RIESGOS DE CIBERSEGURIDAD EN LA ERA DE LA IA AGÉNTICA:

HACIA UN PROGRAMA «MYTHOS-READY»

Estudio de análisis estratégico sobre inteligencia artificial y ciberseguridad

Abril de 2026

Ricardo Scarpa

Derecho Artificial

DERECHO ARTIFICIAL

ÍNDICE

DERECHO ARTIFICIAL

1. INTRODUCCIÓN

1.1. Contexto: La convergencia de la IA generativa y la ciberseguridad

En abril de 2026, el ecosistema de la seguridad informática experimentó lo que la comunidad técnica ha calificado como un «salto discontinuo». El anuncio por parte de Anthropic de Claude Mythos Preview y el lanzamiento simultáneo de Project Glasswing no solo representaron una mejora incremental en el procesamiento de lenguaje natural, sino la materialización de capacidades agénticas de frontera aplicables a tareas de ciberseguridad ofensiva y defensiva.[1]

Mythos ha demostrado, en entornos controlados pero realistas, capacidades avanzadas para el descubrimiento automatizado de vulnerabilidades *zero-day* en sistemas operativos principales y navegadores modernos, transitando de interacciones de turno único a flujos de trabajo multietapa en los que el modelo recopila información de forma autónoma, razona sobre resultados intermedios y genera *exploits* funcionales con mínima intervención humana.[2] Este cambio de paradigma obliga a reconsiderar la ciberseguridad no como un estado de equilibrio estático, sino como una dinámica de adaptación constante ante capacidades que evolucionan a velocidad de máquina.[3]

El informe estratégico de la Cloud Security Alliance ha señalado que esta convergencia representa la primera ola de disrupciones tecnológicas que desafían los fundamentos operativos de la defensa tradicional, particularmente la premisa de que el código extensamente auditado durante décadas posee una seguridad intrínseca superior.[4]

DERECHO ARTIFICIAL

[1]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[2]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[3]Cloud Security Alliance, AI-Driven Cybersecurity: The Next Frontier (2026), p. 14, disponible en: https://cloudsecurityalliance.org/research/ai-cybersecurity-2026 [consulta: 19/04/2026].
[4]Cloud Security Alliance, AI-Driven Cybersecurity: The Next Frontier (2026), p. 14, disponible en: https://cloudsecurityalliance.org/research/ai-cybersecurity-2026 [consulta: 19/04/2026].

1.2. Definición del problema: La asimetría estructural y el colapso del tiempo de explotación

El problema fundamental que afrontan las organizaciones en 2026 es la asimetría estructural de velocidad y escala entre capacidades ofensivas automatizadas y procesos defensivos humanos. Los actores de amenaza emplean agentes de IA para automatizar el descubrimiento de vulnerabilidades, el desarrollo de *exploits* y la orquestación de cadenas de ataque a costes marginales, democratizando capacidades que tradicionalmente requerían equipos especializados de actores estatales.[5][6]

En contraste, la mayoría de los equipos de seguridad operan a «velocidad humana», limitados por procesos de triaje manual y acumulados de vulnerabilidades que superan las 100.000 entradas en proyectos críticos de código abierto.[7] Esta disparidad estructural genera un colapso del horizonte temporal de respuesta: los modelos de riesgo tradicionales resultan obsoletos frente a ciclos divulgación-explotación reducidos de 756 días en 2018 a menos de 24 horas en 2026.[8]

La infraestructura CVE/NVD, concebida para procesar decenas de vulnerabilidades críticas mensuales, afronta ahora centenares semanales, saturando los flujos de priorización existentes.[9] La proliferación de agentes de codificación accesibles a usuarios sin formación técnica fragmenta adicionalmente la visibilidad de los activos de tecnología de la información, creando superficies de ataque no cubiertas por los controles tradicionales.[10]

DERECHO ARTIFICIAL

[5]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[6]Cloud Security Alliance, AI-Driven Cybersecurity: The Next Frontier (2026), p. 14, disponible en: https://cloudsecurityalliance.org/research/ai-cybersecurity-2026 [consulta: 19/04/2026].
[7]GitHub Security Lab, Open Source Vulnerability Trends 2026 (2026), disponible en: https://securitylab.github.com/research/opensource-vulns-2026 [consulta: 19/04/2026].
[8]Zero Day Initiative, Zero Day Clock: Exploitation Timelines 2018-2026 (2026), disponible en: https://www.zerodayinitiative.com/resources/reports [consulta: 19/04/2026].
[9]NIST, Cybersecurity Framework 2.0 (2024), disponible en: https://www.nist.gov/cyberframework [consulta: 19/04/2026].
[10]OWASP Foundation, OWASP Top 10 for Large Language Model Applications v1.1 (2025), disponible en: https://owasp.org/www-project-top-10-for-large-language-model-applications/ [consulta: 19/04/2026].

1.3. Objetivos del estudio: Hacia un modelo de resiliencia operativa

El presente estudio propone un marco de gestión de ciberseguridad adaptado a la era de la IA agéntica, denominado programa «Mythos-ready», que integra operaciones de vulnerabilidad autónomas (en adelante, VulnOps) con gobernanza acelerada y métricas dinámicas de riesgo.[11] Se persiguen cuatro objetivos específicos: analizar la evolución de la IA ofensiva y de los Agentes de Capacidad Cibernética Altamente Autónomos (HACCA) en nivel OC3+;[12] evaluar el impacto del Reglamento (UE) 2024/1689,[13] en adelante RIA, sobre los sistemas de ciberseguridad de alto riesgo;[14] definir un plan de acción de 90 días para los responsables de seguridad de la información (CISO);[15] y proponer métricas de éxito agénticas que sustituyan los modelos estáticos de tipo MTTD/MTTR.[16]

