// ZERO TRUST CONSULTING | ZTC RESEARCH

INVESTIGACIÓN INDEPENDIENTE

LA NUEVA ASIMETRÍA DE LA CIBERSEGURIDAD

La IA acelera el descubrimiento de vulnerabilidades. La arquitectura limita las posibilidades de ataque y la respuesta determina durante cuánto tiempo permanece expuesta una organización.

2024—2026 | EVIDENCIA · ARQUITECTURA · RESILIENCIA

La gravedad de una vulnerabilidad depende también de su accesibilidad, de los privilegios que permite obtener y de los límites que impone la arquitectura.

Informe ZTCR–AI–VM–2026–01 | Versión 1.0 | 31 de octubre de 2026

https://zerotrust.consulting/research

PANORAMA EJECUTIVO / ZTC RESEARCH

La IA vuelve obsoletos los ciclos de respuesta de días: la exposición crítica exige contención en horas.

El descubrimiento asistido por IA y la automatización de técnicas de explotación pueden reducir drásticamente el tiempo disponible para proteger los sistemas. Los plazos de algunos marcos normativos, trasladados sin adaptar a políticas internas, pueden resultar insuficientes ante exposiciones críticas. ZTC Research invita a organismos normativos, reguladores y organizaciones a revisar y actualizar sus criterios de priorización, contención y remediación. La protección inmediata debe activarse en horas; la corrección definitiva requiere un despliegue seguro y una verificación posterior.

72,833	43 días	≤ 4 h
CVE divulgados Enero–septiembre de 2026 [01]	**Mediana de corrección completa** CVE del catálogo KEV; datos de 2025 [29]	**Objetivo de contención** Exposición crítica; propuesta ZTC (§ 9)

Las políticas adquieren valor protector cuando se traducen en controles implementados, probados y medidos: inventarios fiables, segmentación, privilegios mínimos, aislamiento, cambios de emergencia y evidencia de cierre. La aprobación documental es solo el comienzo.

DEL HALLAZGO AL IMPACTO: SEIS ETAPAS



Secuencia 01–06: la arquitectura limita alcance y privilegios; el parche corrige el defecto.

Modelo conceptual. La secuencia ilustra donde puede intervenir la arquitectura: limitar el acceso, los privilegios y el impacto mientras se corrige el defecto original.

DECISIÓN PRIORITARIA

Revisar y actualizar los criterios normativos y las políticas internas; demostrar su aplicación mediante límites de confianza, aislamiento, cambios de emergencia y cierre técnicamente verificado.

La investigación distingue hechos documentados, hipótesis contrastables, propuestas operativas y límites de la evidencia.

SÍNTESIS DE INVESTIGACIÓN / ZTC RESEARCH

Resumen ejecutivo

Los registros públicos muestran una aceleración notable en la divulgación de vulnerabilidades durante 2026. En la serie analizada [01], los CVE publicados entre enero y septiembre pasaron de 35,415 en 2025 a 72,833 en 2026 (+105.7%). En septiembre, el volumen aumento de 4,324 a 14,933 (+245.4% interanual). Son cifras de divulgación: no miden por sí mismas descubrimientos atribuibles a IA ni incidentes confirmados. [01]

La contribución de la IA al descubrimiento de fallas está respaldada por casos técnicos verificables: sistemas autónomos han identificado defectos reproducibles en grandes bases de código; equipos independientes hallaron vulnerabilidades críticas en PHP; y crece el conjunto de CVE con atribución pública a herramientas de IA. Sin embargo, ninguna base global homogenea permite cuantificar cuánto del crecimiento total se debe a esa tecnologia. [06–10]

En su propia cobertura, Google Threat Intelligence Group (GTIG) identificó 141 vulnerabilidades divulgadas con explotación activa entre enero y agosto de 2026, frente a 127 durante todo 2025. En esa muestra, aproximadamente el 0.23% de las divulgaciones presento explotación activa observada. El dato exige una lectura doble: aumenta la presión sobre los equipos de seguridad, pero un CVE publicado no equivale a un incidente. [03] Por otra parte, el DBIR 2026 de Verizon informa una mediana de 43 días hasta la remediación completa de vulnerabilidades KEV en sus datos de 2025; ese período no representa una tasa mundial de compromiso. [29]

HALLAZGO PRINCIPAL

La nueva asimetría no está demostrada por un único gráfico mundial. Surge cuando, dentro de una organización concreta, el flujo de hallazgos relevantes y la rapidez de su posible explotación exceden la capacidad de inventario, priorización, despliegue de parches y comprobación efectiva de cierre.

La arquitectura puede limitar el acceso a componentes vulnerables y reducir los privilegios, el movimiento lateral y la exposición de datos después de una intrusión. Su eficacia depende de configuraciones reales y pruebas negativas. Reduce consecuencias potenciales, pero no elimina el defecto ni garantiza inmunidad. [23–24]

Los marcos vigentes establecen obligaciones de naturaleza distinta. PCI DSS fija un mes para determinadas actualizaciones críticas; CISA BOD 26-04 utiliza ventanas condicionadas de tres, catorce y sesenta días, o la siguiente actualización mayor. ZTC Research propone, para exposiciones de riesgo extremo, activar protecciones comprobables en horas y verificar la corrección definitiva en días. [11, 19–22]

Alcance del estudio. Esta investigación examina fuentes públicas y principios de arquitectura. No constituye un censo mundial de CVE descubiertos por IA ni una evaluación experimental de sistemas en producción. La sección 12 detalla las limitaciones y los criterios de comprobación.

Conclusiones en una página

- **H1 — Aceleración de divulgaciones:** sostenida. Las fuentes publican series diferentes, pero coinciden en la dirección del cambio. [01–03]

- **H2 — IA como causa única del crecimiento:** no demostrada. Existen casos y muestras atribuibles a IA; la evolución agregada también refleja políticas de numeración, automatización, cobertura y ciclos de proveedores. [03, 06, 08]
- **H3 — Riesgo de saturación de remediación:** plausible y medible. Requiere evidencia operacional: exposición, hallazgos aplicables, tiempo a mitigación, capacidad de cierre y deuda pendiente.
- **H4 — Resiliencia por arquitectura:** hipótesis verificable en cada entorno. El aislamiento, la identidad por servicio, el control de egreso y la custodia segregada pueden disminuir alcanzabilidad y radio de impacto sin disminuir el conteo global de CVE. [23–24]
- **Horizonte de 18 meses:** escenario, no pronóstico. La advertencia pública de Jason D. Clinton en octubre de 2026 debe tratarse como una llamada a reforzar fundamentos, no como un calendario confirmado. [15–16]
- **Decisión directiva.** Financiación y gobierno sobre inventario completo, aislamiento, priorización contextual, rutas de emergencia, protección temporal verificada y evidencias de cierre.

Preguntas que guían la investigación

La investigación distingue tres preguntas: cuánto crecen realmente las divulgaciones; qué parte puede atribuirse a la IA; y cómo puede una organización reducir su exposición aun cuando aumente el número de hallazgos relevantes.

1. Pregunta, hipótesis y método

1.1. Hipótesis contrastables

H1. La divulgación de CVE se está acelerando. H2. La IA contribuye a descubrir vulnerabilidades, aunque su efecto agregado no está cuantificado. H3. La exposición crece cuando llegan más hallazgos aplicables de los que pueden contenerse y cerrarse con evidencia. H4. La arquitectura puede limitar rutas de acceso e impacto, siempre que se comprueben aislamiento, privilegios y capacidad de salida de cada activo.

El método combina examen crítico de fuentes públicas, comparación de series y análisis de casos técnicos. No se realizaron pruebas de penetración, auditorías de código, escaneos activos ni peritajes forenses propios. Cada afirmación se clasifica como hecho documentado, declaración de una fuente, inferencia o escenario propuesto.

1.2. Qué mide cada fuente

Fuente	Unidad observada	Uso y límite
CVETodo	CVE publicados por mes; NVD/CVE/KEV	Serie cuantitativa primaria de 2025–2026. No registra de forma universal el método de descubrimiento.
CVE Tools	CVE mensuales y altas mensuales a KEV	Corroboración con resultados cercanos, pero no idénticos. KEV registra otra categoría y otra fecha.
GTIG	Divulgación, riesgo interno y explotación observada	Datos de inteligencia con cobertura y criterios propios; no son una auditoría universal de todos los incidentes.
IBM / Unit 42 / Securin	Muestras y campañas atribuidas a IA	Prueba de capacidad técnica; tienen sesgos de selección, diversidad de definiciones y procedencia.
FIRST	Pronósticos de divulgación	Proyecciones hechas en fechas previas; comparar con observaciones posteriores exige respetar métodos y cortes.

