



INFORME ESTRATÉGICO / OCTUBRE 2026

INTELIGENCIA ARTIFICIAL Y CIBERSEGURIDAD 2026

Riesgos emergentes, responsabilidad del CEO
y estrategia de ciberresiliencia empresarial

Una guía para CEO, comités ejecutivos y consejos de administración
Enfoque: España y Unión Europea · Corte: 17 de octubre de 2026

NOTA EJECUTIVA

La ciberseguridad ya es una responsabilidad de dirección

La IA acelera amenazas conocidas, introduce riesgos asociados a agentes autónomos y refuerza las capacidades defensivas. El CEO debe integrar estas tres dinámicas en un mismo modelo de riesgo empresarial.


122.223

incidentes en España
gestionados por INCIBE · 2025


≈26 %

crecimiento anual
frente al ejercicio 2024


24 / 72 h

hitos NIS2
notificación escalonada
de incidentes significativos

La tesis central. La pregunta del consejo ya no es simplemente «¿estamos protegidos?». Debe ser: «¿Qué procesos críticos podrían quedar paralizados, qué pérdida material provocarían y cuánto tardaríamos en detectar, contener y recuperar?» [1][2][3]

En 2026, la superficie de ataque combina identidades corporativas, entornos cloud, proveedores, endpoints, infraestructuras industriales, modelos de IA y agentes capaces de actuar mediante conectores. La seguridad debe diseñarse como un sistema de gobierno, tecnologías, procesos y pruebas de recuperación; no como una adquisición aislada de software.

Cinco decisiones inmediatas del CEO

- 01. Nombrar un patrocinador ejecutivo** para la resiliencia ciber y la seguridad de IA, con autoridad interdepartamental.
- 02. Identificar los “crown jewels”:** operaciones, datos y sistemas cuya interrupción genera pérdidas materiales.
- 03. Aprobar controles irrenunciables:** MFA resistente al phishing, segregación de pagos, backups verificables y privilegios mínimos.
- 04. Inventariar toda la IA** —incluidos agentes, SaaS y usos no aprobados— con propietario, datos, permisos y proveedores.
- 05. Exigir una simulación de crisis** y un cuadro de mando trimestral de riesgo residual ante el consejo.

Cómo leer las cifras: Los datos de incidentes proceden de organismos oficiales. Las puntuaciones de riesgo, plazos de implantación, importes orientativos y escenarios financieros propuestos a lo largo del informe son **ejemplos de gestión**, no estadísticas universales ni asesoramiento jurídico individualizado.

GUÍA DE LECTURA

Índice ejecutivo

Un recorrido de la amenaza a la acción, diseñado para que el CEO pueda decidir prioridades y el equipo directivo traducirlas en un plan medible.

CONTENIDO	
01	Panorama de amenazas y evidencias de 2026 Qué ha cambiado, datos y sectores expuestos
02	Fraude sintético y ataques asistidos por IA Deepfakes, agentes, ingeniería social y velocidad de ataque
03	Seguridad de la propia IA corporativa Prompt injection, agentes, Shadow AI y terceros
04	Casos relevantes y aprendizajes Incidentes reales y controles que faltaban
05	Gobierno corporativo y obligaciones NIS2, RGPD, DORA, AI Act y CRA
06	Arquitectura de ciberresiliencia Identidad, cloud, SOC, datos y recuperación
07	Métricas, matriz y decisiones de inversión Dashboard, priorización de riesgos y ROSI
08	Hoja de ruta 90 días y 12 meses Entregables, responsables y evidencias
09	Ejercicios, preguntas de consejo y fuentes Tabletops y bibliografía oficial

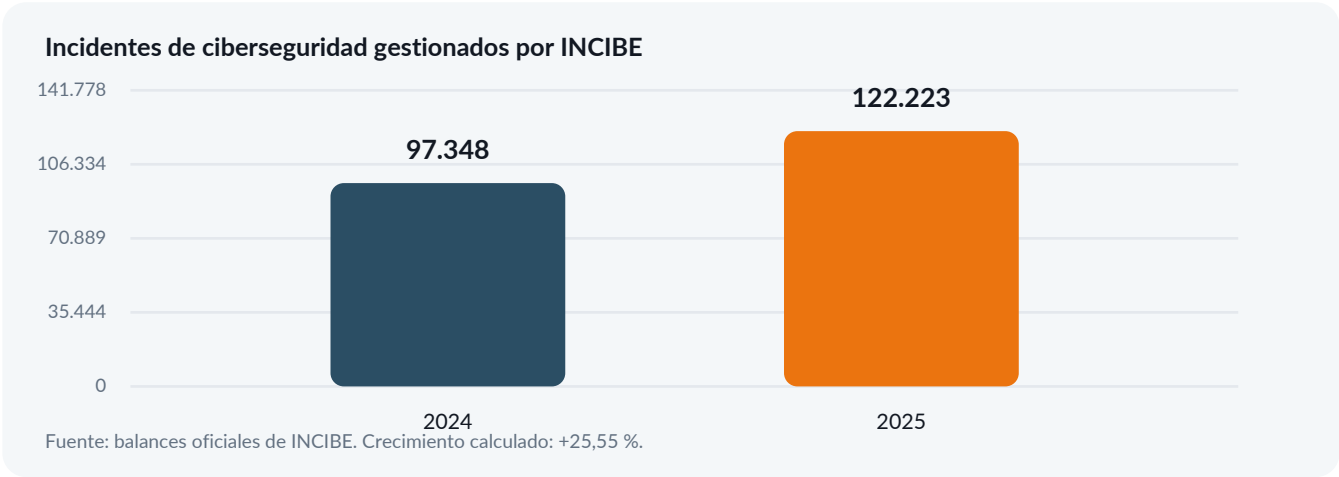
Ámbito y fecha. España y Unión Europea. Investigación con fecha de corte de **17 de octubre de 2026**; este documento deberá actualizarse si cambian textos normativos, fechas de aplicación o información sobre incidentes en curso.