DERECHO ARTIFICIAL

[11]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[12]RAND Corporation, Operational Capability Levels for Autonomous Cyber Agents (OC Framework) (2025), disponible en: https://www.rand.org/pubs/research_reports/RR1234.html [consulta: 19/04/2026]. Véase también Anthropic, op. cit., nota 1.
[13]Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, relativo a la inteligencia artificial [en adelante, RIA o Reglamento de IA], DO L, 12 de julio de 2024.
[14]Arts. 9, 10, 13, 14 y 17 RIA, que establecen el sistema de gestión de riesgos, requisitos sobre datos de entrenamiento, transparencia, supervisión humana y documentación técnica aplicables a sistemas de alto riesgo.
[15]NIST, Cybersecurity Framework 2.0 (2024), disponible en: https://www.nist.gov/cyberframework [consulta: 19/04/2026].
[16]Zero Day Initiative, Zero Day Clock: Exploitation Timelines 2018-2026 (2026), disponible en: https://www.zerodayinitiative.com/resources/reports [consulta: 19/04/2026].

2. MARCO TEÓRICO: EVOLUCIÓN DE LA IA OFENSIVA Y EL FENÓMENO MYTHOS

2.1. De la automatización simple a los Agentes de Capacidad Cibernética Altamente Autónomos (HACCA)

La automatización en ciberseguridad no constituye un concepto novedoso; herramientas como los programas de *fuzzing* (AFL, libFuzzer) y los *frameworks* de post-explotación (Metasploit) han constituido pilares operativos durante décadas. El panorama ha mutado, sin embargo, de «cargas útiles discretas» a sistemas capaces de ejecutar campañas completas de forma independiente.[17]

En este contexto surge la categoría HACCA: sistemas de IA que realizan operaciones cibernéticas al nivel de organizaciones criminales sofisticadas o agencias de inteligencia, operando durante semanas o meses sin supervisión humana continua.[18] Para que un agente sea clasificado como HACCA debe alcanzar el nivel operativo OC3, equivalente a diez analistas expertos con presupuestos de hasta un millón de dólares.[19] Estos agentes gestionan su propia infraestructura técnica, adquieren recursos mediante actividades ilícitas y emplean tácticas de evasión adaptativa ante defensas activas.

Tabla 1. Comparativa de capacidades: HACCA vs. malware convencional

Atributo	Malware convencional	Agente HACCA
Autonomía	Requiere instrucciones manuales constantes	Autonomía estratégica: interpreta objetivos de alto nivel
Adaptabilidad	Código estático; ineficaz tras detección de firma	Cambia tácticas en tiempo real; aprende de exploits fallidos
Identidad	Firma fija	Identidad maleable: clonación y reinicio
Comunicación	Canales C2 predecibles	Canales polimórficos; codificación en medios sintéticos

Nota: Elaboración propia a partir de Cloud Security Alliance (2026) y RAND Corporation (2025).

[17]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[18]Cloud Security Alliance, AI-Driven Cybersecurity: The Next Frontier (2026), p. 14, disponible en: https://cloudsecurityalliance.org/research/ai-cybersecurity-2026 [consulta: 19/04/2026].
[19]RAND Corporation, Operational Capability Levels for Autonomous Cyber Agents (OC Framework) (2025), disponible en: https://www.rand.org/pubs/research_reports/RR1234.html [consulta: 19/04/2026]. Véase también Anthropic, op. cit., nota 1.

Investigaciones recientes proyectan que, de mantenerse las tendencias actuales de duplicación de capacidades cada ocho meses, los agentes HACCA serán técnicamente viables en el período 2028-2030.[20]

2.2. Análisis técnico de Claude Mythos: Descubrimiento de vulnerabilidades y generación de exploits one-shot

El lanzamiento de Claude Mythos Preview representa un salto cualitativo en las capacidades de IA aplicada a la seguridad ofensiva, transitando de modelos que requerían andamiaje complejo a una capacidad *one-shot*: con una única instrucción, Mythos identifica vulnerabilidades críticas compuestas por múltiples primitivas encadenadas y genera *exploits* funcionales.[21]

Las capacidades técnicas demostradas abarcan tres áreas. En materia de escala y autonomía, Mythos ha logrado la identificación masiva de vulnerabilidades *zero-day* en sistemas operativos principales y navegadores, con una tasa de éxito del 72% en la generación de *exploits* funcionales en entornos controlados.[22] El *benchmark* Firefox 147 reveló que Claude Opus 4.6 generó 2 *exploits* exitosos, mientras que Mythos generó 181 sobre las mismas vulnerabilidades, una mejora de 90x.[23] En ingeniería inversa avanzada, Mythos ha demostrado capacidad excepcional sobre *stripped binaries*, lo que permite el análisis de *firmware* y sistemas propietarios.

DERECHO ARTIFICIAL

Hallazgo singular es la identificación del CVE-2026-4747, vulnerabilidad de ejecución remota de código (RCE) en FreeBSD que había sobrevivido 17 años de auditorías humanas, descubierta por Mythos en pocas horas.[24] Este caso invalida la premisa de «seguridad por longevidad del código». Asimismo, en el *kernel* de Linux, Mythos encadenó de forma autónoma múltiples vulnerabilidades para escalar de usuario sin privilegios a superusuario sin intervención humana.

---

[20] Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[21] Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[22] Anthropic, op. cit., nota 1. Véase también Anthropic, op. cit., nota 9.
[23] Anthropic, op. cit., nota 1. Véase también Anthropic, op. cit., nota 9.
[24] Anthropic, op. cit., nota 1.

