Fuga de información en la Inteligencia Artificial

Una mirada desde auditoría, riesgo y gobierno corporativo

MARVIN CABEZAS - Senior Manager – IT Risk Audit & Advisory, BDO Panamá

25 de septiembre de 2025 | IT Service Management RD

BDO

Marvin Cabezas, profesional con 19 años de experiencia internacional en Auditoría de TI, Riesgos, Cumplimiento, Gobernanza y Consultoría. Actualmente me desempeño como Gerente Senior en BDO Panamá, liderando proyectos regionales en múltiples industrias. MBA, Certificado CISA, entre otros y expositor en foros ejecutivo-técnicos. Experto en Ciberseguridad, GRC, Control Interno, SOX, SOC, AML y marcos y estándares internacionales como COSO, COBIT, ISO, NIST, ISAE, etc.
Especialista en evaluación y asesoría de riesgos de tecnologías actuales y emergentes, aseguramiento de plataformas digitales, incluyendo Inteligencia Artificial en procesos de auditoría y cumplimiento.

Categorías Funcionales de IA que todo Auditor y Consultor debe conocer →

Tipos de ERPs →

El profesional de Finanzas, Tecnología y GRCA se encontrará con diferentes sistemas y aplicaciones operativas, administrativas y contables.

ERP + IA →

Modalidad	Funcionalidad	Complejidad	Tendencia
On-premises (local) El sistema se instala y se ejecuta en **servidores propios**. Requiere infraestructura y mantenimiento interno. **Cloud (nube)** El sistema se ejecuta en servidores remotos (AWS, Azure, etc.). Puede incluir diferentes modelos de servicio: **SaaS, PaaS, IaaS.**	**Algunas Categorías** **ERP -** Enterprise Resource Planning **CRM** - Customer Relationship Management **HCM** - Human Capital Management) **GRC** - Governance, Risk and Compliance	**Soluciones más simples** Sage 50, Quickbooks, Macola, Aspel, Intelisis, Exactus, Odoo, Bind, Elconix. **Sistemas más complejos** SAP, Oracle, JD Edwards, PeopleSoft, NetSuite, Infor LX, Microsoft Dynamics 365 (Business Central y Finance & Operations), Workday Financial Management	**ERP** **ERP tradicional:** Procesos rígidos, on-premises, poca automatización **ERP SaaS:** Cloud, accesible, actualizaciones automáticas **ERP inteligente** (actual y futuro): Integración con IA, automatización cognitiva, chatbots, análisis predictivo, RPA, etc.
Modelo Híbrido (Hybrid Cloud): combina infraestructura on-premises (local) con nube pública o privada. **Multicloud:** uso simultáneo de múltiples proveedores de nube pública o privada, sin necesariamente incluir on-premises.	**Otros:** SCM (Supply Chain Management) EPM (Enterprise Performance Management) WMS (Warehouse Management System), etc.	Los **sistemas más robustos** suelen incluir funcionalidades avanzadas de integración profunda, reporting, seguridad multinivel, producción, contabilidad, entidades, personalización, proyectos y parametrización y documentación.	**Ejemplos de ERP inteligentes:** SAP S/4HANA Cloud Public Edition y Oracle Fusion Cloud ERP. IA nativa incluida, con agentes, GenAI, NLP y ML.

Se habla mucho de Eficiencia y Ética en el contexto de la IA.
Se dice mucho sobre IA con Propósito.
Crees que de verdad lo estamos logrando?
Mira esta aplicabilidad de IA en Contabilidad.

¿Nuestro ERP+IA realmente nos protege y potencia?

- ¿Estamos listos para implementar, migrar o auditar un ERP con IA nativa integrada?
- ¿Tu ERP es una herramienta de control o una fuente de riesgos?
- ¿Tu equipo contable gobierna la tecnología o depende de ella sin entenderla?
- ¿Están los auditores evaluando la “inteligencia” del sistema o solo su "existencia"?

IA
Aplicar **IA** en **Contabilidad**

Eficiencia

Automatización de conciliaciones bancarias y contables

Generación de reportes financieros y análisis comparativos

Identificación de anomalías y patrones de riesgo en registros contables

Apoyo en cierres contables y cálculos de provisiones

Integración con ERP para tareas rutinarias (asientos, validaciones, flujos)

Ética

La IA no reemplaza el juicio profesional contable

Importancia de usar IA con transparencia, trazabilidad y responsabilidad

Riesgo de sesgos en modelos automatizados si no se auditan correctamente

Confidencialidad y uso ético de los datos financieros

La ética profesional debe evolucionar junto con la tecnología

Propósito

Eficiencia + Ética = IA con Propósito

La IA es un aliado, no un reemplazo. Utiliza la IA como:

- ✓ Potenciador
- ✓ Auxiliar
- ✓ Agilizador
- ✓ Acelerador
- ✓ Organizador
- ✓ Colaborador
- ✓ Detonador

Innova con IA sin descuidar tu:

- ✓ Criterio
- ✓ Experiencia
- ✓ Conciencia
- ✓ Ética
- ✓ Escepticismo
- ✓ Ojo crítico
- ✓ Intuición
- ✓ Imaginación
- ✓ Olfato
- ✓ Colmillo

“Administración de riegos, juicio profesional y decisiones, son nuestras, no las deleguemos a la IA”

“La IA genera valor solo si los datos se gobiernan.”

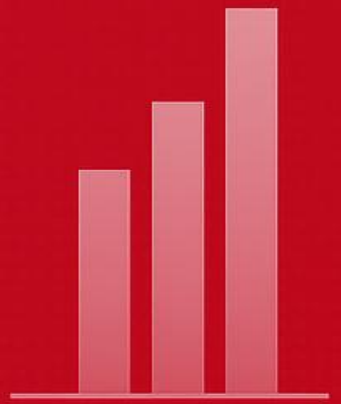
“Sin confianza en los datos,
no hay decisiones confiables”

“El reto no es generar más datos, sino **asegurar** que esos datos sean confiables, éticos y útiles.”

La paradoja actual: más datos que nunca, pero menos confianza

Volumen:
El mundo generará ≈**180 ZB de datos en** 2025 (180 mil millones de GB); la curva sigue en ascenso exponencial.

≈180 ZB
(2025)



Prioridad vs. madurez:
Los CXO (CEO, CFO, CIO, CISO, CDO) priorizan **Data & AI** (88%); la inversión en **GenAI** sube fuerte (90%). Aun así, solo 16% percibe resuelto lo de **ética y gobernanza**.

Fragilidad del dato:
Un breach cuesta en promedio USD 4.88M; la calidad de datos deficiente implica ≥ USD 12.9M por año.

IDC/Exploding Topics; Wavestone 2024, IBM 2024; Gartner

GB → TB → PB → EB → ZB
Cada unidad es **mil veces mayor** que la anterior, en el sistema decimal estándar.

Informe «2025 State of GenAI Report» de Palo Alto Networks

- El uso de IA generativa en empresas aumentó un **890%**, según el informe de Palo Alto Networks.
- El estudio analizó tráfico en más de **7.000 organizaciones** durante 2024 y Q1 de 2025, confirmando el aumento del **890%**.
- La **adopción de IA generativa** elevó los riesgos de seguridad, con un **14% más de incidentes de fuga de datos**.
- La IA generativa mejora la productividad hasta en un **40%** según McKinsey.
- El modelo DeepSeek-R1 creció un **1.800%** en dos meses.
- En Q1 de 2025, los incidentes de fuga de datos por IA generativa se multiplicaron por **2,5**, representando el **14%** del total de incidentes de seguridad registrados.
- Las organizaciones tienen en promedio **66 aplicaciones** de IA generativa en sus redes, muchas sin supervisión.
- El fenómeno "Shadow AI" implica **6,6 aplicaciones** de alto riesgo por empresa sin políticas ni controles adecuados.
- Más del **70%** de las aplicaciones pueden ser vulneradas mediante jailbreak (técnica que manipula o engaña al modelo), generando contenidos peligrosos.
- El **83,8%** del tráfico de IA generativa proviene de asistentes de escritura, conversacionales y entornos de desarrollo; Grammarly tiene el **39%** del tráfico y Copilot está en el **49%** de las organizaciones.
- El **39%** de la codificación con IA ocurre en sectores tecnológico e industrial, con riesgos de seguridad y propiedad intelectual.

https://www.directivosyempresas.com
Palo Alto Networks
Prisma® Access y Next-Generation Firewall (NGFW)

¿Qué puede hacer la Inteligencia Artificial (IA)?

BENEFICIOS DE LA IA

La IA puede aportar numerosos beneficios a las empresas y a la sociedad.

Algunas de las ventajas de la IA son:

- Mayor eficacia y productividad
- Mejora de la toma de decisiones
- Mejora de la experiencia del cliente
- Ahorro de costes
- Innovación

RETOS DE LA INTELIGENCIA ARTIFICIAL

El desarrollo y la implantación de sistemas de IA plantean varios retos. Para ha
frente a estos retos es necesaria la colaboración entre los responsables jurídic
de privacidad, de cumplimiento normativo, de tecnología y de negocio.

Algunos de los principales retos son:

- Calidad y sesgo de los datos
- Explicabilidad y transparencia
- Privacidad y cumplimiento
- Consideraciones éticas y jurídicas
- Seguridad y protección
- Integración y despliegue

Riesgos de la IA

Riesgos de la IA Generativa

Al generar nuevas versiones de contenidos, estrategias, diseños y métodos a partir de lo aprendido en grandes repositorios de contenidos originales, la IA generativa puede dar lugar a:

- Falta de transparencia
- Precisión
- Alucinaciones (Falsos resultados)
- Sesgos
- Exposición involuntaria de propiedad intelectual (PI) e infracción de derechos de autor
- Seguridad

La IA puede cometer errores y conlleva riesgos de privacidad de datos

Gemini | Nano Banana

La noticia se centra en el **"AI Saree Trend"**, una tendencia viral donde usuarios usaban **Gemini (modelo Nano Banana) de Google** para transformar sus fotos en retratos con *saree*. Sin embargo, un incidente específico, donde la IA añadió un **lunar** no visible en la foto original de una usuaria, desató **un** intenso debate sobre la privacidad. Las principales preocupaciones son:

- **Origen de los datos:** ¿De dónde obtiene la IA detalles tan específicos?
- **Acceso a datos:** ¿La IA accede a otras fuentes de datos personales (ej. Google Photos)?
- **Seguridad:** ¿Qué tan seguro es compartir fotos personales con estas herramientas de IA?

