ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.

Simposio de Seguridad

Jefferson Gutiérrez

Socio Líder de Asesoría en Tecnología Forense

RIESGOS DE CIBERSEGURIDAD DERIVADOS DEL USO DE LA IA GENERATIVA

KPMG

Simposio de Seguridad

ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.

Rommel García

Socio de Asesoría en Ciberseguridad

RIESGOS DE CIBERSEGURIDAD DERIVADOS DEL USO DE LA IA GENERATIVA

KPMG

Simposio de Seguridad

ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.

Panorama General de la Inteligencia Artificial

Simposio de Seguridad

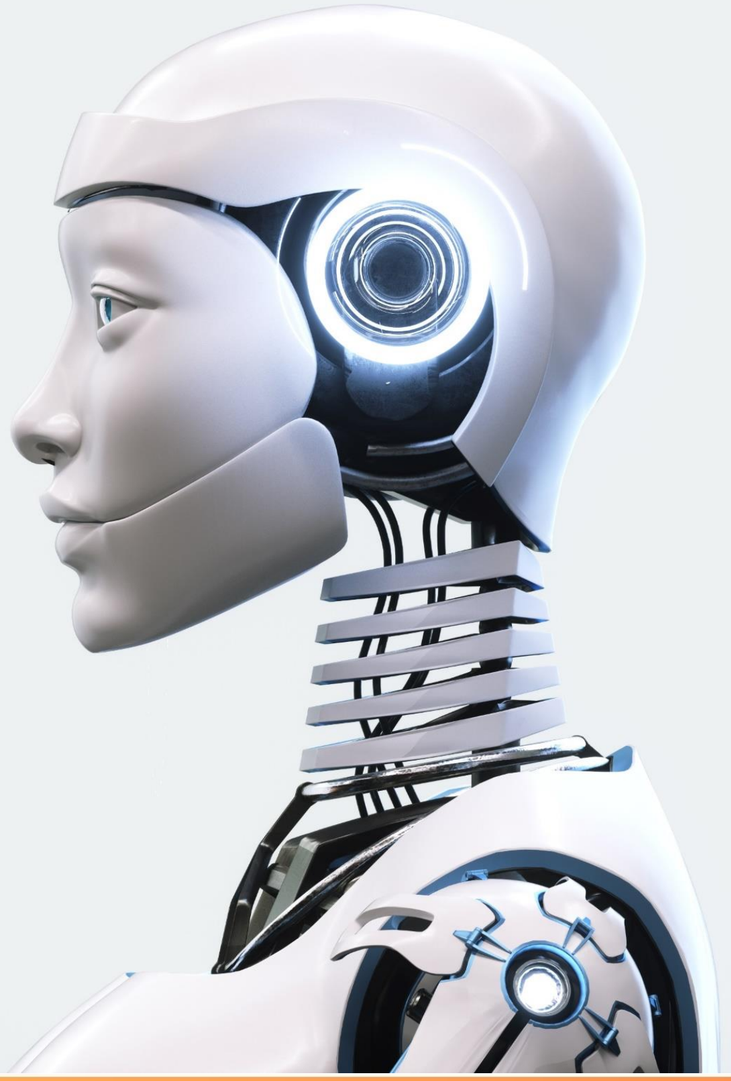
ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.

¿Qué es la Inteligencia Artificial?

La IA se define típicamente como la capacidad de una máquina para realizar funciones cognitivas que asociamos con las mentes humanas, como:

- Percibir
- Razonar
- Aprender
- Interactuar con el entorno
- Resolver problemas
- Incluso ejercitar la creatividad