Criterio de atribución. Se considera "descubierta mediante IA" una vulnerabilidad cuya metodología consta en una fuente primaria o dispone de atribución pública contrastable. Un CVE publicado no demuestra que la IA lo haya descubierto; el ingreso a KEV tampoco indica que su explotación comenzara ese mes.

1.3. Evidencia y niveles de confianza

Nivel	Criterio	Ejemplo
Alto	Advisory del mantenedor, datos y metodología explícitos, registros técnicos de una entidad primaria	PHP Group; GTIG para sus propios datos; CVETodo para su serie
Medio	Estudios propios de proveedor con cifras verificables pero reproducibilidad parcial	NOVA; muestra IBM; tracker Securin
Exploratorio	Opinión, tweet, pronóstico o escenario sin comprobación empírica externa	Horizonte de 18 meses; modelos de saturación

2. El cambio observado en el volumen de vulnerabilidades

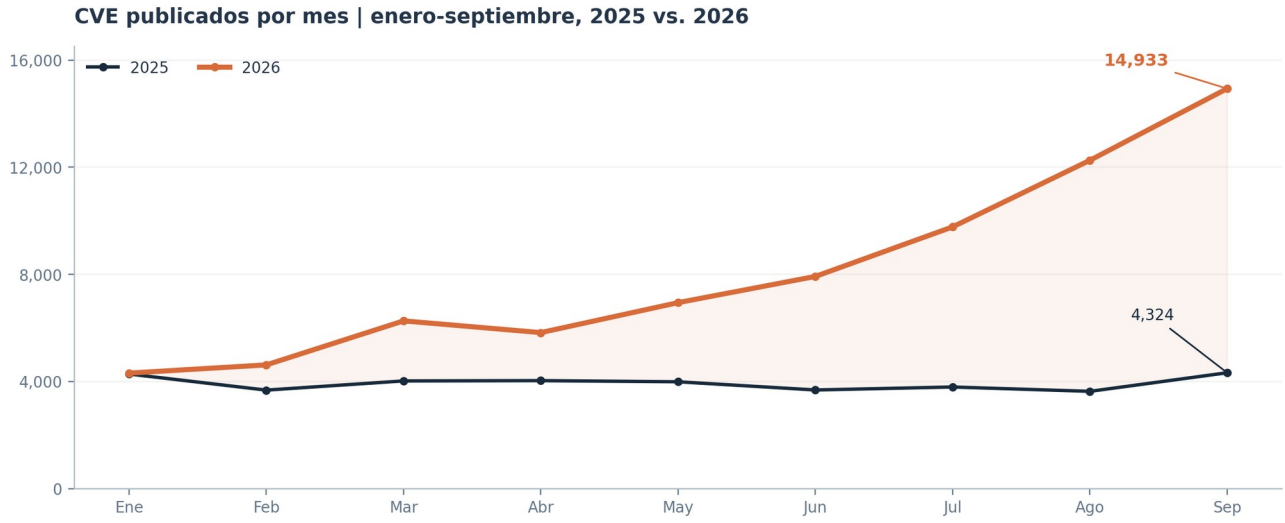


Figura 1. Serie mensual de CVE publicados, enero-septiembre de 2025 y 2026. Elaboración ZTC Research con recuentos públicos de registros CVE [01].

Entre enero y septiembre de 2026, las publicaciones mensuales de la serie pasaron de 4,311 a 14,933, un aumento del 246.4%. Frente a septiembre de 2025, el incremento fue del 245.4%. El acumulado de los primeros nueve meses alcanzó 72,833 registros, frente a 35,415 un año antes (+105.7%). [01]

La variación se produjo en meses, no solo al finalizar el año. Esto cuestiona los ciclos trimestrales aislados de gestión de vulnerabilidades. Para traducir la tendencia en riesgo, cada organización debe identificar los CVE aplicables a sus activos, las rutas de ataque viables y los servicios esenciales que podrían verse afectados.

2.1. Validación cruzada y diferencias legítimas

CVE Tools registra 15,180 CVE para septiembre de 2026 y 4,664 para septiembre de 2025; la serie principal presenta 14,933 y 4,324, respectivamente. Ambas confirman la magnitud y la dirección del cambio, aunque sus

cifras no deben combinarse. Las diferencias responden, entre otros factores posibles, a fechas de consulta, sincronización y reglas de inclusión. [01–02]

En su universo de seguimiento, GTIG observó 5,045 divulgaciones en enero de 2026 y 10,740 en agosto. La serie no es directamente equiparable con la principal. FIRST revisó en junio su pronóstico anual central de unos 59,427 a 66,000 CVE: son estimaciones emitidas entonces, no recuentos definitivos. [03–05]

2.2. Volumen de publicaciones frente a explotación observada

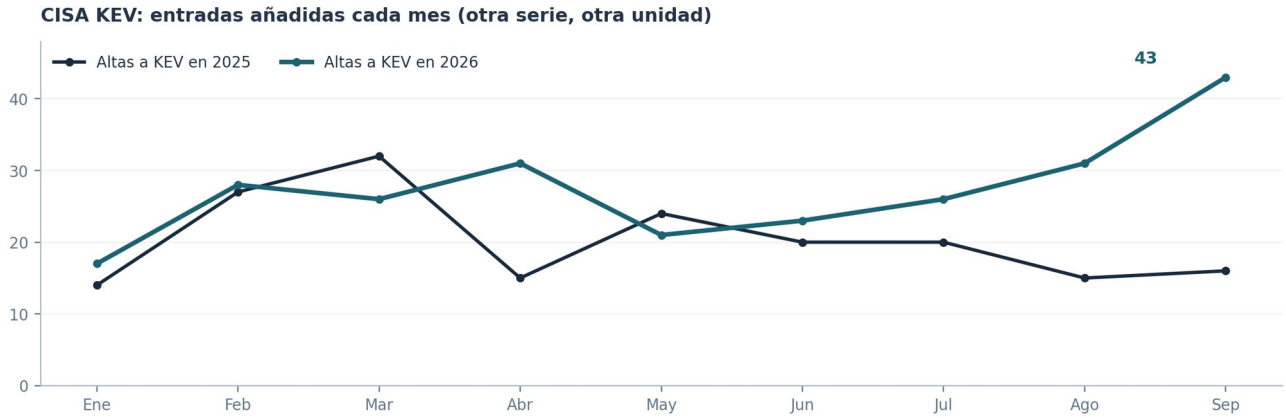


Figura 2. Altas mensuales al catálogo CISA KEV según CVE Tools: enero-septiembre de 2025 y 2026. La fecha corresponde a la inclusión en KEV, no a la publicación original del CVE. [02]

No procede dividir sin más las entradas de KEV por los CVE publicados en el mismo mes y llamar al resultado "tasa real de explotación": ambos flujos pueden corresponder a poblaciones y fechas distintas. El contraste sirve para ilustrar órdenes de magnitud y la importancia de contextualizar. [02]

GTIG, con un método de observación distinto, reportó 141 vulnerabilidades divulgadas y explotadas activamente entre enero y agosto de 2026, unas 18 al mes; en 2025 la media fue de 10.5. Para esa ventana de 2026, la fracción observada explotada fue aproximadamente 0.23% del total divulgado. Es una medición acotada por visibilidad: no demuestra ausencia de explotación en los demás registros. [03]

2.3. El desajuste medido entre riesgo y corrección

El Data Breach Investigations Report 2026 de Verizon informa que, para su conjunto de datos de 2025, únicamente el 26% de las vulnerabilidades incluidas en CISA KEV estaban completamente corregidas; la mediana para la resolución completa fue de 43 días, frente a 32 días en el periodo anterior. No son métricas globales de todas las empresas ni tiempos desde la publicación de cada CVE. Se refieren a la cobertura y unidad de análisis definidas por Verizon. [29]

Ese rezago ayuda a explicar el valor de la arquitectura: una organización que no puede corregir de inmediato necesita reducir alcanzabilidad, accesos y exposición de datos mientras corrige. No sería correcto restar directamente los 43 días de Verizon a los 3 días de una obligación federal: poblaciones, relojes y criterios difieren. La decisión razonable es medir por separado el tiempo a protección, el tiempo a cierre y las instancias que siguen expuestas. [20, 29]