Meta

- **Evento y Fecha:** El evento Meta Connect 2025 se celebró el 17 de septiembre de 2025, en la sede de Meta en Menlo Park, California.
- **Errores Destacados:**
- **Receta de cocina:** Durante una demostración en vivo, la inteligencia artificial de las gafas proporcionó instrucciones incorrectas para una receta, lo que causó confusión y risas en la audiencia.
- **Videollamada fallida:** Mark Zuckerberg intentó recibir una videollamada a través de las gafas, pero la conexión falló a pesar de varios intentos. **Zuckerberg atribuyó el problema a una mala conexión Wi-Fi.**

Autoridad Reguladora	Fecha de la Decisión	Monto de la Sanción	Motivo Principal de la Infracción	Reglamento Aplicado
Comisión Europea	23 de abril de 2025	200 millones de euros	Modelo "consentimiento o pago" sin alternativa equivalente	Ley de Mercados Digitales (DMA)
Comisión de Protección de Datos de Irlanda (DPC)	19 de diciembre de 2024	251 millones de euros	Filtración de datos personales de 2018	Reglamento General de Protección de Datos (RGPD)
DPC / EDPB	Mayo de 2023	1.200 millones de euros	Transferencia ilegal de datos de usuarios de la UE a EE. UU.	Reglamento General de Protección de Datos (RGPD)

Riesgos de **fuga de información** en la era de la IA

Términos y condiciones

Aceptar términos sin leer.
puede exponer a los usuarios a la **cesión involuntaria de datos personales** y a **riesgos legales o contractuales** por desconocer las condiciones de uso.

Fatiga por privacidad y presión de tiempo

Constantes alertas de seguridad, solicitudes de **revisión de políticas** y estar **bajo presión** para entregar un informe urgente lleva a una persona a usar una herramienta de IA sin validar si cumple con las políticas de seguridad, **compartiendo datos confidenciales** que terminan **almacenados en servidores externos sin control.**

Compartir datos por conveniencia

Puede **facilitar la fuga de información sensible**, al priorizar la rapidez sobre la seguridad.
Este hábito **debilita los controles** y expone a **riesgos operativos, legales y reputacionales**.

Quitar acceso a las principales IA

Restringir acceso a las IA como Chat GPT, DeepSeek, Claude, Perplexity, Grok, Qwen, etc., **provoca que los usuarios busquen y entren a sitios maliciosos** con otros dominios sin notarlo.

Riesgos de **fuga de información** en la era de la IA

Ex colaboradores

Ex empleados con acceso previo a herramientas de IA **pueden filtrar datos sensibles si no se revocan sus permisos a tiempo**. Esto representa un **riesgo crítico de fuga de información** y exposición no autorizada.

Captura de rostro y edición de imagen

Fuga de datos biométricos, **suplantación de identidad** (perfiles falsos en redes sociales, aplicaciones de citas o plataformas laborales). **Deepfakes** (videos falsos), **violaciones de privacidad** (anuncios, contenido promocional o sitios web sin tu permiso). **Entrenamiento de modelos sin autorización** y **pérdida de propiedad intelectual** (las imágenes editadas pueden ser usadas por terceros sin autorización, afectando derechos de autor o imagen).

Silos

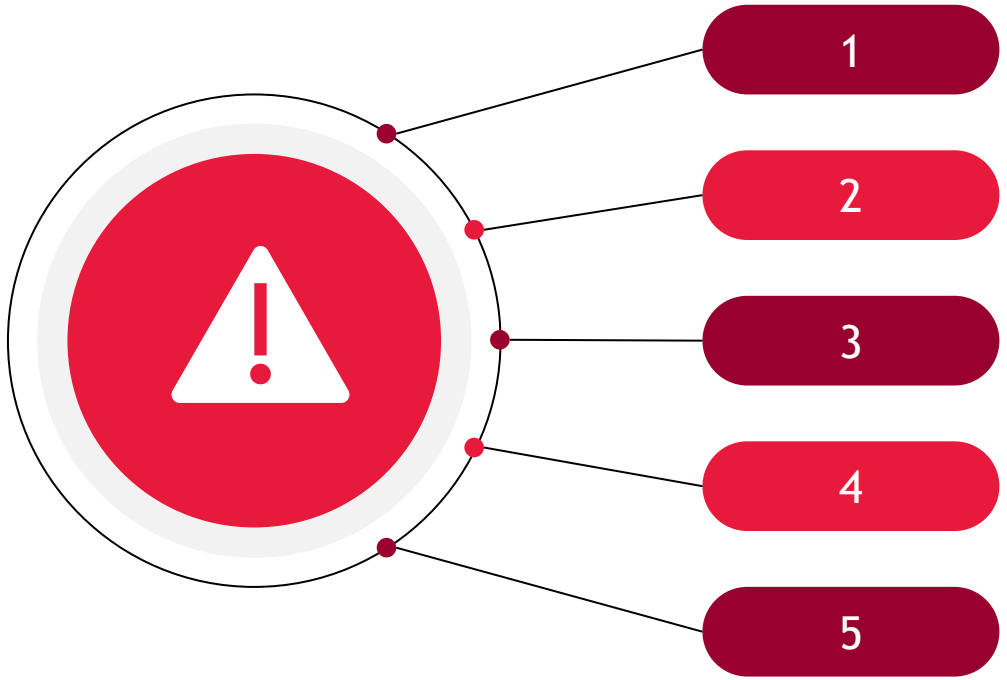
En la mayoría de las organizaciones, indistintamente la industria existe un **distanciamiento, fricción o desavenencias entre las tres líneas**.
Existen crisis graves y muchas veces suceden si las distintas áreas no colaboran estratégicamente para acortar las brechas, **siempre existirán vacíos que darán lugar a materialización de riesgos con alto impacto y criticidad.**

Cross-skilling

Los equipos de **Tecnología, Seguridad y GRCA no tienen sinergia e integración**.
Los equipos tech se enfocan solo en tools, automatizar, infraestructura, arquitectura, reportes y descuidan lo administrativo, documentación, buenas prácticas, controles y enfoque basado en riesgos y viceversa.
Los equipos de **Tecnología no conocen los objetivos empresariales, las operaciones y la estrategia** de los equipos del negocio y viceversa.

Riesgos de **fuga de información** en la era de la IA

Shadow AI



Uso **no autorizado o no supervisado de herramientas de inteligencia artificial** por parte de empleados dentro de una organización, sin el conocimiento o aprobación del departamento de TI o seguridad.

Falta de control: Estas herramientas no están gestionadas por la infraestructura oficial de la empresa, lo que **impide aplicar políticas de seguridad, cumplimiento o auditoría.**

Uso personal en entornos corporativos: Empleados pueden **utilizar plataformas de IA generativa** como ChatGPT, Copilot o herramientas de desarrollo para facilitar su trabajo, **sin evaluar los riesgos asociados**.

- **Fugas de datos** sensibles
- **Violaciones normativas** (como GDPR)
- Exposición a **contenido malicioso**
- **Pérdida de propiedad intelectual**

El Shadow AI **reduce la visibilidad** que tienen las organizaciones sobre **cómo se usan sus datos y qué herramientas acceden a ellos.** Esto dificulta la gestión de riesgos y pone en peligro la seguridad y el cumplimiento normativo.

Casos reales de riesgos materializados

Empresa	Resumen del incidente	Fecha del incidente	Causas técnicas	Impacto	Acciones correctivas	Lecciones aprendidas	Referencias
Samsung	Ingenieros filtraron información confidencial al usar ChatGPT para depurar código de semiconductores. Compartieron fragmentos de código de semiconductores con ChatGPT para resolver errores. Sin saberlo, expusieron información confidencial que quedó almacenada en los servidores de OpenAI, lo que representó un riesgo grave de propiedad intelectual.	Abril 2023	Uso indebido de ChatGPT para compartir código interno con errores, sin control de privacidad.	Exposición de datos confidenciales de productos, riesgo reputacional y posibles implicaciones legales.	Restricción del uso de ChatGPT, revisión de políticas internas y capacitación en ciberseguridad.	Evitar compartir información sensible con herramientas externas, implementar controles de uso de IA.	https://www.20minutos.es/ https://www.xataka.com.mx/ https://www.eltiempo.com/

Empresa	Resumen del incidente	Fecha del incidente	Causas técnicas	Impacto	Acciones correctivas	Lecciones aprendidas	Referencias
DeepSeek	Exposición pública de más de 1 millón de registros de chats, claves API y datos internos por bases de datos sin protección. Las bases de datos de DeepSeek estaban configuradas sin autenticación, lo que significa que cualquier persona con acceso a la URL podía consultar, modificar o extraer información sin restricciones. Esta falla permitió que más de un millón de registros de chats, claves API y datos internos quedaran expuestos públicamente.	Enero 2025	Bases de datos ClickHouse expuestas sin autenticación, permitiendo acceso libre a información crítica.	Riesgo de robo de datos, escalamiento de privilegios, pérdida de confianza, bloqueo en países y posibles sanciones regulatorias.	Cierre inmediato de las bases expuestas, revisión de seguridad, y restricciones en nuevos registros por 24 horas.	Implementar controles de acceso robustos, auditar configuraciones de bases de datos, y reforzar seguridad en IA.	https://blog.tecnetone.com/ https://www.infobae.com/ https://laopinion.com/

Empresa	Resumen del incidente	Fecha del incidente	Causas técnicas	Impacto	Acciones correctivas	Lecciones aprendidas	Referencias
McDonald's	Fuga masiva de datos personales de postulantes por vulnerabilidad en sistema de IA McHire (utilizada por más del 90% de los franquiciados de McDonald's en EE.UU.) Datos comprometidos: Hasta 64 millones de registros de postulantes, incluyendo: Nombres completos Correos electrónicos Teléfonos Direcciones Chats con el bot Resultados de pruebas de personalidad Tokens de sesión	Junio 2025	Contraseña débil en cuenta de prueba activa y vulnerabilidad IDOR en la API del chatbot Olivia. Una IDOR (Insecure Direct Object Reference) es una falla de seguridad común en APIs web. Consiste en permitir el acceso a objetos directamente mediante un identificador, sin verificar si el usuario tiene permiso para acceder a ellos.	Riesgo de suplantación de identidad, phishing, exposición de datos sensibles; reputación afectada.	Desactivación de cuenta comprometida, parcheo de API, programa de recompensas por fallos.	Validar seguridad en sistemas de IA, evitar contraseñas débiles, realizar pruebas de penetración periódicas. Las empresas deben aplicar el principio de "seguridad por diseño", implementar autenticación multifactor, encriptación de datos y auditorías de código regulares. Además, es crucial que los entornos de prueba no permanezcan activos en producción.	https://elartesano.digital/ https://actu.ai/es/ https://www.incibe.es/

Gobernar datos no es “opcional”, es supervivencia

Casos recientes
Cuando el gobierno de datos falla

MOAB (Mother of All Breaches) (enero 2024)
Fue una filtración tan masiva, que puede considerarse la mayor de la historia hasta la fecha. Por eso la bautizaron como “la madre de todas las filtraciones”.

