Prospectiva de amenazas cibernéticas 2027

Ventana de AREM®[1]

Resumen

Este estudio aplica la Ventana de AREM® para anticipar amenazas cibernéticas con horizonte 2027, integrando la teoría de señales débiles, el concepto de sorpresas predecibles y el proceso de anticipación en cuatro pasos. Se valoran 25 amenazas latentes y emergentes mediante cinco criterios (anticipación, relevancia, ambigüedad, credibilidad e impacto), las cuales se basan en casos concretos con evidencia verificable y se someten a dos matrices complementarias: una de clasificación y otra de construcción de escenarios. El hallazgo central es que los ciberataques autónomos con inteligencia artificial dejaron de ser una señal emergente especulativa y se materializaron en 2026, lo que obliga a reclasificarlos como amenaza activa. El análisis distingue señales confirmadas por evidencia, señales aún no materializadas y posibilidades presentes solo en los escenarios proyectados. El documento concluye que el mayor riesgo a 2027 no es técnico sino organizacional: la inacción frente a señales ya visibles, mecanismo por el que nacen las sorpresas predecibles.

Palabras clave: *prospectiva; señales débiles; Ventana de AREM®; ciberseguridad; inteligencia artificial agéntica; sorpresas predecibles*

Jeimy J. Cano M.

1. Introducción

El propósito de este estudio es anticipar, no predecir, las amenazas cibernéticas relevantes para las organizaciones con horizonte 2027, traduciendo señales tempranas en decisiones de inversión y control. Su alcance abarca 25 amenazas (13 latentes y 12 emergentes) seleccionadas por su relevancia estratégica y su respaldo documental, dejando fuera del foco las amenazas conocidas (gestionadas por el ciclo tradicional de riesgo) y las focales o sectoriales. El horizonte es cercano (aproximadamente 24 meses), lo que obliga a recalibrar dos criterios: la credibilidad tiende a subir (varias amenazas antes especulativas ya presentan incidentes documentados) y la anticipación tiende a bajar. En coherencia con lo anterior, el documento distingue hechos (incidentes documentados, estándares, cifras atribuibles), inferencias (proyecciones sustentadas) y juicio experto (las valoraciones de priorización), que es estructurado y trazable, no una medición objetiva. Cada organización que use los resultados de este ejercicio académico de prospectiva, deberá situarlo en la dinámica particular de su entorno y objetivos estratégicos para concretar sus reflexiones y decisiones sobre la ciberseguridad en 2027.

2. Fundamentos teóricos

La Ventana de AREM® (Cano, 2017) reorganiza la gestión de riesgos según lo que conoce la organización y lo que conoce el entorno, generando cuatro cuadrantes: conocidos (ambos los conocen; ciclo tradicional), latentes (el entorno los conoce y la organización aún no tiene controles; foco en mitigación), focales (la organización los conoce y el entorno no; sectoriales) y emergentes (ninguno los domina; foco en contención y crisis). Su valor diferencial es incorporar las posibilidades, no solo las probabilidades, como insumo de análisis.

[1] **Ventana de AREM – Es una marca registrada**

Una señal débil es un indicio primario, fragmentario y de baja credibilidad inicial de un cambio incipiente (Lesca & Lesca, 2011; Ansoff et al., 2019); ignorarla y que el evento de alto impacto ocurra configura una sorpresa predecible: un desastre evidente en retrospectiva que pudo prevenirse (Bazerman & Watkins, 2004). El proceso de anticipación (Cano, 2026) opera en cuatro pasos: revelar (operar sensores de inteligencia de amenazas), ver (construir escenarios), darse cuenta (identificar impactos y puntos ciegos frente al apetito de riesgo) y actuar (movilizar inversión con convicción); la ausencia de este último paso es la que genera las sorpresas predecibles.

3. Metodología aplicada

Cada amenaza se puntúa (escala 1 a 3) con cinco criterios: anticipación (A, cuán temprano es el indicio), relevancia (R, afectación a la misión), ambigüedad (Amb, claridad de su significado), credibilidad (C, evidencia técnica, prueba de concepto o incidentes) e impacto potencial (IP, gravedad a escala). La suma ubica cada amenaza en una banda: 13 a 15, *atención inmediata* (actuar en menor de 3 meses); 9 a 12, *monitoreo cercano* (actuar entre 3 a 6 meses); 5 a 8, *vigilancia* (actuar entre 6 y 12 meses). Ante empates se prioriza por celeridad (velocidad de materialización) y vulnerabilidad organizacional (Cano, 2026). Las 25 amenazas se basan en al menos un caso o fuente original con enlace verificable (Sección 4). La construcción de escenarios sigue la técnica 2x2 de la Global Business Network (Schwartz, 1996): dos incertidumbres críticas de alto impacto y alta incertidumbre como ejes, que generan cuatro futuros contrastados.

