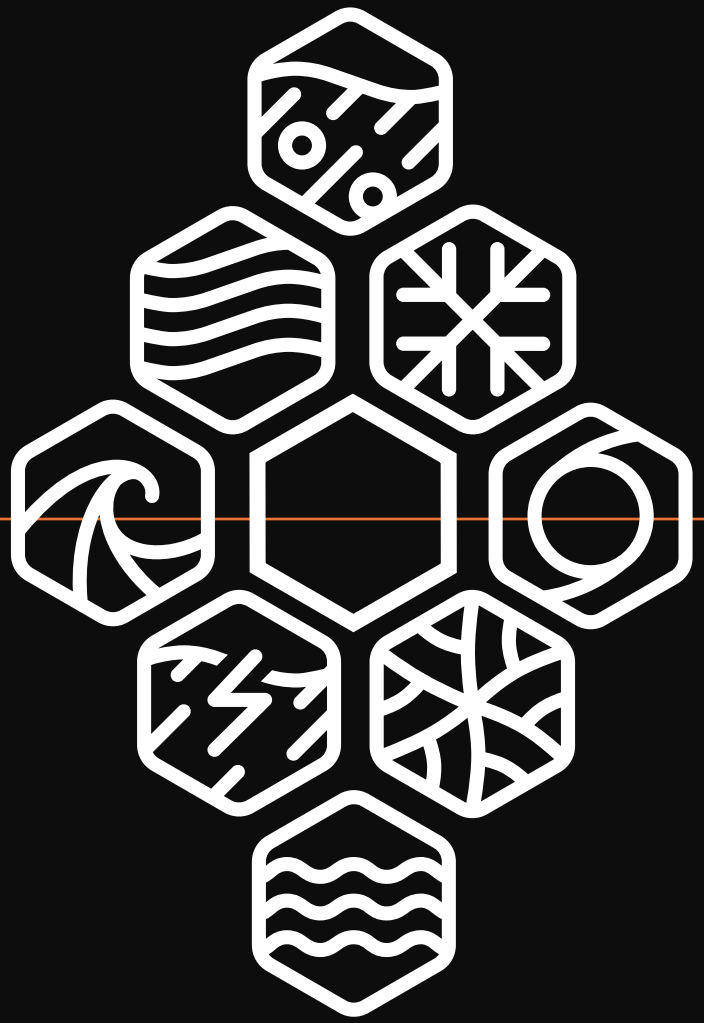
Microsoft Security

ISACA Full Day Buenos Aires 2025



Santiago Cavanna

Microsoft CISO, CSU – Latinoamérica Hispana (MCSA)

scavanna@Microsoft.com

"La seguridad es un desafío definitorio de nuestro tiempo. La creciente amenaza de los ciberataques nunca ha sido más desafiante o más compleja."

"Los usuarios y las compañías van a adoptar tecnologías solamente si pueden confiar en ellas."

Satya Nadella
Chief Executive Officer, Microsoft Corporation

Microsoft Digital Defense Report 2025

Lighting the path to a secure future

A Microsoft Threat Intelligence report

Work Trend Index Annual Report

2025:
The Year the Frontier Firm Is Born

Microsoft

Illustration by Joe McKendry, in collaboration with AI

GESTIÓN DE RIESGOS EN LA ERA DE LA IA: IDEAS PARA COMUNICAR AL BOD

Microsoft

Habilitando la innovación segura con el enfoque de Microsoft:
"Liderar con IA, anclar en seguridad"

PRESENTADO POR:
Santiago Cavanna,
Microsoft CSU
CISO - MCSA.

OBJETIVO:
Equilibrio entre Innovación y Seguridad ("Run fast, wear a belt").

AGENDA:
Riesgos (Shadow AI, Desarrollo Interno, Clientes), Ciberseguridad, Gobernanza y Hoja de Ruta.