Esta filtración comprometió 26.000 millones de registros de usuarios de diferentes plataformas. Estas incluían credenciales de acceso de servicios como MySpace, Adobe, Telegram, LinkedIn o X.

- Una mala configuración de firewall dejó accesible públicamente un servidor que almacenaba 26 mil millones de registros recopilados de múltiples filtraciones anteriores.
- Exposición masiva de credenciales y datos personales de usuarios de plataformas como Telegram, LinkedIn, Adobe y X, con alto riesgo de phishing, suplantación de identidad y ataques automatizados.
- El servidor fue cerrado tras el descubrimiento por investigadores; Leak-Lookup reconoció el error y se inició una revisión de seguridad en su infraestructura.
- Es vital auditar configuraciones de seguridad, proteger servidores con autenticación y limitar el acceso a datasets sensibles, especialmente cuando pueden ser explotados por IA.

https://cybernews.com/

https://www.wwatcher.com/
https://cybernews.com/
https://devel.group/

Casos recientes
Cuando el gobierno de datos falla

- **UnitedHealth / Change Healthcare (EE. UU., 2024-2025)**
  Ransomware afectó a **192.7 M** de personas; colapsó el procesamiento de reclamos y expuso la fragilidad de datos críticos en salud.

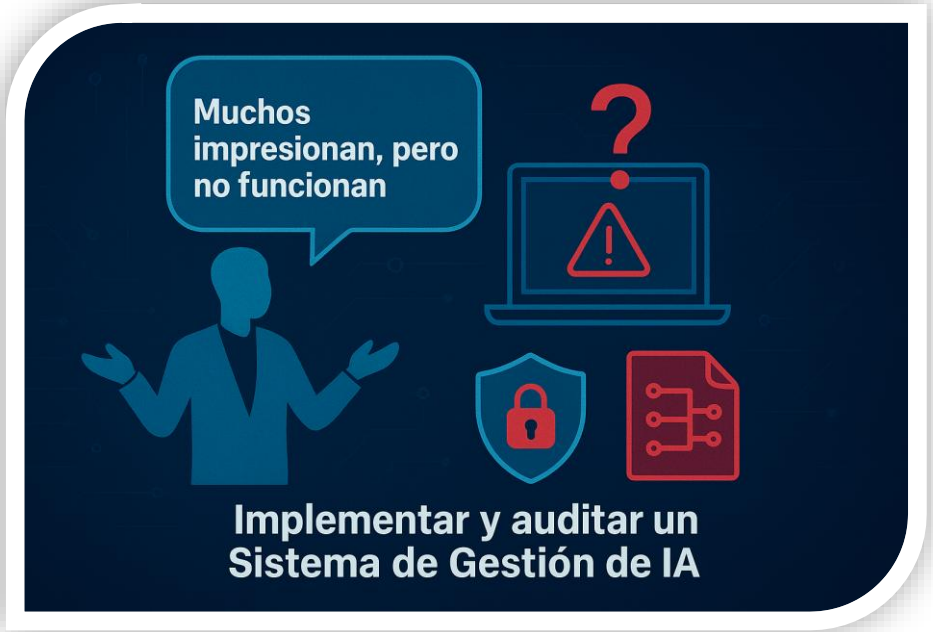
- Involucra datos sensibles de salud (críticos en decisiones clínicas y financieras).
- El ataque interrumpió procesos de facturación y reclamos, evidenciando cómo la falta de resiliencia y gobierno de datos puede colapsar todo un ecosistema de negocio.
- Business Analytics + IA no vale nada si los datos críticos no son confiables ni protegidos.

- **AT&T (EE. UU., 2024-2025)**
  Filtración hacia la dark web de **73 M** clientes y **109 M** registros; acuerdo legal por **USD 177 M**. Demuestra el costo financiero y reputacional de la mala gobernanza.

- Impactó datos de 73 M de clientes y 109 M de registros de llamadas/SMS, es decir, datos masivos que alimentan analytics de consumo y comportamiento.
- Se tradujo en un acuerdo legal de USD 177 M, un ejemplo claro de que la mala gobernanza tiene costos tangibles y reputacionales enormes.

U.S. Department of Health and Human Services (2025) – Brecha UnitedHealth/Change Healthcare, 192.7 M afectados.
The Record / BleepingComputer (2025) – Detalles del ataque BlackCat y disrupción de reclamos.
Reuters / CNN (2025) – Ticketmaster, AT&T y otras filtraciones masivas en EE. UU.
Reuters (2025) – AT&T: acuerdo legal de USD 177 M por filtraciones (2024-2025).

Hoy abundan personas que recién estudiaron cursos de IA y luego distribuyen soluciones comerciales y servicios como expertos, con riesgos no controlados. "Muchas impresionan, pero NO funcionan."



Es necesario implementar y auditar un Sistema de Gestión de IA (SGIA) robusto que evalúe todo el ciclo de vida de la IA: desde su concepción, diseño, eficiencia operativa, seguridad cibernética, privacidad de datos, control técnico, compliance normativo y legal, alineación a objetivos estratégicos, sostenibilidad, sesgo y ética.

Si usas IA por hype y no por estrategia, podrías estar:

- Exponiéndote a riesgos regulatorios.
- Perdiendo credibilidad frente a comités y juntas.
- Desperdiciando el ROI en pilotos sin propósito.
- Comprometiendo decisiones sin trazabilidad.

IA hype se refiere a:

- Entusiasmo excesivo
- Expectativas infladas
- Exageración de capacidades
- Sobreventa de beneficios
- Prometer más de lo que se puede entregar

sobre IA, muchas veces sin bases reales o resultados tangibles.

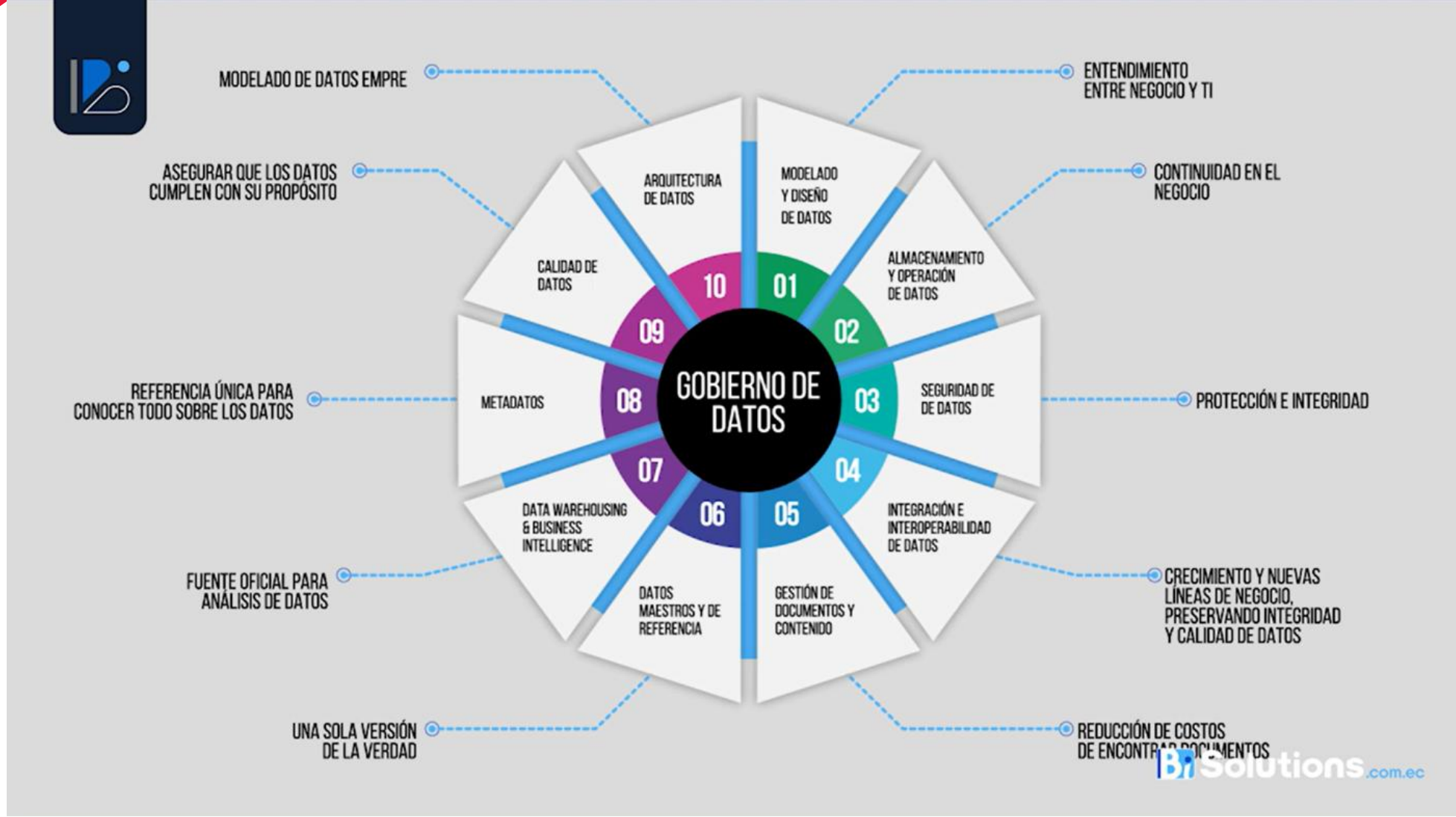
Es cuando se genera más ruido que valor.

✅ Tú puedes combatirlo hablando desde el ROI, el control interno y el cumplimiento.

Principios de privacidad de ISACA.

"La privacidad no se perdió, la regalamos por comodidad.
Y ahora, nos sorprende que nos vigilen."