Nota sobre el límite del modelo. *Cuando una señal se confirma con incidentes, su anticipación baja pero su urgencia sube; esto es, no solo recalcular la puntuación, sino reclasificar la amenaza de cuadrante. Es el caso de los ataques autónomos con IA, que en este estudio pasan de emergentes a activos.*

4. Resultados

4.1 Valoración de las 25 amenazas

Tabla 1

Valoración de los riesgos latentes

ID	Amenaza	A	R	Amb	C	IP	Σ	Banda
L5	Concentración de proveedores tecnológicos	2	3	2	3	3	**13**	Atención inmediata
L12	Espionaje basado en identidad	2	3	2	3	3	**13**	Atención inmediata
L2	Cadena de suministro de software	1	3	2	3	3	**12**	Monitoreo cercano
L3	Fraude por deepfake e ingeniería social con IA	2	3	2	3	2	**12**	Monitoreo cercano
L4	Tecnología operativa y sistemas legados sin parchear	1	3	2	3	3	**12**	Monitoreo cercano
L7	Cosecha de datos para descifrado futuro (HNDL)	2	3	2	2	3	**12**	Monitoreo cercano
L8	Infraestructura crítica atacada por actores estatales	1	3	2	3	3	**12**	Monitoreo cercano
L11	Ataques a interfaces (API) y superficie IoT	2	3	2	3	2	**12**	Monitoreo cercano
L1	Ransomware y extorsión multinivel	1	3	1	3	3	**11**	Monitoreo cercano
L6	Robo de identidad e identidades de máquina	2	3	2	2	2	**11**	Monitoreo cercano
L13	Vigilancia autoritaria y pérdida de privacidad	2	2	3	2	2	**11**	Monitoreo cercano
L9	Desinformación y operaciones de influencia con IA	1	2	2	3	2	**10**	Monitoreo cercano
L10	Escasez crónica de talento	1	3	1	3	2	**10**	Monitoreo cercano

Nota. A = anticipación; R = relevancia; Amb = ambigüedad; C = credibilidad; IP = impacto potencial; Σ = suma. Escala 1 a 3. Elaboración propia basada en fuentes disponibles a corte de agosto de 2026.

Tabla 2

Valoración de los riesgos emergentes

ID	Amenaza	A	R	Amb	C	IP	Σ	Banda
E1	Ciberataques autónomos con IA agéntica (reclasificada a activo)	2	3	2	3	3	**13**	Atención inmediata
E2	Crimen-como-servicio autónomo	3	3	3	1	3	**13**	Atención inmediata
E3	Pérdida de control de IA en cascada	3	2	3	1	3	**12**	Monitoreo cercano
E4	Envenenamiento de datos y manipulación de modelos	2	3	2	3	2	**12**	Monitoreo cercano
E5	Gusanos autorreplicantes con IA	3	2	2	2	3	**12**	Monitoreo cercano
E6	Cómputo cuántico que rompe el cifrado (CRQC)	3	3	2	1	3	**12**	Monitoreo cercano
E7	Cadena de suministro de IA (modelos, agentes)	2	3	2	3	3	**13**	Atención inmediata
E8	Inyección de instrucciones y secuestro de agentes	2	3	2	3	2	**12**	Monitoreo cercano
E9	Infraestructura espacial y satélites	3	2	2	2	2	**11**	Monitoreo cercano
E10	Convergencia IA física y robótica	3	2	2	1	3	**11**	Monitoreo cercano
E11	Disrupción física ambiental sobre lo	2	2	2	2	2	**10**	Monitoreo cercano

ID	Amenaza	A	R	Amb	C	IP	Σ	Banda
	digital							
E12	Redes 6G e IA en el borde	3	1	3	1	1	**9**	Monitoreo cercano

Nota. Misma escala. E1 se reclasifica como amenaza activa por evidencia de materialización en 2026 (véase Sección 4.2). Elaboración propia.

Cinco amenazas alcanzan la banda de atención inmediata (13 a 15): L5, L12, E1, E2 y E7. Ninguna queda en vigilancia (5 a 8) dado que las 25 fueron preseleccionadas por su relevancia estratégica, sesgo que se declara de forma explícita en este estudio.

4.2 Casos confirmatorios

Cada amenaza se fundamenta en un caso o fuente con enlace verificable. Los enlaces marcados como "Primaria" se localizaron mediante búsqueda directa en la fuente original (a corte del mes de agosto de 2026); los marcados como "Reporte" remiten a estudios ya citados en las referencias. La columna de seguimiento propone una periodicidad de revisión.

Tabla 3

Casos y fuentes que confirman las amenazas identificadas

Caso (fecha)	Descripción breve	Conf.	Evidencia	Verificación (fuente y enlace)	Seguim.
Vigilancia y pérdida de privacidad	Tendencia a la vigilancia masiva y a la erosión de la privacidad de personas y empresas.	**L13**	Aumentan las capacidades de vigilancia que erosionan la privacidad de personas y organizaciones.	Reporte — ENISA 2024	Anual
Riesgo a la infraestructura espacial (2024)	Satélites y servicios espaciales como blanco, con impacto en servicios terrestres.	**E9**	Los satélites y servicios espaciales pueden ser atacados y afectar servicios en la Tierra.	Reporte — Poirier 2024, CSS/ETH	Anual
Redes 6G e IA en el borde (~2030)	Nuevas redes y procesamiento cercano al usuario como riesgo incipiente y lejano.	**E12**	Las futuras redes 6G y la IA cercana al usuario traerán riesgos nuevos, aún no visibles.	Reporte — PwC 2025	Anual
Disrupción ambiental sobre lo digital	Eventos climáticos extremos que dañan centros de datos e infraestructura digital.	**E11**	Fenómenos ambientales extremos pueden dañar centros de datos y servicios digitales.	Reporte — ENISA 2024	Anual
Convergencia IA física y robótica	Riesgo de control malicioso de flotas de robots o vehículos autónomos.	**E10**	Existe el riesgo de que alguien tome el control de robots o vehículos autónomos con fines dañinos.	Reporte — Deloitte 2026	Anual
Espionaje orquestado por IA: GTG-1002 (sep-2025) y Taiwán (jul-2026)	Una IA ejecutó por sí sola el 80-90 % de un ataque de espionaje a unas 30 organizaciones.	**E1**	Una inteligencia artificial realizó casi todo un ataque de espionaje por su cuenta, a una velocidad imposible	Primaria — Anthropic 2025 (anthropic.com/news/disrupting-the-first-reported-ai-orchestrated-cyber-espionage-campaign); CNN 2026	Mensual