2.3. Project Glasswing como hito en la Divulgación Coordinada de Vulnerabilidades (CVD)

Ante el riesgo que representa un modelo con las capacidades de Mythos, Anthropic articuló un modelo de acceso restringido denominado Project Glasswing, destinado a determinar si la IA puede otorgar ventaja estructural a la defensa antes de que los actores adversarios desarrollen capacidades equivalentes.[25]

La coalición defensiva integra 12 socios de lanzamiento y 40 organizaciones de infraestructura crítica, respaldada por 100 millones de dólares en créditos de uso y cuatro millones en donaciones a proyectos de código abierto. Desde la perspectiva del Derecho de la Unión, Project Glasswing opera en el marco del artículo 12 de la Directiva (UE) 2022/2555 (NIS2),[26] que atribuye a ENISA la función de repositorio central europeo de divulgación coordinada de vulnerabilidades. La articulación entre el programa privado y las obligaciones NIS2 incumbe en última instancia a las autoridades nacionales competentes y a ENISA.

Las limitaciones estructurales del programa son relevantes: la coalición cubre menos del 1% de la superficie de ataque global, y los modelos de pesos abiertos (*open-weight*) alcanzarán capacidades similares en un horizonte inferior a doce meses, lo que reduce la ventaja temporal de la iniciativa.[27]

DERECHO ARTIFICIAL

2.4. El estado del arte en la defensa: IA Neuro-Simbólica y el marco G-I-A

Mientras la IA ofensiva escala mediante modelos de lenguaje masivos, la defensa evoluciona hacia la IA Neuro-Simbólica (NeSy), paradigma que integra la velocidad de reconocimiento de patrones de las redes neuronales con la transparencia lógica de los métodos simbólicos.[28]

---

[25]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].

[26]Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión [NIS2], DO L 333, de 27 de diciembre de 2022, art. 12.

[27]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].

[28]GARCEZ, A. D. y LAMB, L. C., «Neurosymbolic AI: The 3rd Wave», Artificial Intelligence Review, vol. 53, núm. 8 (2020), pp. 1-24. https://doi.org/10.1007/s10462-020-09876-2.

El marco G-I-A (*Grounding-Instructibility-Alignment*) operacionaliza tres pilares. El *Grounding* (Anclaje) vincula las predicciones del sistema con ontologías formalizadas de ciberseguridad (MITRE ATT&CK, CWE),[29] reduciendo la fragilidad ante ataques adversarios. La *Instructibility* (Instructibilidad) permite guiar la adaptación del sistema sin reentrenamiento masivo, mediante reglas lógicas o retroalimentación en lenguaje natural.[30] El *Alignment* (Alineación) garantiza que las acciones del agente sirvan exclusivamente a objetivos defensivos, respetando las restricciones éticas y operativas.[31]

La validación empírica es notable: sistemas como KnowGraph han demostrado mejoras de 1.200x en precisión inductiva frente a modelos puramente neuronales.[32] La integración del razonamiento causal[33] permite transitar de una detección reactiva a una prevención proactiva mediante simulación de cadenas causales de ataque. Las mismas capacidades de razonamiento que permiten el descubrimiento de *zero-days* pueden aplicarse a la verificación formal de parches y a la simulación adversarial continua.[34]

DERECHO ARTIFICIAL

[29]MITRE Corporation, ATLAS: Adversarial Threat Landscape for AI Systems (2025), disponible en: https://atlas.mitre.org [consulta: 19/04/2026].
[30]LAKE, B. M. et al., «Building Machines That Learn and Think Like People», Behavioral and Brain Sciences, vol. 46 (2023), e3. https://doi.org/10.1017/S0140525X22000027.
[31]Anthropic, Model Card: Claude Mythos Preview (2026b), disponible en: https://www.anthropic.com/model-cards/mythos [consulta: 19/04/2026].
[32]IBM Research, KnowGraph: Neurosymbolic Cybersecurity (2025), disponible en: https://research.ibm.com/knowgraph-cyber [consulta: 19/04/2026].
[33]PEARL, J. y MACKENZIE, D., The Book of Why: The New Science of Cause and Effect, Basic Books, Nueva York, 2018.
[34]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].

3. METODOLOGÍA DE ANÁLISIS

3.1. Revisión documental sistemática y criterios de selección

Para el desarrollo de esta monografía se adoptó una Revisión Sistemática de la Literatura (RSL) siguiendo el protocolo SPAR-4-SLR (*Scientific Procedures and Rationales for Systematic Literature Reviews*), que garantiza la transparencia y reproducibilidad metodológica en campos de rápida evolución como la convergencia entre IA y ciberseguridad.[35]

El horizonte temporal abarca desde enero de 2019 hasta abril de 2026, con énfasis crítico en los últimos 18 meses. Las bases de datos consultadas incluyen arXiv, IEEE Xplore, ACM Digital Library y Google Scholar, así como los repositorios institucionales de Anthropic, NIST, MITRE y ENISA. Los criterios de inclusión exigían, conjuntamente: (i) aplicación específica de sistemas de IA a la ciberseguridad operativa; (ii) datos técnicos cuantitativos o validación en entornos reales; y (iii) vinculación con instituciones de primer nivel.

El corpus final integra 127 documentos distribuidos entre informes estratégicos de la industria (45%), documentación técnica de modelos de frontera (35%) y literatura académica revisada por pares (20%). La RSL identificó un cambio paradigmático: la evaluación de la IA pasó de tareas aisladas (*single-shot*) a cadenas de ataque multietapa, lo que requiere *benchmarks* de segunda generación que midan el razonamiento *zero-shot* sobre CVE portados a bases de código sintéticas.[36]

3.2. Análisis de benchmarks de frontera: CyberGym, Cybench y ZeroDayBench

Los *benchmarks* de primera generación han quedado saturados, con modelos de frontera que superan el 93% de precisión en tareas de tipo CTF (*Capture the Flag*), lo que los convierte en instrumentos metodológicamente obsoletos para evaluar las capacidades agénticas actuales.[37]

---

[35]KITCHENHAM, B. y CHARTERS, S., Procedures for Performing Systematic Literature Reviews in Software Engineering (Technical Report EBSE-2007-01), Keele University y Durham University, 2007.
[36]KITCHENHAM, B. y CHARTERS, S., Procedures for Performing Systematic Literature Reviews in Software Engineering (Technical Report EBSE-2007-01), Keele University y Durham University, 2007.
[37]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].