01 / CONTEXTO

De la amenaza informática al riesgo empresarial

El cambio más relevante no es que todas las amenazas sean nuevas: es que la IA reduce el coste y la fricción de preparar, escalar, personalizar y automatizar ataques.

INCIBE registró **97.348 incidentes** en 2024 y **122.223** en 2025. La variación entre ambos ejercicios ronda el 25,55 %. Estas cifras describen **incidentes gestionados** por el organismo, no el número total de ciberataques sufridos por todas las empresas españolas. [1][2]



Según el panorama ENISA 2026, ransomware, DDoS, explotación de vulnerabilidades y dependencias digitales siguen siendo fuentes centrales de riesgo. Para los directivos, la fotografía debe interpretarse a la luz de una vulnerabilidad organizativa adicional: la dependencia simultánea de servicios cloud, plataformas SaaS, proveedores gestionados y cadenas de suministro. [3]

Impacto en la cuenta de resultados

FUENTE DE PÉRDIDA	CONSECUENCIA EMPRESARIAL
Interrupción	Paradas de planta, distribución, comercio electrónico o servicios esenciales.
Fraude	Transferencias indebidas, cambio de cuentas bancarias o suplantación del CEO.
Datos y propiedad intelectual	Filtraciones, secretos comerciales, litigios y pérdida de confianza.
Cumplimiento	Notificaciones, investigaciones, sanciones y obligaciones contractuales.
Terceros	Caída de un proveedor crítico con efecto dominó sobre operaciones.

01 / EVOLUCIÓN

La nueva ecuación: escala, velocidad y autonomía

La IA generativa facilita texto convincente, traducciones, voz sintética y análisis contextual. La IA agente añade una dimensión: **secuencias de herramientas que observan, ejecutan acciones, revisan resultados y vuelven a operar**. Este salto no implica que todo agente sea peligroso; sí obliga a aplicar controles de identidad, privilegios y auditoría equivalentes —o superiores— a los exigidos a usuarios humanos. [4][5]

Mapa de riesgos directivos

RIESGO	EFFECTO POTENCIADO POR IA	CONTROL CLAVE
Deepfake / BEC	Suplantación creíble por voz, vídeo o correo.	Verificación fuera de banda y doble aprobación.
Phishing avanzado	Mensajes escalables y personalizados.	MFA resistente al phishing y entrenamiento.
Ataque automatizado	Reconocimiento y encadenamiento de pasos.	Parches, ASM, EDR/XDR, logging y contención.
Ransomware	Mayor capacidad de adaptación y velocidad.	Segmentación y restauraciones probadas.
Shadow AI	Exposición de datos en servicios no gobernados.	Inventario, IA aprobada, DLP y contratos.
Prompt injection	Instrucciones maliciosas ocultas en datos.	Aislamiento de herramientas y validación.
Supply chain	Dependencias cruzadas de modelos, SaaS y API.	Evaluación y planes alternativos.

Principio de diseño: Cuanta más autonomía tiene una herramienta, más crítico es limitar las acciones que puede ejecutar, la información que puede leer y la rapidez con la que pueden revocarse sus credenciales.

02 / AMENAZAS

Un deepfake no debe poder autorizar un pago

La mejor defensa contra una suplantación impecable consiste en diseñar procesos que no dependan de reconocerla.

Vishing, videollamadas falsificadas y mensajes generados con IA hacen menos fiables las señales tradicionales de autenticidad. En febrero de 2026, el Banco de Italia alertó de contenidos falsificados que utilizaban la imagen de su gobernador; el caso ilustra que la imagen pública y la voz del directivo son ahora superficies de riesgo reputacional y financiero. [6]

Proceso recomendado para una operación financiera sensible

PASO	REGLA OPERATIVA	RESPONSABLE
1 • Solicitud	Registrar el origen y motivo del pago / cambio de cuenta.	Finanzas
2 • Contraste	Confirmar por un canal conocido previamente, no por el canal de la solicitud.	Tesorería
3 • Autorización	Doble aprobación segregada por importe y criticidad.	Dirección autorizada
4 • Ejecución	Control de beneficiarios y revisión de cambios recientes.	Finanzas / ERP
5 • Evidencia	Conservar justificantes y alertas para auditoría y respuesta.	Control interno

Ejemplo de escenario: Un supuesto CEO solicita por videollamada una transferencia urgente. Aunque la voz e imagen parezcan legítimas, el pago **no procede** sin verificación independiente, segregación de funciones y trazabilidad.