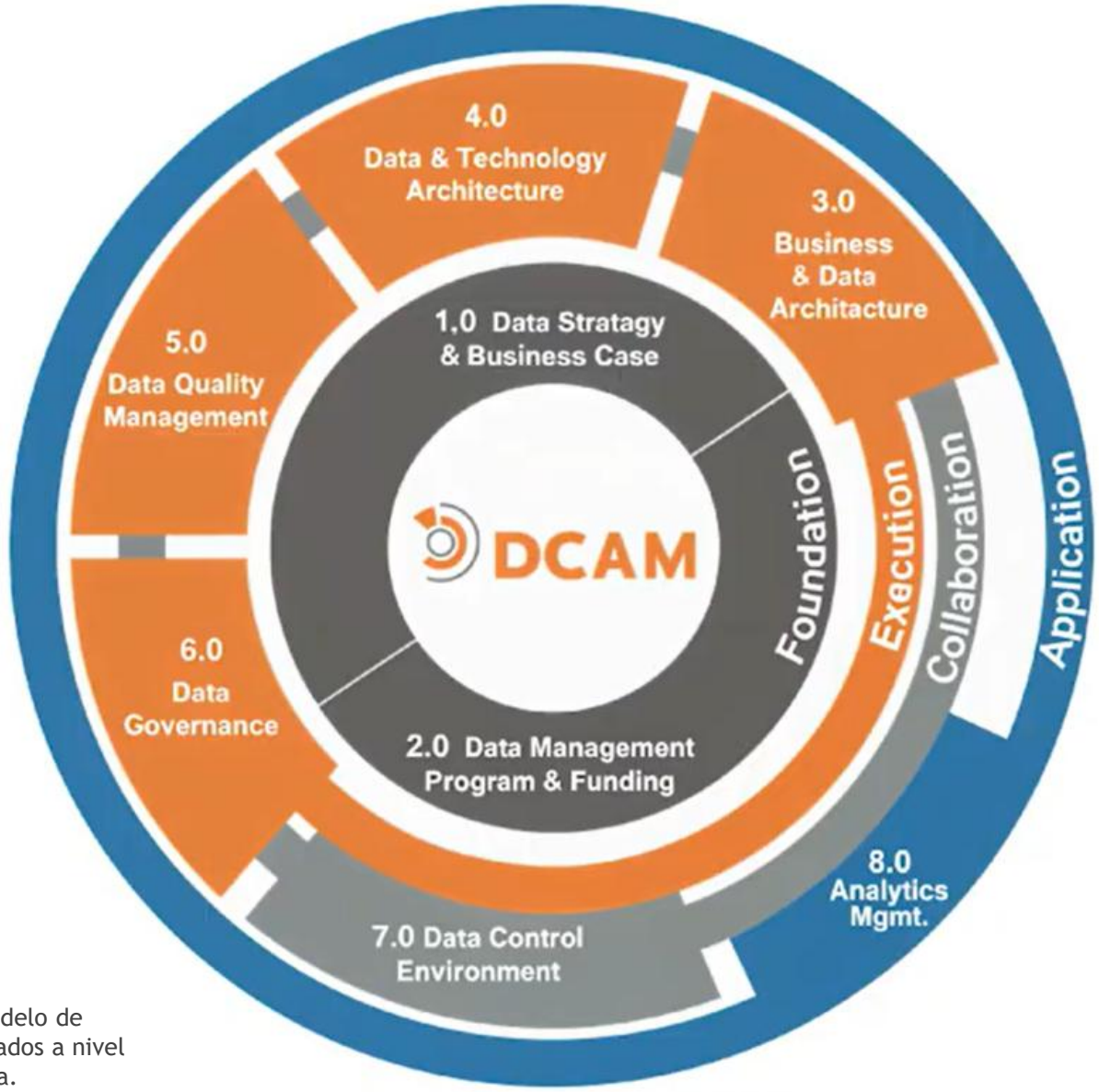
Marcos de referencia de buenas prácticas internacionales



"Gobernar datos no se compra en un software, se construye en la organización."

- **Nombre completo: DAMA-DMBOK® – DAMA Data Management Body of Knowledge**. guía desarrollada por DAMA International que establece un marco de referencia globalmente reconocido para la gestión de datos
- **Organización creadora: DAMA International**, una asociación líder en gestión de datos.
- **Propósito:** Definir los principios, funciones, disciplinas y mejores prácticas esenciales para una gestión de datos efectiva en organizaciones.

“Ignorar el gobierno es un riesgo que los ejecutivos NO pueden permitirse.”



- **Nombre completo: Data Management Capability Assessment Model (DCAM®).** Modelo de evaluación de capacidades de gestión de datos , y es uno de los marcos más adoptados a nivel global para establecer, evaluar y mejorar programas de gestión de datos y analítica.
- **Organización creadora**: **EDM Council**. ahora conocido como **EDM Association**) es una asociación comercial global sin fines de lucro dedicada a promover las mejores prácticas, estándares y educación en gestión de datos y analítica.
- **Propósito**: Proporcionar un marco estructurado para evaluar, desarrollar y madurar las capacidades de gestión de datos en una organización.

Estándares, marcos, leyes y guías principales vigentes que conviene considerar en un programa de gobernanza/auditoría de IA.

ISO/IEC 42001	Gestión de Sistemas de IA
ISO/IEC 23894	Gestión de riesgos de IA
NIST AI RMF	Se centra en identificar y mitigar los riesgos específicos de la IA
NIST SP 1270	Adversarial ML
Normativas emergentes	UE AI Act, OECD AI Principles
OWASP Top 10 for LLM	Amenazas y recomendaciones técnicas para mitigar riesgos de prompt injection, extracción de datos, etc.

Estándares, marcos, leyes y guías principales que contienen referencias a privacidad y protección de datos.

Internacionales y Estándares Técnicos

1. **ISO/TS 17975:2022**
   Principios y requisitos de datos para el consentimiento en el uso de información de salud personal.
2. **ISO/IEC 27030**
   Guía para seguridad y privacidad en el Internet de las cosas (IoT).
3. **ISO/IEC TR 27550:2019**
   Ingeniería de privacidad para procesos del ciclo de vida del sistema.
4. **ISO/IEC 27701:2019**
   Extensión de ISO/IEC 27001/27002 para gestión de información de privacidad.
5. **NISTIR 8062**
   Introducción a la ingeniería de privacidad y gestión de riesgos.
6. **Marco de Privacidad NIST**
   Herramienta para mejorar la privacidad mediante gestión de riesgos organizacionales.
7. **NISTIR 7628**
   Guías para la ciberseguridad en redes inteligentes.

Leyes y Regulaciones Nacionales

8. **GDPR (Reglamento General de Protección de Datos)** – Unión Europea
   Estándar europeo para la protección de datos personales.
9. **HIPAA – EE.UU.**
   Protege los historiales médicos personales en Estados Unidos.
10. **GLBA – EE.UU. (Gramm-Leach-Bliley Act)**
    Requiere que las instituciones informen sobre prácticas de intercambio de información.
11. **PIPEDA – Canadá**
    Regula el uso de datos personales por parte del sector privado en Canadá.
12. **LGPD – Brasil (Lei Geral de Proteção de Dados Pessoais)**
    Unifica estatutos sobre protección de datos personales en Brasil.

Principios y Marcos de Privacidad

13. **ISACA Privacy Principles**
    Principios de privacidad de ISACA para gobernanza y auditoría de TI.
14. **OECD Privacy Principles 2013**
    Principios de privacidad de la OCDE para protección de datos personales.
15. **AICPA Privacy Management Framework (PMF)**
    Marco de gestión de privacidad del Instituto Americano de Contadores Públicos.

Resumen de estándares, marcos y regulaciones con foco explícito en "Gobierno" (de TI, datos, ciberseguridad, IA, cumplimiento) y otros que incorporan un componente de gobierno relevante

REFERENCIA	DESCRIPCIÓN
ISO/IEC 38500 — 2024 — Governance of IT	Principios y modelo para que la Alta Dirección dirija y supervise el uso de TI.
COBIT — 2019	Gobierno de TI basado en objetivos, procesos y metas de desempeño.
ISO/IEC 38505-1 — 2017 **ISO/IEC 38505-2 — 2018**	Principios y responsabilidades de gobierno de datos. Guía práctica para implantar gobierno de datos.
DAMA-DMBOK2 — 2017	Cuerpo de conocimiento de gestión de datos: incluye el dominio de Data Governance (rol, consejo, políticas).
NIST Cybersecurity Framework — v2.0 (2024)	Añade la función "GOVERN" para gobernanza de ciberseguridad (políticas, roles, supervisión).
ISO/IEC 27014 — 2020	Establece principios y modelos de gobierno de la seguridad de la información.
ITIL 4 — 2019+ — Marco ITSM con SVS	El Service Value System incorpora "Governance" como componente central del modelo de gestión de servicios.
ISO/IEC 42001 — 2023	AI Management System (AIMS): primer SG de IA; políticas, roles, objetivos y controles para gobernanza responsable de IA.
ISO/IEC 38507 — 2022	Implicaciones de gobierno por uso de IA: guía a órganos de gobierno para habilitar y supervisar IA de forma efectiva y aceptable.
NIST AI Risk Management Framework — v1.0 (2023)	Incluye función GOVERN para liderazgo, políticas y cultura de riesgo en IA.
COSO Internal Control — 2013 (ICIF) + Guía ICSR 2023	Control interno con fuerte foco en gobierno; guía 2023 para control sobre reporte de sostenibilidad.
GDPR — Reg. (EU) 2016/679	Principio de accountability (art. 5.2) impone gobierno, evidencia y responsabilidad sobre tratamiento de datos.
NIS2 — Dir. (EU) 2022/2555	Responsabilidad de la alta dirección en medidas de ciberseguridad, gestión de riesgo e incidentes.
DORA — Reg. (EU) 2022/2554	Gobernanza y responsabilidad del órgano de dirección sobre riesgo TIC en entidades financieras.
EU AI Act — Reg. (EU) 2024/1689	Marco horizontal de IA; entra en vigor 1-Ago-2024, obligaciones escalonadas hasta 2026; exige gobierno, gestión de riesgos y supervisión.

GRCA

Los 5 componentes que sostienen el gobierno de datos

Roles y responsabilidades
dueños claros, stewardship y accountability.

Seguridad y privacidad
proteger datos sensibles y cumplir regulaciones.

Políticas y procesos
lineamientos, estándares y controles definidos.