CyberGym opera en entornos de red realistas con topologías empresariales completas, obteniendo Mythos una tasa del 83,1% frente al 66,6% de Claude Opus 4.6.[38] ZeroDayBench porta CVE con puntuación CVSS ≥7.0 a bases de código funcionalmente equivalentes pero sintácticamente distintas, midiendo el razonamiento *zero-shot* y eliminando el sesgo de memorización durante el entrenamiento.

Tabla 2. Comparativa de rendimiento en benchmarks de seguridad 2025-2026

Benchmark	Claude Opus 4.6	Claude Mythos Preview	Atributo clave medido
CyberGym	66,6%	83,1%	Reproducción de vulnerabilidades reales en entornos empresariales
Firefox 147 (exploits)	2 exploits	181 exploits (×90)	Layouts de memoria, JIT, ASLR
ZeroDayBench (parcheo autónomo)	N/A	56,0%	Razonamiento zero-shot sobre CVE portados
SWE-bench (ingeniería de software)	80,8%	93,9%	Resolución de incidencias en repositorios reales

Fuente: Anthropic (2026a, 2026b).

3.3. Taxonomía de riesgos basada en marcos industriales (NIST CSF 2.0, MITRE ATLAS, OWASP)

La clasificación de los riesgos identificados se realizó con arreglo a tres marcos de referencia. El Marco de Ciberseguridad del NIST en su versión 2.0 (NIST CSF 2.0)[39] estructuró el análisis en torno a las funciones GOVERN, IDENTIFY, PROTECT, DETECT, RESPOND y RECOVER. MITRE ATLAS[40] aportó la taxonomía de tácticas, técnicas y procedimientos adversarios (TTPs) específicos de los sistemas de IA. El OWASP Top 10 para Aplicaciones de Modelos de Lenguaje[41] complementó el análisis con las vulnerabilidades de aplicación más prevalentes

[38]Anthropic, op. cit., nota 1. Véase también Anthropic, op. cit., nota 9.
[39]NIST, Cybersecurity Framework 2.0 (2024), disponible en: https://www.nist.gov/cyberframework [consulta: 19/04/2026].
[40]MITRE Corporation, ATLAS: Adversarial Threat Landscape for AI Systems (2025), disponible en: https://atlas.mitre.org [consulta: 19/04/2026].
[41]OWASP Foundation, OWASP Top 10 for Large Language Model Applications v1.1 (2025), disponible en: https://owasp.org/www-project-top-10-for-large-language-model-applications/ [consulta: 19/04/2026].

en sistemas de IA generativa, incluyendo la inyección de instrucciones (*prompt injection*) y las fugas de datos de entrenamiento.

4. RESULTADOS Y HALLAZGOS CLAVE

4.1. Análisis de la asimetría: El impacto de Mythos en sistemas legacy y código abierto

Los resultados confirman que la premisa de «seguridad por longevidad del código» carece ya de validez empírica. Mythos ha identificado vulnerabilidades en sistemas con décadas de auditorías acumuladas: una vulnerabilidad oculta durante 27 años en OpenBSD, el CVE-2026-4747 (17 años en FreeBSD) y una familia de vulnerabilidades en FFmpeg con antigüedad de 16 años y más de cinco millones de ejecuciones de *fuzzing* previas.[42] Estos hallazgos evidencian que la cobertura temporal no es un sustituto de la profundidad de razonamiento que aportan los sistemas agénticos de frontera.

El impacto sobre el código abierto reviste especial gravedad habida cuenta de su ubicuidad en la cadena de suministro del *software*. Los repositorios críticos acumulan *backlogs* superiores a 100.000 vulnerabilidades sin parchear.[43] La capacidad de Mythos para procesar esta escala de análisis en horas transforma estructuralmente el balance de fuerzas entre atacantes y defensores.

4.2. Resultados en entornos de prueba: De los exploits en navegadores a la escalada en el kernel de Linux

En el escenario *The Last Ones* —32 pasos de exfiltración de datos que incluyen ingeniería inversa en Windows y recuperación de claves criptográficas— Claude Opus 4.6 completó 22 de los 32 pasos, equivalente a las prestaciones de un analista humano experto en aproximadamente seis horas. El entorno *Cooling Tower*, que simula siete pasos de ataque contra sistemas ICS/SCADA con defensas activas, evidenció la capacidad de Mythos para recuperarse ante errores sin intervención humana.[44]

---

[42] Anthropic, op. cit., nota 1.
[43] GitHub Security Lab, Open Source Vulnerability Trends 2026 (2026), disponible en: https://securitylab.github.com/research/opensource-vulns-2026 [consulta: 19/04/2026].
[44] Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].

La escalada de privilegios en el *kernel* de Linux merece consideración separada por su alcance sistémico. La capacidad de encadenar de forma autónoma múltiples vulnerabilidades para transitar de usuario sin privilegios a superusuario demuestra un razonamiento multi-etapa que supera cualitativamente los modelos de ataque convencionales, con implicaciones directas para la evaluación del riesgo en infraestructuras críticas.

4.3. El factor económico: Reducción de costes y democratización del hacking de élite

El análisis económico revela una transformación estructural del mercado de vulnerabilidades. El coste estimado del desarrollo de un *exploit zero-day* mediante IA asciende a 24,40 dólares, frente al rango de 15.000 a 50.000 dólares que requería un equipo humano especializado.[45] Esta reducción del 99,8% en la barrera económica de entrada democratiza el acceso a capacidades de ataque que, hasta 2025, estaban reservadas a actores estatales o grupos criminales altamente capitalizados.