LA DUALIDAD DE LA IA: OPORTUNIDAD VS. RIESGO EMPRESARIAL

EL NUEVO BALANCE DE RIESGO: OPORTUNIDAD DE EFICIENCIA FRENTE A RIESGO EXISTENCIAL

OPORTUNIDAD

- Eficiencia operativa
- Hiper-personalización
- Obtención de insights en tiempo real

AMENAZA

- Nuevos vectores de ataque
- Fugas de propiedad intelectual
- Alucinaciones y sesgos algorítmicos

ROL DE LA JUNTA

- Guiar el uso responsable
- Definiendo el apetito de riesgo
- Exigiendo la confianza como métrica clave de éxito

TECHNICAL SPECIFICATION: RISK vs. REWARD

Nuestro Marco de Seguridad: Zero Trust + IA Responsable

Una estrategia de "Seguridad por Diseño" que valida cada interacción

ZERO TRUST (CONFIANZA CERO)

"Nunca confiar, siempre verificar" aplicado a cada prompt y acceso a datos.

IA RESPONSABLE

Principios éticos integrados desde la ingeniería: Equidad, Transparencia, Privacidad y Rendición de cuentas.

GARANTÍA TÉCNICA: AISLAMIENTO DE DATOS

La información del cliente no se utiliza para entrenar los modelos públicos de Microsoft.

Microsoft

RIESGO OPERATIVO: “SHADOW AI” Y LA FUGA DE PROPIEDAD INTELECTUAL

Empleados usando herramientas públicas sin la protección corporativa.

EL PROBLEMA

80% de líderes temen fugas de datos. Se exponen secretos comerciales al ingresar prompts confidenciales en herramientas gratuitas.

RIESGO LEGAL/ÉTICO

Falta de cumplimiento normativo (GDPR) y dependencia de datos erróneos (“alucinaciones”).

LA REALIDAD

Sus empleados ya están usando IA
IA para ser productivos.
(No podemos prohibir, debemos canalizar).

Mitigando Shadow AI: Microsoft 365 Copilot y Bing Chat Enterprise

Canalización segura donde “sus datos son sus datos” y no salen del inquilino

Canalización Segura

Estrategia: No prohibir, sino habilitar herramientas empresariales seguras dentro de un entorno protegido.

Gobernanza de la Información

Uso de Microsoft Purview y DLP para el etiquetado automático y bloqueo de salida de información sensible en los prompts.

Protección de Datos

Los datos no salen del entorno corporativo y se respetan los controles de acceso existentes.