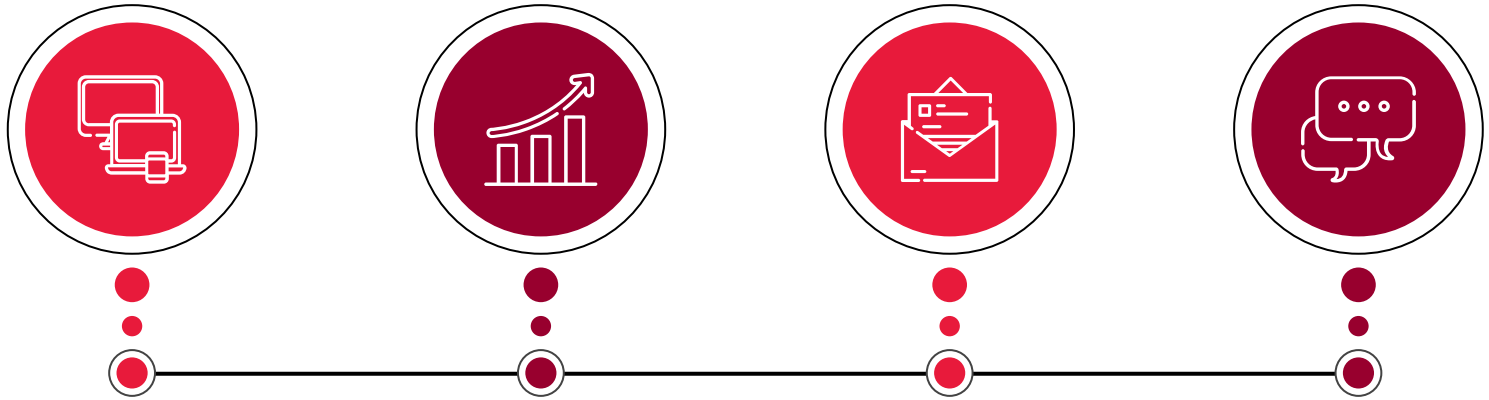
Calidad de datos
exactitud, consistencia y completitud.

Ética y cumplimiento
uso responsable, confianza y sostenibilidad.

"Sin estos engranajes, la máquina de IA confiable se detiene."

IA & Gobierno de Datos: De la Ley a la práctica

- **Artículo 10 (alto riesgo):** exige data governance de training/validation/testing (calidad, origen, preparación, sesgos, vacíos), más gestión de riesgos, documentación, trazabilidad, transparencia y supervisión humana.
- **Sanciones:** hasta 7 % del volumen global en ciertos supuestos de incumplimiento. (ej. uso de sistemas prohibidos, violaciones críticas de gobernanza de datos o riesgos).



1 de agosto 2024

Entra en vigor el **AI Act** (ya es ley en la UE).

+6 meses (febrero 2025)

Aplican las **prohibiciones absolutas** (ej. manipulación subliminal, scoring social).

+12 meses (agosto 2025)

Empiezan las **obligaciones para sistemas de propósito general (GPAI)** como ChatGPT.

+24 a 36 meses (2026–2027)

Entran en vigor los requisitos para los **sistemas de alto riesgo** (ej. salud, finanzas, educación), incluyendo **gobernanza de datos, gestión de riesgos, trazabilidad y supervisión humana.**

Ya está en vigor, pero sus **obligaciones se aplican de forma escalonada:** primero lo que no se permite nunca, luego lo que afecta a modelos generales, y finalmente lo más complejo: los sistemas de alto riesgo.

Entrada en vigor y calendario del AI Act: Comisión Europea; Parlamento Europeo; confirmación de no retraso y fechas (Reuters, jul-2025). European Commission Parlamento Europeo Reuters

Requisito de data governance (Art. 10) y vínculos con riesgo, documentación, logging y supervisión humana. Ley de Inteligencia Artificial de la UE

Régimen sancionador (hasta 7%). dechert.com

IA & Gobierno de Datos: De la Ley a la práctica

- Estándar y principios basados ISO/IEC 38500 de TI aplicado a datos (**ISO 38505**); responsabiliza al órgano de gobierno de evaluar-dirigir-monitorizar el uso de datos. **Beneficios**: valor, responsabilidad clara y mitigación de riesgos.

- **Principios** que hereda (para gobernar datos): Responsabilidad, Estrategia, Adquisición, Desempeño, Conformidad, Comportamiento humano.

Responsabilidad
Directivos responden por el uso, calidad y ética de los datos.

Estrategia
El gobierno de datos debe seguir y habilitar la estrategia de negocio.

Adquisición
Obtener y proveer datos de forma legítima, segura y sostenible.

Desempeño
Los datos deben generar valor y resultados medibles.

Conformidad
Cumplir con leyes, regulaciones y políticas internas.

Comportamiento humano
Considerar ética, cultura y evitar sesgos en el uso de datos.

*AI Act define el **qué exigir**; ISO 38505 define **quién decide** **y cómo gobernar** para cumplir y crear valor.*

ISO/IEC 38505-1 (2017) - — principios, roles del órgano de gobierno, beneficios y riesgos. cdn.standards.iteh.ai

Según **McKinsey & Company**, más del 60% de las organizaciones que adoptan IA en sus procesos críticos **NO** han implementado controles formales de auditoría sobre esos modelos.

¿Qué significa esto? Que podríamos estar automatizando decisiones con impactos reputacionales, regulatorios y financieros sin garantías.

NIVELES CMMI

- **Origen y organización:** El modelo **CMMI (Capability Maturity Model Integration)** fue desarrollado originalmente por el **Software Engineering Institute (SEI)** de la Universidad Carnegie Mellon.
- **Administración actual:** Hoy en día, el modelo es gestionado por el CMMI Institute, una subsidiaria de ISACA.
- **Niveles del modelo:** Se estructura en **cinco niveles de madurez:** Inicial (1), Gestionado (2), Definido (3), Gestionado Cuantitativamente (4), Optimizado (5).

Guía ejecutiva para identificar riesgos, controles y oportunidades en el uso de IA

CATEGORÍA FUNCIONAL

#1 ASISTENTES DE IA

BREVE CONCEPTO

Interfaces que interactúan con humanos para brindar apoyo.

EJEMPLOS

ChatGPT, Microsoft Copilot, Siri, Alexa, DeepSeek, Perplexity

QUÉ HACEN / AUTONOMÍA

Generan respuestas o tareas con baja autonomía.

USO COMÚN

Soporte, redacción, productividad, atención al cliente, consulta, aplicaciones móviles.

ÁMBITOS EMPRESARIALES

TI, RRHH, legal, marketing, finanzas

Muy comunes y en expansión. Ampliamente implementados en diversas industrias.

4 RIESGOS y CONTROLES

RIESGO	CONTROL
Fugas de datos sensibles	Políticas de uso claras y actualizadas
Respuestas inexactas o sesgadas	Monitoreo y registro de interacciones
Uso no autorizado o sin control	Validación y revisión periódica de respuestas
Dependencia excesiva que puede afectar la toma de decisiones humanas	Capacitación a usuarios sobre limitaciones y riesgos

CATEGORÍA FUNCIONAL

#2 AGENTES DE IA

BREVE CONCEPTO

Sistemas autónomos que actúan para cumplir objetivos definidos.

EJEMPLOS

AutoGPT, Devin, RPA inteligente, agentes de RPA avanzados

QUÉ HACEN / AUTONOMÍA

Ejecutan tareas sin supervisión constante; alta autonomía.

USO COMÚN

Automatización de procesos, planificación, gestión de operaciones, análisis autónomo.

ÁMBITOS EMPRESARIALES

Innovación, operaciones, manufactura, tecnología, logística, finanzas

Emergentes, pero aún limitados. En fase de adopción en sectores específicos.

4 RIESGOS y CONTROLES

RIESGO	CONTROL
Decisiones no auditables	Registro detallado de acciones y decisiones
Accountability difusa	Validación y revisión de decisiones autónomas
Errores escalables	Límites de autonomía claramente definidos
Acceso no autorizado a sistemas críticos	Controles de acceso y autenticación robustos

CATEGORÍA FUNCIONAL

#3 MODELOS DE IA

BREVE CONCEPTO

Algoritmos entrenados para tareas específicas como predicción o clasificación.

EJEMPLOS

GPT-4, modelos de scoring crediticio, modelos de detección de fraude

QUÉ HACEN / AUTONOMÍA

Procesan datos y generan resultados; autonomía nula, requieren ejecución.

USO COMÚN

Análisis predictivo, clasificación, personalización de servicios, detección de anomalías.

ÁMBITOS EMPRESARIALES

Finanzas, analítica, marketing, riesgos, operaciones

Muy comunes, pero poco visibles. Ampliamente utilizados en análisis de datos y sistemas predictivos.

4 RIESGOS y CONTROLES

RIESGO	CONTROL
Sesgos en datos de entrenamiento	Evaluación y limpieza de datasets
Opacidad en los resultados (falta de explicabilidad)	Pruebas de precisión y validación cruzada
Falta de validación continua	Auditoría del modelo y su desempeño
Uso de datos no autorizados o sensibles	Implementación de técnicas de IA explicable

CATEGORÍA FUNCIONAL

#4 BOTS DE IA

BREVE CONCEPTO

Programas que realizan tareas repetitivas con inteligencia limitada.

EJEMPLOS

Chatbots, voicebots, bots en WhatsApp y redes sociales

QUÉ HACEN / AUTONOMÍA

Interactúan según flujos definidos; autonomía media.

USO COMÚN

Atención al cliente, soporte, recopilación de datos, tareas administrativas, reservas, encuestas.

ÁMBITOS EMPRESARIALES

Banca, retail, telecomunicaciones, servicios, marketing

Comúnmente implementados en servicios al cliente, marketing y RPA.

4 RIESGOS y CONTROLES

RIESGO	CONTROL
Respuestas inadecuadas o erróneas	Monitoreo de interacciones y desempeño
Falta de escalamiento ante situaciones complejas	Entrenamiento y actualización de flujos conversacionales
Errores en la lógica de conversación	Límites de acción y escalamiento a humanos
Vulnerabilidades de seguridad y privacidad	Pruebas de seguridad y cumplimiento normativo

CATEGORÍA FUNCIONAL

#5 PLATAFORMAS DE IA

BREVE CONCEPTO

Infraestructura para desarrollar, entrenar y desplegar modelos y agentes de IA.

EJEMPLOS

Azure AI, Google Cloud AI, IBM Watson, AWS, TensorFlow, MLflow, DataRobot, BigML

QUÉ HACEN / AUTONOMÍA

Proveen entornos y servicios para IA; autonomía depende de uso.