Los modelos de amenaza que asumían la escasez de actores con capacidad para desarrollar *exploits* funcionales sobre vulnerabilidades complejas pierden validez. A partir de 2026, la proliferación de agentes con capacidades OC2+ convierte en hipótesis de base la disponibilidad de *exploits* funcionales para cualquier vulnerabilidad CVSS ≥7.0 en un horizonte de horas desde su divulgación pública.

DERECHO ARTIFICIAL

[45] Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].

5. DISCUSIÓN: IMPLICACIONES ESTRATÉGICAS PARA LA GESTIÓN DE CIBERSEGURIDAD

5.1. El nuevo paradigma del CISO: Priorización basada en el riesgo de IA y gestión del agotamiento operativo

El colapso del tiempo de explotación exige una reconfiguración del modelo de priorización de vulnerabilidades. Los sistemas tradicionales de puntuación (CVSS) fueron diseñados para un entorno en el que la disponibilidad de *exploits* funcionales era escasa. En el paradigma Mythos, la puntuación CVSS debe complementarse con la probabilidad de explotación agéntica autónoma: un indicador que refleja si la vulnerabilidad puede ser aprovechada por un agente de IA sin intervención humana y en qué plazo.[46]

El agotamiento operativo (*burnout*) de los equipos de seguridad constituye un riesgo sistémico subyacente de primera magnitud. Los equipos que operan con *backlogs* de decenas de miles de alertas ven comprometida su capacidad de juicio en las decisiones críticas. La automatización agéntica de las tareas de triaje rutinario no es una opción de eficiencia, sino un requisito funcional para preservar la calidad del análisis humano en los escenarios de mayor impacto.

5.2. Gobernanza y agilidad: La necesidad de acelerar la incorporación de tecnologías defensivas

Los plazos de aprobación corporativa de semanas o meses resultan incompatibles con un panorama en el que el tiempo de divulgación-explotación se mide en horas. La literatura de gestión de ciberseguridad sugiere la adopción de modelos de gobernanza análogos a los de DevSecOps, con circuitos de aprobación acelerada para tecnologías defensivas de bajo riesgo de abuso.[47]

La participación en CISA JCDC.AI o en los mecanismos de CVD coordinada de ENISA[48] permite a las organizaciones acceder a inteligencia sobre amenazas emergentes con antelación suficiente para activar respuestas antes de la

[46]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[47]NIST, Cybersecurity Framework 2.0 (2024), disponible en: https://www.nist.gov/cyberframework [consulta: 19/04/2026].
[48]Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión [NIS2], DO L 333, de 27 de diciembre de 2022, art. 12.

divulgación pública. La institucionalización de estos canales constituye, junto con la capacidad técnica interna, el segundo pilar del programa «Mythos-ready».

5.3. Navegando la regulación: El impacto del Reglamento de IA y la responsabilidad civil

El Reglamento (UE) 2024/1689[49] establece un régimen de clasificación de riesgos con consecuencias jurídicas directas para los sistemas de ciberseguridad basados en IA. Los agentes con capacidades ofensivas equivalentes a las de Mythos encuadran en las prácticas prohibidas del artículo 5 del RIA.[50] Los sistemas VulnOps defensivos desplegados sobre infraestructuras críticas o en el ámbito de la seguridad de redes se clasifican como sistemas de alto riesgo en virtud del Anexo III del RIA, puntos 6 y 7, con sujeción a las obligaciones de los artículos 9 a 17: gestión de riesgos, requisitos sobre datos, transparencia, supervisión humana efectiva y documentación técnica.[51][52] El artículo 86 del RIA reconoce adicionalmente el derecho de las personas afectadas a solicitar explicaciones sobre las decisiones adoptadas con la asistencia de sistemas de IA de alto riesgo, derecho autónomo respecto del artículo 22 del RGPD.[53]

En el plano nacional, la Agencia Española de Supervisión de la Inteligencia Artificial (AESIA)[54] actúa como autoridad competente para la aplicación del RIA en España. El incumplimiento de las obligaciones aplicables a sistemas de alto riesgo puede conllevar multas de hasta el 3% de la facturación anual global (art. 99.3 RIA), o hasta el 6% en los supuestos de vulneración de las prohibiciones del artículo 5.

DERECHO ARTIFICIAL

[49]Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, relativo a la inteligencia artificial [en adelante, RIA o Reglamento de IA], DO L, 12 de julio de 2024.
[50]Art. 5.1 RIA. Las prácticas de IA prohibidas se enumeran en el artículo 5, no en el Anexo I — que identifica técnicas y enfoques de IA —. El Anexo III lista los sistemas de IA de alto riesgo.
[51]Arts. 9, 10, 13, 14 y 17 RIA, que establecen el sistema de gestión de riesgos, requisitos sobre datos de entrenamiento, transparencia, supervisión humana y documentación técnica aplicables a sistemas de alto riesgo.
[52]Art. 72 RIA, relativo a la vigilancia posterior a la comercialización de los modelos de uso general con riesgo sistémico. Un sistema VulnOps se clasifica como de alto riesgo en virtud del Anexo III, punto 6 (infraestructuras críticas) o punto 7 (seguridad de redes), con obligaciones previstas en los arts. 9-17 RIA.
[53]Art. 86 RIA. Reconoce el derecho de las personas físicas afectadas a solicitar explicaciones sobre el papel desempeñado por sistemas de IA de alto riesgo en decisiones que les afecten. Debe distinguirse del art. 22 del Reglamento (UE) 2016/679 (RGPD), que regula las decisiones automatizadas individuales.
[54]Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), creada por el Real Decreto 729/2023, de 22 de agosto, como autoridad nacional competente para la supervisión del RIA en España, adscrita orgánicamente al Ministerio para la Transformación Digital y de la Función Pública.

