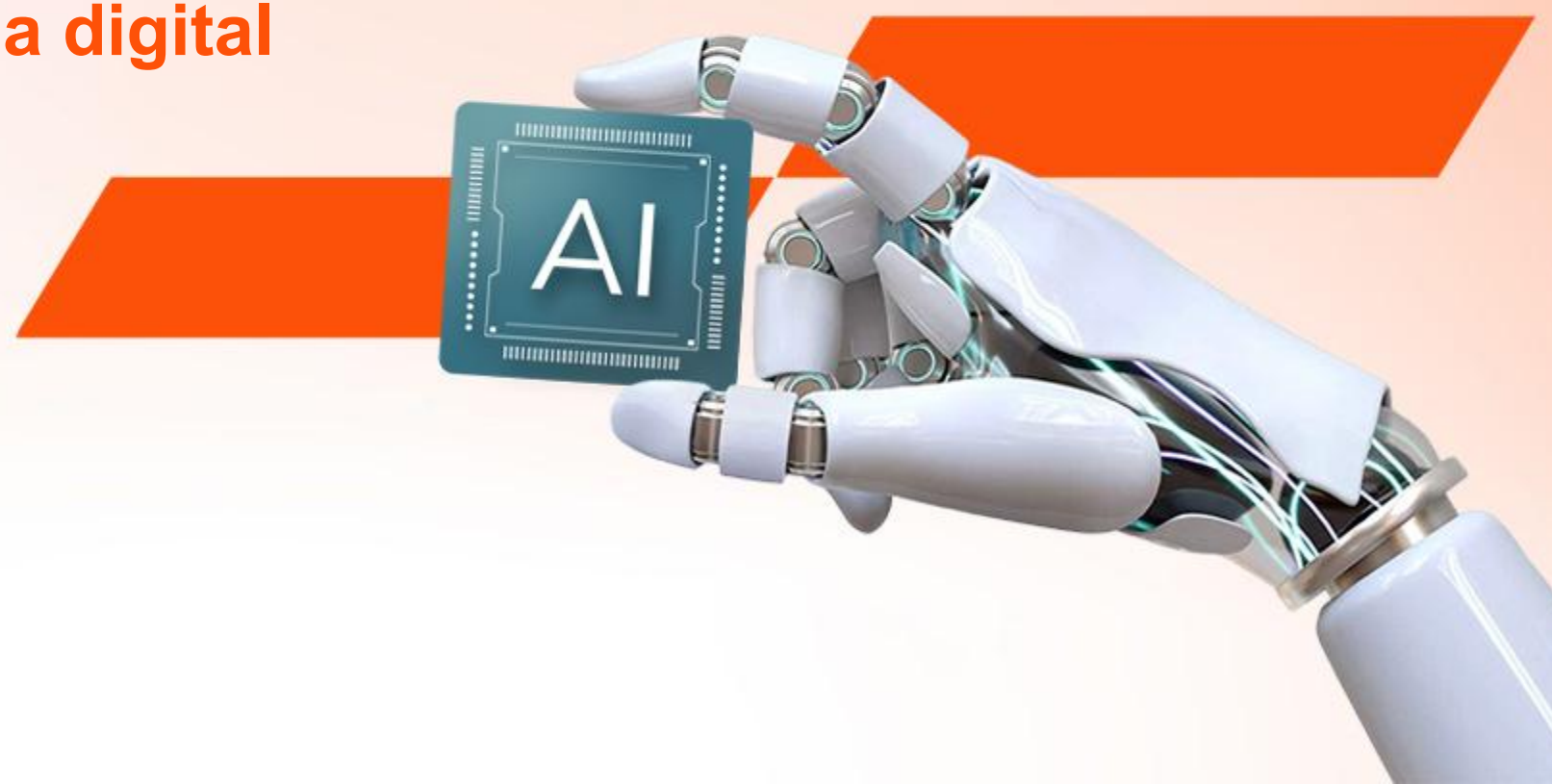
Ciberseguridad en la era de la Inteligencia Artificial