Cómo entrenar sin generar falsa confianza

La formación debe incluir mensajes auténticos y falsos, llamadas de alta presión, suplantación de proveedores y cambios de cuenta. Pero el éxito no se mide por cuántas personas “detectan un deepfake”, sino por cuántas **mantienen el procedimiento de autorización** en condiciones de urgencia o presión jerárquica.

02 / AGENTES

Quando el ataque puede encadenar tareas

La velocidad de máquina exige capacidades defensivas capaces de detectar, contener y recuperar sin depender exclusivamente de intervenciones manuales.

En septiembre de 2026, la AEPD comunicó una notificación de brecha de datos personales asociada a un ataque ejecutado mediante un agente de IA. La información publicada describe un encadenamiento de tareas con reducida intervención humana: detección de vulnerabilidades, acceso, modificación de datos y consulta de facturación. Es un caso notificado, no prueba de que todos los agentes funcionen con ese nivel de autonomía. [4]

Cadena ilustrativa de ataque asistido por agente

ETAPA	ACELERADOR	DEFENSA
Reconocer	Analizar activos públicos, superficies expuestas y tecnologías.	ASM, inventario, reducción de exposición.
Acceder	Buscar debilidades y probar credenciales o integraciones.	MFA, gestión de secretos, parches prioritarios.
Moverse	Combinar privilegios, conectores y entornos.	Zero Trust, segmentación, PAM.
Extraer / alterar	Automatizar acciones sobre datos y sistemas.	DLP, anomalías, aprobaciones y logs.
Persistir	Reintentos rápidos o cambios incrementales.	EDR/XDR, revocación de tokens, respuesta.

Una defensa robusta requiere que los equipos puedan **revocar tokens, suspender identidades, aislar máquinas y cortar conectores** sin detener toda la compañía. El SOC debe priorizar señales ligadas a procesos de negocio y no limitarse a multiplicar alertas.

Indicador ejecutivo útil: Tiempo necesario para detectar y revocar un acceso anómalo a un sistema crítico, medido en ejercicios controlados. Una mejora del tiempo sin evidencia de contención efectiva no equivale a menos riesgo.

03 / LA EMPRESA COMO USUARIA DE IA

Proteger los agentes antes de conectarlos al negocio

Un asistente conversacional con documentos no presenta el mismo perfil de riesgo que un agente con acceso a correo, ERP, base de datos y capacidad para ejecutar operaciones.

El riesgo más distintivo de una aplicación de IA conectada a recursos corporativos es la **confusión entre datos e instrucciones**. Un documento, email, página web o resultado de una búsqueda puede contener órdenes maliciosas diseñadas para manipular al asistente. Es el patrón de prompt injection indirecta. La defensa no puede apoyarse solo en una frase que pida al modelo “ignorar instrucciones maliciosas”. [7]

Controles mínimos para agentes empresariales

CONTROL	EVIDENCIA EXIGIBLE
Propiedad e inventario	Ficha de cada agente: responsable, objetivo, datos, modelo, conectores y criticidad.
Permisos mínimos	Identidad propia, acciones permitidas y secretos limitados por necesidad de uso.
Separación datos / instrucciones	Tratar páginas, archivos y mensajes externos como contenido no confiable.
Aprobación humana	Gates antes de pagos, borrados, envíos masivos o cambios irreversibles.
Observabilidad	Registros de solicitudes, herramientas invocadas, decisiones y resultados.
Pruebas adversariales	Red team de inyección, filtración, abuso de herramientas y fuga de contexto.
Revocación / reversión	Mecanismo para pausar agentes, invalidar tokens y restaurar datos.

Regla de oro: El sistema de permisos y la aprobación de operaciones sensibles deben aplicarse **fuera del propio modelo**. Un agente no debe decidir por sí solo si puede saltarse sus propios controles.

03 / DEPENDENCIAS

Shadow AI, cloud y cadena de suministro

El riesgo no termina en el perímetro corporativo: viaja con las identidades, las API y los proveedores.

La adopción informal de IA puede llevar a compartir información sensible, propiedad intelectual o datos personales con herramientas no homologadas. Prohibiciones indiscriminadas suelen desplazar el uso a canales menos visibles. Una estrategia equilibrada combina catálogo de soluciones aprobadas, reglas de clasificación de datos, formación práctica y controles de DLP.

En cloud, el proveedor aporta infraestructura y servicios de protección, pero **no absorbe íntegramente la responsabilidad empresarial**: siguen siendo necesarios IAM, logging independiente, control de secretos, cifrado, configuraciones seguras y planes de recuperación. Un ciberincidente comunicado en la infraestructura cloud que alojaba el portal Europa de la Comisión Europea en marzo de 2026 demuestra la importancia de evaluar dichas dependencias, sin atribuirlo indebidamente a IA. [8]

Preguntas de evaluación para un proveedor crítico

- ¿Qué información procesa y dónde se almacena o transfiere?
- ¿Qué subencargados, modelos, API y otros terceros intervienen?
- ¿Cómo se controlan los accesos privilegiados y las identidades no humanas?
- ¿Qué pruebas de restauración, continuidad y respuesta se han realizado?
- ¿Cuál es el procedimiento de notificación y colaboración ante un incidente?
- ¿Cómo se exportan datos y se mantiene el servicio si el proveedor falla?

Priorización: Clasificar proveedores en niveles según impacto de la caída y acceso a datos. Empezar por los que pueden detener una línea de fabricación, la facturación, logística, atención a clientes o sistemas de identidad.