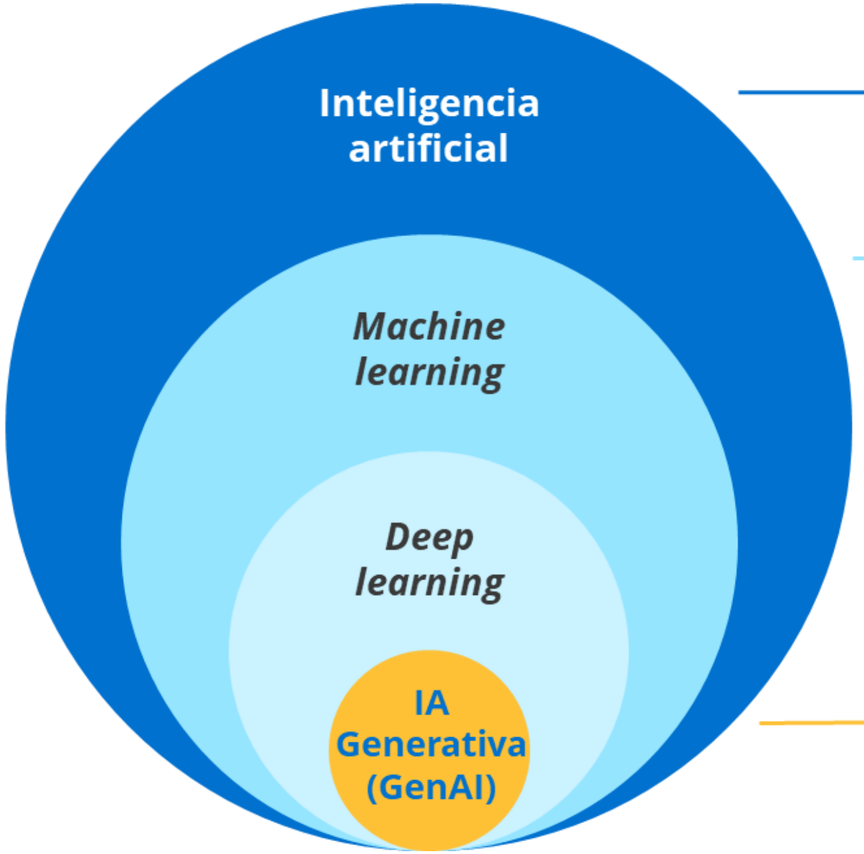
Ejemplos de tecnologías:
Robótica, vehículos autónomos, visión por computadora, lenguajes (Python), agentes virtuales y aprendizaje automático



Evolución de la Inteligencia Artificial

Drivers de la evolución

- Evolución de algoritmos
- Explosión de datos
- Incremento del poder de cómputo



Habilidad de una máquina para imitar el comportamiento humano.

Capacidad de detectar patrones, aprender y mejorar con la experiencia.

Habilidad para reconocer patrones complejos en imágenes, texto, sonidos y otros datos.

Capacidad de crear contenido nuevo, como texto, imágenes, videos, música o código.

Los ejecutivos están entusiasmados, pero aún hay cautela

Cuando se les pidió que **identificaran los principales beneficios de implementar GenAI** en su organización, los CEO citaron:

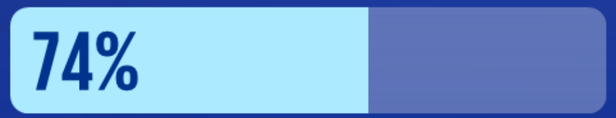
- **Aumento de la eficiencia y la productividad mediante la automatización de las operaciones rutinarias**
- **Mejorar las habilidades de la fuerza laboral para la preparación futura**
- **Aumento de la innovación**

Los CEOs identifican con mayor frecuencia estas tres áreas funcionales principales en las que su organización realizará inversiones en GenAI durante los próximos tres años:

Finanzas y contabilidad

Ventas y marketing

Tecnología de la información



Cuando se les pidió que identificaran los mayores desafíos cuando se trata de implementar GenAI, los CEO identificaron:

Desafíos éticos

Falta de regulación

Fuente: Perspectivas del CEO de KPMG para 2024

Simposio de Seguridad

ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.

¿La IA está en control?