3. El papel comprobable de la inteligencia artificial

3.1. Evidencia de capacidad y límites de extrapolacion

NOVA (Unit 42). El grupo informó haber examinado 3,915 proyectos abiertos y confirmado 14,090 hallazgos en dos meses, de los cuales 99.4% no estaban reportados previamente. El 39.7% obtuvo clasificación alta o crítica bajo CVSS v4.0. Son hallazgos de una campaña de investigación y validación, no 14,090 CVE emitidos ni 14,090 incidentes. El estudio procede del desarrollador del sistema; la validación externa completa de la muestra no está publicada en la misma extensión. [07]

Muestra IBM X-Force. En 223 vulnerabilidades públicamente reportadas como descubiertas con modelos de frontera, IBM observó 55.2% de severidad alta/crítica, 91% accesibles remotamente y 25.6% compatibles con ejecución de código. Las cifras describen esa muestra seleccionada; no estiman las proporciones del universo de CVE ni prueban que la IA encuentre exclusivamente fallas graves. [06]

Tracker Securin. Reúne 597 CVE relacionados con actores o sistemas de IA, pero diferencia la calidad de la atribución: 356 tienen evidencia pública explícita de descubrimiento con IA (166 autónomos y 190 con asistencia humana); otros 178 sólo están ligados a personas de un laboratorio, 14 carecen de método declarado y 49 no presentaban evidencia pública de atribución localizada. La precisión de esta separación es tan importante como el recuento bruto. [08]

Antecedente verificable. Google Project Zero y DeepMind publicaron en 2024 un hallazgo real identificado mediante Big Sleep en SQLite; en 2025 se documentó CVE-2025-6965 con apoyo del mismo programa. El avance técnico precede, por tanto, a la aceleración de registros de 2026. [13, 17]

3.2. Las causas del aumento no se reducen a una sola variable

La cantidad de CVE también puede aumentar por la ampliación del programa de numeración, cambios en autoridades emisoras, automatización de registros, mayor transparencia, revisiones de código heredado y divulgaciones masivas de proveedores. GTIG identificó cerca de 5,000 CVE asociados a descripciones del kernel de Linux entre enero y agosto de 2026, sin observar zero-days explotados activamente en ese subconjunto. [03]

Aislar la contribución causal de la IA exigiría conocer el método de descubrimiento de cada falla, eliminar duplicados y controlar los cambios de cobertura, proveedores y ciclos de divulgación. Las fuentes públicas todavía no permiten realizar esa estimación con suficiente robustez. El aumento observado es verificable; su atribución completa a la IA no lo es.

4. PHP2Shell: evidencia técnica y alcance real

El 21 de octubre de 2026, PHP anunció correcciones de CVE-2026-109618 y CVE-2026-109619. Ambas pueden permitir ejecución remota de código sin autenticación en contextos afectados. La primera involucra manejo inadecuado de memoria ante agotamiento; la segunda, estado interno conservado después de liberar memoria. El proyecto señaló mayor riesgo para mod_php y advirtió que la actualización es la protección completa disponible. [09]

OpenAI informó que Calif identificó una de las fallas con un modelo de frontera y que Wiz descubrió y notificó ambas de forma independiente. Los mantenedores de PHP encabezaron la corrección y coordinaron la respuesta con otros actores. La evidencia documenta una contribución real de herramientas de IA, con investigación humana y mantenimiento de software en el centro del proceso. [10]

Familia PHP	Versión corregida mínima	Lectura operativa
8.2	8.2.35	Planificar migración antes de fin de soporte; verificar binario desplegado.
8.3	8.3.36	Actualizar y comprobar procesos efectivos.
8.4	8.4.27	Actualizar y verificar todas las instancias y contenedores.
8.5	8.5.12	Actualizar; en Windows validar reconstrucciones posteriores cuando corresponda.

Alcance de la exposición. El uso extendido de PHP no demuestra que todos esos sitios ejecuten versiones vulnerables o puedan ser explotados. La evaluación exige confirmar versión, forma de despliegue, configuración, ruta de acceso, controles operativos y, cuando corresponda, evidencia forense. [09]

El 23 de octubre, PHP describió además un escenario específico vinculado a `session.upload_progress.enabled` y cambios dinámicos de `session.serialize_handler`, bajo el identificador CVE-2026-115560. Lo trató como una combinación inusual pendiente de corrección en el ciclo regular, con mitigación de configuración recomendada. No debe confundirse con los dos CVE principales. [09]

Implicación arquitectónica. Una falla del entorno de ejecución puede activarse antes de que la aplicación compruebe la identidad del usuario. Por ello, la autenticación de una API debe complementarse con aislamiento, privilegios mínimos, segmentación, controles de salida y actualizaciones oportunas. Cuando el fabricante advierte que las mitigaciones son incompletas, la corrección del software conserva prioridad.

5. Dieciocho meses de presión: un escenario que exige preparación

5.1. Quién es Jason D. Clinton y por qué importa su diagnóstico

Jason D. Clinton ocupa el cargo de Deputy Chief Information Security Officer en Anthropic y fue el primer responsable principal de seguridad de la empresa. Antes dirigió la seguridad de la infraestructura de Google Chrome, con experiencia en ChromeOS, Android Pay y amenazas persistentes avanzadas. Su trayectoria, documentada por RSAC Conference, aporta contexto a sus advertencias; no convierte una opinión prospectiva en una proyección cientifica. [25]

En DEF CON 34, en agosto de 2026, Clinton describió la posibilidad de un período especialmente exigente de 18 a 24 meses. El 22 de octubre situó su referencia en torno a 18 meses e insistio en reforzar los controles básicos. También ha defendido aislar sistemas críticos cuando la velocidad del descubrimiento supera la capacidad de parcheo. Son declaraciones del especialista, no resultados de un modelo predictivo. [15, 26]

Su razonamiento introduce una tensión importante: los mismos sistemas capaces de encontrar errores podrían abaratar su eliminación en el futuro. La idea de que modelos más accesibles terminarán reduciendo la introducción de defectos es una expectativa de Clinton, todavía no un resultado probado. [26]

La advertencia del 22 de octubre invita a fortalecer los fundamentos antes de aumentar la cartera de herramientas. Los 18 meses constituyen un horizonte útil para poner a prueba planes de resiliencia, no una fecha demostrada de crisis generalizada. [15, 26]

5.2. Tres mecanismos plausibles de presión

- **Entrada más rápida:** los agentes de investigación amplían cobertura de repositorios, generan hipótesis, verifican casos y reducen el costo de encontrar fallas difíciles.
- **Ventana ofensiva potencialmente más corta:** la publicación de detalles técnicos y la automatización pueden facilitar la reproducción de ataques. El efecto varia según la vulnerabilidad; aún no existe una medida universal del cambio en el tiempo hasta explotación.
- **Remediación condicionada por dependencias:** incluso una falla aplicable exige identificar propietarios, probar versiones, coordinar proveedores, gestionar cambios y comprobar el software que realmente opera en producción.

5.3. Modelo ilustrativo de saturación

Definimos B_t como el número de hallazgos aplicables pendientes al término del mes t; λ_t representa entradas nuevas verificadas y μ_t cierres verificados. La identidad operativa básica es $B_t = \max(0, B_{t-1} + \lambda_t - \mu_t)$. No incluye severidad, riesgo ni dependencias, que deben medirse por separado. No es un pronóstico de ZTC ni una cifra tomada del mercado.

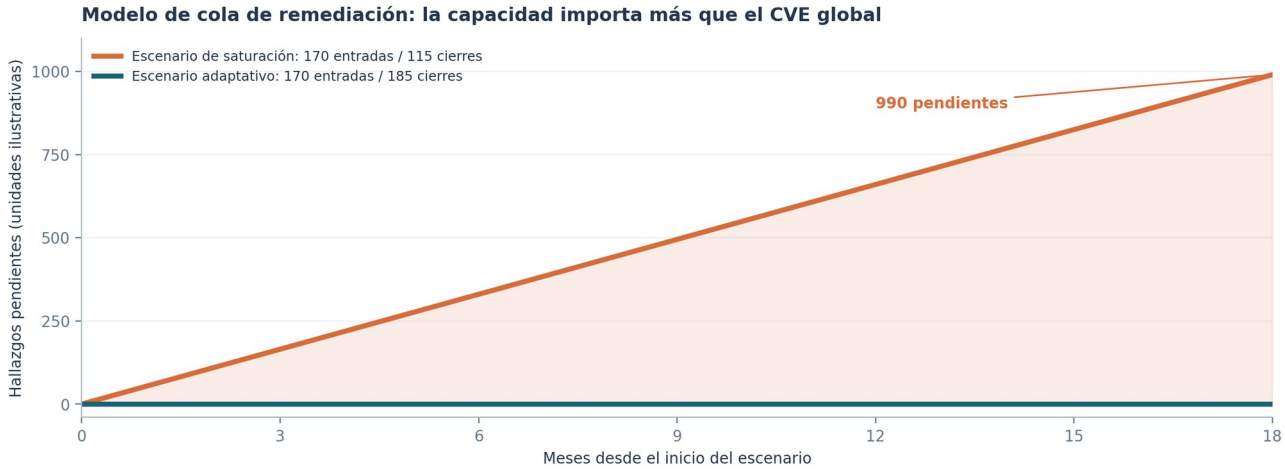


Figura 3. Simulación didáctica de 18 meses con datos hipotéticos, no observados. En saturación, la deuda crece 55 casos al mes y alcanza 990; con capacidad suficiente no acumula pendientes.

