

Seguridad, gobierno e inteligencia aplicada a la IA

Neural Ghost ayuda a organizaciones que utilizan inteligencia artificial a conocer su exposición, probar sus sistemas, proteger los datos y mantener decisiones trazables.

01

Gobierno de IA

Inventario, riesgo y política

02

AI red team

Pruebas controladas y evidencias

03

Supervisión gestionada

Seguimiento y mejora continua

04

Fraude y confianza

Señales explicables y revisión humana

§01 EL CONTEXTO

La IA ya forma parte de la empresa, aunque nadie la haya inventariado

Copilotos, asistentes, agentes y herramientas con funciones generativas introducen nuevos flujos de datos y nuevas decisiones. La organización necesita saber qué utiliza, quién accede y qué puede fallar.

01

EXPOSICIÓN

Datos fuera de control

Información personal, contratos, credenciales o código pueden terminar en proveedores y extensiones que la organización no ha aprobado.

03

GOBIERNO

Decisiones sin trazabilidad

Una recomendación que no conserva fuente, versión y contexto resulta difícil de revisar, corregir o defender ante dirección, auditoría o un cliente.

02

ATAQUE

Sistemas fáciles de manipular

Una instrucción maliciosa dentro de un correo, una web o un documento puede alterar el comportamiento de un asistente o de un agente conectado a herramientas.

04

OPERACIÓN

Fraude más automatizado

Los atacantes combinan cuentas comprometidas, identidades sintéticas, automatización y campañas de suplantación para escalar el abuso.

OBJETIVO

Convertir la adopción de IA en un proceso visible, evaluable y seguro, con controles proporcionales y evidencias que permitan actuar.

§02 LA PROPUESTA

Tres servicios que cubren el ciclo completo

Cada servicio puede contratarse por separado. La combinación permite pasar del diagnóstico inicial a una operación continua y medible.

01

Evaluación de seguridad y gobierno de IA

Inventario de herramientas y casos de uso. Clasificación de riesgos, evaluación de proveedores, pruebas técnicas, política corporativa y plan corrector.

Resultado: mapa de IA, riesgos priorizados y hoja de ruta

02

Supervisión gestionada de IA

Revisión de nuevas herramientas, reevaluación de proveedores, pruebas periódicas, gestión de incidentes y seguimiento para dirección.

Resultado: control continuo y evidencias de mejora

03

Inteligencia antifraude y confianza digital

Correlación de señales operativas con cuentas comprometidas, dominios, campañas criminales y patrones anómalos, siempre con explicación y revisión humana.

Resultado: casos priorizados y decisiones mejor informadas

§03 SEGURIDAD Y GOBIERNO

Visibilidad sobre cada sistema, proveedor y flujo de datos

La evaluación combina gobierno, arquitectura, privacidad y pruebas técnicas. El alcance se adapta a las herramientas que la organización ya utiliza o quiere implantar.

QUÉ REVISAMOS

- Herramientas corporativas y usos no autorizados.
- Datos que entran, datos que salen y condiciones del proveedor.
- Modelos, asistentes, agentes, conectores y sistemas RAG.
- Permisos, registros, retención, secretos y control de acceso.
- Impacto sobre personas, clientes y procesos internos.

QUÉ ENTREGAMOS

- Inventario corporativo de sistemas y casos de uso de IA.
- Matriz de riesgo con responsables y prioridades.
- Política de uso y procedimiento de aprobación.
- Informe técnico con evidencias reproducibles.
- Plan corrector y resumen ejecutivo para dirección.

PRINCIPIO DE TRABAJO

La organización conserva el control sobre sus datos y sobre las decisiones. Neural Ghost documenta el alcance, las limitaciones y las evidencias de cada conclusión.

Los datos del cliente no se utilizan para entrenar modelos compartidos por defecto. Cualquier uso adicional requiere autorización expresa y un alcance contractual específico.

Pruebas controladas sobre LLM, RAG y agentes

Reproducimos ataques realistas dentro de un entorno autorizado. Cada hallazgo incluye evidencia, impacto, reproducción y recomendación de corrección.

01

Prompt injection

Ataques directos e indirectos desde documentos, webs, correos o contenido recuperado.

02

Divulgación de información

Extracción de datos sensibles, instrucciones internas, contexto y secretos.

03

RAG y embeddings

Poisoning, documentos maliciosos, permisos deficientes y recuperación entre usuarios.

04

Cadena de suministro

Modelos, librerías, prompts, conectores y componentes externos.

05

Tratamiento de salidas

Respuestas que activan código, consultas, scripts o acciones sin validación suficiente.

06

Agencia excesiva

Herramientas, acciones o conectores con más permisos de los necesarios.

07

Integridad y consumo

Manipulación de resultados, desinformación, abuso de recursos y denegación de servicio.

08

Reprueba

Validación de las correcciones y cierre documentado de los hallazgos.