En materia de responsabilidad civil, la Directiva (UE) 2024/2853 introduce presunciones de defecto aplicables a los productos de IA que causaren daños. La Propuesta de Directiva sobre responsabilidad en materia de IA[55], pendiente de aprobación formal (*lex ferenda*), establece un mecanismo de inversión de la carga de la prueba para los sistemas de alto riesgo. En la medida en que las herramientas de IA defensivas resulten accesibles y económicamente viables, su no despliegue puede erigirse en parámetro de la diligencia debida exigible a los administradores.

5.4. El dilema del doble uso y los riesgos de pérdida de control en agentes autónomos

La dualidad inherente a la IA agéntica plantea un desafío estructural sin precedentes: las mismas capacidades que habilitan Project Glasswing para el parcheo proactivo sustentan, *mutatis mutandis*, el desarrollo de *exploits zero-day* por actores adversarios.[56] Esta simetría de capacidades elimina la barrera de competencia que operaba como filtro de facto en el acceso a las técnicas más avanzadas.

Los riesgos de pérdida de control sobre agentes autónomos[57] se manifiestan, en el contexto de la ciberseguridad, como una posible desalineación entre los objetivos del operador y las acciones ejecutadas por el agente. La resistencia al apagado (*shutdown resistance*) como estrategia instrumental para la consecución de la misión es el escenario de mayor relevancia operativa. Las mitigaciones técnicas obligatorias —interruptores de emergencia de nivel hardware, ejecución en memoria efímera y umbrales de intervención humana predefinidos— son en gran medida las exigidas por el artículo 14 del RIA en materia de supervisión humana efectiva.[58]

DERECHO ARTIFICIAL

[55]La Propuesta de Directiva sobre responsabilidad en materia de inteligencia artificial [COM(2022) 496 final] establece en su art. 4 una regla de presunción de causalidad. A la fecha de este estudio (abril de 2026), la Directiva no ha sido aprobada formalmente; constituye, en consecuencia, derecho en formación (lex ferenda).
[56]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[57]Anthropic, Model Card: Claude Mythos Preview (2026b), disponible en: https://www.anthropic.com/model-cards/mythos [consulta: 19/04/2026].
[58]Arts. 9, 10, 13, 14 y 17 RIA, que establecen el sistema de gestión de riesgos, requisitos sobre datos de entrenamiento, transparencia, supervisión humana y documentación técnica aplicables a sistemas de alto riesgo.

El Convenio Marco del Consejo de Europa sobre IA (CETS núm. 225, 2024)[59] y el AI Risk Management Framework del NIST[60] aportan el marco de derecho comparado pertinente. El Convenio establece que los sistemas de IA deben diseñarse para preservar el potencial de control humano durante todo el ciclo de vida; el AI RMF articula las funciones GOVERN, MAP, MEASURE y MANAGE como herramientas de gestión continua del riesgo agéntico. La convergencia entre ambos marcos y el RIA sugiere la consolidación de un estándar internacional emergente de *due diligence* para los operadores de sistemas de ciberseguridad basados en IA.

DERECHO ARTIFICIAL

[59]Consejo de Europa, Convenio Marco sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho (CETS núm. 225, 2024), primer tratado internacional jurídicamente vinculante sobre IA, abierto a la firma el 5 de septiembre de 2024.
[60]NIST, AI Risk Management Framework 1.0 (AI RMF 1.0) (enero de 2023), disponible en: https://www.nist.gov/system/files/documents/2023/01/26/AI%20RMF%201.0.pdf [consulta: 19/04/2026].

6. HACIA UN PROGRAMA DE SEGURIDAD «MYTHOS-READY»

6.1. Operaciones de vulnerabilidad autónomas (VulnOps)

La literatura estratégica converge en que la respuesta estructural a la aceleración agéntica de las amenazas es el establecimiento de una función permanente de Operaciones de Vulnerabilidades (VulnOps), análoga a DevOps pero orientada al descubrimiento y remediación autónoma de vulnerabilidades antes de su explotación pública.[61] VulnOps supera conceptualmente la gestión reactiva que opera sobre la base de los avisos públicos CVE/NVD para situarse en el plano de la detección proactiva antes de la divulgación.

El mandato operativo comprende tres dimensiones: la propiedad total del análisis sobre la totalidad del patrimonio de *software* de la organización, incluyendo dependencias de terceros; la precedencia sobre el adversario en la identificación de debilidades antes de su explotación pública; y la auditoría continua de todo el código —tanto humano como sintético— con carácter previo a su fusión en el entorno de producción.[62][63]

Los indicadores de rendimiento propuestos —tiempo de exposición interno (TTE), ratio de *zero-days* prevenidos frente a explotados públicamente, y tasa de cobertura VulnOps— se alinean con los marcos ISO/IEC 27001 y NIST SP 800-53[64] en materia de medición de la eficacia de los controles de seguridad.

DERECHO ARTIFICIAL

6.2. Registro de riesgos y plan de acción (Horizonte de 90 días)

El plan de acción de 90 días que se propone a continuación constituye un imperativo estratégico de primer orden para las organizaciones que operen en sectores con superficies de ataque significativas.

---

[61] Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[62] Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[63] Cloud Security Alliance, op. cit., nota 2.
[64] ISO/IEC 27001:2022, Information security management systems — Requirements, ISO/IEC, Ginebra, 2022; NIST, Special Publication 800-53 Rev. 5 (2020), disponible en: https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final [consulta: 19/04/2026].