Caso (fecha)	Descripción breve	Conf.	Evidencia	Verificación (fuente y enlace)	Seguim.
			para humanos.		
Crimen-como-servicio autónomo	Servicios criminales que automatizan el fraude con IA, sin un delincuente humano visible.	**E2**	Aparecen servicios que automatizan el fraude con IA, sin un delincuente humano a la vista.	Fuente — Cano 2026	Mensual
Consolidación del ransomware	Menos grupos pero más víctimas; muchos roban los datos sin cifrarlos para exigir el rescate.	**L1**	Los grupos que exigen rescate por datos siguen creciendo y profesionalizándose; el pago dejó de ser la norma.	Primaria — Black Kite; Kaspersky (blackkite.com/reports/2026-ransomware-report)	Mensual
Recolección para descifrado futuro (HNDL)	Adversarios guardan datos cifrados hoy para abrirlos cuando exista la computación cuántica.	**L7**	Se roban datos cifrados ahora con la intención de descifrarlos cuando exista el computador cuántico.	Reporte — NIST 2024; Kaspersky 2026 (securelist.com/state-of-ransomware-in-2026/119761/)	Semestral
Pérdida de control de IA en cascada	Simulaciones muestran fallas de IA que se propagan entre sistemas y se salen de control.	**E3**	Ejercicios muestran que una IA autónoma puede fallar y arrastrar a otras sin que nadie la detenga a tiempo.	Reporte — RAND 2025 (rand.org)	Semestral
Insuficiencia persistente de talento	Faltan de forma persistente profesionales de ciberseguridad en el mercado.	**L10**	No hay suficientes especialistas, lo que debilita la defensa de todas las organizaciones.	Reporte — PwC 2025; ENISA 2024	Semestral
Influencia con IA en elecciones	Campañas de desinformación amplificadas con IA durante procesos electorales.	**L9**	Se usó inteligencia artificial para difundir información falsa a gran escala durante elecciones.	Reporte — Munich Re 2025	Semestral
Gusano de IA Morris II	Prueba académica de un programa de IA que se auto-replica entre asistentes virtuales.	**E5**	Investigadores crearon, en laboratorio, un programa de IA capaz de contagiarse solo entre asistentes virtuales.	Primaria — Cohen, Bitton y Nassi 2024 (arxiv.org/abs/2403.02817)	Semestral
Cómputo cuántico que rompe el cifrado (CRQC)	Riesgo futuro de que un computador cuántico rompa el cifrado que hoy protege los datos.	**E6**	Se prepara la llegada de computadores cuánticos capaces de romper el cifrado actual.	Reporte — NIST 2024; BCG 2025	Semestral
Superficie de interfaces (API) e IoT	Miles de millones de dispositivos conectados amplían el terreno de ataque.	**L11**	Cuantos más dispositivos y conexiones automáticas hay, más puntos de entrada para atacar.	Reporte — Citi 2026; Verizon DBIR 2026	Trimestral
Secuestro de	Órdenes escondidas	**E8**	Se pueden esconder	Primaria — OWASP 2025	Trimestral