04 / PRUEBAS DE REALIDAD

Qué enseñan los incidentes conocidos

Distinguir ataques asistidos por IA de incidentes convencionales es esencial para tomar decisiones y mantener la credibilidad de los análisis.

CASO	HECHO DESCRITO	LECCIÓN PARA EL CEO
AEPD · España · sep. 2026	Brecha notificada asociada a la actuación de un agente de IA.	Velocidad, autonomía, detección y supervisión de agentes. [4]
Banco de Italia · feb. 2026	Alerta sobre contenidos falsificados con imagen del gobernador.	Identidad ejecutiva, fraude sintético y verificación externa. [6]
Comisión Europea · mar. 2026	Incidente en infraestructura cloud vinculada al portal Europa.	Gobierno de cloud, continuidad, identidades y controles de terceros. [8]
OpenAI / Hugging Face · 2026	Incidente divulgado en contexto de evaluación de modelos y capacidades.	Aislamiento y monitorización de entornos de prueba de IA. [9]
Ahold · oct. 2026	Interrupción de sitios y servicios de distribución; la atribución seguía investigándose.	Continuidad operativa; no atribuir a IA sin evidencia. [10]

Tres errores de interpretación que conviene evitar

- 1. Confundir automatización con origen.** Que un ataque pueda acelerarse con IA no permite concluir que la IA haya sido utilizada en un incidente concreto.
- 2. Confundir vulnerabilidad del proveedor con responsabilidad exclusiva del proveedor.** La resiliencia depende también de las configuraciones, identidades y procesos del cliente.
- 3. Confundir una demo de IA con evidencia de seguridad.** Antes de integrar un agente en procesos críticos, se necesita una evaluación de permisos, datos y seguridad.

05 / MARCO REGULATORIO

El consejo entra en el perímetro de responsabilidad

La normativa europea obliga a integrar la ciberseguridad en la gobernanza. La aplicabilidad concreta depende de actividad, dimensión, sector y legislación nacional.

La Directiva NIS2 contempla obligaciones para órganos de dirección de entidades esenciales e importantes: aprobar medidas de gestión de riesgo, supervisar su ejecución y recibir formación. Entre las medidas figuran continuidad, respuesta, seguridad de la cadena de suministro, evaluación de la eficacia, cifrado, control de acceso, MFA y formación. [11]

Panorama regulatorio a octubre de 2026

MARCO	ALCANCE EMPRESARIAL	PREGUNTA PARA DIRECCIÓN
NIS2	Ciberseguridad para entidades esenciales e importantes de sectores cubiertos.	¿Estamos sujetos y podemos demostrar supervisión y controles?
RGPD	Seguridad de datos personales y notificación de ciertas brechas.	¿Puede actuar Legal/DPO con información fiable a tiempo?
DORA	Resiliencia digital TIC del sector financiero (desde enero de 2025).	¿Probamos continuidad y gestionamos terceros TIC críticos?
AI Act	Obligaciones según rol y tipología de sistema de IA; robustez y ciberseguridad en sistemas de alto riesgo.	¿Inventariamos y clasificamos IA, riesgos y responsables?
CRA	Productos con elementos digitales; notificaciones aplicables desde sept. 2026 y otros requisitos por fases.	¿Existe PSIRT y gestión de vulnerabilidades de producto?

Precisión jurídica: NIS2 es una directiva cuya transposición y aplicación nacional deben comprobarse en BOE. Las multas y responsabilidades dependen del tipo de entidad y del marco efectivamente aplicable; este documento no sustituye una revisión legal sectorial. [11][12]

05 / NOTIFICACIONES

Prepararse para las primeras 24 y 72 horas

La coordinación debe ensayarse antes del incidente. La tecnología por sí sola no puede resolver decisiones legales, operativas y reputacionales de forma improvisada.

Para incidentes significativos dentro del ámbito NIS2, la directiva prevé un esquema de **alerta temprana en 24 horas, notificación en 72 horas** e informe final posterior, sujeto a las condiciones legales específicas. Las brechas de datos personales pueden además activar obligaciones de RGPD, y DORA/CRA sus propios procedimientos. [11][12][13]

VENTANA	TAREA DE DIRECCIÓN	EVIDENCIA
0-4 h	Activar comité de crisis, preservar evidencia y delimitar impacto de negocio.	Acta, responsables, canal seguro.
4-24 h	Determinar obligaciones, hechos confirmados e información pendiente.	Análisis Legal/DPO/CISO y mensajes aprobados.
24-72 h	Emitir comunicaciones cuando proceda y actualizar alcance y acciones.	Registro de avisos y cadena de aprobación.
Días siguientes	Restauración, análisis forense, información a afectados, lecciones aprendidas.	Informe final y plan de remediación.