§05 OPERACIÓN CONTINUA

La adopción cambia cada semana. El control también debe cambiar

La supervisión gestionada mantiene actualizado el inventario, revisa nuevos proveedores y convierte los hallazgos en acciones con responsables y plazos.

01

Entrada

Solicitud de nueva herramienta, cambio de proveedor, nuevo agente o incidente.

02

Evaluación

Revisión de datos, permisos, arquitectura, contrato y riesgo del caso de uso.

03

Decisión

Aprobación, condiciones de uso, controles compensatorios o rechazo motivado.

04

Seguimiento

Pruebas periódicas, incidentes, cambios de modelo y evidencias para dirección.

ENTREGABLES RECURRENTES

Registro vivo de IA · comité de seguimiento · informe de riesgos · control de acciones · apoyo ante incidentes · formación y actualización documental

Inteligencia explicable para priorizar operaciones sospechosas

La correlación combina señales internas autorizadas con inteligencia de amenazas. El objetivo consiste en reducir investigación manual y detectar relaciones que una regla aislada no muestra.

01 SEÑALES INTERNAS

Reservas, accesos, dispositivos, cambios de cuenta, incidencias y resultados conocidos.

02 INTELIGENCIA EXTERNA

Credenciales, dominios, campañas, infraestructura criminal, fuentes abiertas y dark web.

03 CORRELACIÓN

Reglas, modelos y contexto convierten eventos dispersos en un caso comprensible.

04 REVISIÓN HUMANA

El equipo valida el caso antes de una decisión con impacto sobre una persona u operación.

CASOS DE USO

Cuentas comprometidas

Cambios, accesos y relaciones con credenciales expuestas.

Suplantación

Dominios, perfiles y campañas que imitan a la organización.

Automatización abusiva

Patrones repetitivos, bots y consumo anómalo de recursos.

Investigación reforzada

Contexto dark web, OSINT y blockchain cuando resulte pertinente.

S07 LA PLATAFORMA

Ghostnet convierte señales dispersas en evidencia utilizable

La arquitectura desacopla fuentes, análisis y entrega. Cada cliente recibe un entorno delimitado por alcance, permisos, retención y requisitos de tratamiento.

01

Fuentes autorizadas

Datos del cliente, fuentes abiertas, inteligencia de amenazas y telemetría prevista en el contrato.

02

Normalización y aislamiento

Validación de formato, control de acceso, trazabilidad y separación lógica por cliente.

03

Ghostnet

Correlación, enriquecimiento, relaciones, scoring y contexto operativo.

04

DarkForensic-7B

Análisis especializado en castellano con apoyo a investigación y priorización.

05

Evidencia y entrega

Hallazgos, fuentes, hash, informe, API y salidas para los sistemas del cliente.

508 CÓMO TRABAJAMOS

Un proceso con alcance, evidencias y cierre

Cada intervención comienza con autorización y termina con resultados priorizados, responsables y una revisión de las medidas implantadas.

01	02	03	04	05
Alcance	Descubrimiento	Evaluación	Corrección	Validación
Sistemas, datos, personas, objetivos y límites.	Inventario, arquitectura, proveedores y dependencias.	Riesgo, pruebas técnicas y evidencia reproducible.	Controles, responsables, prioridades y plazos.	Reprueba, cierre e informe para dirección.

ENTREGABLES

- Inventario de sistemas y casos de uso.
 - Mapa de arquitectura y flujo de datos.
 - Matriz de riesgos y plan corrector.
 - Informe técnico con evidencias y reproducción.
- Política corporativa y procedimiento de aprobación.
 - Protocolo de incidentes relacionados con IA.
 - Resumen ejecutivo para dirección.
 - Sesión de cierre y hoja de ruta priorizada.

MODALIDADES

Diagnóstico inicial	Piloto controlado	Supervisión continua
Punto de partida y priorización.	Validación sobre un alcance concreto.	Seguimiento, pruebas y gobierno.

Tecnología propia, responsabilidad definida

PRINCIPIOS

- Separación por cliente y mínimo privilegio.
- Uso de datos limitado al alcance contratado.
- Evidencias con fuente, fecha, versión e integridad.
- Supervisión humana ante decisiones de impacto.
- Retención, exportación y borrado definidos.

MARCOS DE REFERENCIA

- Reglamento europeo de inteligencia artificial.
- RGPD y LOPDGDD.
- ISO/IEC 42001 e ISO/IEC 23894.
- ISO/IEC 27001 y NIST AI RMF.
- OWASP Top 10 para LLM e IA generativa.

La alineación con estos marcos no implica certificación. El alcance normativo se determina para cada organización y caso de uso.

¿Quieres saber qué IA utiliza tu organización y qué riesgo introduce?

Empezamos con una conversación de alcance. Identificamos el primer caso que merece evaluación y planteamos un diagnóstico o piloto controlado.

neural@neural-ghost.com

neural-ghost.com