender Research	https://www.microsoft.com/en-us/security/blog/2026/02/10/ai-recommendation-poisoning/	01/09/2026	Validada
C31 — Marimo	Sysdig Threat Research Team	https://www.sysdig.com/blog/ai-agent-at-the-wheel-how-an-attacker-used-llms-to-move-from-a-cve-to-an-internal-database-in-4-pivots	01/09/2026	Validada
C32 — Toolkit ransomware / EDR	Sophos	https://www.sophos.com/en-us/blog/pointing-a-cursor-at-evading-detection	01/09/2026	Validada
C33 — Meta AI Support	Meta	https://www.bleepingcomputer.com/news/security/meta-ai-support-data-breach-affects-20-000-instagram-accounts/	01/09/2026	Reserva: fuente pública secundaria
C34 — Outsider Enterprise	Google / FBI legal disruption	https://blog.google/innovation-and-ai/technology/safety-security/combatting-ai-scams/	01/09/2026	Validada
C35 — Crypto-clipper	Check Point Research	https://research.checkpoint.com/2026/from-stars-to-upvotes-fake-reputation-fueling-a-crypto-clipboard-hijacker/	01/09/2026	Validada
C36 — Phantom Squatting	Palo Alto Networks Unit 42	https://unit42.paloaltonetworks.com/phantom-squatting-hallucinated-web-domains/	01/09/2026	Validada
C37 — JadePuffer / EncForge	Sysdig Threat Research Team	https://www.sysdig.com/blog/jadepuffer-evolves-the-agentic-threat-actor-deploys-ransomware-built-to-destroy-ai-models	01/09/2026	Validada
C38 — Forg365	ZeroBEC	https://zerobec.com/blog/inside-forg365-telegram-distributed-sneaky2fa-style-phaas	01/09/2026	Validada
C39 — PowerShell / AD	Huntress	https://www.huntress.com/blog/ai-coded-malware-vibe-coding-active-directory	01/09/2026	Validada
C40 — TuxBot v3	Palo Alto Networks Unit 42	https://unit42.paloaltonetworks.com/tuxbot-v3-evolution-iot-botnet/	01/09/2026	Validada
C41 — Gemini CLI	TrendAI / Trend Micro Research	https://www.trendmicro.com/en_us/research/26/g/actor-behind-patriot-bait-used-ai-to-deploy-c2-botnet.html	01/09/2026	Validada
C42 — FakeGit / AgentBaiting	Island Security Research	https://www.island.io/blog/agentbaiting-how-800-fake-ai-skills-and-mcp-servers-delivered-malware	01/09/2026	Validada
C43 — Dolphin X	Varonis Threat Labs	https://www.varonis.com/blog/dolphin-x-stealer	01/09/2026	Validada
C44 — Hermes / Tailandia	Hunt.io	https://hunt.io/blog/thailand-ministry-finance-targeted-with-hermes-ai-agent	01/09/2026	Validada
C45 — DeepSeek vía Hermes	Palo Alto Networks Unit 42	https://unit42.paloaltonetworks.com/autonomous-ai-cyber-attack-campaign/	01/09/2026	Validada
C46 — Prompt injection cripto	Zscaler ThreatLabz	https://www.zscaler.com/blogs/security-research/indirect-prompt-injection-web-content-targets-ai-agents	01/09/2026	Validada
C47 — Ollama / LiteLLM	Sysdig Threat Research Team	https://www.sysdig.com/blog/llmjacking-evolved-attackers-are-using-stolen-ai-compute-to-build-offensive-agentic-tools	01/09/2026	Validada
C48 — AWS - 72 horas	Sygnia	https://www.sygnia.co/blog/inside-an-ai-assisted-cloud-attack/	01/09/2026	Validada
C49 — Text salting	Barracuda	https://blog.barracuda.com/2026/07/16/text-salting-ai-email-security	01/09/2026	Validada
C50 — Trim / AI Pentest Checker	Cato CTRL	https://www.catonetworks.com/blog/cato-ctrl-how-one-threat-actor-turned-frontier-ai-into-an-offensive-	01/09/2026	Validada

Caso asociado	Fuente	URL completa	Consulta	Estado de validación
		platform/		
C51 — Sandworm_Mode	Socket Research	https://socket.dev/blog/sandworm-mode-npm-worm-ai-toolchain-poisoning	01/09/2026	Validada
C52 — DeepSeek proxyjacking	Jesta Security	https://jesta.ai/blog/darkreasoning	01/09/2026	Validada
C53 — Poipet	OpenAI	https://openai.com/index/disrupting-malicious-uses-of-ai-criminal-scam-operation/	01/09/2026	Validada
C54 — ProKYC	Cato CTRL / MITRE ATLAS	https://www.catonetworks.com/blog/prokyc-selling-deepfake-tool-for-account-fraud-attacks/	01/09/2026	Validada; fuente base anterior a 2026
C55 — Siemens S7	NSA + CISA/FBI/DOE/EPA	https://www.cisa.gov/news-events/cybersecurity-advisories/aa26-231a	01/09/2026	Validada
C56 — RedC2 4.0	TrendAI Security	https://www.trendaisecurity.com/en-us/resources-insights/trendai-security-blog/redc2-ai-powered-linux-implant	01/09/2026	Validada
C57 — UAT-10147	Cisco Talos	https://blog.talosintelligence.com/uat-10147-chinese-speaking-adversary-integrates-agentic-ai-into-post-compromise-operations/	01/09/2026	Validada
C58 — AnonyMousKIT	SOCRadar	https://socradar.io/blog/anonymouskit-ai-phaas-supply-chain/	01/09/2026	Validada

Anexo D. Notas de alcance y control de cierre

- C04 SmartLoader se conserva como caso de ataque a la supply chain de IA/MCP, pero queda NC en U/M/T porque la fuente no demuestra uso material de IA por el atacante.
- C19 Bubble se conserva como registro de phishing en infraestructura legítima, pero queda NC en U/M/T porque no se demuestra el uso material del componente de IA por el adversario.
- C29 Hextrike/Strix se mantiene por relevancia técnica comparada, con advertencia de probable nexo estatal o de espionaje y sin confundirlo con criminalidad financiera probada.
- C54 ProKYC se conserva como antecedente validado; la fuente primaria base es anterior a 2026 y no se computa como prueba independiente de una campaña nueva en 2026.
- C07 Arkanix y C17 TroyDen se mantienen con reservas explícitas porque no se localizó una fuente primaria canónica que pruebe de manera directa el papel atribuido a la IA.
- C16 PhantomRaven queda NC en U/M/T: la referencia localizada documenta dependencias npm invisibles, pero no sustenta la atribución a slopsquatting ni un papel material de IA.
- U/M/T se calculó caso por caso sobre la matriz de 58 registros. Hay 55 registros clasificables y tres NC; cada eje es independiente y no se suma en un puntaje compuesto.
- El manifiesto bibliográfico incorpora URL completa, fecha de consulta, caso asociado y estado de validación para los 58 registros.
- El corte sustantivo del corpus permanece en el 29/08/2026; la consulta de enlaces realizada el 01/09/2026 no agrega casos posteriores al corte.