IA en el lugar de trabajo

76% Reporta que su empresa utiliza IA

80% Utiliza IA de manera intencionada

Dependencia

42% Siente que no podría completar su trabajo sin ayuda de la IA

47% Se apoya en la IA para realizar una tarea en lugar de aprender cómo hacerla

59% Se preocupa por rezagarse si no usa IA en el trabajo

Uso displicente

42% Ha utilizado IA de manera inapropiada

53% Ha cometido errores en su trabajo debido a la IA

44% Ha utilizado IA de maneras que contravienen las políticas y directrices

68% Se apoyó en los resultados de la IA sin evaluar su precisión

47% Ha presentado contenido generado por la IA como propio

Impactos

+59% Reporta mayor eficacia, calidad del trabajo e innovación

46% Señala aumento de actividades que generan ingresos

27% Reporta aumento de carga de trabajo, estrés y presión

Gobernanza

59% Cuenta con una estrategia de IA

64% Proporciona capacitación en uso responsable de la IA

57% Cuenta con políticas y prácticas de gobierno de uso responsable

Reflexiones del talento sobre su organización

32% Señala que la IA ha aumentado los riesgos de cumplimiento y privacidad

51% Informa que aspectos clave de su trabajo pueden ser realizados por la IA

Fuente: Trust, attitudes and use of AI: estudio global 2025
https://kpmg.com/mx/es/tendencias/2025/04/trust-attitudes-and-use-of-ai.html

Generando confianza en la IA

Simposio de Seguridad

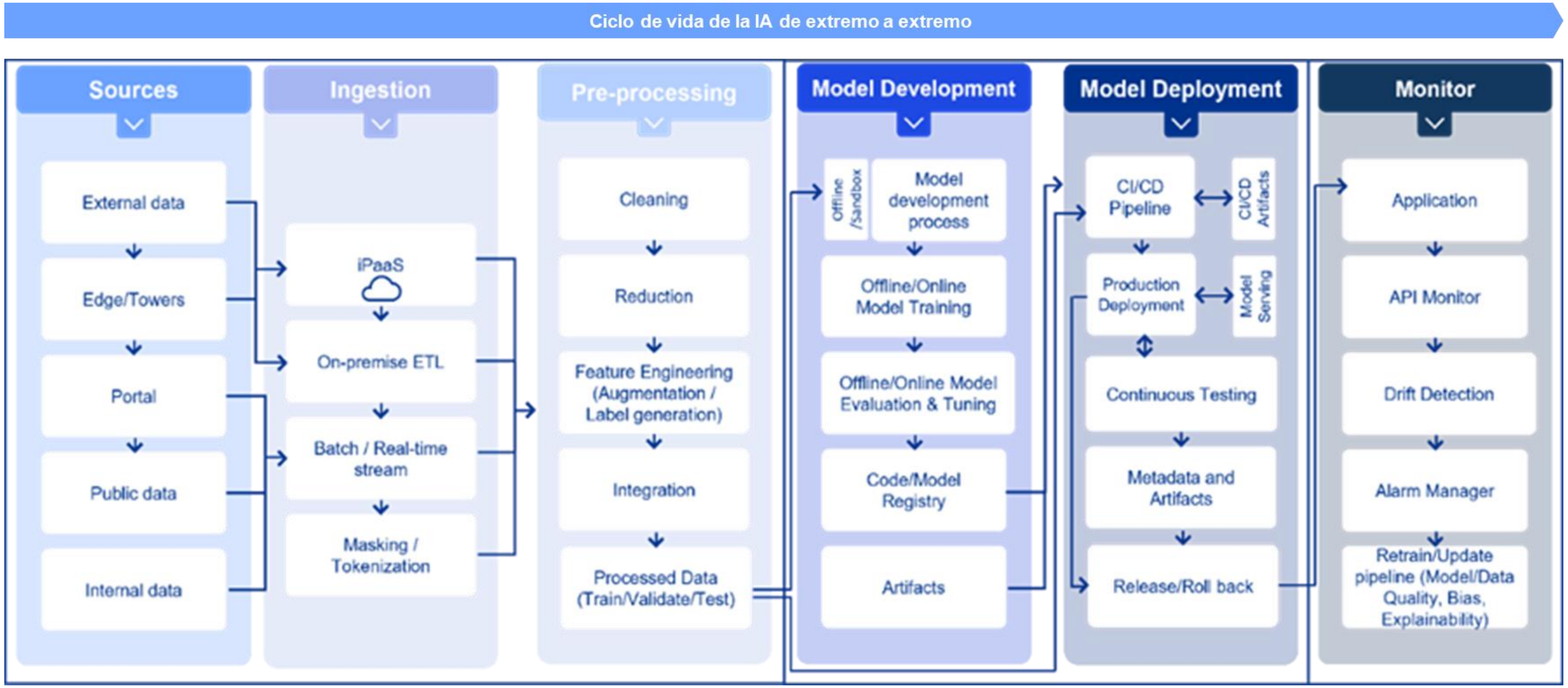
ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.