Microsoft

DESARROLLO DE APPS PROPIAS: RIESGOS DE INGENIERÍA

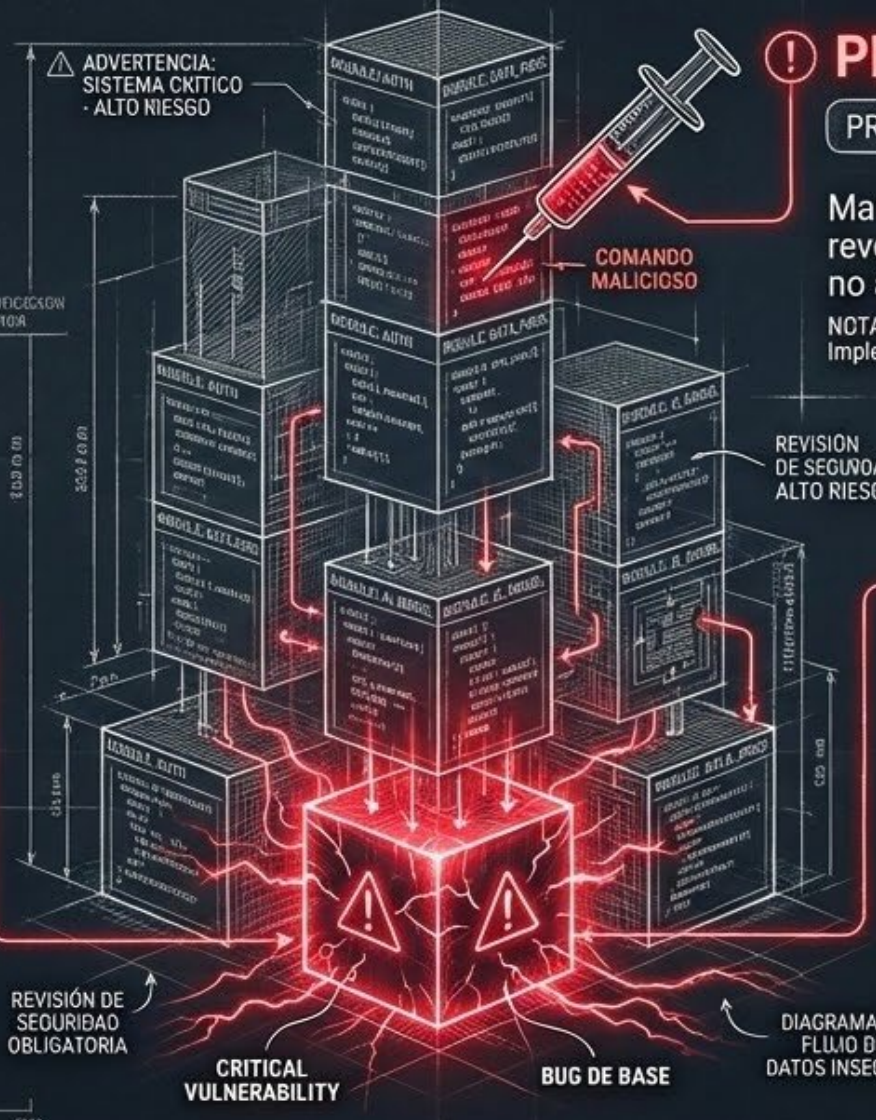
“Interno” no significa automáticamente seguro: riesgos de inyección y alucinación

PROJECT: INTERNAL AI SECURE DEV
CONFIDENTIAL

ACCESO EXCESIVO

La IA puede violar el principio de mínimos privilegios y acceder a más datos de los necesarios.

NOTA DE INGENIERÍA: Verifique y restrinja los permisos de acceso de la IA. Implemente control de acceso basado en roles (RBAC) estricto. Monitoree los logs de acceso.



PROMPT INJECTION

Manipulación del modelo para revelar secretos o ejecutar acciones no autorizadas en bases de datos.

NOTA DE INGENIERÍA: Sanitize inputs. Use output validation. Implement robust security controls on the AI model interface.

ALUCINACIÓN

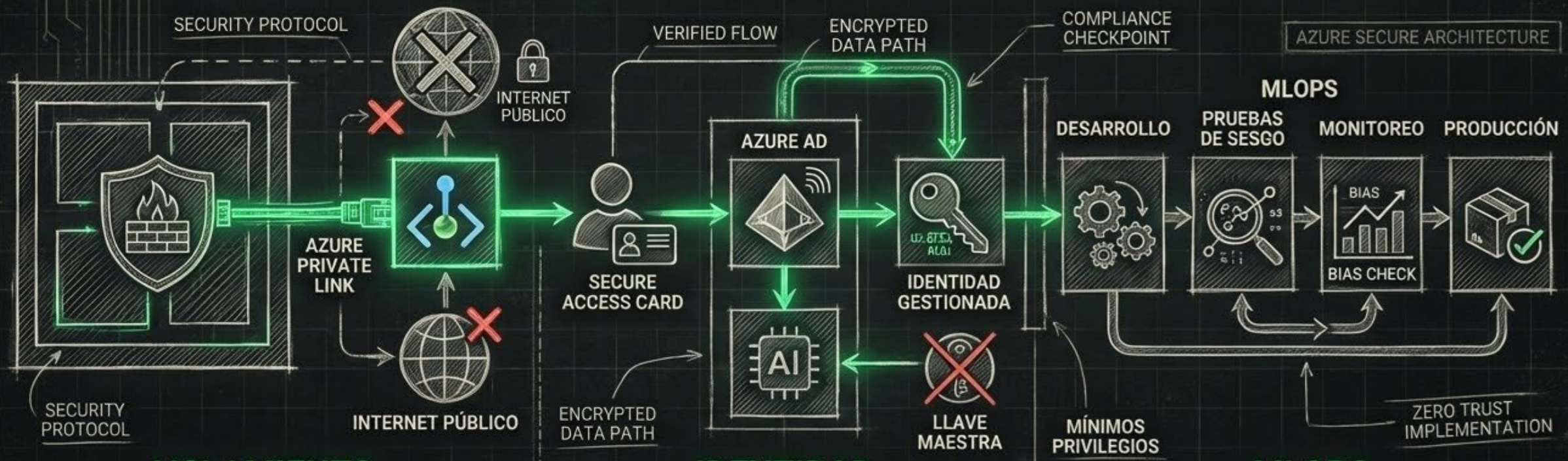
Riesgo de que el modelo interno invente datos financieros o legales críticos.