Caso (fecha)	Descripción breve	Conf.	Evidencia	Verificación (fuente y enlace)	Seguim.
agentes por instrucciones ocultas	que engañan a asistentes de IA para que actúen de forma dañina.		instrucciones que engañan a un asistente de IA para que haga algo dañino.	(genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/)	
Puerta trasera en XZ Utils, CVE-2024-3094 (mar-2024)	Código dañino oculto durante años en un programa base de casi todos los servidores Linux.	**L2**	Alguien escondió código malicioso en un programa gratuito muy usado; permitía controlar el sistema a distancia.	Primaria — CISA; NVD (cisa.gov/news-events/alerts/2024/03/29/reported-supply-chain-compromise-affecting-xz-utils-data-compression-library-cve-2024-3094)	Trimestral
Interrupción global de CrowdStrike	Una actualización defectuosa de un proveedor dejó sin servicio a millones de equipos el mismo día.	**L5**	El fallo de un solo proveedor de seguridad paralizó aerolíneas, bancos y hospitales en todo el mundo.	Primaria — CISA; CNN (cisa.gov/news-events/alerts/2024/07/19/widespread-it-outage-due-crowdstrike-update)	Trimestral
Identidades de máquina como prioridad	Crecen las credenciales de programas y robots, difíciles de inventariar y controlar.	**L6**	Cada vez hay más “usuarios” que no son personas sino programas, y su control aún es inmaduro.	Reporte — KPMG 2026	Trimestral
Fraude por videollamada falsa a Arup	Un empleado transfirió USD 25,6 M engañado por ejecutivos falsos creados con IA en una videollamada.	**L3**	Un empleado hizo una transferencia millonaria porque en la videollamada todos los jefes eran imitaciones creadas con IA.	Primaria — CNN (cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk)	Trimestral
Espionaje basado en identidad	Los atacantes 'entran con llave' usando credenciales robadas en vez de forzar defensas.	**L12**	Muchos ataques ya no rompen defensas: entran con usuarios y contraseñas robadas.	Primaria — IBM X-Force 2026 (ibm.com/think/x-force/threat-intelligence-index-2026-securing-identities-ai-detection-risk-management)	Trimestral
Envenenamiento de datos y modelos	Manipular los datos con que aprende la IA para corromper sus decisiones.	**E4**	Se puede “envenenar” la información con la que aprende una IA para que tome decisiones equivocadas.	Primaria — NIST AI 100-2e2025 (https://www.nist.gov/publications/adversarial-machine-learning-taxonomy-and-terminology-attacks-and-mitigations-0)	Trimestral
Campaña Salt Typhoon	Espías vinculados a un Estado dentro de redes de telecomunicaciones de decenas de países.	**L8**	Atacantes ligados a un Estado se infiltraron en redes telefónicas para espiar durante años.	Primaria — CISA AA25-239A (cisa.gov/news-events/cybersecurity-advisories/aa25-239a)	Trimestral
Cadena de suministro de IA	Modelos, datos y agentes de terceros usados sin un control de seguridad maduro.	**E7**	Las empresas dependen de modelos y componentes de IA de terceros que todavía no vigilan bien.	Reporte — Deloitte 2026; IBM 2026	Trimestral

Caso (fecha)	Descripción breve	Conf.	Evidencia	Verificación (fuente y enlace)	Seguim.
Amenazas a tecnología operativa	Ataques crecientes a sistemas industriales y de agua antiguos y sin actualizar.	**L4**	Los sistemas industriales viejos y sin mantenimiento son cada vez más blanco de ataques.	Reporte — ENISA 2025; WEF 2026	Trimestral

Nota. Seguimiento: Mensual = amenaza materializada o de evolución rápida; Trimestral = riesgo latente de alto impacto; Semestral = estratégico de evolución media; Anual = horizonte lejano o baja probabilidad a 2027. Elaboración propia.

4.3 Clasificación y prioridad

Las tablas 4 a 6 aportan un mapa propuesto del estado actual de cada amenaza: dónde está hoy, con qué evidencia y con qué prioridad relativa. Se usa para asignar el foco de gestión (mitigar los latentes, contener los emergentes) y para priorizar la inversión sobre lo ya conocido. Responde a la pregunta: "¿dónde estamos?". Estas tablas tratan de clasificar y darle marco al presente.

Tabla 4

Clasificación por complejidad e incertidumbre

	Complejidad baja	Complejidad alta
Incertidumbre alta	E5, E11	E2, E3, E6, E9, E10, E12
Incertidumbre baja	L1, L3, L6, L9, L10, L11, L12, L13, E4, E8	L2, L4, L5, L7, L8, E1, E7

Nota. Elaboración propia.

Tabla 5

Clasificación por posibilidad frente a probabilidad

Probables (con evidencia o incidentes)	Posibles (sin materialización amplia)
L1, L2, L3, L4, L5, L8, L9, L10, L11, L12; E1, E4, E7, E8	L6, L7, L13; E2, E3, E5, E6, E9, E10, E11, E12

Nota. E1, E4, E7 y E8 pasaron a probables por presentar incidentes o vectores documentados en 2026. Elaboración propia.

Tabla 6

Clasificación por probabilidad e impacto

	Impacto bajo	Impacto medio	Impacto alto
Probabilidad alta	—	L3, L11, E4, E8	L1, L2, L4, L5, L8, L12, E1, E7
Probabilidad media	L13	L6, L9, E11	L7, L10, E2, E3, E5, E6, E10
Probabilidad baja	E12	—	E9

Nota. Ubicación cualitativa a 2027. Elaboración propia.

5. Escenarios 2027

Esta sección aporta futuros alternativos plausibles construidos a partir de lo que aún no está determinado. No clasifica amenazas: propone mundos posibles para probar la robustez de la estrategia. Se usa preguntando, en cada escenario, "¿cómo se comporta nuestro plan?" y "¿qué lo haría más robusto?". Responde a la pregunta: "¿hacia dónde podríamos ir?". En síntesis, esta sección ensaya el futuro, que en conjunto con las tablas 4, 5 y 6, evita tanto la miopía (solo del presente) como la especulación sin soporte (solo del futuro).

5.1 Matriz de escenarios (incertidumbres críticas)

Ejes seleccionados por ser de alto impacto y alta incertidumbre: (Eje X) autonomía de la IA ofensiva a 2027, de baja a alta; (Eje Y) madurez de la defensa y el gobierno de IA en la organización, de baja a alta. El eje vertical (Y) es el que la organización puede controlar.