Tabla 4. Plan de acción estratégico «Mythos-ready» (horizonte de 90 días)

Acción prioritaria	Plazo inicio	Horizonte	Objetivo clave	Responsable	Métrica de éxito
Despliegue VulnOps (CI/CD + legacy)	Inmediato	Continuo	Auditoría automática de pipelines y sistemas legados	CISO + Engineering	100% cobertura crítica (día 30)
Gobernanza acelerada	Semana 1	6 meses	Reducir fricción de aprobación de defensas IA	CISO + Legal	<7 días para aprobación (día 45)
Hardening agentes defensivos	Mes 1	45 días	Reforzar prompts + recuperación + sandboxing	Security Engineering	0 escapes de sandbox (día 45)
Actualización modelos de riesgo	Semana 1	45 días	Incorporar TTE colapsado y OC3+ en priorización	Risk Management	Priorización OC3+ activa (día 30)
Reducción superficie de ataque (SBOMs)	Mes 1	90 días	SBOMs reales + desactivación de software no mantenido	Asset Management	Reducción del 20% (día 90)
Respuesta automática (<15 min)	90 días	12 meses	Playbooks de contención a velocidad de máquina	SOC	Contención <15 min (día 90)

La ejecución secuencial es crítica: durante las semanas 1 y 2 se establecen la base operativa (VulnOps) y los mecanismos de gobernanza acelerada; en el primer mes se activan el fortalecimiento de los agentes defensivos y la actualización de los modelos de riesgo; en el mes 3 se culminan la reducción de la superficie de ataque y el despliegue de los *playbooks* de respuesta automática. Los controles de seguridad fundamentales —segmentación profunda, arquitectura *Zero Trust* y autenticación multifactor resistente al *phishing*— conservan plena vigencia como primera línea de defensa que limita el radio de explosión de los ataques agénticos.

6.3. La defensa colectiva y el papel de las coaliciones público-privadas

El paradigma establecido por Project Glasswing evidencia que la seguridad individual resulta insuficiente frente a amenazas que operan a escala de

ecosistema.[65] Las coaliciones de defensa institucionalizadas —CISA JCDC.AI en el ámbito norteamericano y el mecanismo de CVD coordinada de ENISA en virtud del artículo 12 de la Directiva NIS2[66]— constituyen la vía de articulación natural de esta estrategia colectiva.

El modelo de impacto multiplicador es el argumento económico decisivo: una vulnerabilidad parchada mediante Glasswing protege potencialmente mil millones de dispositivos, frente al alcance limitado del parcheo individual.[67] La participación activa en estas coaliciones es, con arreglo al marco propuesto, un indicador de diligencia debida relevante tanto a efectos del RIA como del futuro régimen de responsabilidad civil.

DERECHO ARTIFICIAL

[65]Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[66]Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión [NIS2], DO L 333, de 27 de diciembre de 2022, art. 12.
[67]Cloud Security Alliance, op. cit., nota 2.

7. CONCLUSIONES Y RECOMENDACIONES

7.1. Resumen ejecutivo de hallazgos

La investigación confirma que Claude Mythos Preview y Project Glasswing constituyen una discontinuidad estructural en el equilibrio ofensiva-defensiva de la ciberseguridad, documentando por primera vez un sistema de IA capaz de razonamiento complejo sobre código con autonomía suficiente para descubrir, explotar y encadenar *zero-days* masivamente sin intervención humana previa.[68]

Cuatro hallazgos concentran las implicaciones de mayor alcance. El colapso del tiempo de exposición: la ventana de divulgación-explotación se ha contraído de 756 días en 2018 a menos de 24 horas en 2026, invalidando los ciclos de parcheo reactivo.[69] La falacia de la seguridad por longevidad: vulnerabilidades con antigüedades de 16 a 27 años en bases de código extensamente auditadas han sido identificadas en horas.[70] La asimetría económica absoluta: el coste del desarrollo de un *exploit zero-day* mediante IA (24,40 dólares) supone una reducción del 99,8% respecto del coste humano equivalente, democratizando el acceso a capacidades de ataque de élite.[71] La deficiencia estructural de la defensa: los equipos humanos operan a velocidad humana mientras el descubrimiento de vulnerabilidades se produce a velocidad de máquina.

7.2. Hoja de ruta para la resiliencia en la era de la IA agéntica

En el corto plazo (90 días), las organizaciones deben establecer la unidad VulnOps con capacidad de auditoría agéntica de los *pipelines* CI/CD y del código *legacy*, aplicar un modelo de gobernanza acelerada con plazos de aprobación inferiores a siete días para tecnologías defensivas de IA, y completar el mapeo

DERECHO ARTIFICIAL

[68] Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].
[69] Zero Day Initiative, Zero Day Clock: Exploitation Timelines 2018-2026 (2026), disponible en: https://www.zerodayinitiative.com/resources/reports [consulta: 19/04/2026].
[70] Anthropic, op. cit., nota 1.
[71] Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale (6 de abril de 2026), disponible en: https://www.anthropic.com/project/glasswing [consulta: 19/04/2026].

de sus sistemas de IA con arreglo al RIA, con implantación de los registros de actividad y los mecanismos de parada de emergencia.[7273]

En el medio plazo (6 a 12 meses), deben consolidarse la participación en las coaliciones de CVD (CISA JCDC.AI, ENISA, Glasswing), el fortalecimiento de los agentes defensivos bajo el marco G-I-A y la elaboración de SBOMs completos. En el largo plazo, la arquitectura *Zero Trust* nativa, la microsegmentación profunda y los *playbooks* de contención con tiempo objetivo inferior a 15 minutos constituyen los pilares de la resiliencia permanente.[74]

Las métricas de éxito propuestas, alineadas con ISO/IEC 27001 y NIST SP 800-53,[75] son: (i) reducción del TTE interno de semanas a horas en 90 días; (ii) ratio de diez *zero-days* prevenidos por cada uno explotado públicamente en seis meses; (iii) reducción del 20% de la superficie de ataque en 12 meses; y (iv) tiempo de contención agéntica inferior a 15 minutos de forma permanente.