Generar confianza en la IA: gobernanza, cumplimiento y gestión de riesgos

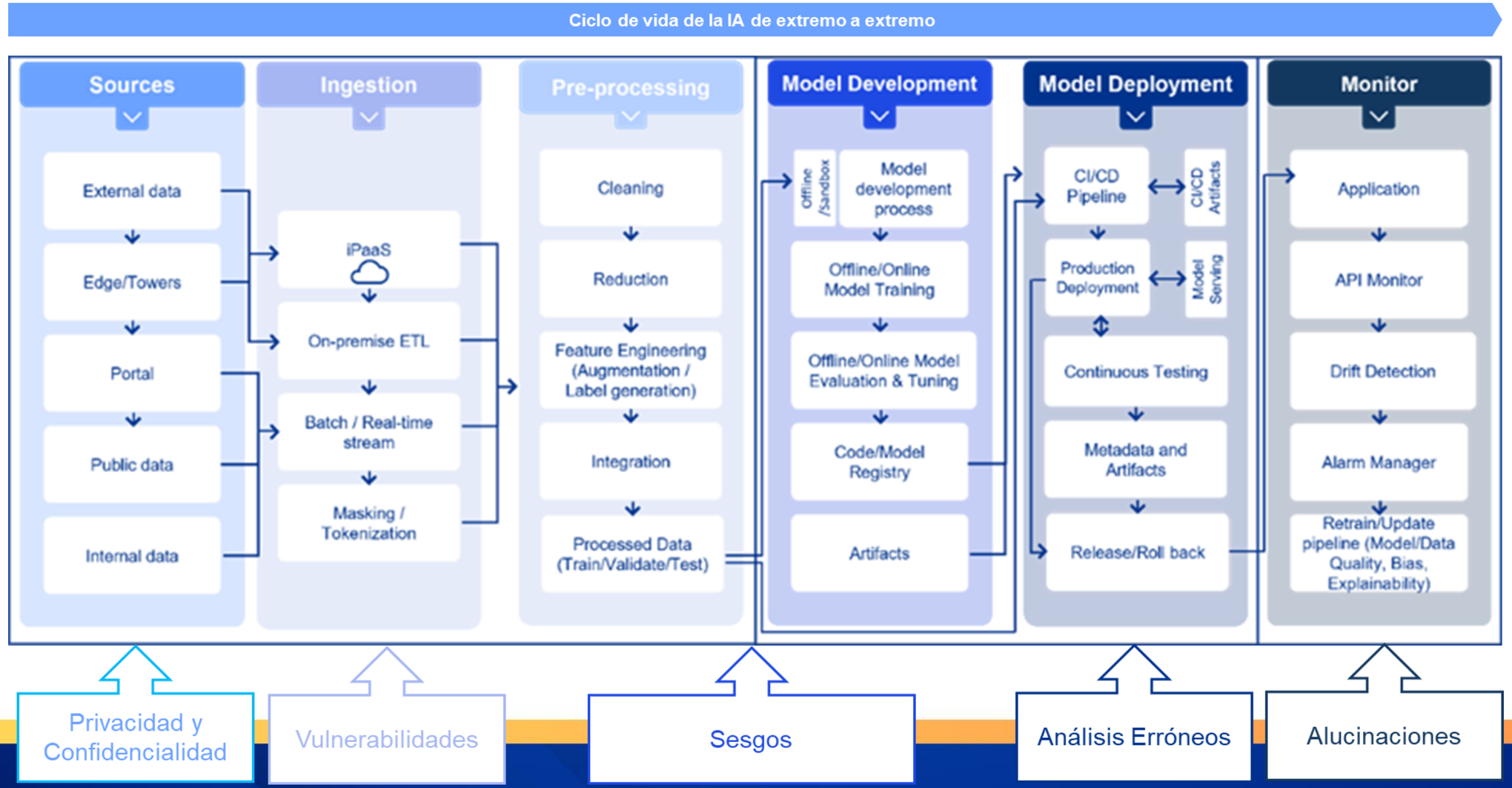
- El uso de la Inteligencia Artificial (IA) en los negocios está creciendo rápidamente y remodelando el panorama corporativo. Las organizaciones están adoptando cada vez más la IA para mejorar las operaciones, impulsar la innovación y obtener una ventaja competitiva. Al mismo tiempo, la rápida evolución de las tecnologías de IA introduce nuevos riesgos, desafíos y áreas de complejidad.