El modelo permite enfocar las decisiones: ampliar la capacidad de cierre, reducir la exposición, automatizar la comprobación y descartar hallazgos no aplicables con evidencia. Incorporar herramientas que solo multiplican alertas puede agrandar la deuda sin mejorar la protección.

5.4. Indicadores de alerta temprana

Métrica	Definición operacional	Señal relevante
Cobertura de inventario	% de activos expuestos con propietario, versión, dependencia y criticidad	Crecimiento de activos sin propietario o sin versión.
Tiempo a clasificación	Horas desde alerta hasta decisión de aplicabilidad y prioridad	Se prolonga aun para servicios expuestos.
Tiempo a protección	Horas hasta comprobar un control temporal o un parche efectivo.	Supera la ventana de actividad adversaria observada.

Métrica	Definición operacional	Señal relevante
Tiempo a cierre verificado	Horas/días hasta comprobar versión y ausencia de exposición	Cierre administrativo sin prueba técnica.
Backlog relevante	Casos aplicables abiertos en activos críticos, ponderados por edad	Entradas > cierres durante varios ciclos.
Cobertura de terceros	Servicios y librerías sujetos a un proveedor con tiempos de respuesta verificados.	Sin evidencia de actualización ni aislamiento alterno.

5.5. Casos extremos y tendencias: distinguir las métricas

Un ataque demostrado en pocos minutos acredita lo sucedido bajo condiciones concretas; no determina el tiempo medio de explotación, la velocidad habitual de escalada ni el impacto en todas las organizaciones. Para interpretarlo son indispensables el entorno, el punto de partida, la población examinada, el reloj utilizado y el criterio de éxito. [29–30]

En la muestra de casos investigados por Mandiant, la diferencia media entre explotación inicial y divulgación relevante fue de -7 días: algunas fallas ya se explotaban antes de hacerse públicas. Esa cifra tampoco representa el tiempo de compromiso de cualquier empresa. Agrupar medidas diferentes en una sola curva crearia una precisión ilusoria. [30]

6. Un modelo de defensa centrado en reducir la exposición

6.1. Secuencia de decisión por vulnerabilidad

La respuesta debe considerar aplicabilidad, acceso real, señales de explotación, impacto en el negocio y riesgos del cambio. CVSS describe severidad técnica; EPSS estima la probabilidad de explotación observada durante los siguientes 30 días conforme a su modelo; KEV reúne vulnerabilidades con explotación confirmada según CISA. Ninguno reemplaza la evaluación del entorno. [11–12]

Etapa	Evidencia mínima para darla por terminada
1. Descubrir	Inventario y versión comprobada; SBOM o evidencia equivalente si aplica.
2. Determinar alcance	Rutas de entrada, exposición pública, privilegios, dependencia y propietario del activo.
3. Decidir respuesta	Registro de KEV/EPSS/CVSS, impacto, excepciones, medida y responsable.
4. Contener	Aislamiento/limitación de acceso/salida o regla temporal con pruebas de efectividad y efecto colateral.
5. Corregir	Nueva imagen/paquete publicado, despliegue verificable, regresión funcional y rollback preparado.
6. Comprobar	Lectura de versión efectiva, test negativo reproducible, telemetría y acta de cierre.

6.2. Controles que reducen la asimetría

- **Inventario y mapeo de dependencias:** activos, contenedores, runtimes, API, componentes de terceros, cuentas de servicio, versión ejecutada y propietario; incorporar SBOM y VEX cuando existan datos confiables.
- **Aislamiento arquitectónico:** privilegio mínimo, separación de administración y datos, microsegmentación, redes privadas y salidas controladas. Comprobar que una integración secundaria no restablezca accesos indebidos.
- **Preparación de cambios:** entornos representativos, despliegues graduales, automatización, reversibilidad y pruebas funcionales. La protección debe comprobarse sobre la versión efectivamente ejecutada.
- **Telemetría de explotación:** registro de actividad útil para identificar intentos y confirmar compromisos; activar análisis forense cuando haya señales de intrusión previa al parche.
- **Inteligencia y priorización:** KEV, EPSS, superficie accesible, criticidad de procesos, datos sensibles, capacidad de movimiento lateral y dependencia operativa.
- **Responsabilidad de terceros:** plazos contractuales de respuesta, inventario de dependencias, mantenimiento de componentes sin soporte y evidencia de actualización.

6.3. Priorización en cuatro decisiones

Decisión	Disparadores indicativos	Respuesta
EMERGENCIA	Explotación confirmada y ruta accesible hacia activo crítico; o RCE remota crítica con señal confiable de actividad.	Contener de inmediato, investigar exposición histórica, parchear y verificar.
ACELERADA	Alta probabilidad y exposición real, dependencia esencial o impacto alto.	Despliegue preferente con tiempo objetivo aprobado y pruebas.
PROGRAMADA	Aplicable pero con restricciones demostrables y controles compensatorios eficaces.	Cambio planificado con vencimiento, aceptación explícita y seguimiento.
NO APLICABLE	Producto ausente o condición de ataque descartada con evidencia verificable.	Documentar descartes; reabrir si cambia arquitectura o inteligencia.

Los estados siguientes son una propuesta de gestión interna; no constituyen categorias regulatorias de PCI DSS, ISO 27001 o CISA. Cada organización debe verificar sus obligaciones legales, contractuales y normativas según el sector, la jurisdiccion y el momento.

6.4. Arquitectura resiliente: limitar el daño aunque aumenten los CVE

El número de CVE pertenece al ecosistema; la exposición efectiva pertenece a la arquitectura de cada organización. NIST SP 800-207 desplaza el foco desde confiar en la ubicación de red hacia proteger recursos con autenticación y autorización explícitas; NIST SP 800-160 v2 define resiliencia como la capacidad de anticipar, resistir, recuperarse y adaptarse aun bajo compromiso. [23–24]

Hipótesis arquitectónica (H4). Ante una misma versión vulnerable, un sistema con accesos públicos limitados, identidades independientes, privilegios mínimos, segmentación funcional y datos segregados debería ofrecer menos rutas de ataque hacia activos críticos. La hipótesis requiere pruebas negativas y telemetría; los controles deben demostrar su eficacia en la operación. [23–24]

6.4.1. Una cadena de intrusión tiene varios puntos de ruptura

Para transformar un defecto de software en un daño real suele ser necesario encadenar varias condiciones: acceder a la función vulnerable, activarla, obtener permisos útiles, alcanzar nuevos recursos y ejecutar una acción perjudicial. Los controles independientes deben interrumpir distintas etapas. Un firewall de aplicaciones o una red privada no bastan cuando la explotación nace de un proceso que ya tiene acceso autorizado. [23, 24, 27]

6.4.2. Controles comprobables y pruebas de fallo seguro

Capa	Decisión arquitectónica	Evidencia verificable
Exposición	Inventario de servicios y dependencias; cierre de rutas de administración pública.	Reconocimiento externo autorizado: sin ruta directa a consola o runtime interno.
Identidad	Identidad independiente por carga, credenciales efímeras, autorización por recurso.	Comprometer una cuenta de servicio no permite acceder a otras funciones o privilegios.
Aislamiento	Segmentación este-oeste, políticas denegación por defecto, aislamiento entre clientes.	Pruebas negativas entre servicios, redes y tenants; registrar rechazos y bypass.
Ejecución	Procesos sin privilegios, sandbox, sistemas de archivos de solo lectura y endurecimiento.	Un RCE de ensayo queda restringido al contenedor, sin acceso al host o secretos.
Datos	Minimización, tokenización y custodia externa cuando proceda; control de uso de claves.	El servicio periférico no puede leer datos originales ni ampliar su autorización.
Egreso	Salidas de red permitidas por destino, protocolo y finalidad; vigilancia de anomalías.	Intento de exfiltración hacia un destino no autorizado es rechazado.
Resiliencia	Despliegue inmutable, canary/blue-green, reversión y restauración probada.	Simulacro permite contener, actualizar y restaurar sin comprometer la integridad.