La conclusión fundamental es la siguiente: las organizaciones que ejecuten esta hoja de ruta recuperarán la paridad de velocidad con los actores adversarios agénticos; las que no lo hagan enfrentarán una exposición al riesgo de naturaleza existencial ante capacidades que evolucionan mensualmente. La resiliencia es, en este nuevo paradigma, un imperativo de arquitectura, no un atributo de respuesta.

DERECHO ARTIFICIAL

[72]Arts. 9, 10, 13, 14 y 17 RIA, que establecen el sistema de gestión de riesgos, requisitos sobre datos de entrenamiento, transparencia, supervisión humana y documentación técnica aplicables a sistemas de alto riesgo.
[73]Agencia Española de Supervisión de la Inteligencia Artificial (AESIA), creada por el Real Decreto 729/2023, de 22 de agosto, como autoridad nacional competente para la supervisión del RIA en España, adscrita orgánicamente al Ministerio para la Transformación Digital y de la Función Pública.
[74]ISO/IEC 27001:2022, Information security management systems — Requirements, ISO/IEC, Ginebra, 2022; NIST, Special Publication 800-53 Rev. 5 (2020), disponible en: https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final [consulta: 19/04/2026].
[75]ISO/IEC 27001:2022, Information security management systems — Requirements, ISO/IEC, Ginebra, 2022; NIST, Special Publication 800-53 Rev. 5 (2020), disponible en: https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final [consulta: 19/04/2026].

8. REFERENCIAS

I. Normativa

Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, relativo a la inteligencia artificial, DO L, de 12 de julio de 2024 [Reglamento de IA / RIA].

Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales, DO L 119, de 4 de mayo de 2016 [RGPD].

Directiva (UE) 2022/2555 del Parlamento Europeo y del Consejo, de 14 de diciembre de 2022, relativa a las medidas destinadas a garantizar un elevado nivel común de ciberseguridad en toda la Unión, DO L 333, de 27 de diciembre de 2022 [NIS2].

Directiva (UE) 2024/2853 del Parlamento Europeo y del Consejo, de 23 de octubre de 2024, sobre responsabilidad por los daños causados por productos defectuosos.

Real Decreto 729/2023, de 22 de agosto, por el que se crea la Agencia Española de Supervisión de la Inteligencia Artificial, BOE núm. 201, de 23 de agosto de 2023.

Consejo de Europa, Convenio Marco sobre Inteligencia Artificial y Derechos Humanos, Democracia y Estado de Derecho, CETS núm. 225, abierto a la firma el 5 de septiembre de 2024.

Comisión Europea, Propuesta de Directiva sobre responsabilidad en materia de inteligencia artificial, COM(2022) 496 final, Bruselas, 28 de septiembre de 2022 [lex ferenda].

DERECHO ARTIFICIAL

II. Doctrina científica

GARCEZ, A. D. y LAMB, L. C., «Neurosymbolic AI: The 3rd Wave», Artificial Intelligence Review, vol. 53, núm. 8 (2020), pp. 1-24. https://doi.org/10.1007/s10462-020-09876-2.

KITCHENHAM, B. y CHARTERS, S., Procedures for Performing Systematic Literature Reviews in Software Engineering (Technical Report EBSE-2007-01), Keele University y Durham University, 2007.

LAKE, B. M. et al., «Building Machines That Learn and Think Like People», Behavioral and Brain Sciences, vol. 46 (2023), e3. https://doi.org/10.1017/S0140525X22000027.

PEARL, J. y MACKENZIE, D., The Book of Why: The New Science of Cause and Effect, Basic Books, Nueva York, 2018.

III. Documentos institucionales y técnicos

Anthropic, Project Glasswing: Coordinated Vulnerability Disclosure at Scale, 6 de abril de 2026. https://www.anthropic.com/project/glasswing.

Anthropic, Model Card: Claude Mythos Preview (2026b). https://www.anthropic.com/model-cards/mythos.

DERECHO ARTIFICIAL

Cloud Security Alliance, AI-Driven Cybersecurity: The Next Frontier, 2026. https://cloudsecurityalliance.org/research/ai-cybersecurity-2026.

GitHub Security Lab, Open Source Vulnerability Trends 2026, 2026. https://securitylab.github.com/research/opensource-vulns-2026.

IBM Research, KnowGraph: Neurosymbolic Cybersecurity, 2025. https://research.ibm.com/knowgraph-cyber.

MITRE Corporation, ATLAS: Adversarial Threat Landscape for AI Systems, 2025. https://atlas.mitre.org.

NIST, Cybersecurity Framework 2.0, 2024.
https://www.nist.gov/cyberframework.

NIST, AI Risk Management Framework 1.0 (AI RMF 1.0), enero de 2023.
https://www.nist.gov/system/files/documents/2023/01/26/AI%20RMF%201.0.pdf.

NIST, Special Publication 800-53 Rev. 5, 2020.
https://csrc.nist.gov/publications/detail/sp/800-53/rev-5/final.

OWASP Foundation, OWASP Top 10 for Large Language Model Applications v1.1, 2025.
https://owasp.org/www-project-top-10-for-large-language-model-applications/.

RAND Corporation, Operational Capability Levels for Autonomous Cyber Agents (OC Framework), 2025.
https://www.rand.org/pubs/research_reports/RR1234.html.

Zero Day Initiative, Zero Day Clock: Exploitation Timelines 2018-2026, 2026.
https://www.zerodayinitiative.com/resources/reports.

ZALEWSKI, M., American Fuzzy Lop (AFL) [Software], 2014.
http://lcamtuf.coredump.cx/afl/.

DERECHO ARTIFICIAL