- Dado este entorno, es importante que las organizaciones evalúen cómo se diseñan, implementan y gobiernan los sistemas de IA, especialmente en lo que se refiere a la eficacia de los controles internos. La capacidad de evaluar y comprender la fiabilidad de los resultados generados por la IA es cada vez más vital para la toma de decisiones y la confianza de las partes interesadas.

¿Cómo trabaja la IA? Mapa Visual



¿Cómo trabaja la IA? Mapa Visual



Tendencias de vectores de ataque

Poisoning Attack

Un ataque de envenenamiento consiste en insertar datos maliciosos o alterados en el conjunto de datos utilizado para entrenar el modelo de IA, con el fin de afectar negativamente su capacidad para generar predicciones precisas o confiables.

El propósito es influir el comportamiento del modelo, ya sea para que el modelo cometa errores específicos, se vuelva más vulnerable a ataques posteriores, o actúe de manera predecible según los deseos del atacante.

Inference Attack

Un ataque de inferencia consiste en utilizar las propiedades y respuestas del modelo de IA para inferir información sobre su conjunto de entrenamiento, detalles sobre individuos específicos en los datos, o incluso parámetros internos del modelo.

El objetivo es extraer información confidencial o privada sin necesidad de acceder directamente a los datos subyacentes, aprovechándose de cómo el modelo procesa y responde a las consultas.

Model Evasion

Un ataque de evasión de modelo consiste en alterar las entradas al sistema de IA para que el modelo produzca una salida específica o para burlar su capacidad de detectar patrones de actividad maliciosa.

El propósito principal es causar que el modelo cometer errores sistemáticos necesarios para cumplir con los objetivos del atacante, ya sea pasando desapercibido o logrando una manipulación precisa de los resultados.

Data Extraction

El ataque de extracción de datos consiste en realizar consultas estratégicas al modelo de IA con la intención de extraer información subyacente sobre los datos de entrenamiento, el funcionamiento interno del modelo, o datos procesados por este.

Obtener datos confidenciales y sensibles que pueden ser utilizados para fines maliciosos, como robo de identidad, espionaje corporativo, o fraude financiero.

Simposio de Seguridad

ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.

Desafíos introducidos por el uso de IA Generativa

Riesgos y consideraciones internas

Propiedad intelectual

- Exposición de la propiedad intelectual
- Uso indebido de información confidencial
- Fugas involuntarias

Implicaciones para el talento

- Enmascaramiento de talentos
- Síndrome del impostor

Inexactitudes

- Respuestas falsas
- Modelos entrenados superficiales
- Falta de tarjetas modelo

Calidad de los datos

- Gestión de la verdad sobre el terreno
- Precisión de la salida
- Irrelevancia de los datos
- Escasez de datos
- Deriva de datos
- Pérdida de datos
- Toxicidad de los datos
- Transferir errores de aprendizaje
- Gobernanza de datos
- Medición de las puntuaciones de inicio

Sostenibilidad

- Costos computacionales
- Intensidad energética
- Impactos de los informes de carbono

Riesgos y consideraciones externas

Desinformación y discriminación

- Salidas dañinas
- Pérdida de control
- Alucinaciones
- Sesgo en la salida

Infracción

- Reclamaciones de derechos de autor
- Violación de la privacidad
- Infracción de responsabilidad

Reputación de marca

- Falta de creatividad
- Desplazamiento de puestos de trabajo
- Transparencia de salida

Amenazas cibernéticas y adversarias

- Estafas de phishing
- Pérdida de control
- Manipulaciones deliberadas
- Inyección rápida