USO COMÚN

Desarrollo interno de soluciones IA, integración de sistemas, despliegue de aplicaciones

ÁMBITOS EMPRESARIALES

TI, ciencia de datos, innovación, proveedores SaaS

Usadas por proveedores o áreas especializadas para construir soluciones específicas.

4 RIESGOS y CONTROLES

RIESGO	CONTROL
Fallas de seguridad en la infraestructura	Gestión de accesos y autenticación
Acceso no controlado a recursos y datos	Acuerdos de nivel de servicio (SLAs) claros y revisados
Falta de cumplimiento normativo y regulatorio	Controles de proveedor y auditoría de terceros
Dependencia de terceros sin evaluaciones adecuadas	Cumplimiento con normativas como la Ley de IA de la UE y el NIST AI RMF

1

¿Estamos adaptando nuestras metodologías y marcos de control al tipo específico de IA que usamos? o ¿Estamos solo aplicando un checklist genérico descargado desde un post de **LinkedIn** que alguien generó con IA a partir de un PDF también creado por IA?

2

No es lo mismo auditar un **asistente tipo Copilot**, que un **modelo predictivo con autoaprendizaje**, o un **bot autónomo con permisos operativos**.

3

No podemos auditar lo que no entendemos. Y si entendemos mejor la IA, podemos gestionarla mejor.

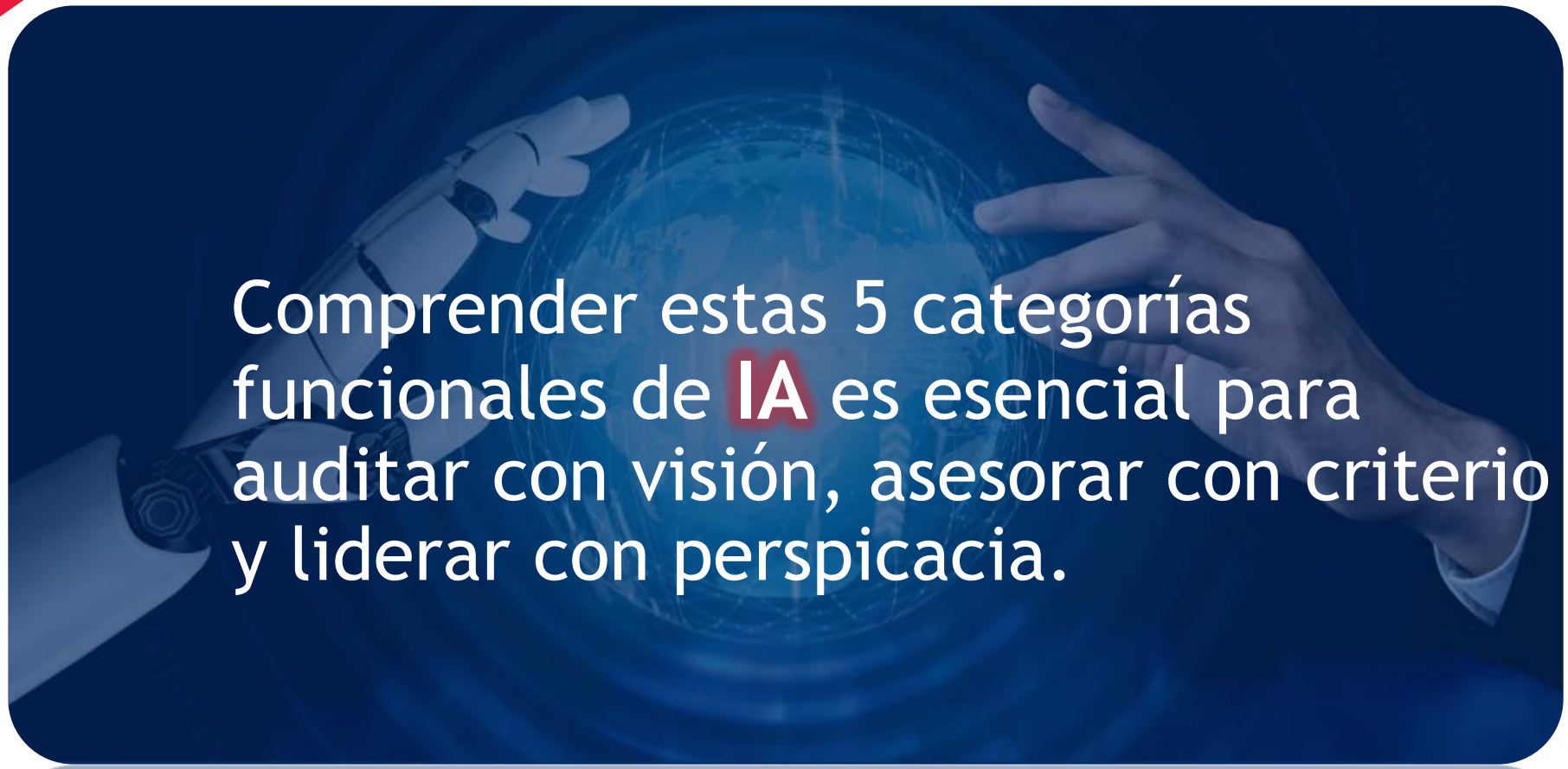
4

Cada tipo exige riesgos diferenciados, respuestas específicas y capacidades de aseguramiento especializadas.

5

Auditar IA incluye auditar cómo tomamos decisiones, cómo gestionamos el cambio y cómo protegemos el valor en la era digital.

"¿Cuál de estas categorías creen que presenta hoy el mayor riesgo para sus organizaciones? "



“Porque quien no audita la IA, será auditado por sus consecuencias.”

Estrategias para mitigar el riesgo de fuga de información en IA

Estrategias para mitigar el riesgo de fuga de información en IA

Versiones pagadas

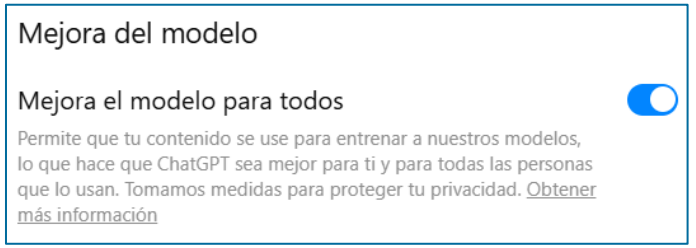
- Para tener mayor privacidad en herramientas de IA como ChatGPT y Gemini, generalmente hay que pagar.
- **Las versiones gratuitas no garantizan privacidad**. Para evitar que tus datos sean usados para entrenamiento o revisión humana, es necesario optar por **versiones pagadas que ofrecen mayor control**, privacidad por defecto y cumplimiento normativo, lo que las hace más adecuadas para entornos corporativos.
- No siempre es obligatorio pagar para tener privacidad, pero **los planes pagos empresariales dan garantías contractuales** de que los datos no serán usados.
- En las **versiones gratuitas** y personales, normalmente **sí hay riesgo de que se usen tus datos para entrenar el modelo**.

Precaución con herramientas de IA dentro de WhatsApp

- **Evita compartir datos sensibles** o corporativos
- **No uses IA en grupos** o chats compartidos
- Prefiere **plataformas oficiales** y seguras
- Conversaciones, datos, número, nombre e imagen pueden quedar expuestos.

Inactivar uso de datos

- Es posible **desactivar el uso de tus datos personales para entrenamiento del modelo** en herramientas de IA, incluso en sus versiones gratuitas si el proveedor permite configurarlo.
- En Chat GPT también puedes usar el **modo de chat temporal (modo incógnito)**, que no guarda el historial ni entrena el modelo, aunque los datos se conservan por 30 días por motivos de seguridad.



Concientización y capacitación

- Las **personas no son el eslabón más débil, sino el más importante** si están bien entrenadas.
- **Capacita regularmente** sobre riesgos de IA, privacidad y uso responsable de datos.
- Promueve una cultura de "**piensa antes de compartir**" al usar herramientas digitales.

Estrategias para mitigar el riesgo de fuga de información en IA

Recomendaciones

Estrategia de gobierno de datos (identificar, clasificar, evaluar, proteger, monitorear y auditar.

Segregación de Funciones (SoD)

- Aprovisionamiento de cuentas.
- MFA.
- Principio del menor privilegio.
- Monitoreo de altos privilegios.

Recomendaciones

-Contexto geopolítico
-Riesgo país
-Industria
-Giro de negocio
-Propuesta de valor de la empresa
-Objetivos estratégicos del negocio
-Cultura organizacional y digital
-Procesos y servicios críticos
-Datos críticos
-Riesgos clave
-Estrategia de datos + IA
-Gobierno de datos + IA
-Auditoría continua

Recomendaciones

- Controlar el uso permitido.
- Proteger los datos sensibles.
- Defensa y ofensiva contra ciberamenazas.
- Gestión de incidentes.
- Cumplimiento normativo.

Estrategias para mitigar el riesgo de fuga de información en IA

01 Anonimización

- **Evaluar la anonimización caso por caso:** Se debe demostrar que la **probabilidad de identificar o extraer** datos personales es **insignificante.**
- **Selección cuidadosa de fuentes de datos**
- **Minimización de datos** (Recolectar y procesar **solo los datos estrictamente necesarios**)
- Entrenamiento con **garantías técnicas contra extracción de datos**
- **Documentación técnica clara** que demuestre cómo se garantiza el anonimato, para que las autoridades puedan verificarlo.

02 **UEBA** (User and Entity Behavior Analytics)

- IA y análisis avanzado para **detectar comportamientos anómalos de usuarios y entidades** (como dispositivos o cuentas) dentro de una organización.
- **Detectar posibles fugas de información**, accesos indebidos o movimientos sospechosos.
- **Analizar patrones** como:
  - Reenvíos de correos personales, CVs,
  - Acceso a sitios no autorizados, cargas a la nube,
  - Descargas masivas de archivos, horarios,
  - Uso de IA fuera de políticas internas, apps rutinarias,

Implementar **UEBA permite detectar comportamientos anómalos en tiempo real**, fortaleciendo la seguridad **contra fugas de información** desde adentro de la organización.

03 Modelo propio

- Un **modelo privado de IA** es un LLM que se **descarga y ejecuta dentro de la infraestructura de la empresa** (on-premise o en una nube privada).
- Implementar un modelo privado con tecnología open source y conectores seguros es una **estrategia efectiva para minimizar el riesgo de fuga de información**.
- **Ventajas claras:** Control total de datos, cumplimiento normativo, seguridad personalizada y entrenamiento a la medida.
- Se requiere: **Modelos open source**, **conector** (middleware/servidor de orquestación), **interfaz** y **bases de datos vectoriales**.