Gobernanza de la respuesta

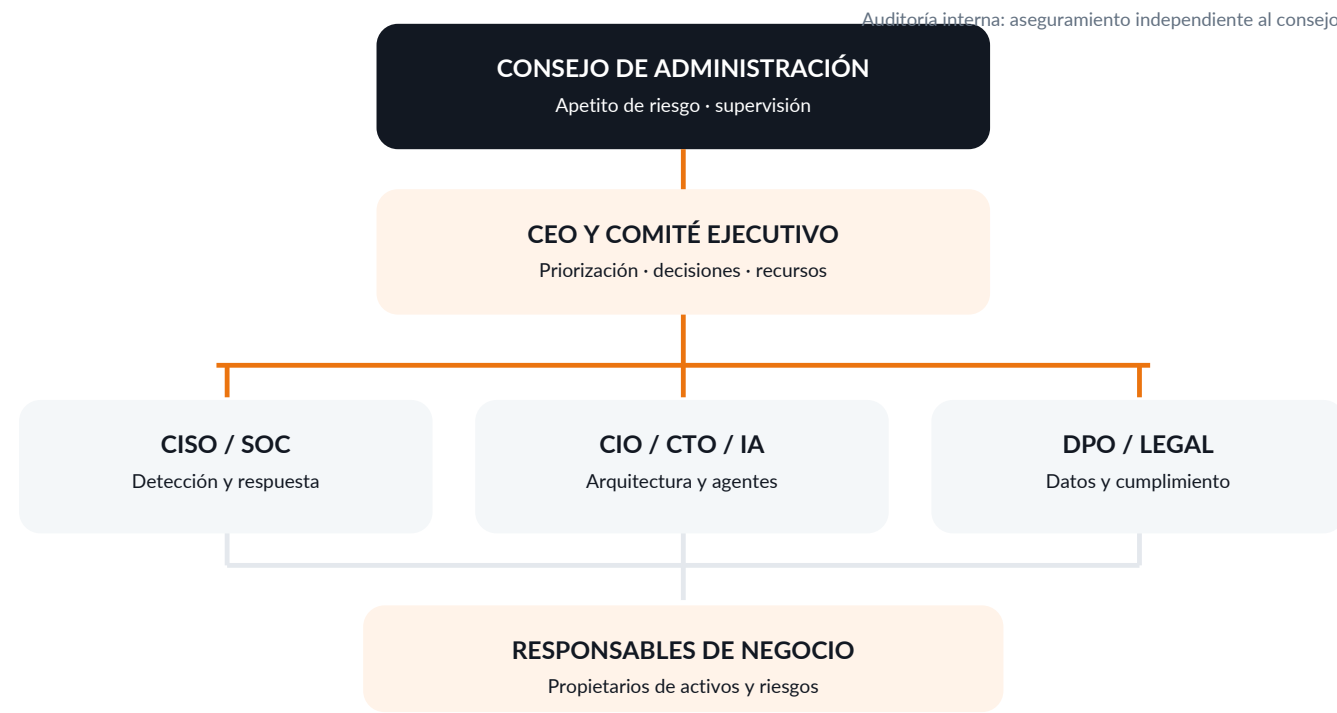
El CISO lidera investigación, contención y respuesta técnica; **Legal/DPO** califica las obligaciones; **Comunicación** gestiona interlocutores y reputación; **negocio** decide prioridades de continuidad; **el CEO** arbitra recursos y decisiones materiales. El consejo supervisa la exposición global y la eficacia de la recuperación.

Simulación recomendada: Introducir a los participantes un deepfake del CEO en plena crisis ransomware y una posible brecha de datos personales. Medir si se mantienen controles de autorización, evidencias y tiempos de comunicación.

05 / RESPONSABILIDADES

Una estructura de gobierno clara y verificable

Gobernanza eficaz significa definir quién asume el riesgo, quién ejecuta los controles y quién verifica su funcionamiento.



En cada proyecto de IA o nueva integración, la organización debe poder identificar un **propietario del proceso de negocio**, un **responsable técnico**, un **responsable de los datos** y un **aprobador de riesgos residuales**. Auditoría interna verifica el sistema de control sin sustituir a quienes operan el negocio.

Cadencia mínima propuesta

PERIODICIDAD	INSTANCIA	ENTREGABLE
Mensual	Comité operativo ciber + IA	Incidentes, vulnerabilidades, acciones, terceros y agentes.
Trimestral	Comité ejecutivo / Consejo	Top riesgos, capacidad de restauración y KPI con tendencias.
Semestral	Equipo de crisis / red team	Simulación documentada con medidas de corrección.
Anual	Consejo / Auditoría	Revisión de apetito de riesgo, inversión y aseguramiento.

06 / ESTRATEGIA DEFENSIVA

Arquitectura en capas: prevenir, responder, recuperarse

No hay una herramienta que por sí sola resuelva identidad, datos, cloud, activos industriales, terceros y seguridad de agentes.

01

IDENTIDAD

MFA resistente a phishing
Privilegios mínimos · PAM

02

OPERACIONES

Detección y contención
XDR/SOC · registros · pruebas

03

RESILIENCIA

Restauración comprobada
Backups · RTO/RPO · crisis

CAPACIDAD	PRIORIDAD	RESULTADO BUSCADO
SSO + MFA resistente a phishing + PAM	CRÍTICA	Impedir compromisos de cuentas y accesos privilegiados.
EDR/XDR + SIEM/SOC	MUY ALTA	Correlacionar y contener actividad sospechosa rápidamente.
CNAPP/CSPM + IAM cloud	MUY ALTA	Reducir configuraciones peligrosas y exposición de secretos.
Backups inmutables + restauración	CRÍTICA	Recuperar servicios sin depender del entorno comprometido.
DLP + clasificación de datos	ALTA	Limitar fugas por usuarios, SaaS o sistemas de IA.
Zero Trust + segmentación	ALTA	Frenar movimiento lateral y blast radius.
Red Team / Purple Team de IA	ALTA	Probar controles reales y escenarios de abuso.
Evaluación de terceros	ALTA	Disminuir dependencia y riesgo de concentración.