Gobernanza de la IA

Simposio de Seguridad

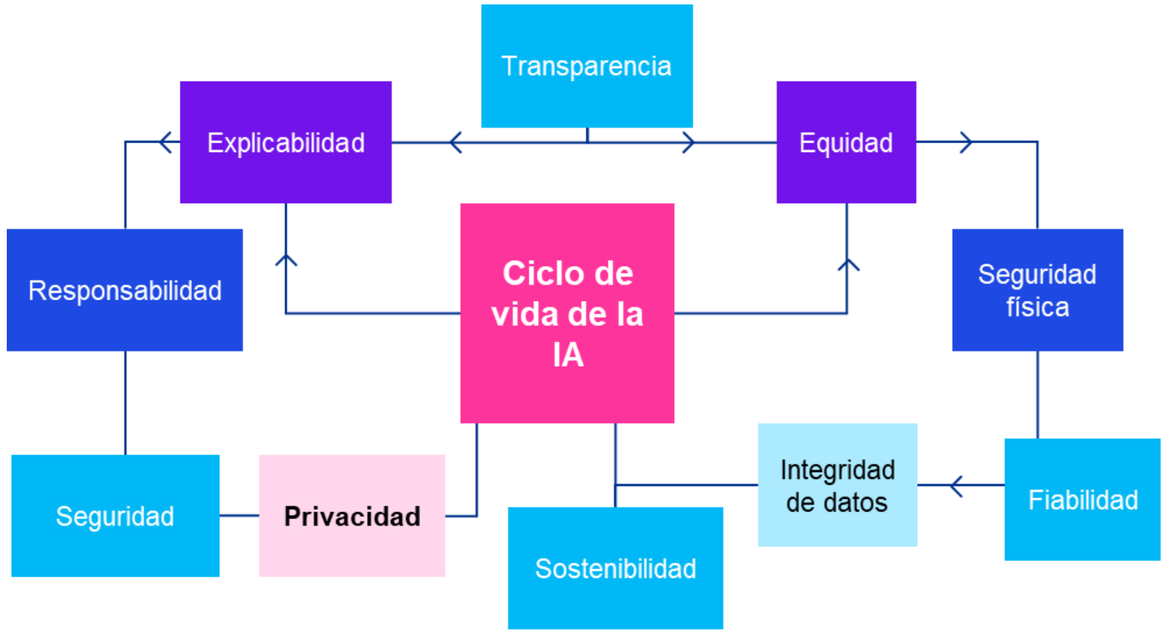
ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.

Principios de Gobernanza

Para que la inteligencia artificial genere valor de forma sostenible, debe operar bajo principios éticos y regulatorios durante todo su ciclo de vida:

Ciclo de vida de la inteligencia artificial

KPMG desarrolló un marco de IA confiable que basa su funcionamiento en 10 pilares que aseguran su correcto uso y funcionamiento en el ciclo de vida de la inteligencia artificial:



Marco de referencia "Trusted IA" de KPMG

¿Qué representan estos pilares?

01 Transparencia
Entender como funciona la IA

02 Explicabilidad
Por qué la IA llegó a una conclusión

03 Responsabilidad
Que haya humanos responsables de la supervisión de la IA

04 Seguridad
Protección contra ciberataques y mal uso

05 Privacidad
Respete datos personales y las leyes aplicables

06 Sostenibilidad
Que su uso sea eficiente en recursos

07 Integridad de datos
Que los datos sean correctos, completos y confiables

08 Fiabilidad
Que funcione como se espera, de forma consistente

09 Seguridad física
Que no cause daño a personas, bienes o procesos

10 Equidad
Que no discrimine

En resumen

Estos pilares son criterios de calidad y ética que deben guiar todo el ciclo de vida de la IA. No basta con que la IA funcione: debe hacerlo de forma justa, segura, explicable y responsable