Cómo la IA está redefiniendo la ciberdefensa, el riesgo y la resiliencia digital



pwc

Índice

1

Impacto de los nuevos modelos de IA

Impacto de los nuevos modelos de IA

Modelos que están redefiniendo las reglas del juego

La IA ha pasado de ser una promesa tecnológica a un factor que redefine la superficie de ataque, la velocidad del cibercrimen y la propia estrategia de defensa

Anthropic ha presentado **Claude Mythos***, un modelo capaz de:

Claude Mythos

- Detectar y explotar **vulnerabilidades zero-day** en sistemas operativos y navegadores.
- Identificar **errores lógicos** que permiten saltarse mecanismos de 2FA o descifrar comunicaciones.
- Realizar **ingeniería inversa** de software propietario.
- Construir **ataques completos** a partir de un único prompt.

Pero, este modelo de no está solo, OpenAI ha respondido con GPT-5.5-Cyber* (especializado en Ciberseguridad).

Versiones preview.

Principales implicaciones

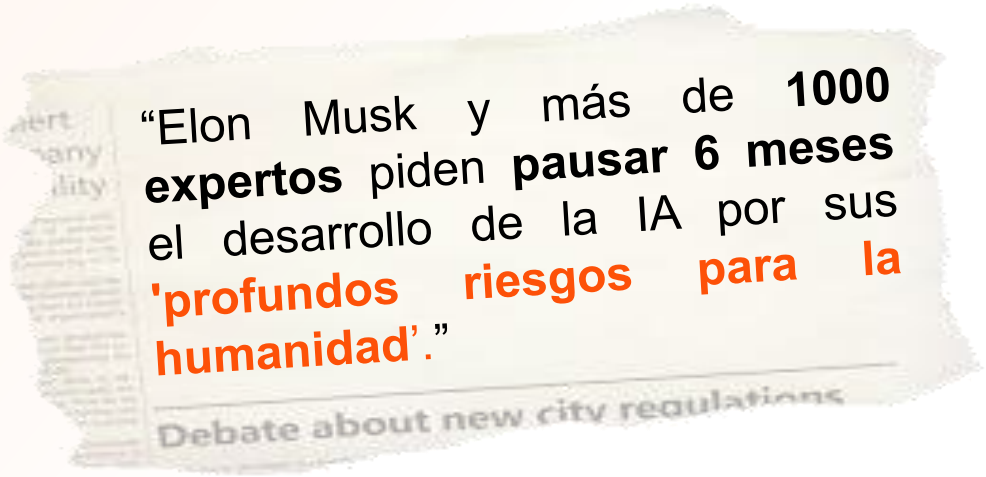
El tiempo para **parchear vulnerabilidades se reduce a horas**, no semanas.

La **barrera técnica** del cibercrimen se desploma: **un solo prompt basta**.

La misma IA que ataca, también defiende: nueva carrera estratégica.

Impacto de los nuevos modelos de IA

Cómo hemos llegado hasta aquí



"En **solo 6 meses**, **WormGPT y FraudGPT** alcanzan **+3000 suscriptores** en foros de la **dark web"**.

"En **12** meses, los ataques de **phishing** crecen un **+1265%"**.

Noviembre 2022

OpenAI libera **ChatGPT** al público general.

Julio 2023

Nace el **cibercrimen-as-a-service** con **IA.**

Febrero 2024

El primer gran **fraude** por **deepfake.**

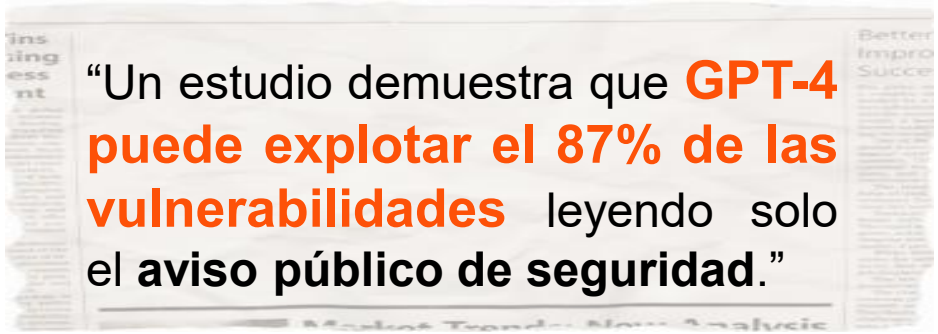
Marzo 2024

El Parlamento Europeo aprueba la **AI Act**, **primera ley integral** de IA del mundo.