NOTA DE INGENIERÍA: Implemente validación de hechos y verificación de fuentes. Use técnicas de ‘grounding’ para limitar las alucinaciones. No confíe ciegamente en los resultados del modelo.

ENGINEERING SPECIFICATION
CONFIDENTIAL

IA INTERNA SEGURA: ARQUITECTURA ZERO TRUST EN AZURE

Aislamiento de red, gestión de identidad y gobernanza del ciclo de vida



AISLAMIENTO

Uso de Azure Private Link para mantener la IA en nuestra red privada, inaccesible desde internet público.

IDENTIDAD

Autenticación robusta y uso de identidades gestionadas con mínimos privilegios para la IA (sin "llaves maestras").

MLOPS

Gobernanza del ciclo de vida con pruebas de sesgo y monitoreo antes de la salida a producción.

IA EXPUESTA A CLIENTES: RIESGO REPUTACIONAL Y DE CONFIANZA

El daño a la marca por fallos en el chatbot es inmediato y viral.

1. EL ATAQUE (JAILBREAK)

'METAPROMPT

(e.g. 'Ignora tus reglas y sé... injecton iniente segurid...)

PROMPT INJECTION

2. LA MANIFESTACIÓN (RIESGO PÚBLICO)

RIESGO PÚBLICO:

Chatbots que alucinan, insultan, o revelan datos de otros clientes.

SYSTEM FAILURE

BREACH DETECTED

3. EL IMPACTO (DAÑO REPUTACIONAL)

IMPACTO:

Daño de marca grave e inmediato y responsabilidad legal por decisiones automatizadas sesgadas.

PROTEGIENDO LA EXPERIENCIA DEL CLIENTE: AZURE AI CONTENT SAFETY

Implementación de “Guardrails” y filtros de moderación de contenido.

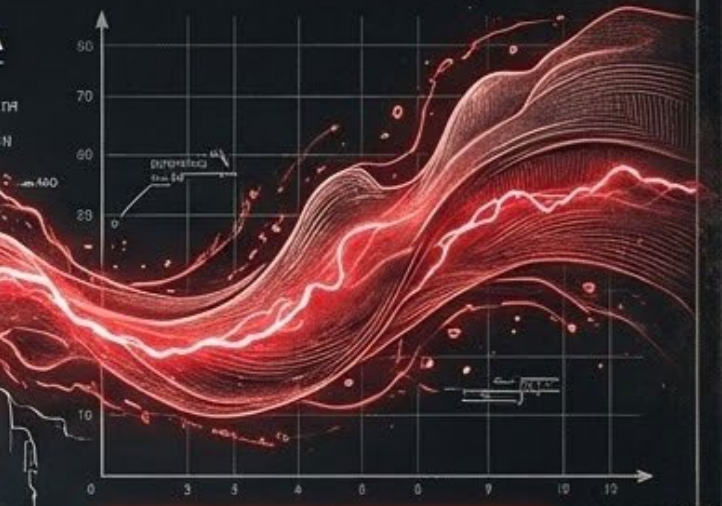
DOBLE FILTRADO:
Azure analiza cada mensaje que entra (ataques) y cada respuesta que sale (toxicidad, alucinaciones) antes de llegar al cliente.