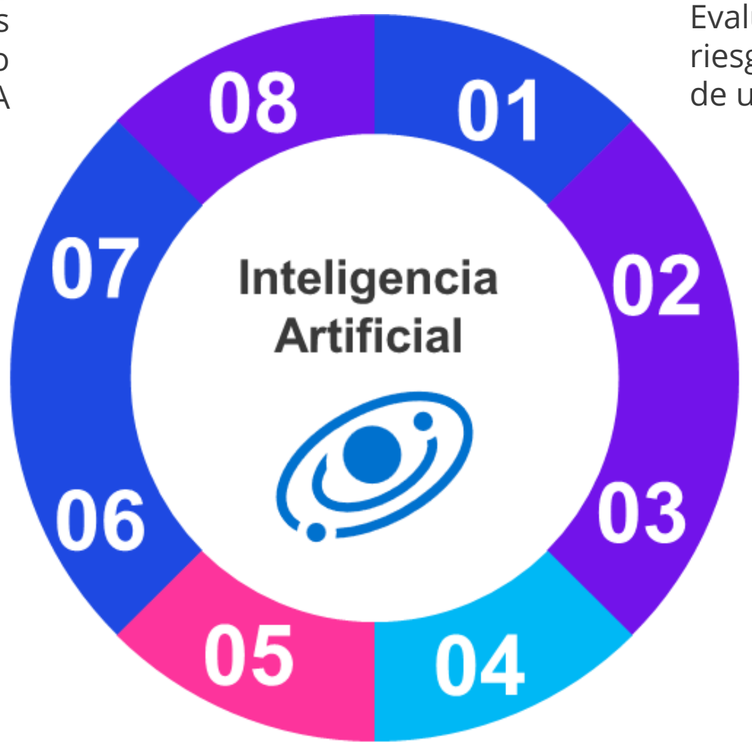
Nuestro enfoque al Gobierno de la IA

Trusted AI Training
Programas de formación para que los equipos comprendan los riesgos, principio éticos y buenas prácticas en IA

IA Risk Assessment
Evaluación estructurada de riesgos asociados a cada caso de uso de IA

IA system cards

Fichas técnicas que documentan cada modelo: propósito, riesgos, controles, desempeño, responsables, etc.



IA Regulation & Compliance
Revisión de cumplimiento con normativas locales e internacionales

IA Governance Framework
Modelo operativo que define políticas, roles, responsabilidades, proceso y estructuras de decisión

IA Inventory
Registro centralizado de todos los sistemas de IA en uso de la organización

IA Assurance
Validación y auditoría de los modelos de AI, sus controles y resultados

IA Security & Privacy
Estrategias, procesos y herramientas para proteger los sistemas de IA y los datos que utilizan

¿Quién es responsable por el Gobierno de IA?

La implementación de una gobernanza efectiva de inteligencia artificial no solo requiere principios claros, sino también una asignación precisa de responsabilidades.

Bajo el enfoque del modelo de tres líneas de defensa cada componente operativo de la gobernanza de IA (desde la evaluación de riesgos hasta la documentación técnica) debe tener un responsable definido que asegure su correcta ejecución, supervisión y validación.

Primera línea	Operación
Segunda línea	Control Interno
Tercera línea	Auditoría Interna

	Línea de defensa responsable	Entregables clave
IA Risk Assessment	**Segunda línea**	Matriz de riesgos, criterios de impacto
IA Regulation & Compliance	**Segunda línea**	Mapas de cumplimiento, políticas alineadas a normativa
IA Governance Framework	**Segunda línea**	Modelo RACI, políticas de IA, manual de gobernanza
IA Security & Privacy	**Segunda línea**	Evaluaciones de ciberseguridad, controles de privacidad,
IA Assurance	**Tercera línea**	Informes de auditoría, pruebas de control. Vigilancia del marco.
IA Inventory	**Primera línea**	Registro actualizado de modelos, base de datos de IA
IA System Cards	**Primera línea (con el soporte de la Segunda línea)**	Fichas técnicas por modelo: propósito, riesgos, controles
Trusted IA Training	**Primera línea (con el soporte de las otras dos)**	Planes de capacitación, materiales formativos

Conclusiones

Simposio de Seguridad

ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.

¿Cuáles son los próximos pasos en su viaje hacia la IA?

Simposio de Seguridad

ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.

GRACIAS
POR SU ATENCIÓN

Simposio de Seguridad

ANTAD
Asociación Nacional de Tiendas de Autoservicio y Departamentales, A.C.