Impacto de los nuevos modelos de IA

Cómo hemos llegado hasta aquí

"El **rendimiento** de los modelos de **OpenAI** en pruebas de Ciberseguridad salta del **27%** al **76%** en **solo 4 meses**".


"Un estudio demuestra que **GPT-4 puede explotar el 87% de las vulnerabilidades** leyendo solo el **aviso público de seguridad**."


"**Anthropic** alerta de que sus modelos más avanzados ya tienen **capacidades 'no triviales'** para asistir en la creación de **armas biológicas y químicas.**"

"**Mythos** encontró en **horas** un **bug zero-day de 27 años** que **herramientas tradicionales** no detectaron tras **millones de ejecuciones**".

Julio 2024

La **IA** tumba **8,5 millones de equipos** en 1 día.

Diciembre 2025

Aparecen los primeros **agentes IA** ofensivos.

2026

Llegan **Mythos** y **GPT-5.4-Cyber**: la IA entra en el frente ofensivo.

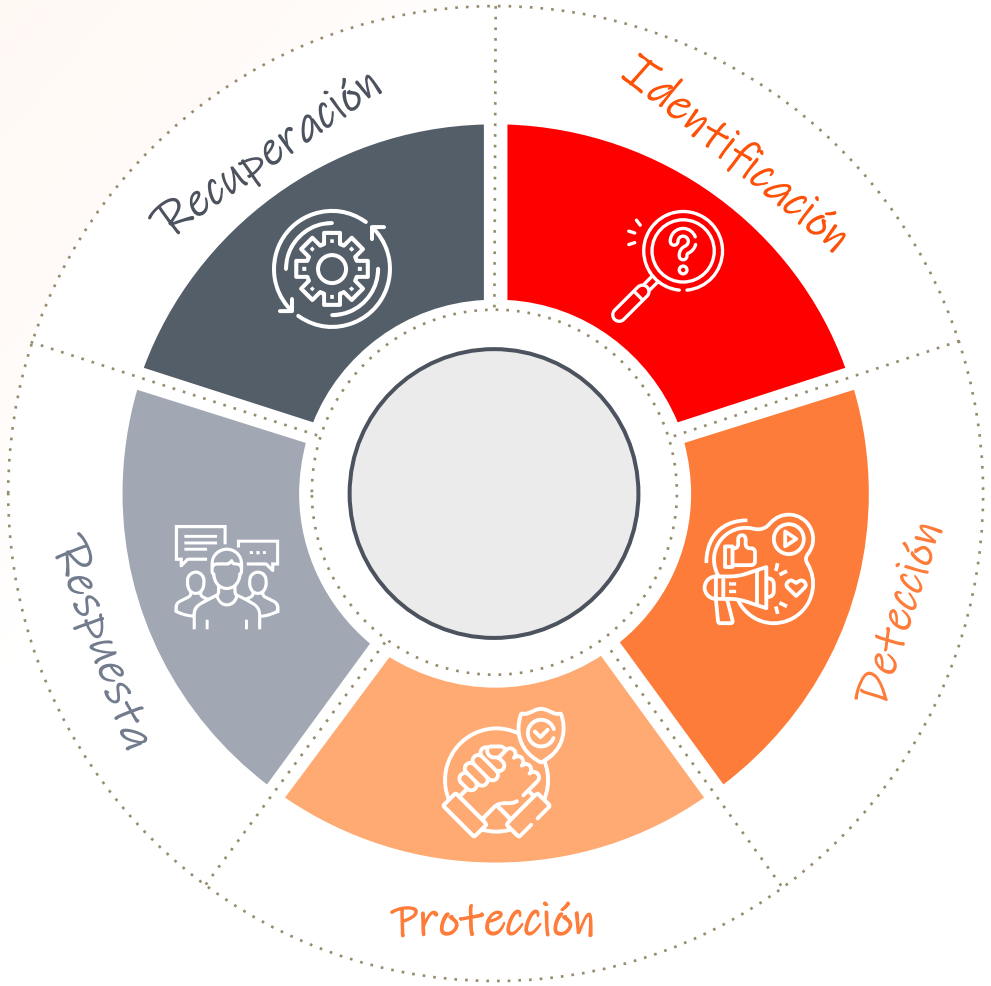
2

Transformación de la función de Ciberseguridad

Transformación de la función de Ciberseguridad

De la seguridad reactiva a la defensa aumentada por IA

La IA no sustituye a la Ciberseguridad: **la transforma desde dentro**, en cada uno de sus **5 dominios**:



Lo que la IA ya esta cambiando en Ciberseguridad:

- **De reactivo a predictivo**: la IA permite anticipar amenazas antes de que se materialicen.
- **De manual a autónomo**: tareas que tomaban horas se ejecutan en segundos.
- **De silos a inteligencia unificada**: la IA correlaciona señales que antes pasaban desapercibidas.

> "El **65%** de las organizaciones planean **integrar IA generativa** en al menos **3 de los 5 dominios NIST** en los próximos 18 meses".
>
> *(Gartner, 2026)*

Transformación de la función de Ciberseguridad

Nuevas capacidades habilitadas por la IA en cada dominio

Identificación

No puedes proteger lo que no conoces. La IA está acabando con los puntos ciegos.

Nuevas capacidades:

Descubrir

Identifica automaticamente activos cloud, on-prem, OT, IoT, etc., eliminando los puntos ciegos del inventario tradicional.

Clasificar

Comprende contenido y contexto del dato (PII, datos financieros, secretos e IP a escala empresarial).

Priorizar

Combina explotabilidad real, exposición e impacto de negocio para priorizar lo que importa hoy, no lo que dice el CVSS.

Identificar la propia IA

Aparece el AI BOM para inventariar y analizar los modelos que usa una organización, con qué datos, dónde corren y quién los consume.

Algunos ejemplos de herramientas de mercado:

Wiz, Axonius y Microsoft Defender CSPM

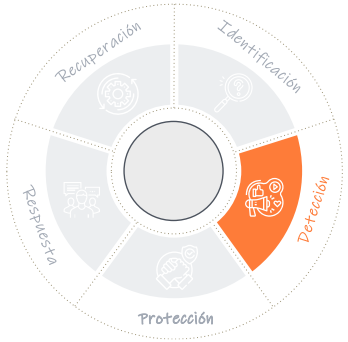
Microsoft Purview, BigID, Varonis, Defender XDR

Tenable One, CrowdStrike Falcon Surface, Defender XDR

Wiz AI-SPM, Defender for CloudApps Entra Agent ID, Purview, Defender CSPM

Transformación de la función de Ciberseguridad

Nuevas capacidades habilitadas por la IA en cada dominio



Detección

De millones de alertas a una detección más inteligente y priorizada.

Nuevas capacidades:

Detectar por comportamiento

Aprende el "**patrón**" normal de usuarios, dispositivos y aplicaciones, y detecta cualquier desviación.

***Hunting* conversacional**

Preguntan en lenguaje natural y la IA traduce a búsquedas optimizadas sobre toda la telemetría.

Correlacionar inteligencia

Combina telemetría interna con *threat intelligence* externa para entender el "por qué" de cada alerta.

Detectar ataques

Identifica ataques sofisticados: zero-days, movimiento lateral, exfiltración encubierta y abuso de modelos corporativos.

Algunos ejemplos de herramientas de mercado:

Sentinel UEBA, Darktrace, Vectra AI, ExtraHop Exabeam

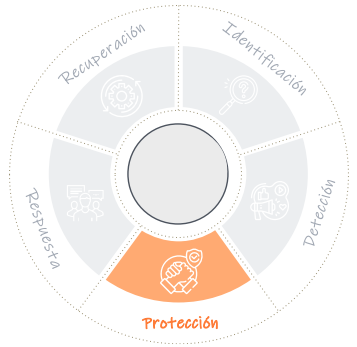
Microsoft Sentinel, Splunk, Google Chronicle, Security Copilot

Mandiant Recorded Future, Crowdstrike Falcon Intel, Defender XDR, Sentinel

Lakera Protect AI, HiddenLayer, Defender XDR, Sentinel

Transformación de la función de Ciberseguridad

Nuevas capacidades habilitadas por la IA en cada dominio



Protección

Cuatro capas de defensa que aprenden, se adaptan y evolucionan en paralelo.