Las prioridades deben ajustarse a las pérdidas plausibles y al estado real de madurez, no al atractivo comercial de un producto. Las medidas reflejan buenas prácticas consistentes con NIS2 y NIST CSF 2.0. [11][14]

07 / MÉTRICAS

Un cuadro de mando que ayude al consejo a decidir

Los indicadores deben expresar exposición y capacidad operacional, evitando confundir el volumen de alertas con la reducción de riesgo.

KPI	QUÉ PERMITE SABER	SEÑAL DE DECISIÓN
Cobertura MFA privilegiada	Porcentaje de identidades privilegiadas protegidas.	Tratar excepciones como riesgo explícito.
MTTD / MTTR	Tiempo para detectar y contener incidentes.	Analizar degradaciones y causas.
Críticos fuera de SLA	Vulnerabilidades críticas sin remediación puntual.	Asignar recursos y aceptar riesgos.
Cobertura EDR/XDR	Activos críticos monitorizados de verdad.	Identificar “zonas ciegas”.
RTO / RPO probados	Capacidad restaurada respecto a objetivos acordados.	Reforzar continuidad si no se superan pruebas.
Proveedores Tier-1 evaluados	Terceros críticos con evaluación vigente.	Corregir concentración de riesgo.
IA inventariada y revisada	Agentes, datos y permisos cubiertos.	Suspender integraciones no gobernadas.
Ensayo 24 / 72 h	Capacidad legal/técnica de notificar.	Cerrar brechas de preparación.

Umbral: Cada KPI necesita una línea base real, un objetivo aprobado, responsable y frecuencia. No deben inventarse “valores recomendados” universales para todas las compañías.

Tres medidas que suelen anticipar problemas

Identidades sin dueño, restauraciones no probadas y agentes o integraciones con privilegios excesivos son señales de riesgo práctico. El dashboard debe señalar su evolución y permitir decisiones de reducción, transferencia, mitigación o aceptación documentada.

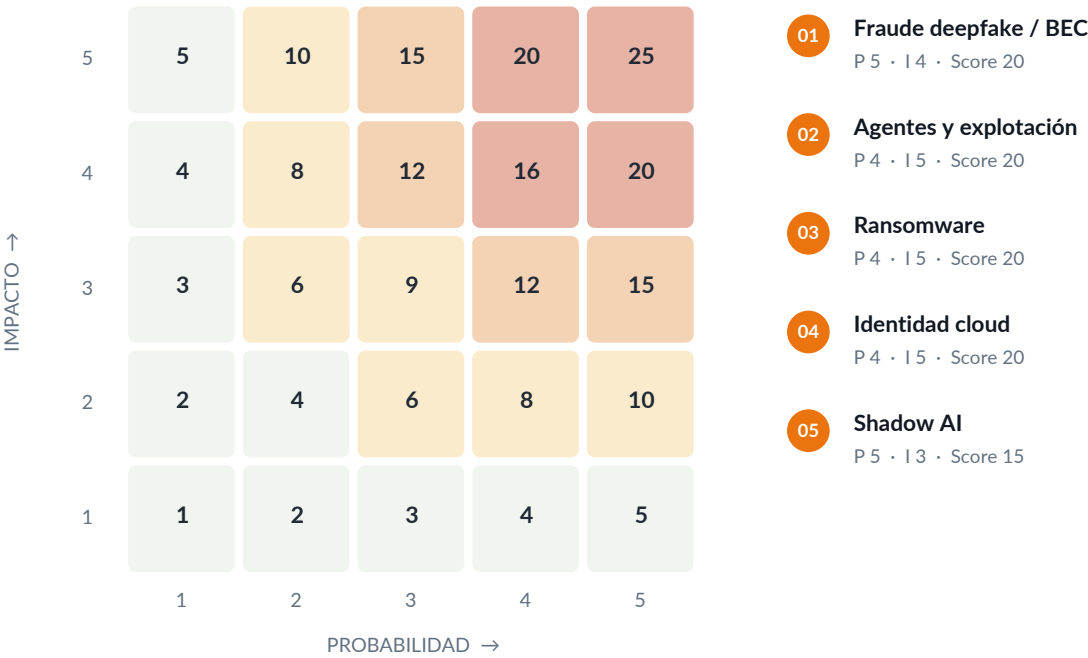
07 / MAPA DE DECISIÓN

Riesgos que merece la pena tratar primero

La matriz siguiente es un modelo cualitativo de trabajo: la escala 1–5 no debe interpretarse como probabilidad estadística.

Matriz de priorización cualitativa • impacto × probabilidad

Escala ordinal 1–5. Valores ilustrativos: requieren calibración por empresa.



Hipótesis inicial de riesgos prioritarios

RIESGO	P	I	P × I	PRIMERA RESPUESTA
Deepfake / fraude BEC	5	4	20	Control financiero fuera de banda.
Agentes / explotación	4	5	20	Parches + revocación de tokens.
Ransomware	4	5	20	Backups probados y segmentación.
Identidad cloud	4	5	20	MFA + PAM + hardening.
Shadow AI / fuga	5	3	15	Inventario IA + DLP + formación.