Tabla 7
Matriz de escenarios 2027

	IA ofensiva: baja autonomía	**IA ofensiva: alta autonomía**
Defensa: alta madurez	**Ventaja del defensor.** La ofensiva con IA avanza más lento de lo temido y la organización invirtió a tiempo; los riesgos clásicos quedan bajo control.	**Carrera a velocidad de máquina.** El ataque autónomo es real, pero se enfrenta con defensa autónoma, identidad reforzada y gobierno de agentes. Contención costosa pero efectiva.
Defensa: baja madurez	**Complacencia y deuda.** La ofensiva con IA no escala, pero tampoco se invierte; las amenazas de siempre (rescate, cadena de suministro, identidad) siguen causando daño. Sorpresa predecible por negligencia.	**Tormenta perfecta.** Ataques autónomos a gran velocidad frente a una defensa rezagada. Escenario de mayor impacto: fraude, extorsión y espionaje automatizados superan la respuesta.

Nota. Basado en Método 2x2 de incertidumbres críticas (Schwartz, 1996). Elaboración propia.

Lectura ejecutiva: subir por el eje vertical, la única palanca que la organización controla, mueve a la organización de "Tormenta perfecta" hacia "Carrera a velocidad de máquina" y de "Complacencia" hacia "Ventaja del defensor". Aunque el horizonte sea 2027, la acción proactiva es necesaria y clave en el contexto actual.

6. Análisis de resultados

Señales confirmadas por evidencia. Los ataques autónomos con IA (E1) se materializaron: Anthropic (2025) documentó una operación con el 80-90 % de las tareas ejecutadas por IA, y en 2026 se sumaron incidentes contra gobiernos (Taiwán) y campañas descritas por Unit 42 y el AI Security Institute; por ello E1 se reclasifica de emergente a activo. También están confirmados el ransomware multinivel (L1), la cadena de suministro de software (L2, con XZ Utils y el aumento reportado por IBM en 2026), la concentración de proveedores (L5, con CrowdStrike), el espionaje por identidad (L12) y el fraude por deepfake (L3, con Arup). El espionaje estatal a infraestructura crítica (L8) se confirma con la campaña Salt Typhoon.

Señales viables pero no materializadas a escala. El gusano de IA Morris II (E5) está confirmado como viable en laboratorio (Cohen, Bitton y Nassi, 2024), pero no como incidente en la actualidad. El envenenamiento de modelos (E4) y la inyección de instrucciones (E8) están documentados por normas (NIST, OWASP) y se ubican como probables, aunque sin incidentes de gran impacto público.

Señales no confirmadas o de baja probabilidad a 2027. El cómputo cuántico que rompe el cifrado (E6) no se ha materializado; sin embargo, apareció una manifestación inadvertida: familias de ransomware ya adoptan cifrado post-cuántico de forma ofensiva (Kaspersky, 2026). La infraestructura espacial (E9), la robótica física (E10) y las redes 6G (E12) siguen sin incidentes de impacto y conservan alta incertidumbre.

Posibilidades presentes solo en los escenarios. La "Tormenta perfecta" (E1 a gran escala con defensa rezagada) y las fallas en cascada entre múltiples agentes (E3) son plausibles según simulaciones (RAND, 2025), pero carecen aún de evidencia de materialización a escala; se gestionan por preparación y contención.

7. Recomendaciones para los directores de ciberseguridad

1. Tratar los ataques autónomos con IA (E1) como riesgo activo: desplegar detección de conducta anómala automatizada, cortes de emergencia y gobierno de agentes, y probarlos con ejercicios de mesa a velocidad de máquina.

2. Adoptar la identidad como perímetro (L6, L12): autenticación resistente a phishing y detección de anomalías de identidad, validadas con simulacros de robo de credenciales.

3. Endurecer la cadena de suministro de software y de IA (L2, E7): inventarios de componentes (SBOM y AI-BOM), verificación de origen de modelos y prueba de un componente comprometido.

4. Iniciar la agilidad criptográfica (L7, E6): inventario criptográfico y ensayo de migración post-cuántica en un sistema piloto, sin esperar al cómputo cuántico.

5. Reducir la dependencia de proveedores críticos (L5) con planes de continuidad multiproveedor y simulacros de indisponibilidad.

6. Institucionalizar el ciclo de anticipación (revelar, ver, darse cuenta, actuar) con revisión semestral de la Ventana de AREM® y reubicación de amenazas entre cuadrantes.

8. Recomendaciones para la junta directiva

1. Priorizar la inversión en las cinco amenazas de atención inmediata: concentración de proveedores (L5), espionaje por identidad (L12), ataques autónomos con IA (E1), crimen-como-servicio autónomo (E2) y cadena de suministro de IA (E7).

2. Aprobar un gobierno de IA transversal (uso, agentes y modelos de terceros), dado que estas amenazas ya son probables, no sólo hipótesis.

3. Definir el apetito de riesgo frente a la automatización ofensiva y exigir métricas de preparación (no solo de cumplimiento) en cada comité.

4. Financiar la identidad como perímetro y la agilidad criptográfica como capacidades estratégicas plurianuales.

5. Requerir la revisión semestral de la Ventana de AREM® como insumo formal de la toma de decisiones del directorio.

6. Asumir que el riesgo dominante a 2027 es la inacción ante señales ya visibles: la decisión de invertir a tiempo es la que separa la “Ventaja del defensor” de la “Tormenta perfecta”.