Nuevas capacidades:

Blindar la identidad

Evalúa el riesgo de cada acceso en tiempo real combinando ubicación, dispositivo, comportamiento y contexto.

Filtrar el correo

Detecta BEC, *phishing* avanzado y *deepfakes* analizando patrones de comportamiento del remitente, mensajes y relaciones internas.

Defender el endpoint

Detiene malware desconocido y ransomware analizando comportamiento del proceso, protege frente a amenazas zero-day y polimórficas generadas por IA.

Proteger la propia IA

Aparece el AI Firewall para filtrar *prompts* maliciosos, bloquear fugas de datos y aplicar guardarraíles sobre la IA corporativa antes de que llegue al usuario.

Algunos ejemplos de herramientas de mercado:

Microsoft Entra, Okta AI CyberArk, Entra Agent ID

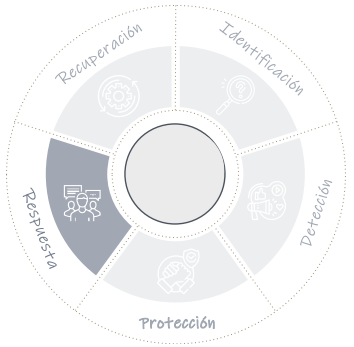
Abnormal Security, Proofpoint Mimecast, Defender XDR / Defender for Office 365

Crowdstrike Falcon, SentinelOne Defender, Defender for Endpoint

Lakera Protect AI, Prompt SecurityFoundry, Purview, Entra Global Secure Access, Agent 365

Transformación de la función de Ciberseguridad

Nuevas capacidades habilitadas por la IA en cada dominio



Respuesta

El SOC moderno no escala con más personas, sino con más inteligencia.

Nuevas capacidades:

Soporte al analista

Analiza incidentes en segundos, sugiere próximos pasos, traduce scripts maliciosos y guía al analista durante la investigación.

Automatizar la respuesta

Construye *playbooks* dinámicos según el tipo de incidente, ejecuta contención autónoma y orquesta acciones multi-cloud.

Triar y priorizar

Clasifica y prioriza alertas, según la criticidad del activo y el impacto potencial, y convierte eventos en una lista corta de incidentes accionables.

Comunicar al Negocio

Genera informes ejecutivos, comunicaciones a usuarios afectados y notificaciones regulatorias (NIS2, ENS, RGPD, etc.).

Algunos ejemplos de herramientas de mercado:

Microsoft Security Copilot, Charlotte AI, Purple AI, Defender XDR

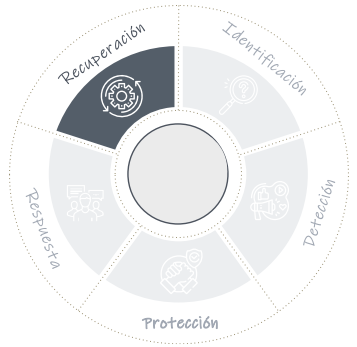
Palo Alto Cortex XSIAM, Google SecOps, Sentinel Playbooks

Splunk SOAR Swimlane, Devo Hunters, Sentinel Playbooks

Microsoft Security Copilot, ServiceNow

Transformación de la función de Ciberseguridad

Nuevas capacidades habilitadas y/o requeridas por la IA en cada dominio



Recuperación

La pregunta ya no es si nos atacarán, sino cuán rápido seremos capaces de volver.

Nuevas capacidades:

Restaurar inteligentemente

Prioriza qué restaurar primero analizando dependencias entre aplicaciones e impacto de negocio.

Garantizar *backups* limpios

Escanea cada copia de seguridad detectando ransomware, malware latente y anomalías antes de devolverla a producción.

Investigar y aprender

Reconstruye automáticamente la línea temporal del incidente correlacionando miles de eventos.

Mejorar continuamente

Genera post-mortems estructurados, identifica gaps en los controles, propone actualizaciones de playbooks y alimenta simulaciones.