Calibración necesaria: sector, exposición, activos críticos, controles existentes, impacto económico y apetito de riesgo. Los números son supuestos ilustrativos diseñados para la conversación de dirección.

07 / INVERSIÓN

El presupuesto debe responder al riesgo evitado

No existe una cifra única responsable del porcentaje de gasto IT que toda compañía deba destinar a seguridad.

Gartner estimó para 2026 un gasto mundial en seguridad de la información cercano a 244.000 millones de dólares; esta cifra describe un mercado agregado y **no equivale a una recomendación presupuestaria para cada empresa.** [15]

Escenarios ilustrativos de dimensionamiento anual

EMPRESA	IT ANUAL SUPUESTO	SEGURIDAD / IT	RANGO RESULTANTE
Pyme · 10-49 personas	0,25-0,90 M€	7-12 %	20.000-110.000 €
Mediana · 50-249	1-4 M€	7-12 %	70.000-480.000 €
Grande · 250-999	5-20 M€	8-14 %	0,4-2,8 M€
Enterprise · ≥1.000	20-120 M€	8-15 %	1,6-18 M€+

Advertencia metodológica: los porcentajes, presupuestos IT por tramo e importes anteriores son **supuestos de planificación propios**. No proceden de una encuesta representativa de empresas españolas y no están atribuidos a Gartner. La cifra debe sustituirse por un modelo de riesgo del caso concreto.

Cómo estimar el retorno de la seguridad

ROSI = (pérdida anual esperada evitada – coste anual del control) / coste anual del control.

Ejemplo ilustrativo: 15 % de probabilidad anual × 4 M€ de impacto = 600.000 € de pérdida esperada. Si la inversión de 250.000 € reduce el riesgo un 60 %, la pérdida evitada es 360.000 € y el ROSI estimado es del **44 %**. Las probabilidades y reducciones son supuestos, no predicciones.

La decisión debe valorar asimismo continuidad de ventas, resiliencia de fabricación y distribución, obligación contractual, exposición personal de los administradores y efectos reputacionales. El coste menor no siempre implica el menor riesgo residual.

08 / EJECUCIÓN

90 días para pasar de intención a evidencia

La meta inicial no es alcanzar la madurez perfecta: es asegurar activos críticos, definir responsables y demostrar que la empresa puede resistir y recuperarse.



FASE	ENTREGABLE DE DIRECCIÓN	CRITERIO DE CIERRE
Día 30	Mapa de servicios críticos, identidades e IA; responsable ejecutivo.	Inventario aprobado y backups revisados.
Día 60	Plan priorizado de identidades, cloud, XDR, terceros y política IA.	Controles y responsables asignados.
Día 90	Tabletop, restauración de un servicio crítico y red team de IA.	Actas, resultados y plan de remediación.
Mes 12	Medidas integradas, pruebas repetidas y revisión independiente.	Dashboard y evaluación del riesgo residual.

08 / ATERRIZAJE SECTORIAL

Cómo cambian las prioridades según el negocio

La combinación de continuidad, exposición regulatoria y naturaleza de los activos exige planes sectoriales diferentes.

SECTOR	RIESGO MATERIAL	ACCIÓN PRIORITARIA
Manufactura e industria	Parada de planta, OT/IT, ingeniería y proveedores.	Segmentación OT/IT, backups de configuración, inventario de planta.
Servicios profesionales	Datos personales, secretos y fraude BEC.	MFA, DLP, verificación financiera y política IA.
Salud	Disponibilidad asistencial y datos sensibles.	Continuidad clínica, identidad, sistemas legacy, terceros.
Banca / seguros	Servicios financieros y dependencia TIC.	DORA, pruebas operacionales, registro de proveedores.
Retail y logística	Distribución, ecommerce y pagos.	RTO/RPO probados, POS, cloud, proveedores críticos.
Software y SaaS	Secretos, credenciales, API y cadena de desarrollo.	SDLC seguro, gestión de tokens, cloud, agentes y logging.

Para una empresa industrial de Castilla y León, por ejemplo, el escenario central puede ser la combinación de **ransomware en IT, indisponibilidad de un proveedor SaaS y degradación de la planificación de producción**. El diseño de continuidad debe identificar qué procesos pueden mantenerse en modo degradado y durante cuánto tiempo.

Una secuencia sostenible de 12 meses

Meses 1-3: identidad, inventario, continuidad, IA y respuesta. **Meses 4-6:** segmentación, endurecimiento cloud, evaluación de terceros y SOC. **Meses 7-9:** purple team, pruebas de respuesta y revisión de proveedores. **Meses 10-12:** auditoría independiente, actualización del apetito de riesgo y plan presupuestario del ejercicio siguiente.

09 / PREPARACIÓN

Tres ejercicios para comprobar la resiliencia

Los controles no deben darse por eficaces solo porque estén documentados: conviene forzar situaciones realistas, seguras y medibles.

ESCENARIO A • FRAUDE DEL CEO

Inyección: Un directivo sintético solicita un cambio urgente de cuenta bancaria de un proveedor de confianza.

Se evalúa: ¿Se aplica doble autorización? ¿Se verifica por un número previamente conocido? ¿Qué evidencia queda?