6.4.3. Escenario de prueba: RCE en un servicio público

Arquitectura con confianza amplia: una ejecución remota de código en un servicio web puede facilitar el movimiento lateral si comparte credenciales y acceso a bases de datos. Arquitectura segmentada: el mismo defecto puede comprometer el proceso inicial, mientras identidades acotadas y custodia separada limitan su alcance. Es una comparación conceptual, no el resultado de una prueba en un entorno real.

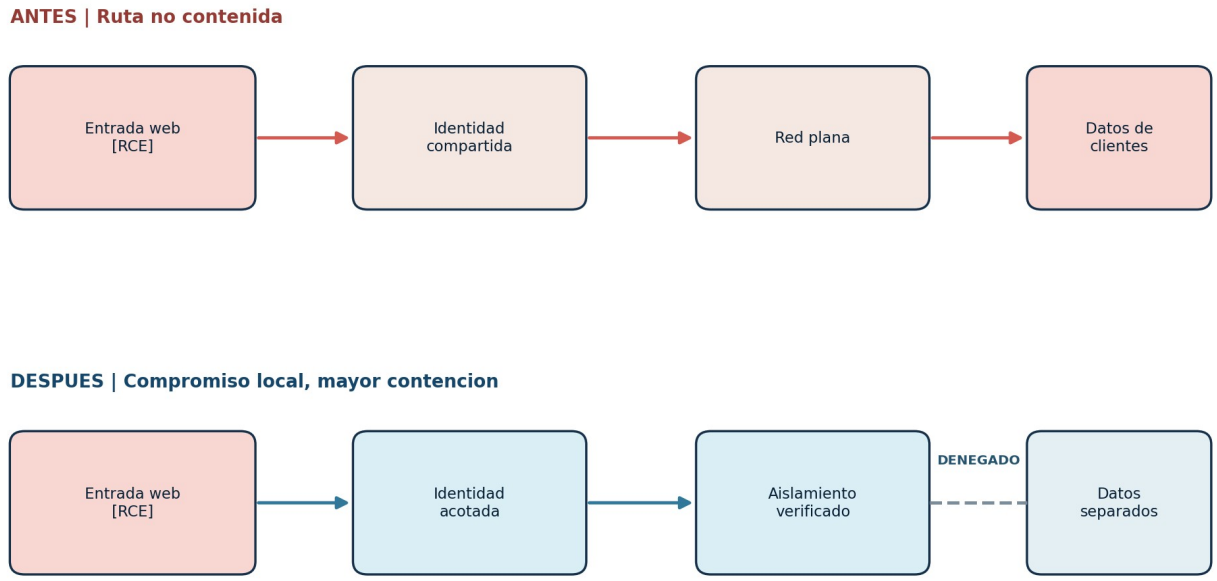


Figura 4. Dos topologías ante una misma vulnerabilidad. La arquitectura limita el alcance del compromiso; la actualización corrige el defecto.

6.4.4. Medición de eficacia con evidencia verificable

Para cada activo conviene registrar la versión afectada, su accesibilidad, los permisos disponibles tras un compromiso, las rutas laterales, los datos alcanzables, las salidas permitidas y el tiempo hasta una protección comprobable. Las comparaciones deben realizarse mediante escenarios repetibles. Una cifra universal de reducción del riesgo careceria de fundamento sin mediciones de explotabilidad e impacto.

Indicadores útiles: cobertura de pruebas negativas; activos públicos sin administración directa; rutas laterales autorizadas; servicios que comparten identidades; horas de exposición de activos esenciales, ponderadas por criticidad; y percentil 90 del tiempo hasta contención y cierre. Los CVE no aplicables deben justificarse y conservarse en el registro.

Límites del modelo. Un compromiso del sistema de identidad, la salida de datos por canales autorizados, las configuraciones incorrectas o los accesos concedidos a terceros pueden desbordar barreras bien disenadas. La criptografía y la tokenización protegen información dentro de su alcance, pero no corrigen fallas de memoria ni reemplazan los controles de acceso o el parcheo. [23–24, 27]

6.4.5. Cuatro límites de resiliencia medibles

La resiliencia arquitectónica puede describirse mediante cuatro límites sometidos a pruebas de fallo seguro. El propósito es comprobar que determinados recursos siguen inaccesibles aún cuando un componente periférico se considera comprometido. [23, 24, 31]

Límite	Diseño	Prueba negativa
Entrada	Acceso público mínimo y aislamiento administrativo	RCE en un servicio periférico no abre consola ni planos internos
Identidad	Privilegio por carga, caducidad, autorización por recurso	Token comprometido no permite elevar permisos ni cruzar clientes

Límite	Diseño	Prueba negativa
Datos	Custodia separada, minimización y control de uso	Servicio web comprometido no recupera originales fuera de su necesidad
Salida y recuperación	Egreso explícito, reversión, restauración con evidencia	Exfiltración no autorizada bloqueada; cambio reversible y verificado

Las pruebas deben cubrir credenciales autenticas con permisos insuficientes, movimiento entre servicios, aislamiento entre clientes, salidas no autorizadas y fallos de despliegue. Ni un firewall, ni un diagrama de red, ni la presencia de un token acreditan por si mismos que las barreras funcionan. [23–24, 31]

7. Implicaciones para servicios financieros y proveedores críticos

En banca, pagos y plataformas con datos personales, el riesgo de una vulnerabilidad depende de los límites de confianza que puede atravesar. Importa si compromete identidades, alcanza información sensible o aprovecha permisos persistentes de terceros. Una evaluación rigurosa reconstruye esas rutas completas en lugar de limitarse a marcar un CVE como presente.

- **Plano público:** API gateways, portales, reverse proxies, runtimes y librerías expuestas; separar explotación de aplicación de compromiso del entorno de ejecución.
- **Plano de identidad:** credenciales efímeras, autorización por transacción, MFA resistente al phishing donde aplique, permisos de servicio y segregación de tareas.
- **Plano de datos:** limitar exposición de PAN/PII y datos sensibles, reducir copias, usar tokenización cuando sea apropiado y probar restricciones en lectura y salida.
- **Plano de proveedores:** registrar bibliotecas críticas, responsabilidades de actualización, acuerdos de divulgación y límites de confianza en integraciones.
- **Plano de resiliencia:** pruebas negativas de aislamiento, rollback operativo, restauración comprobada, retest y análisis de registros de explotación antes/después del cierre.

Perspectiva de cumplimiento. La conformidad con un estándar no implica protección frente a vulnerabilidades futuras. Del mismo modo, un listado de hallazgos no sustituye la validación de controles y de su alcance. La continuidad del negocio requiere integrar la gestión de vulnerabilidades con la arquitectura y la operación.

7.1. Preguntas de gobierno que deberían tener respuesta documentada

- ¿Cuántos servicios expuestos carecen hoy de propietario, versión de runtime y dependencias identificadas?
- ¿Cuáles de los CVE nuevos afectan activos nuestros y cuáles disponen de camino de ataque y privilegios suficientes para producir impacto?
- ¿Cuál es la mediana y percentil 90 entre la alerta, el aislamiento, el despliegue efectivo y el cierre técnicamente verificado?
- ¿Qué ocurre si un proveedor esencial tiene una vulnerabilidad crítica y no entrega parche durante días?
- ¿Qué pruebas demuestran que los límites entre clientes, servicios y redes sobreviven al compromiso de un componente periférico?