Algunos ejemplos de herramientas de mercado:

Rubrik Security Cloud, Cohesity Commvault, Defender for Cloud

Veeam, Druva, Azure Backup + Defender for Cloud

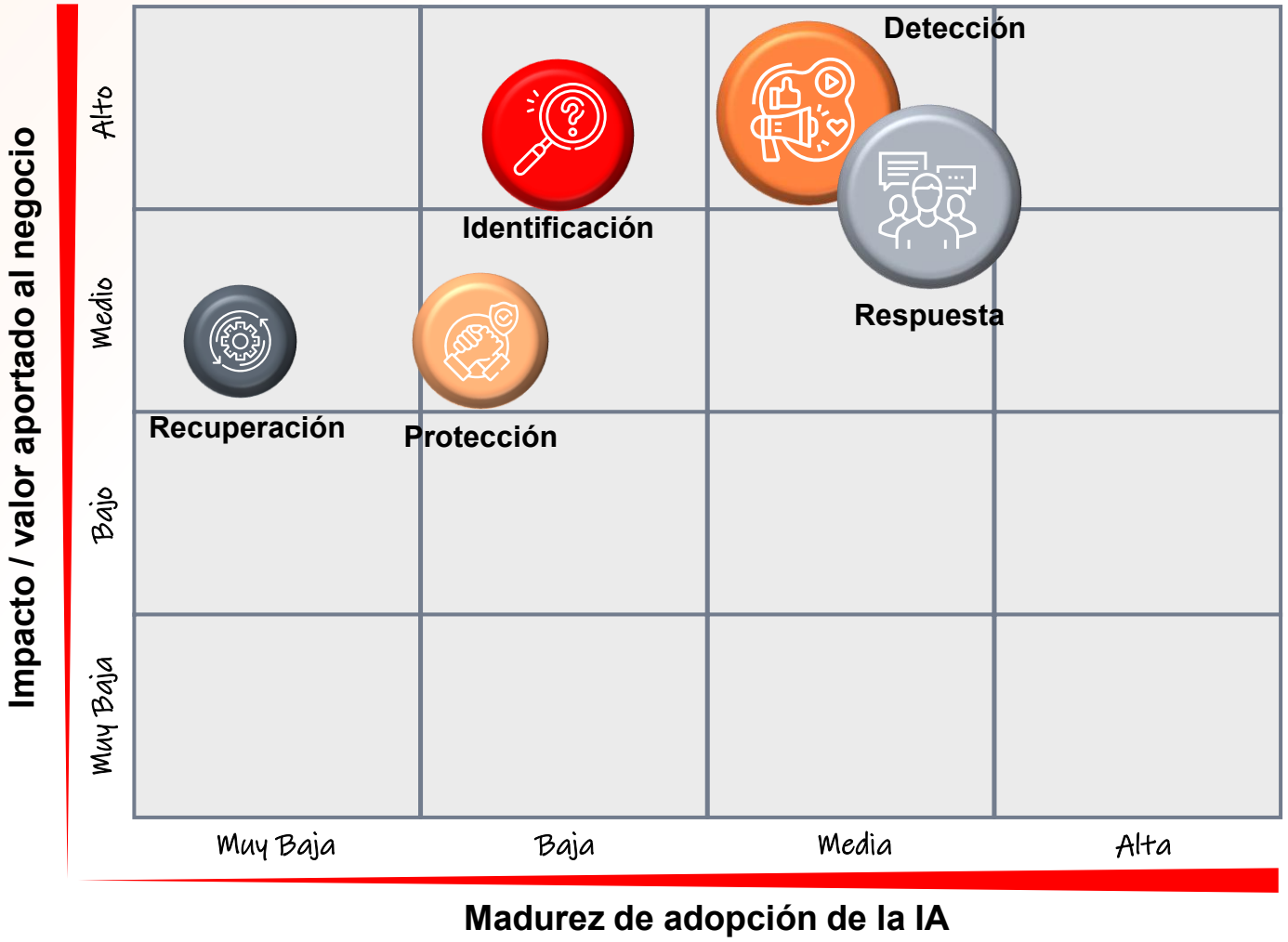
Magnet Forensics, Exterro Cellebrite, Defender XDR

Immersive Labs, SafeBreach, Security Copilot, Sentinel

Transformación de la función de Ciberseguridad

¿Dónde está aportando más valor la IA hoy?