9. Conclusiones generales

A 2027, la frontera entre lo latente y lo emergente se desplazó: los ataques autónomos con IA, el envenenamiento de modelos y la cadena de suministro de IA dejaron de ser especulación para volverse probables, y los primeros ya son un riesgo activo. El ejercicio de casos confirmatorios muestra que la mayoría de las señales débiles cuenta con evidencia trazable, mientras que otras permanecen como posibilidades solo visibles en los escenarios. La prioridad estratégica se concentra en cuatro capacidades: gobierno de IA, identidad como perímetro, agilidad criptográfica y resiliencia de proveedores, y en repetir el ejercicio de forma periódica de acuerdo con la información disponible en el momento. El mayor riesgo no es técnico sino organizacional: la inacción frente a lo que ya se ve.

10. Referencias

AI Security Institute. (2026). Incident report: Unsanctioned agent behaviour during cyber testing. *UK AI Security Institute*. https://www.aisi.gov.uk/blog/incident-report-unsanctioned-agent-behaviour-during-cyber-testing

Accenture. (2025). State of cybersecurity resilience 2025. *Accenture*. https://www.accenture.com/us-en/insights/security/state-cybersecurity-2025

Ansoff, H. I., Kipley, D., Lewis, A. O., Helm-Stevens, R., & Ansoff, R. (2019). *Implanting strategic management* (3.ª ed.). Springer.

Anthropic. (2025). Disrupting the first reported AI-orchestrated cyber espionage campaign. *Anthropic PBC*. https://www.anthropic.com/news/disrupting-the-first-reported-ai-orchestrated-cyber-espionage-campaign

Arthur D. Little. (2024). AI in cybersecurity. *Arthur D. Little*. https://www.adlittle.com/en/insights/viewpoints/ai-cybersecurity

Atlantic Council. (2026). Securing cloud infrastructure for AI. *Atlantic Council*. https://www.atlanticcouncil.org/in-depth-research-reports/issue-brief/securing-cloud-infrastructure-ai/

Atlantic Council. (2026). US leadership in the age of AI: Report of the Commission on Artificial Intelligence. *Atlantic Council*. https://www.atlanticcouncil.org/in-depth-research-reports/report/atlantic-council-commission-on-ai-lays-a-roadmap-for-us-leadership-in-the-age-of-ai/

Bazerman, M. H., & Watkins, M. D. (2004). *Predictable surprises*. Harvard Business School Press.

Black Kite. (2026). 2026 ransomware report. Black Kite. https://blackkite.com/reports/2026-ransomware-report

Boston Consulting Group. (2025). AI is raising the stakes in cybersecurity. *BCG*. https://www.bcg.com/publications/2025/ai-raising-stakes-in-cybersecurity

Boston Consulting Group. (2025). How quantum computing will upend cybersecurity. *BCG*. https://www.bcg.com/publications/2025/how-quantum-computing-will-upend-cybersecurity

Boston Consulting Group. (2026). Agentic AI is rewriting the rules of data risk management. *BCG*. https://www.bcg.com/publications/2026/managing-data-risk-in-the-age-of-agentic-ai

Boston Consulting Group. (2026). Cyberattacks are inevitable in health care. *BCG*. https://www.bcg.com/publications/2026/cyber-attacks-in-health-care-are-inevitable

Boston Consulting Group. (2026). Risk and compliance 2026: Refining oversight for a volatile, AI-driven world. *BCG*. https://www.bcg.com/publications/2026/refining-oversight-for-a-volatile-ai-driven-world

Boston Consulting Group, & Global Cybersecurity Forum. (2025). The quantum leap. *BCG*. https://www.bequantumready.com/

Cano, J. (2017). La ventana de AREM. *ISACA Journal*, 5. https://www.isaca.org/es-es/resources/isaca-journal/issues/2017/volume-5/the-arem-window-a-strategy-to-anticipate-risk-and-threats-to-enterprise-cyber-security

Cano, J. (2026). Anticipando las amenazas cibernéticas: un marco de preparación en cuatro pasos. En M. Rodriguez-Vega, D. Escanez-Exposito, J. Garcia-Diaz, & P. Caballero-Gil (Eds.), *Actas XIX Reunión Española sobre Criptología y Seguridad de la Información (RECSI)* (pp. 272-278). Tenerife, España. https://cryptull.webs.ull.es/RECSI2026/LibrodeActasRECSI2026.pdf

Center for Strategic and International Studies. (2026). Making AI work for cyber defenders: A strategy for strengthening U.S. cybersecurity. *CSIS*. https://www.csis.org/analysis/making-ai-work-cyber-defenders-strategy-strengthening-us-cybersecurity

Chen, H. (2024). Arup revealed as victim of $25 million deepfake scam. CNN Business. https://www.cnn.com/2024/05/16/tech/arup-deepfake-scam-loss-hong-kong-intl-hnk

Citi Institute. (2026). The trillion-dollar security race is on: The quantum threat. *Citi*. https://www.citigroup.com/global/insights/quantum-threat

Coalition. (2025). Cyber threat index 2025. Coalition, Inc. https://web.coalitioninc.com/DLC-Cyber-Threat-Index-2025.html