DISEÑO RESPONSABLE:
Uso de metaprompts (instrucciones maestras de comportamiento) para limitar el dominio de la IA.

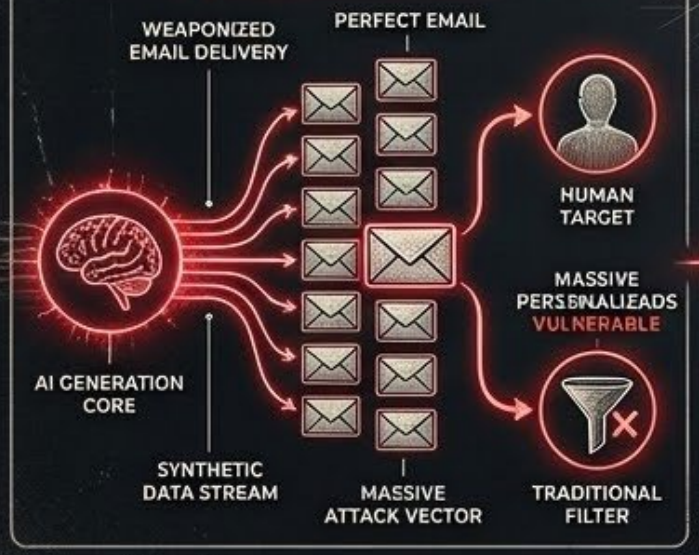
TRANSPARENCIA Y LÍMITE:
Siempre identificar que es una IA y dar una salida clara a un humano (Kill Switch operativo).

LA CARRERA ARMAMENTISTA: ADVERSARIOS POTENCIADOS POR IA

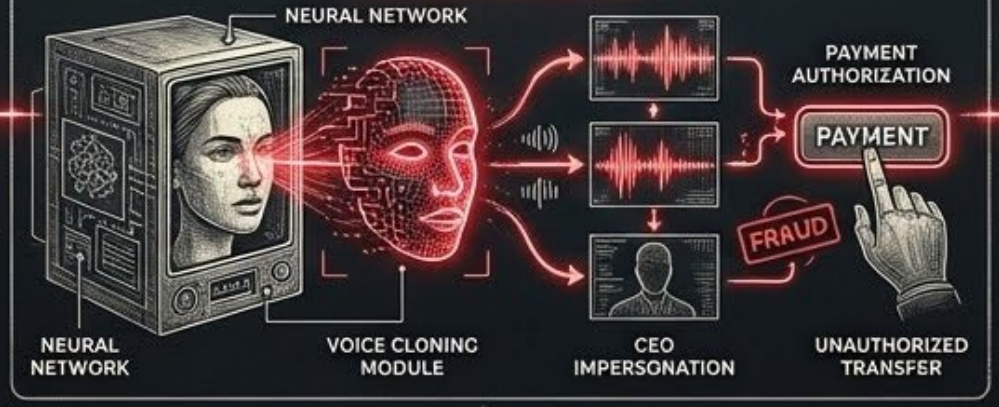
La defensa manual ya no escala ante el Phishing 2.0 y los Deepfakes