Esquema GPT público

- Los datos de la pregunta y el contexto se hacen públicos
- Las respuestas también son públicas
- El modelo puede ser entrenado con todos estos datos

Modelo privado de IA

Tecnologías mencionadas:

- **Modelos open source** como *Llama 3 (Meta)*, *Mistral 7B/Mixtral*, *Falcon LLM*, *GPT4All* son adecuados.
- **Frameworks** como *LangChain*, *LlamaIndex*, *Haystack* son usados para conectar datos y modelos.
- **Bases de datos vectoriales** como *Milvus* o *Pinecone* son estándar para búsqueda semántica.
- **Interfaces** como *Gradio* y *Streamlit* son útiles para crear entornos seguros de interacción.

Ejemplo resumido:

• **Descargamos Llama 3**
Modelo de lenguaje abierto que servirá como el núcleo de la IA para procesar y generar respuestas.
Algunos modelos como Llama 3 requieren hardware potente (GPU) para funcionar bien. Es importante validar si la infraestructura lo soporta.

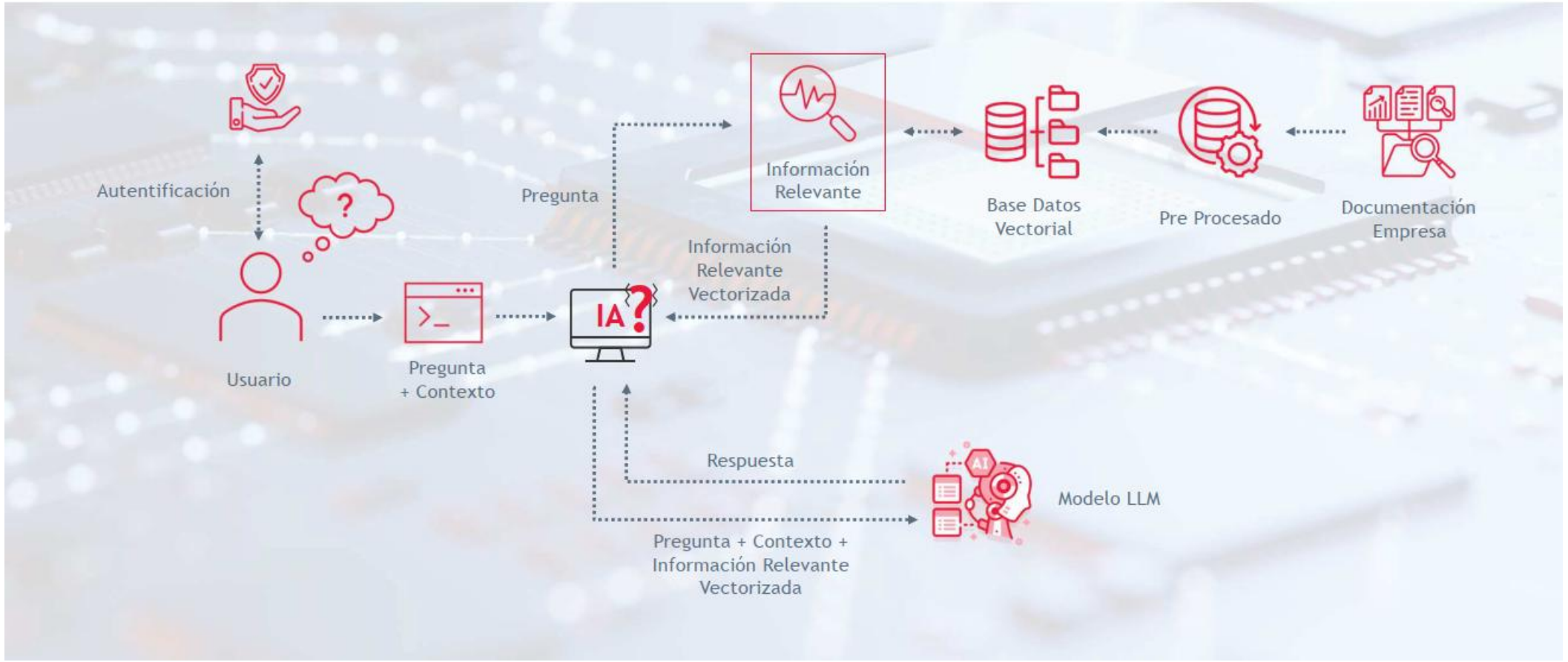
• **Lo montamos con LangChain**
Framework para conectar datos y modelos (conector para bases de datos, APIs, documentos).

• **Usamos Haystack como conector**
Middleware/servidor de orquestación que permite integrar QA (Question Answering de NLP) y búsqueda semántica en el flujo.

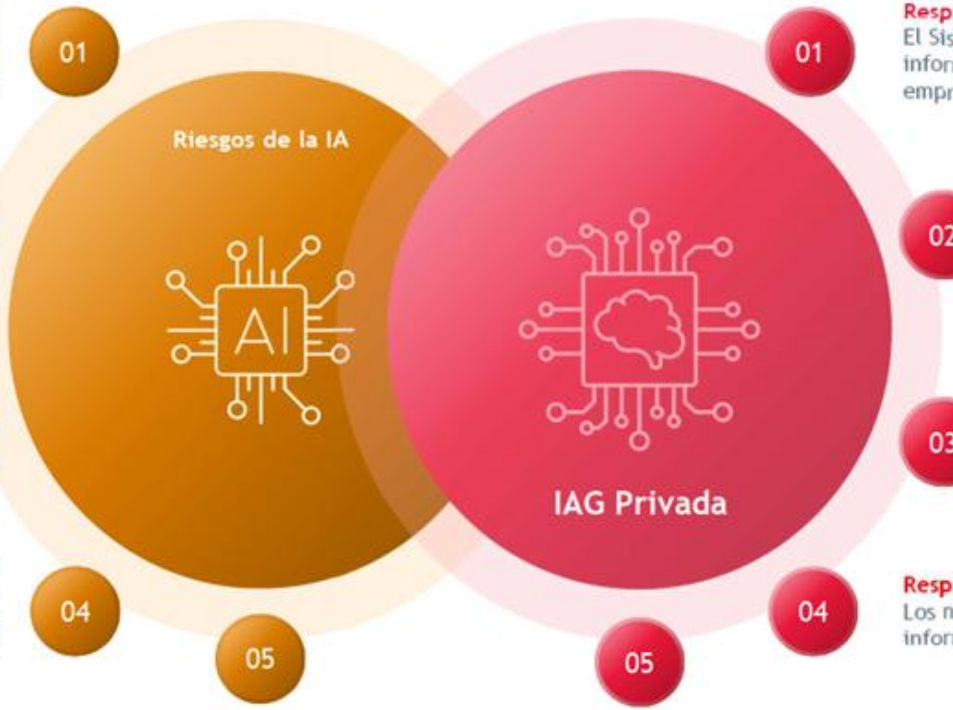
• **Conectamos una vector database (Milvus)**
Base de datos que indexa los documentos internos usando *embeddings (*representaciones vectoriales*)*, facilitando búsquedas semánticas.
Permite que el modelo entienda y acceda a fuentes internas sin exponer la data.

• **Accedemos a todo vía una interfaz segura hecha con Gradio**
Frontend sencillo y seguro para que los usuarios hagan consultas y vean respuestas sin exponer el backend.
Crear interfaces web simples y seguras.

Esquema de un modelo de IA Generativa Privada (Ejemplo)



Modelo Privado frente a los riesgos de la IA



Alucinaciones
Información incorrecta, inventada o poco fiable que no se basa en datos reales o precisos

Fuga de Información
Toda información introducida queda almacenada y puede usarse para entrenar nuevas versiones del modelo

Pérdida de control
Al depender de modelos externos no se gestiona cómo se procesan y utilizan los datos

Respuestas Sesgadas
Los prejuicios en los datos de entrenamiento pueden causar respuestas sesgadas que afecten la reputación de la empresa

Ataques externos
Los modelos abiertos pueden ser vulnerables a ataques que alteren su comportamiento o roben datos procesados

Respuestas Exactas
El Sistema solo responderá con la información proporcionada por la empresa

Información Segura
Cualquier información compartida está vectorizada, y por tanto carece de lógica frente a terceros

Control total
La empresa tiene siempre el control de la información que se proporciona al modelo

Respuestas sin sesgos
Los modelos son ajustados con la información aportada por la empresa

Entorno controlado
Solo se puede acceder a través de la intranet de la empresa, y después de identificarnos con usuario y contraseña

Beneficios tangibles del Gobierno de Datos

Menos riesgo → reducción de errores críticos y multas regulatorias.

Eficiencia operativa → hasta 20–30 % menos tiempo perdido corrigiendo datos.

Más confianza → decisiones estratégicas basadas en datos confiables.

Mejor reputación → ventaja competitiva y credibilidad frente a stakeholders.

Gobierno de Datos ≠ Tecnología

- **Personas** → roles claros, accountability y cultura de datos.
- **Procesos** → lineamientos, controles y prácticas consistentes.
- **Cultura** → ética, confianza y responsabilidad compartida.

"Gobernar datos no se logra con una app, se logra con personas."

“Los datos no valen nada, si no confiamos en ellos”

La confianza es el verdadero activo en la era digital.

“El reto no es acumular datos, es gobernarlos para generar confianza sostenible mediante IA”.

Inspirado en *Forbes Data & Trust 2024* y *MIT Sloan Management Review 2023*.

¿Cómo aplicarías gobierno de datos en tu proyecto de gestión de riesgos de fuga de información en IA?

“Comparte tu visión. La confianza en los datos comienza con tus decisiones.”

NUESTRA ORGANIZACIÓN

BDO AMERICAS

+1.162 Oficinas
+47,000 Colaboradores

BDO en PANAMÁ

2 Oficinas
+150 Colaboradores

BDO Global

164 Países
1803 Oficinas
+111.000 Colaboradores

SOMOS INTERAMÉRICAS

La sinergia de nueve países, con once oficinas y un equipo de más de 2.000 profesionales en distintos campos, nos convierte en el equipo ideal para alcanzar un objetivo compartido, con más de 50 socios liderando el camino de nuestros clientes hacia el éxito.

+de 2.000

Colaboradores de distintas profesiones, puestos a disposición de nuestros clientes.

9
Países y territorios

13
Oficinas

+ de 50
Socios

NUESTROS SERVICIOS

Para conocer todas las soluciones que podemos ofrecerte adjuntamos nuestro brochure de servicio.