8. Plazos vigentes de gestión de vulnerabilidades y parches

Los marcos vigentes establecen obligaciones y orientaciones distintas: algunos fijan plazos máximos por criticidad; otros, frecuencias de evaluación o condiciones por activo. La frecuencia de escaneo, el plazo de

corrección y una recomendación técnica son conceptos diferentes. Distinguirlos evita conclusiones regulatorias incorrectas. [11, 19–22]

8.1. Comparación de requisitos y orientaciones vigentes

Marco	Plazo o frecuencia documentada	Lectura correcta
PCI DSS 4.0.1	Req. 6.3.3: parches/actualizaciones para vulnerabilidades críticas dentro de un mes de publicación. Los demás, según riesgo definido por la entidad.	Obligación en sistemas dentro del alcance PCI; no significa esperar todo el mes. [19]
CISA BOD 26-04	Máximos condicionales de 3, 14 o 60 días naturales, o en siguiente actualización mayor; nivel de 3 días puede exigir triage forense. Cuatro factores por activo.	Mandato a agencias federales cubiertas. Fase de aplicación plena desde 7 dic 2026. [11, 20]
CIS Controls v8.1	7.2 revisión de estrategia de remediación mensual o más frecuente; 7.3/7.4 parcheo automatizado mensual o más frecuente; 7.7 remediación mensual o más frecuente.	Frecuencias recomendadas; no son plazo universal concedido a cada CVE. [21]
DORA / RTS 2024/1774	Art. 10: escaneo al menos semanal de activos que soportan funciones críticas/importantes; exige procedimientos de emergencia y plazos de instalación y escalamiento definidos.	No fija en ese artículo un SLA universal de 3, 7 o 30 días para todas las fallas. [22]
NIST SP 800-40 Rev. 4	Parches como mantenimiento preventivo: identificar, priorizar, adquirir, instalar y verificar; diseño de estrategia empresarial de riesgo.	Guía técnica, sin fecha universal por CVE. [28]
ISO/IEC 27001:2022 + ISO/IEC 27002:2022	Anexo A, control 8.8: obtener información de vulnerabilidades técnicas, evaluar la exposición y tomar medidas apropiadas. La norma no fija un plazo universal de parcheo en días.	Exige gestión del riesgo y tratamiento demostrable en el SGSI; la organización debe justificar prioridades y tiempos. ISO 27002 aporta orientación de implementación. [33–34]

Cómputo de plazos. En PCI DSS, el mes previsto para actualizaciones críticas se cuenta desde la publicación del parche aplicable. En CISA BOD 26-04, el plazo surge de las condiciones previstas por la directiva, incluida la incorporación a KEV o la detección y reporte de la falla en activos federales. Un cambio en la exposición o en la información de amenazas puede exigir actuar antes. [19, 20]

ISO/IEC 27001 y 27002. La gestión de vulnerabilidades técnicas (A.8.8) se relaciona con inteligencia de amenazas (A.5.7), configuración (A.8.9), gestión de cambios (A.8.32) y preparación de las TIC para continuidad (A.5.30). ISO/IEC 27001 requiere tratar el riesgo, pero no fija un número universal de días para cada parche. Los plazos internos deben justificarse según exposición, criticidad y capacidad de respuesta. [33–34]

8.2. Como deben evolucionar los marcos ante el descubrimiento asistido por IA

Primera propuesta: vincular los plazos a la explotación confirmada, la automatización posible, la exposición pública, los permisos alcanzables y el impacto sobre datos y servicios esenciales, además de la severidad. Cada criterio debe poder comprobarse y auditarse.

Segunda propuesta: distinguir dos tiempos. El primero transcurre desde la alerta confiable hasta la protección efectiva; el segundo termina con el parche o la eliminación definitiva verificados. Una medida temporal mantiene abierto el caso hasta completar la corrección.

Tercera propuesta: comprobar periódicamente el aislamiento entre servicios y clientes, la seguridad de las identidades y las restricciones de salida. Cuarta: establecer rutas de cambio urgente, mecanismos de reversibilidad y obligaciones de respuesta verificables para proveedores críticos. [23–24, 27]

ZTC Research invita a los organismos normativos a revisar sus criterios y a las organizaciones a actualizar las políticas que los aplican. El riesgo debe evaluarse por exposición, explotación confirmada, capacidad de automatización y consecuencias sobre identidades y datos. CISA BOD 26-04 ilustra un enfoque contextual; este estudio propone, además, contención verificada en horas, pruebas de aislamiento y reporte del riesgo residual. Se trata de una recomendación técnica, no de un mandato vigente. [20, 32]

9. Propuesta ZTC Research: protección en horas, corrección verificable en días

Los objetivos siguientes son una propuesta voluntaria para activos relevantes. Deben incorporarse a políticas internas aprobadas y materializarse en controles operativos, probados y medibles, con responsables identificados, evidencias y excepciones temporales. No reemplazan los requisitos legales, contractuales o normativos aplicables; siempre prevalece el plazo obligatorio más exigente. Contener una ruta de ataque no equivale a eliminar la vulnerabilidad.

9.1. Umbrales propuestos por riesgo contextual

Nivel	Disparador contextual	Objetivo interno propuesto
A / Emergencia	Explotación activa aplicable; KEV alcanzable; RCE remota automatizable con alto impacto y ruta externa.	Contención ≤ 4 h; evaluación forense inmediata; cierre objetivo ≤ 72 h si existe corrección segura.
B / Alta exposición	Ruta externa demostrada y explotación plausible / prueba funcional / servicio crítico.	Protección ≤ 24 h; cierre objetivo ≤ 7 días.
C / Expuesta sin explotación	Alta/critica aplicable, acceso público, defensas verificadas y sin actividad confirmada.	Protección ≤ 72 h; corrección objetivo ≤ 14 días.
D / Interna acotada	No expuesta ni alcanzable desde zonas no confiables, con verificación y alcance limitado.	Revisar en ≤ 7 días; corregir objetivo ≤ 30 días.
E / Baja relevancia	No crítica, no explotada, inventario y ruta documentados, sin señales de cambio.	Ventana según análisis de riesgo; revisión continua, no cierre por silencio.

En el nivel A, si aún no existe parche o su instalación inmediata compromete la continuidad, debe evaluarse retirar el servicio de exposición, restringir su uso o aislarlo temporalmente con prueba de eficacia. La excepción requiere escalamiento inmediato, medidas de continuidad y preservacion de evidencia forense. El caso permanecera abierto hasta su remediación definitiva. [20, 23–24]

9.2. Métricas que pueden auditarse

Tiempo a protección (MTP): horas desde la alerta confiable hasta comprobar que un control interrumpe la ruta de ataque. Tiempo a cierre (MTC): horas o días hasta verificar la corrección definitiva. Exposición de activos esenciales (EAEC): horas acumuladas de activos vulnerables con rutas de ataque relevantes, ponderadas por criticidad. Cobertura de barreras: proporción de pruebas negativas aprobadas en límites de identidad, red, datos, clientes y salida.

Las métricas deben desglosarse por gravedad y percentiles, indicando poblaciones y exclusiones. El total de CVE no representa por si solo el esfuerzo operativo: una falla presente en cientos de sistemas puede producir mayor exposición que cientos de CVE sin aplicabilidad en el inventario.

9.3. Límites de costo y continuidad

Reducir el tiempo de exposición requiere invertir también en pruebas automatizadas, despliegue seguro, capacidad de respuesta, redundancia y reversibilidad. Una política de actualización acelerada sin validaciones puede aumentar interrupciones. El análisis económico debe comparar horas de exposición, costos de control, frecuencia de intervencion y resultados de continuidad, sin contabilizar beneficios hipotéticos como ahorros realizados.

9.4. Prioridad de inversión y rendición de cuentas

La prioridad de inversión comienza por barreras arquitectonicas que limiten el impacto, continua con actualizaciones y pruebas automatizadas y se consolida mediante observabilidad y respuesta sostenibles. Cada iniciativa necesita responsable, costo total, horizonte de implantación, riesgo residual y prueba de eficacia. La reducción de incidentes esperados debe presentarse como beneficio potencial, no como ahorro financiero ya obtenido.

Antes de exigir ventanas de despliegue más breves conviene ensayar aislamiento, reversión, restauración y comprobación de integridad. Una respuesta más rápida debe mejorar la resiliencia sin introducir interrupciones desproporcionadas.

9.5. Tres relojes y un criterio de cierre

Reloj 1, reconocimiento: desde la alerta confiable hasta la decisión registrada de aplicabilidad. Reloj 2, protección: hasta demostrar que el control interrumpe la ruta de ataque. Reloj 3, cierre: hasta desplegar y validar la corrección en los activos afectados, con pruebas funcionales y de seguridad.