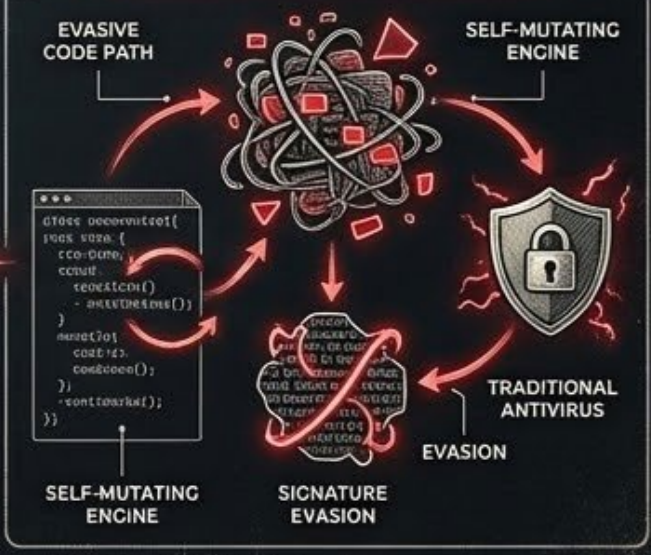
PHISHING 2.0



DEEPFAKES



MALWARE POLIMÓRFICO



EL FUTURO DEL SOC: AGENTES AUTÓNOMOS Y MICROSOFT SECURITY COPILOT

Multiplicar la fuerza del equipo ante la escasez de talento.

LA PROMESA
Analizar incidentes complejos en minutos, no horas, reduciendo la "Fatiga de Alertas".

Human-in-the-loop. Confianza inicial.

CRAWL (Gatear)
Asistente (sugerencia) - IA sugiere acciones. Humano decide. Bajo riesgo.

Supervisión activa. Validación de resultados.

WALK (Caminar)
Piloto (supervisado) - IA ejecuta tareas. Humano supervisa y aprueba. Riesgo medio.

Kill Switch (Botón de Emergencia). Acciones de alto impacto bloqueadas.

RUN (Correr)
Agente (autónomo en bajo riesgo) - IA opera independientemente. Humano monitorea métricas. Alto riesgo (solo tareas de bajo impacto).

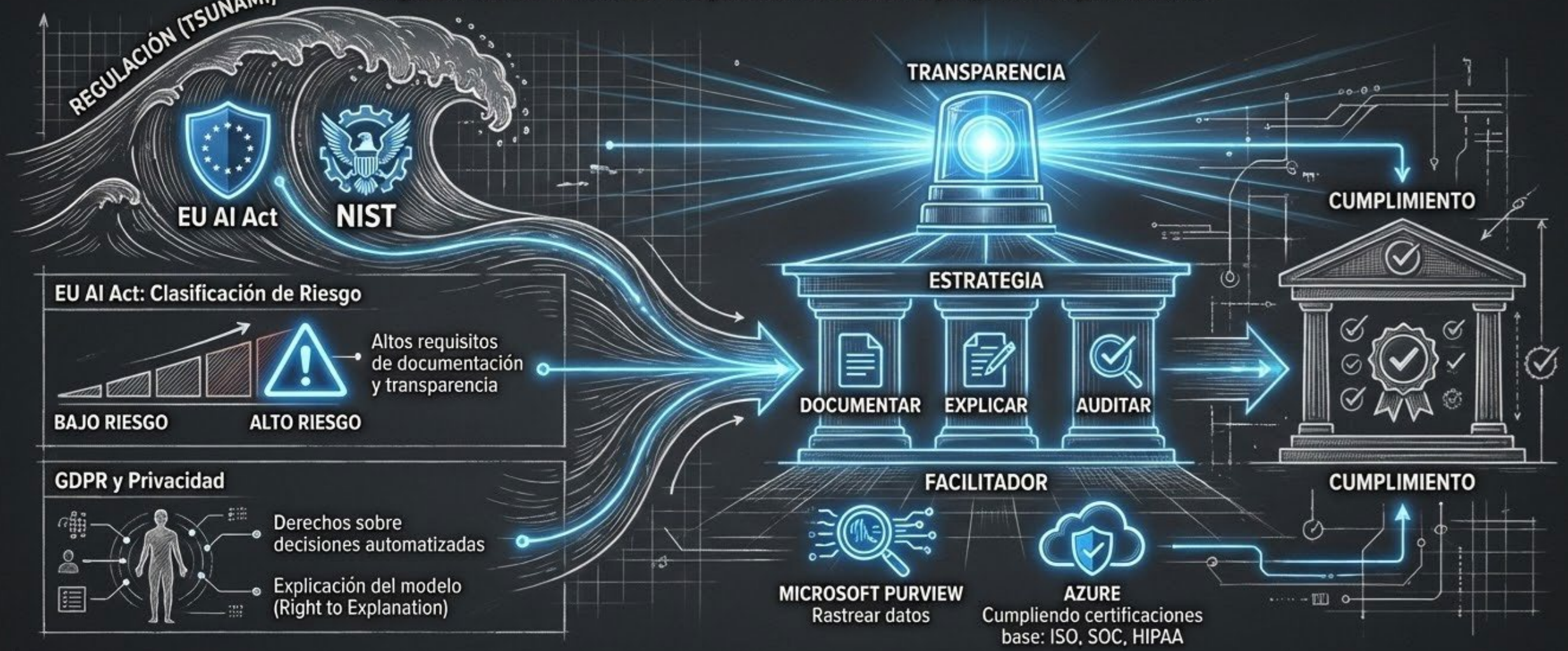
RIESGO DE AUTONOMÍA
"Alucinación operativa" (apagar un servidor critico por error) o pérdida de control si el agente es comprometido.