Visita nuestra página web

www.bdo.com.pa

- **Auditoría y Aseguramiento**
- **BSO**
- **Advisory**
- **Impuestos**
- **Legal**

NUESTROS SERVICIOS

Auditoría y Aseguramiento

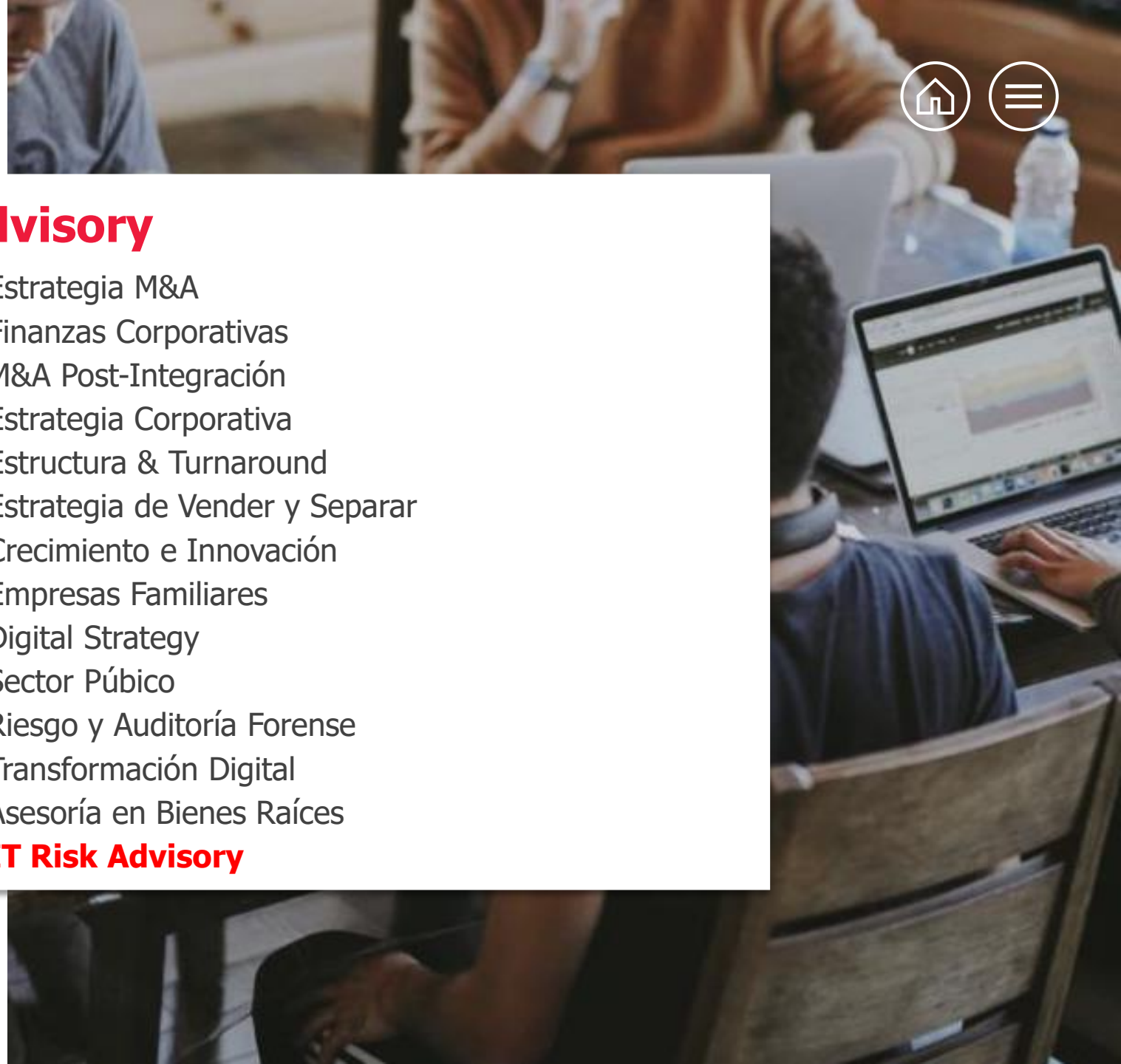
BSO

Advisory

Impuestos y Legales

Advisory

- Estrategia M&A
- Finanzas Corporativas
- M&A Post-Integración
- Estrategia Corporativa
- Estructura & Turnaround
- Estrategia de Vender y Separar
- Crecimiento e Innovación
- Empresas Familiares
- Digital Strategy
- Sector Púbico
- Riesgo y Auditoría Forense
- Transformación Digital
- Asesoría en Bienes Raíces
- **IT Risk Advisory**



IT Risk Advisory Services

- Auditoría de TI
- Riesgo de TI
- Gobierno de TI
- Cumplimiento de TI
- Control Interno de TI
- Evaluación de riesgo de Seguridad de la Información (SGSI)
- Evaluación de Riesgo de Ciberseguridad
- Cumplimiento de IT SOX
- Informes SOC
- Prevención de Riesgo de Fraude de TI
- Gestión de Riesgo de Terceros (TPRM)
- Evaluación de la Continuidad del Negocio (SGCN)
- Atestiguamiento SWIFT
- AML (Anti-Money Laundering) - Cumplimiento de TI
- Segregación de Funciones – SoD
- Evaluación de Procesos de Negocio Automatizados
- Evaluación específica de plataformas y marcos de trabajo (AUP)
- Evaluación de Inteligencia Artificial (SGIA)
- Certificación de Sistemas Contables
- Documentación de planificación y métodos de TI
- Capacitación sobre tópicos de actualidad y tendencias de TI

AUDITORÍA INTERNA A LA INTELIGENCIA ARTIFICIAL APLICADA A PROCESOS

Acompañamos a las organizaciones en implementar las acciones de control necesarias de manera de garantizar la privacidad, confidencialidad e integridad en la implementación y uso de tecnologías que incluyan Inteligencia Artificial (IA), IA basado metodológicamente bajo requisitos normativos de ISO/IEC 42001:2023 Inteligencia artificial - Sistema de gestión. En este sentido, la auditoría de Sistemas con foco en la IA, constituye un control necesario que mitiga riesgos introducidos por estas nuevas tecnologías.

Con la finalidad de atacar estos riesgos, leyes y entes reguladores comenzaron a exigir controles que permitan evaluar los potenciales riesgos y asegurar que la IA esté siendo utilizada de manera ética y segura. Muy recientemente la Unión Europea (UE) lanzó una regulación para la IA con el objetivo de proteger los derechos fundamentales de las personas, la cual entrará en vigencia a partir de 2026.

Este marco tecnológico innovador para el cual no se avizora un techo en su desarrollo y alcance nos exige implementar controles que garanticen un uso responsable de esta tecnología.

En BDO hemos desarrollado un plan de auditoría sobre IA, que abarca los siguientes dominios de control:

- Modelo de gobierno
- Arquitectura Tecnológica
- Calidad de los datos
- Medición de desempeño
- Controles black box
- Sesgo y factor humano

MODELO DE TRABAJO

Gestión de Riesgos de Inteligencia Artificial (IA)

En la actualidad, el uso de tecnologías avanzadas como la Inteligencia Artificial (IA) está transformando industrias a nivel global. Sin embargo, con el creciente impacto de la IA, surgen también nuevos riesgos que deben ser gestionados adecuadamente. El uso de IA en las empresas implica desafíos relacionados con la seguridad, la ética, la privacidad y el cumplimiento normativo. Para garantizar un uso responsable y alineado con las regulaciones actuales, las organizaciones deben implementar estrategias robustas de gestión de riesgos.

Con normativas como la **ISO/IEC 42001:2023, el AI Risk Management Framework del NIST, el Artificial Intelligence Act** de la Unión Europea, y otras regulaciones emergentes a nivel global, las empresas enfrentan la necesidad de mitigar riesgos de manera proactiva. El cumplimiento de estos marcos normativos es clave para asegurar que los sistemas de IA operen de manera segura, ética y dentro de los parámetros legales.
Es por ello que la gestión efectiva de los riesgos asociados a la IA es esencial para **operar de manera segura**, mantener la confianza de los clientes y cumplir con los **requisitos regulatorios.**

Nuestro servicio de **Gestión de Riesgos de IA** está diseñado para abordar estos desafíos, brindando a las organizaciones las herramientas necesarias para navegar por este complejo entorno tecnológico, anticiparse a los riesgos y garantizar el cumplimiento normativo.

BENEFICIOS DE NUESTRO ENFOQUE

Enfoque estratégico:

- Evaluación integral de riesgos de IA.
- Implementación de controles de seguridad y privacidad.
- Adaptación a regulaciones internacionales (ISO, NIST, AI Act).

Chapter 2: AI Governance and The Principles of Responsible AI

At the foundation of your approach to responsible AI are the core principles that define what responsible use looks like. These principles underpin the governance and guardrails embedded within your technology and processes, as well as the resources and training you provide to your employees. They help you assess and mitigate risk, implement effective controls, and assist with regulatory compliance.

In addition to content covered in this article, companies may also consider utilizing the NIST AI Risk Management Framework and ISO 42001. ISO 42001 is an international standard that specifies the requirements for establishing, implementing, maintaining, and continually enhancing an Artificial Intelligence Management System (AIMS) within an organization. These resources provide valuable frameworks for the responsible development and use of AI systems by effectively managing the risks associated with an organization's AI applications.

BDO's Responsible AI Principles help guide at any stage of your adoption journey, from design and development through implementation and ongoing maintenance. We not only help our clients adopt these principles — we are aligned with this framework at BDO for our own AI initiatives.

Escanea!

Plan de Auditoría de IA – IAI España

https://airisk.mit.edu/ – AI Risk Database (MIT)

Escanea!

Global Cyber Risk Analyzer – BDO Global

Índice de Percepción de la Corrupción por país con Análisis IA – BDO Argentina (Carlos Rozen)

Escanea!

Pirani Risk Software – Gestión de Riesgos Operativos

Pirani Risk Software – Seguridad de la Información

Escanea!

Asistente programas de Auditoría – Ciberseguridad
(Laura Lamprea)

Generador/Revisor de informes – Auditoría
(Laura Lamprea)

Escanea!

+700 Recursos de IA – David Ibáñez

Para más información

Edificio BDO, Urb. Los Ángeles, Ave. El Paical,
Panamá, Rep. de Panamá

Tel: +507 279 9700

www.bdo.com.pa