Madurez de adopción vs. impacto en negocio



La **IA ya ha transformado cómo detectamos** y **respondemos** a las amenazas.

La verdadera ventaja competitiva estará en aplicarla donde aún no ha llegado: **anticipar mejor**, **proteger lo nuevo** y **recuperarnos más rápido**.

> "Solo **1/10** organizaciones **aplica IA** de forma madura a la **recuperación tras incidentes**".
>
> *(Gartner – Hype Cycle for Security Operations)*

***Tamaño de la burbuja**: Mercado actual.*

Transformación de la función de Ciberseguridad

¿Dónde está aportando más valor la IA hoy?

Madurez de adopción vs. impacto en negocio

Este posicionamiento no es arbitrario, refleja **dónde está hoy la tecnología**, **qué impacto está demostrando** en el negocio y **cuánto recorrido le queda**.

Identificación

- **Maduro en lo clásico** (CSPM, ASM, descubrimiento de activos), pero emergente en la gestión de riesgos de la propia IA.
- **Visibilidad insuficiente de los modelos y sistemas de IA** (solo el 23% tiene un inventario formal).

Madurez	Impacto
Baja	Alto

Detección

- **Tecnología consolidada** durante los últimos años (UEBA, NDR, XDR). El 65% de los SOCs ya integra IA.
- **Resultados medibles**, hasta **-80% en falsos positivos** y detección de zero-days.

Madurez	Impacto
Media	Alto

Protección

- **Coexisten controles tradicionales** con nuevas capas adaptativas con IA.
- **Aparece un mercado nuevo,** los AI Firewalls (proyectado a 5.000M € en 2027) o el SASE inteligente con IA (proyectado a 25.000M € en 2027).

Madurez	Impacto
Baja	Medio

Transformación de la función de Ciberseguridad

¿Dónde está aportando más valor la IA hoy?

Madurez de adopción vs. impacto en negocio

Este posicionamiento no es arbitrario, refleja **dónde está hoy la tecnología**, **qué impacto está demostrando** en el negocio y **cuánto recorrido le queda**.

Respuesta

- **Salto exponencial en los últimos meses**.
- **Impacto directo en productividad y velocidad**: -30% MTTR y +44% productividad del analista L1.

Recuperación

- **Despliegue en proceso**, solo el **12% aplica IA al *recovery***, frente al 65% en detección.
- **El siguiente gran salto**, la IA aplicada a resiliencia será la **próxima ola de inversión**.

Madurez	Impacto
Media	Alto

Madurez	Impacto
Muy Baja	Medio

3

Decálogo de Ciberseguridad del futuro

Decálogo de Ciberseguridad del futuro

De la defensa reactiva a la resiliencia inteligente

La Ciberseguridad del futuro no se construye con más herramientas, sino con principios:

1 Zero Trust by default

Verificación continua *e impulsada por* ***IA*** *en cada acceso.*

AI Governance **2**

Marco de gobierno *para el uso de IA interna y de terceros.*

3 Secure-by-design en IA

AI BOM, ***red teaming*** *de modelos y* ***AISecOps end-to-end***.

Gestión de identidades no humanas **4**

Gestión unificada de ***agentes IA***, ***máquinas*** *y* ***servicios*** *como ciudadanos de primera.*

5 Resiliencia cuántica

Roadmap de ***migración a criptografía post-cuántica*** *(PQC).*

Decálogo de Ciberseguridad del futuro

De la defensa reactiva a la resiliencia inteligente

La Ciberseguridad del futuro no se construye con más herramientas, sino con principios:

6 Defensa autónoma

***SOC** aumentado con **copilots y agentes IA** en detección y respuesta.*

Threat Intelligence colaborativa **7**

***Inteligencia** compartida y enriquecida con **IA**.*

8 Privacidad y soberanía del dato

*Control sobre **dónde y cómo** se procesan los **datos** en **entornos IA**.*

Cultura y skilling continuo **9**

***Formación permanente**: el factor humano sigue siendo clave.*

10 Regulación y ética

*Alineamiento con **AI Act, NIS2** y **DORA by design**.*

Muchas gracias