Cuando una vulnerabilidad se explota antes de divulgarse, el reloj interno no refleja toda la exposición histórica. Se requiere una revisión retrospectiva de eventos, signos de intrusiones previas y períodos de conservación de registros. Un cierre administrativo sin pruebas técnicas puede ocultar compromiso residual. [29–30]

10. Interpretaciones alternativas y alcance de las conclusiones

Primera objeción: más CVE no implica necesariamente peor software. Una auditoría más amplia puede revelar fallas antiguas sin aumentar la tasa de defectos nuevos. La mayor visibilidad puede favorecer la defensa, aunque también presiona la capacidad de corrección. [03–05]

Segunda objeción: la IA también fortalece la defensa. La coordinación en PHP y el trabajo de Big Sleep muestran que encontrar y comunicar fallas antes de su explotación puede dar tiempo adicional al defensor. El beneficio depende del despliegue efectivo de las correcciones. [10, 13, 17]

Tercera objeción: algunos recuentos pueden sobrestimar el problema. Es necesario distinguir hallazgo técnico, CVE, instancia afectada, prueba de explotación, intento de ataque y compromiso confirmado. Los resultados de NOVA y las muestras asistidas por IA no deben sumarse al total global como si fueran CVE nuevos independientes. [06–08]

Cuarta objeción: los 18 meses son una estimación subjetiva. Se trata de un horizonte útil para ejercicios de resiliencia, no de una fecha demostrada de crisis o recuperación. La investigación evita convertir ese juicio en un pronóstico estadístico. [15–16]

Quinta objeción: la explotación observada representa una fracción reducida. Esa fracción puede concentrar daños severos cuando afecta plataformas muy extendidas. En cambio, numerosos CVE no aplicables pueden carecer de efecto operativo para una empresa. El riesgo surge de la combinación de criticidad, exposición y velocidad de respuesta. [03, 18]

11. Plan de actuacion: primeras 72 horas, 14 días y mejora continua

Fase	Objetivo	Evidencia de resultado
0–4 horas	Nombrar equipo de guardia; localizar activos críticos de acceso público y activar contención de emergencia.	Decisión registrada, responsables y ruta de aislamiento comprobable.
4–24 horas	Validar CVE aplicables y vías de ataque; aplicar protecciones temporales y vigilar actividad.	Registro de alcance y pruebas de bloqueo / compromisos previos.
24–72 horas	Aplicar correcciones seguras, pruebas posteriores, despliegue gradual y plan de reversión; comprobar accesos laterales.	Versiones efectivas, pruebas negativas y cierre de casos máximos.
Días 4–7	Generalizar política de privilegio mínimo, inventario de dependencias y restricciones de egreso en sistemas esenciales.	Matriz de rutas y propietarios; pruebas de aislamiento entre servicios/tenants.
Días 8–14	Ensayar explotación controlada y recuperación; formalizar objetivos por riesgo.	Cobertura de verificación, p90 de protección, excepciones con vencimiento.
Días 15–30	Automatizar cadena de actualización y pruebas; contratos de respuesta con terceros.	Balance entradas/cierres y evidencia de reducción de exposición.
Continuo: 1–18 meses	Sostener rediseño por límites de confianza; medir tendencias y validar independientemente.	Escenarios de compromiso contenidos y evidencia recurrente de resiliencia.

Gobernanza de respuesta breve. Durante los primeros catorce días, las alertas sobre activos esenciales deben revisarse a diario. Después, la vigilancia continua se complementa con informes semanales de excepciones y tendencias. La dirección debe conocer, como mínimo, cuatro indicadores: activos críticos sin cobertura, hallazgos de alto impacto aún expuestos, percentil 90 del tiempo de protección y relacion entre nuevos casos y cierres. La inversión en IA defensiva se justifica por resultados medibles, no por el volumen de alertas generadas.

12. Alcance y limitaciones de la investigación

Este estudio es documental, comparativo y arquitectónico. Diferencia observaciones públicas, inferencias razonadas y medidas propuestas que requieren validación en entornos concretos. Definir estos límites permite reproducir el análisis y establecer que evidencias confirmarian o refutarian cada hipótesis.

12.1. Límites metodológicos y medidas de control

Dimensión	Limitación	Control de validez
Series CVE y KEV	Los repositorios difieren por ingesta, fechas, cambios retroactivos y criterios de inclusión. Publicación no equivale a descubrimiento.	Se comparan tendencias dentro de cada fuente, con fecha de corte; se evita construir un total mundial único o una tasa KEV/CVE artificial. [01–05]
Atribución a IA	No existe un campo universal verificable que indique cómo se descubrió cada CVE. Las muestras divulgadas por proveedores pueden tener sesgos de selección.	Se distinguen casos documentados de series agregadas. No se estima un porcentaje causal mundial atribuible a IA. [06–10]
Explotación real	CISA KEV y la telemetría de empresas de seguridad captan universos distintos; ausencia de observación no demuestra ausencia de ataques.	Se separan explotación observada, exploit disponible, intentos y compromiso. Se declara el denominador de cada afirmación. [02–03]
Plazos de corrección	Las normas expresan obligaciones o frecuencias para poblaciones distintas; los tiempos de campo usan relojes, activos y muestras no siempre comparables.	Se diferencian reconocimiento, protección y cierre; las medianas observadas no se confunden con un plazo regulatorio universal. [19–22, 29]
Eficacia arquitectónica	El modelo de límites de confianza se apoya en principios técnicos; no se ejecutaron pruebas activas ni comparaciones controladas en infraestructuras de clientes.	Se presentan mecanismos y pruebas negativas propuestas, sin atribuir porcentajes de reducción de incidentes ni inmunidad. [23–24, 31]
Alcance sectorial y económico	Las fuentes internacionales no constituyen una muestra representativa de México o Latinoamérica. No se dispone de costos e impactos medidos por entidad.	Se evita extrapolar tasas sectoriales y retornos financieros; se proponen métricas de exposición, continuidad y costo para validación local.
Horizonte y vigencia	La advertencia de dieciocho meses no es una predicción estadística. Los registros y la dinámica de capacidades de IA pueden cambiar tras el corte de 2026.	El horizonte se usa como escenario de estrés, no como fecha cierta; las cifras deberán reconfirmarse en futuras ediciones. [15–16]

12.2. Qué permiten afirmar estos resultados

La evidencia permite sostener tres conclusiones: las divulgaciones aumentaron en las series examinadas; existen hallazgos relevantes descubiertos con apoyo de IA; y una arquitectura con límites de confianza puede reducir rutas de ataque si su funcionamiento se comprueba. La magnitud de este efecto depende de los activos, la configuración y la persistencia de los controles.

Los datos disponibles no permiten atribuir a la IA todo el crecimiento mundial de CVE, considerar explotable cada registro, trasladar un mismo plazo a todas las organizaciones ni asignar porcentajes universales de reducción del riesgo arquitectónico. Los objetivos de respuesta son propuestas que deben probarse; las simulaciones no describen resultados obtenidos en sistemas reales.

12.3. Evidencia adicional necesaria para avanzar

- **Reproducibilidad temporal:** publicar capturas fechadas y consultas de cada serie, reglas de deduplicación y registro de correcciones posteriores.
- **Atribución independiente:** documentar para cada hallazgo el papel concreto de la IA, la revisión humana, la prueba de concepto y el momento de divulgación.
- **Validación arquitectónica:** comparar, en laboratorio autorizado y con datos sintéticos, el mismo componente vulnerable bajo diseños diferentes; medir alcanzabilidad, escalada, acceso cruzado entre clientes y egreso.
- **Seguimiento longitudinal:** registrar por activo los tres relojes (reconocimiento, protección y cierre), percentiles, fallas de contención, efectos sobre disponibilidad y costos verificables.

12.4. Criterio de revisión y posible refutación

La hipótesis arquitectónica debe revisarse si una prueba controlada muestra que es posible atravesar límites declarados, consultar datos fuera del permiso concedido o mantener salidas no autorizadas tras comprometer un componente. Los tiempos propuestos también deben ajustarse si la contención no puede verificarse o si sus efectos operativos superan el riesgo evitado. Estas condiciones de refutación forman parte del método.