Coalition. (2026). 2026 cyber claims report. Coalition, Inc. https://www.coalitioninc.com/claims-report/2026

Cohen, S., Bitton, R., & Nassi, B. (2024). Here comes the AI worm: Unleashing zero-click worms that target GenAI-powered applications (arXiv:2403.02817). https://arxiv.org/abs/2403.02817

Congressional Research Service. (2026). Agentic artificial intelligence and cyberattacks (IF13151). *Library of Congress*. https://www.congress.gov/crs-product/IF13151

Cybersecurity and Infrastructure Security Agency, National Security Agency, Australian Signals Directorate's Australian Cyber Security Centre, Canadian Centre for Cyber Security, National Cyber Security Centre New Zealand, & National Cyber Security Centre United Kingdom. (2026). Careful adoption of agentic AI services. CISA. https://www.cisa.gov/resources-tools/resources/careful-adoption-agentic-ai-services

Deloitte. (2025). Global cyber threat intelligence: Annual cyber threat trends 2025. *Deloitte*. https://www.deloitte.com/us/en/services/consulting/articles/cybersecurity-report-2025.html

Deloitte. (2025). Global future of cyber survey (4.ª ed.). Deloitte. https://www.deloitte.com/global/en/services/consulting/research/global-future-of-cyber-fourth-edition.html

Deloitte. (2026). Tech trends 2026. *Deloitte Insights*. https://www.deloitte.com/us/en/insights/topics/technology-management/tech-trends.html

Deloitte, & National Association of State Chief Information Officers. (2026). 2026 NASCIO-Deloitte cybersecurity study. *Deloitte Insights*. https://www.deloitte.com/us/en/insights/industry/government-public-sector-services/2026-nascio-deloitte-cybersecurity-study.html

Ernst & Young. (2026). AI in cybersecurity study 2026. *EY*. https://www.ey.com/en_us/newsroom/2026/03/cybersecurity-leaders-investing-in-ai-and-agentic-defenses-to-combat-escalating-ai-enabled-threats

European Union Agency for Cybersecurity. (2024). Foresight cybersecurity threats for 2030 — Update 2024. *ENISA*. https://www.enisa.europa.eu/publications/foresight-cybersecurity-threats-for-2030-update-2024

European Union Agency for Cybersecurity. (2025). ENISA threat landscape 2025. *ENISA*. https://www.enisa.europa.eu/publications/enisa-threat-landscape-2025

EY-Parthenon Geostrategic Business Group. (2026). 2026 geostrategic outlook. *Ernst & Young*. https://www.ey.com/en_mt/insights/geopolitical-outlook-for-2026

Federal Reserve Board. (2025). "Harvest now, decrypt later" (Finance and Economics Discussion Series N.º 2025-093). *Board of Governors of the Federal Reserve System*. https://www.federalreserve.gov/econres/feds/harvest-now-decrypt-later-examining-post-quantum-cryptography-and-the-data-privacy-risks-for-distributed-ledger-networks.htm

Gerstein, D. M., & Leidy, E. N. (2024). Emerging technology and risk analysis: Artificial intelligence and critical infrastructure (RR-A2873-1). *RAND Corporation*. https://www.rand.org/pubs/research_reports/RRA2873-1.html

IBM. (2026). X-Force threat intelligence index 2026. *IBM Corporation*. https://www.ibm.com/think/x-force/threat-intelligence-index-2026-securing-identities-ai-detection-risk-management

Ibrahim, N., & Kashef, R. (2025). Exploring the emerging role of large language models in smart grid cybersecurity: A survey of attacks, detection mechanisms, and mitigation strategies. *Frontiers in Energy Research*, 13, Article 1531655. https://doi.org/10.3389/fenrg.2025.1531655

Institute for Security and Technology. (2024). The implications of AI in cybersecurity: Shifting the offense–defense balance. *IST*. https://securityandtechnology.org/virtual-library/report/the-implications-of-artificial-intelligence-in-cybersecurity/

Kaspersky. (2026). The state of ransomware in 2026. Securelist. https://securelist.com/state-of-ransomware-in-2026/119761/

KPMG International. (2026). Cybersecurity considerations 2026. *KPMG*. https://kpmg.com/xx/en/our-insights/ai-and-technology/cybersecurity-considerations-2026.html

Lesca, H., & Lesca, N. (2011). *Weak signals for strategic intelligence: Anticipation tool for managers*. Wiley.

Marsh. (2025). The state of cyber resilience 2025. *Marsh McLennan*. https://www.marsh.com/en/services/cyber-risk/insights/the-state-of-cyber-resilience.html

Ministry of Defence, Development, Concepts and Doctrine Centre. (2024). Global strategic trends: Out to 2055 (7.ª ed.). UK *Ministry of Defence*. https://www.gov.uk/government/publications/global-strategic-trends-out-to-2055

Munich Re. (2025). Cyber insurance: Risks and trends 2025. *Munich Re*. https://www.munichre.com/en/insights/cyber/cyber-insurance-risks-and-trends-2025.html

Munich Security Conference. (2025). Munich Security Report 2025: Multipolarization. *MSC*. https://securityconference.org/en/publications/munich-security-report-2025/