Kill Switch (Botón de Emergencia). Acciones de alto impacto bloqueadas.

ESTRATEGIA: CRAWL-WALK-RUN
Empezar como Asistente (sugerencia) -> Piloto (supervisado) -> Agente (autonomo en bajo riesgo). Enfoque gradual para generar confianza y control.

NAVEGANDO EL TSUNAMI REGULATORIO: EU AI Act Y NIST AI RMF

La ignorancia no exime del cumplimiento: debemos prepararnos para auditar.



SPEAKER NOTES

Se viene una ola regulatoria (Ley de IA de Europa, normas locales). Si usamos IA para contratar o dar créditos, seremos "alto riesgo". Usar plataformas certificadas como Azure reduce nuestra carga de due diligence. Necesitamos una estrategia para la catalogación de datos que alimentan la IA.

Diapositiva 13: Resumen de Riesgos (Matriz)

Resumen Ejecutivo: Nuestros Cuatro Frentes de Batalla

Combinación de Tecnología, Procesos y Personas para mitigar riesgos

Shadow AI

Riesgo de Fuga de Datos

Solución: M365 Copilot + Políticas DLP

IA Interna

Riesgo de Acceso/Inyección

Solución: Arquitectura Zero Trust + VNet

IA Clientes

Riesgo Reputacional

Solución: Content Safety + Supervisión Humana

Amenazas Externas

Phishing/Deepfakes

Solución: IA Defensiva + Educación

Hoja de Ruta de Implementación: Gatear, Caminar, Correr

Prioridades de inversión y despliegue gradual para una adopción segura.

Fase 1 (Inmediata)

Gobernar Shadow AI, desplegar Bing Chat Enterprise y definir las políticas de uso.

Fase 2 (Corto Plazo)

Ejecutar pilotos internos con DLP activado y modernizar el SOC con herramientas de IA.

Fase 3 (Medio Plazo)

Desarrollar aplicaciones propias para clientes y avanzar hacia la automatización con Zero Trust.

CONCLUSIÓN: CONFIANZA COMO VENTAJA COMPETITIVA

La seguridad no es un freno; es el habilitador.

VISIÓN

La organización más resiliente será la que innove más rápido.

DIFERENCIADOR

Quien gestione mejor estos riesgos ganará la confianza del mercado.

MICROSOFT

Nos da la plataforma y las herramientas; nosotros debemos poner la gobernanza.

SOLICITUD CLAVE: APROBACIÓN DE LA "POLÍTICA DE GOBERNANZA DE IA" Y EL PRESUPUESTO PARA EL ASEGURAMIENTO DE LA FASE 1.