12.5. Estado de las hipótesis y próximo criterio de comprobación

Hipótesis	Estado de la evidencia	Prueba decisiva pendiente
H1 · CVE	Crecimiento observado en las series examinadas.	Reconstruir capturas de los repositorios e identificar efectos de cambios de registro. [01–05]
H2 · IA	Casos reales demostrados; causalidad global no cuantificada.	Construir una cohorte con atribución del método de descubrimiento y contraste independiente. [06–10]
H3 · Respuesta	Desajuste plausible; magnitud depende del entorno.	Medir nuevas entradas, cierre definitivo, tiempo a protección y deuda por activo esencial. [29]
H4 · Arquitectura	Mecanismo técnicamente sustentado; efecto empírico no medido aquí.	Enfrentar una misma falla en dos diseños y documentar rutas bloqueadas y rutas que persisten. [23–24,31]

13. Conclusiones

Primera. El crecimiento de divulgaciones de vulnerabilidades durante 2026 es verificable y exige reconsiderar ciclos de gestión mensuales o trimestrales. Los recuentos varian según fuente y fecha; la tendencia de aceleración es consistente. [01–03]

Segunda. La IA ya ha contribuido al descubrimiento de defectos relevantes, incluidos fallos antiguos en software de uso extendido. PHP2Shell y NOVA ilustran esa capacidad, aunque sus resultados no representan todos los CVE divulgados. [07, 09–10]

Tercera. El crecimiento de publicaciones no demuestra por si mismo que la IA sea su única causa ni que exista una evolución proporcional de ataques exitosos. Divulgación, severidad, probabilidad de explotación y daño requieren mediciones diferenciadas. [03, 06, 08]

Cuarta. La arquitectura constituye una defensa estructural: puede reducir accesibilidad, privilegios y alcance del compromiso aun cuando subsista una falla. Su eficacia debe comprobarse sobre los límites de confianza, pues una misma versión vulnerable puede representar riesgos diferentes según el diseño. [23–24]

Quinta. Organismos normativos y organizaciones deben revisar y actualizar criterios y políticas para responder a la velocidad de las amenazas asistidas por IA. La mejora exige controles implementados, probados y medidos; el indicador principal es el tiempo hasta protección efectiva, no la cantidad de parches informados.

Sexta. Los siguientes 18 meses constituyen un escenario de presión plausible, no una fecha cierta de crisis. Ante explotación activa y exposición crítica, la contención debe funcionar en horas y el cierre definitivo verificarse lo antes posible. Los plazos de PCI DSS y CISA tienen alcances y condiciones propias; ISO/IEC 27001 exige tratar el riesgo sin fijar un número universal de días para parchar. [11, 19–20, 33–34]

CONCLUSIÓN CENTRAL · ZTC RESEARCH

La divulgación de vulnerabilidades se ha acelerado y los sistemas de IA amplían la capacidad de identificar fallas relevantes. La medida de resiliencia más útil es la exposición comprobada de cada organización frente a su capacidad efectiva de detectar, aislar, corregir y verificar rutas de ataque hacia activos críticos.

Implicación central. La resiliencia se demuestra cuando la organización identifica sus rutas de ataque, contiene los accesos críticos, limita privilegios, corrige los defectos y acredita la recuperación con evidencia.

Anexo A. Datos utilizados en las figuras

Mes	CVE 2025	CVE 2026	Var. anual	KEV 2025	KEV 2026
Ene	4,276	4,311	+0.8%	14	17
Feb	3,676	4,617	+25.6%	27	28
Mar	4,019	6,261	+55.8%	32	26
Abr	4,030	5,823	+44.5%	15	31
May	3,987	6,942	+74.1%	24	21
Jun	3,681	7,920	+115.2%	20	23
Jul	3,792	9,770	+157.6%	20	26
Ago	3,630	12,256	+237.6%	15	31
Sep	4,324	14,933	+245.4%	16	43

Las series de CVE y KEV provienen de repositorios con fechas y criterios distintos [01–02]. Se incluyen para comparar tendencias, sin construir una tasa artificial de explotación a partir de ambas. Los recuentos pueden variar por actualizaciones retroactivas o diferencias de sincronización; la fecha de publicación de un CVE y su ingreso a KEV no tienen por que coincidir.

Anexo B. Glosario funcional

Término	Definición de trabajo
CVE	Identificador de una vulnerabilidad pública; no es un incidente ni prueba de que un activo esté expuesto.
CWE	Categoría de debilidad de software; agrupa clases de defectos.
KEV	Catálogo de vulnerabilidades cuya explotación en el mundo real está confirmada según los criterios de CISA.
EPSS	Estimación estadística de probabilidad de explotación observada en los siguientes 30 días; no calcula el impacto empresarial.
CVD	Divulgación coordinada para dar espacio a correcciones y medidas de protección antes de exposición masiva.
RCE	Ejecución remota de código; efecto potencial cuyo contexto, condiciones y privilegios deben investigarse.
SBOM / VEX	Inventario de componentes de software / declaración de afectación o no afectación con evidencia.
Cierre verificado	Comprobación técnica de que una medida reduce o elimina la exposición en el sistema realmente desplegado.

Fuentes y trazabilidad

Corte documental: 31 de octubre de 2026. Las referencias permiten rastrear las afirmaciones publicadas. El documento sintetiza fuentes y presenta un análisis independiente; no incluye resultados de pruebas activas propias.

[01] CVETodo — CVE Statistics 2026; serie mensual 2025-2026. Consulta: 31 oct 2026. Abrir fuente

[02] CVE Tools — Monthly CVE trends; serie de snapshots y adiciones a CISA KEV. Consulta: 31 oct 2026. Abrir fuente

[03] Google Threat Intelligence Group — Vulnerability Discovery and Exploitation Trends in the AI Era, 30 sep 2026. Abrir fuente

[04] FIRST — 2026 Vulnerability Forecast, 11 feb 2026. Abrir fuente

[05] FIRST — Mid-Year Vulnerability Forecast, 15 jun 2026. Abrir fuente

[06] IBM X-Force — AI can find serious vulnerabilities at unprecedented speed. Can defenders keep up?, oct 2026. Abrir fuente

[07] Palo Alto Networks Unit 42 — The Frontier AI Vulnerability Burst: Industrializing Autonomous Zero-Day Discovery, 4 ago 2026. Abrir fuente

[08] Securin — AI-Discovered CVE Tracker; base atribuida con niveles de evidencia, 27 oct 2026. Abrir fuente

[09] PHP Group — Avisos de seguridad y versiones corregidas de PHP2Shell, 21 y 23 oct 2026. Abrir fuente

[10] OpenAI — Patch the Planet: global response coordination as agents uncover critical vulnerabilities, 21 oct 2026. Abrir fuente

[11] CISA — BOD 26-04: Prioritizing Security Updates Based on Risk, jun 2026. Abrir fuente

[12] FIRST — EPSS Frequently Asked Questions: alcance, usos y limitaciones. Abrir fuente

[13] Google Project Zero — From Naptime to Big Sleep, 1 nov 2024. Abrir fuente

[14] Jon Hernández — La IA está rompiendo la seguridad de Internet, 26 oct 2026. Abrir fuente

[15] Jason D. Clinton — Publicación en X, 22 oct 2026; contenido corroborado mediante reproducción de terceros. Abrir fuente

[16] Mantic — Black October and the future of cybersecurity, 27 oct 2026. Abrir fuente

[17] Google — Big Sleep y descubrimiento de CVE-2025-6965, 2025. Abrir fuente

[18] CISA — Vulnerability Review: controles básicos, priorización y explotación, 2026. Abrir fuente

[19] **PCI Security Standards Council —** FAQ 1597, distinción de 6.3.3 y 11.3; actualizada mayo 2025 Abrir fuente

[20] **CISA —** BOD 26-04, Directiva oficial, 10 jun 2026; plazos condicionales Abrir fuente

[21] **Center for Internet Security —** CIS Controls v8.1, Control 7, frecuencias de proceso y remediación Abrir fuente

[22] **Comisión Europea —** Reglamento delegado (UE) 2024/1774, artículo 10, DORA Abrir fuente

[23] **NIST —** SP 800-207, Zero Trust Architecture, ed. 2020 Abrir fuente

[24] **NIST —** SP 800-160 Vol. 2 Rev. 1, cyber resiliency engineering, dic 2021 Abrir fuente

[25] **RSAC Conference —** Jason Clinton, perfil biográfico y trayectoria verificada Abrir fuente

[26] **Clarín / DEF CON 34 —** Declaración de Jason Clinton sobre periodo de 18 a 24 meses, ago 2026 Abrir fuente

[27] **Microsoft Learn —** Zero Trust, límites de confianza y defensa en profundidad; guía de arquitectura Abrir fuente

[28] **NIST —** SP 800-40 Rev. 4, planificación de gestión de parches, abr 2022 Abrir fuente

[29] **Verizon —** 2026 Data Breach Investigations Report; cobertura KEV y mediana de cierre (datos de 2025). Abrir fuente

[30] **Google Cloud / Mandiant —** M-Trends 2026; tiempo a explotación estimado de -7 días en la muestra. Abrir fuente

[31] **NIST —** SP 800-207A, identidad y controles de acceso por servicio en cloud-native. Abrir fuente

[32] **FedRAMP —** Respuesta institucional a CISA BOD 26-04 y calendario de implementación. Abrir fuente

[33] ISO — ISO/IEC 27001:2022, requisitos del sistema de gestión de seguridad de la información. Abrir fuente

[34] ISO — ISO/IEC 27002:2022, guía de controles de seguridad de la información. Abrir fuente