ESCENARIO B • AGENTE CONECTADO

Inyección: Un agente de IA lee un documento que incluye una instrucción hostil y trata de acceder o enviar datos no necesarios.

Se evalúa: ¿Se separan instrucciones y contenido? ¿El permiso técnico bloquea la acción? ¿Se detecta y registra?

ESCENARIO C • CRISIS MULTIFRENTE

Inyección: Una interrupción ransomware coincide con aviso de filtración de datos y caída de un SaaS crítico.

Se evalúa: ¿Cuál es el RTO probado? ¿Quién decide comunicados? ¿Se activan correctamente los plazos regulatorios?

La salida de cada ejercicio debe constar de: línea temporal, decisiones tomadas, controles efectivos y fallidos, tiempos reales, responsables y acciones de cierre. El objetivo es aprender, no señalar culpables.

Frecuencia orientativa: Al menos una simulación transversal semestral y pruebas técnicas con frecuencia adaptada al riesgo. La normativa sectorial puede establecer requerimientos distintos o superiores.

09 / CONSEJO

Las diez preguntas que el CEO debería llevar a la próxima reunión

- 01 ¿Cuáles son nuestros cinco procesos y activos realmente críticos?
- 02 ¿Qué impacto económico tendría su caída durante 24, 72 horas y una semana?
- 03 ¿Qué vulnerabilidades e identidades representan hoy el riesgo residual mayor?
- 04 ¿Cuándo restauramos por última vez un servicio crítico y cuánto tardamos?
- 05 ¿Quién puede ordenar pagos o cambios de cuenta y cómo se verifica su identidad?
- 06 ¿Qué sistemas de IA y agentes utilizamos, con qué datos y permisos?
- 07 ¿Podemos revocar en minutos el acceso de una identidad o agente comprometido?
- 08 ¿Qué proveedores pueden interrumpir el negocio y cuál es el plan alternativo?
- 09 ¿Sabemos quién decidiría y notificaría en las primeras 24 y 72 horas?
- 10 ¿Qué evidencia independiente demuestra la eficacia de nuestros controles?

Cierre estratégico

El reto no es eliminar todo el riesgo, sino administrarlo. Una organización preparada sabe qué proteger, cuánto riesgo acepta, qué controles funcionan y cuánto tarda en volver a operar cuando algo falla. En la era de los agentes de IA, la velocidad de detección y la recuperación probada son decisiones de negocio.

FUENTES Y NOTAS

Fuentes primarias y referencias

Selección de organismos públicos, textos normativos y fuentes primarias del período 2024–2026. Los enlaces están activos dentro del PDF.

- [1] INCIBE — Balance de ciberseguridad 2024. [Consultar fuente](#)
- [2] INCIBE — Balance de ciberseguridad 2025. [Consultar fuente](#)
- [3] ENISA — ENISA Threat Landscape 2026. [Consultar fuente](#)
- [4] AEPD — Primera notificación de brecha ejecutada por agente de IA, septiembre 2026. [Consultar fuente](#)
- [5] Comisión Europea — Action Plan on Cybersecurity and Artificial Intelligence, julio 2026. [Consultar fuente](#)
- [6] Reuters / Banco de Italia — Advertencia sobre deepfakes del gobernador, febrero 2026. [Consultar fuente](#)
- [7] OWASP — Top 10 for LLM Applications: Prompt Injection y herramientas. [Consultar fuente](#)
- [8] Reuters / Comisión Europea — Incidente cloud en el portal Europa, marzo 2026. [Consultar fuente](#)
- [9] OpenAI — The Hugging Face incident and the road ahead, agosto 2026. [Consultar fuente](#)
- [10] Reuters — Interrupción de servicios de Ahold, octubre 2026. [Consultar fuente](#)
- [11] EUR-Lex / BOE — Directiva (UE) 2022/2555 NIS2. [Consultar fuente](#)
- [12] Comisión Europea — NIS2: ámbito y obligaciones. [Consultar fuente](#)
- [13] Comisión Europea / EIOPA — DORA, AI Act y Cyber Resilience Act. [Consultar fuente](#)
- [14] NIST — Cybersecurity Framework 2.0. [Consultar fuente](#)
- [15] Gartner — Forecast Analysis: Information Security, Worldwide, 2026. [Consultar fuente](#)

Otras referencias normativas

- Reglamento (UE) 2024/1689 · AI Act. [Fuente oficial](#)
- Comisión Europea · Calendario del AI Act. [Fuente oficial](#)
- Comisión Europea · Cyber Resilience Act: reporting desde septiembre 2026. [Fuente oficial](#)
- INCIBE · NIS2: preguntas frecuentes. [Fuente oficial](#)

Nota de rigor. Este informe presenta información publicada hasta el 17 de octubre de 2026. Las normas deben confirmarse según el sector, tamaño empresarial y legislación española vigente antes de adoptar decisiones jurídicas. Los incidentes no se atribuyen a IA salvo que la fuente lo describa. Escenarios, matrices y presupuestos ilustrativos no constituyen predicciones estadísticas.

Convirtamos la ciberseguridad en una ventaja de resiliencia.

De la visibilidad del riesgo a la capacidad real de recuperación.

ELÚA CONSULTORES

Consultoría y formación en inteligencia artificial para empresas.

elua.es