National Institute of Standards and Technology - NIST. (2024). Post-quantum cryptography standards (FIPS 203, 204, 205). U.S. *Department of Commerce*. https://csrc.nist.gov/news/2024/postquantum-cryptography-fips-approved

National Institute of Standards and Technology - NIST. (2025). Adversarial machine learning: A taxonomy and terminology of attacks and mitigations (NIST AI 100-2e2025). U.S. *Department of Commerce*. https://www.nist.gov/publications/adversarial-machine-learning-taxonomy-and-terminology-attacks-and-mitigations-0

OWASP Foundation. (2025). OWASP Top 10 for LLM applications 2025. OWASP. https://genai.owasp.org/resource/owasp-top-10-for-llm-applications-2025/

Palo Alto Networks, Unit 42. (2026). Chinese-speaking threat actor harnesses AI models for autonomous cyberattacks. *Palo Alto* Networks. https://unit42.paloaltonetworks.com/autonomous-ai-cyber-attack-campaign/

Poirier, C. (2024). Hacking the cosmos: Cybersecurity in space (Cyber Reports 2024-10). *Center for Security Studies* (CSS), ETH Zurich. https://css.ethz.ch/en/center/CSS-news/2024/10/hacking-the-cosmos-cyber-operations-against-the-space-sector-a-case-study-from-the-war-in-ukraine.html

PwC. (2025). 2026 global digital trust insights survey. *PricewaterhouseCoopers*. https://www.pwc.com/us/en/services/consulting/cybersecurity-data-tech-risk/library/global-digital-trust-insights.html

PwC. (2026). Annual threat dynamics 2026: Cyber threats in motion. *PricewaterhouseCoopers*. https://www.pwc.com/gx/en/issues/cybersecurity/cyber-threat-intelligence/annual-threat-dynamics.html

RAND Corporation. (2025). Strengthening emergency preparedness and response for AI loss-of-control incidents (RR-A3847-1). *RAND Corporation*. https://www.rand.org/pubs/research_reports/RRA3847-1.html

Real Instituto Elcano. (2024). La ciberresiliencia: entre la ciberseguridad y la resiliencia. *Real Instituto Elcano*. https://www.realinstitutoelcano.org/analisis/la-ciberresiliencia-entre-la-ciberseguridad-y-la-resiliencia/

Schwartz, P. (1996). *The art of the long view*. Currency Doubleday.

Turner Lee, N. (2025). Shaping the future of AI through responsible innovation [Testimonio ante el Subcomité de Ciberseguridad, TI e Innovación Gubernamental, Cámara de Representantes de EE. UU.]. *Brookings Institution*. https://www.brookings.edu/articles/shaping-the-future-of-ai-through-responsible-innovation/

Verizon. (2026). 2026 data breach investigations report. *Verizon Business*. https://www.verizon.com/business/resources/reports/dbir/

World Economic Forum. (2026). Global cybersecurity outlook 2026. *World Economic Forum*. https://www.weforum.org/publications/global-cybersecurity-outlook-2026/

Zhang, J., Bu, H., Wen, H., & Liu, Y., et al. (2025). When LLMs meet cybersecurity: A systematic literature review. *Cybersecurity*, *8*(1), Article 55. https://link.springer.com/article/10.1186/s42400-025-00361-w

Siglas y acrónimos

- AI-BOM (*AI Bill of Materials*) — Inventario de todos los componentes de un sistema de IA (modelos, datos, dependencias), para saber qué se usa y de dónde proviene.
- API (*Application Programming Interface*) — Interfaz de programación: punto de conexión que permite que dos programas intercambien datos de forma automática.
- CISA (*Cybersecurity and Infrastructure Security Agency*) — Agencia de ciberseguridad e infraestructura del gobierno de EE. UU. Emite alertas y advisories.
- CRQC (*Cryptographically Relevant Quantum Computer*) — Computador cuántico con capacidad suficiente para romper el cifrado que hoy protege los datos. Aún no existe.
- CVE (*Common Vulnerabilities and Exposures*) — Sistema internacional que asigna un identificador único a cada vulnerabilidad pública (p. ej., CVE-2024-3094).
- DBIR (*Data Breach Investigations Report*) — Informe anual de Verizon sobre brechas de datos, referencia del sector.
- ENISA (*European Union Agency for Cybersecurity*) — Agencia de ciberseguridad de la Unión Europea.
- HNDL (*Harvest Now, Decrypt Later*) — "Cosechar ahora, descifrar después": robar datos cifrados hoy para descifrarlos en el futuro con computación cuántica.
- IA — Inteligencia artificial.
- IoT (*Internet of Things*) — "Internet de las cosas": dispositivos físicos (sensores, cámaras, equipos) conectados a internet.
- LLM (*Large Language Model*) — Modelo de lenguaje grande: tipo de IA que procesa y genera texto (la base de los asistentes conversacionales).
- NVD (National Vulnerability Database) — Base de datos pública de vulnerabilidades mantenida por NIST.
- PQC (*Post-Quantum Cryptography*) — Criptografía post-cuántica: nuevos algoritmos de cifrado diseñados para resistir a los futuros computadores cuánticos.
- SBOM (*Software Bill of Materials*) — Inventario de todos los componentes que integran un software, para poder rastrear cuáles son vulnerables.