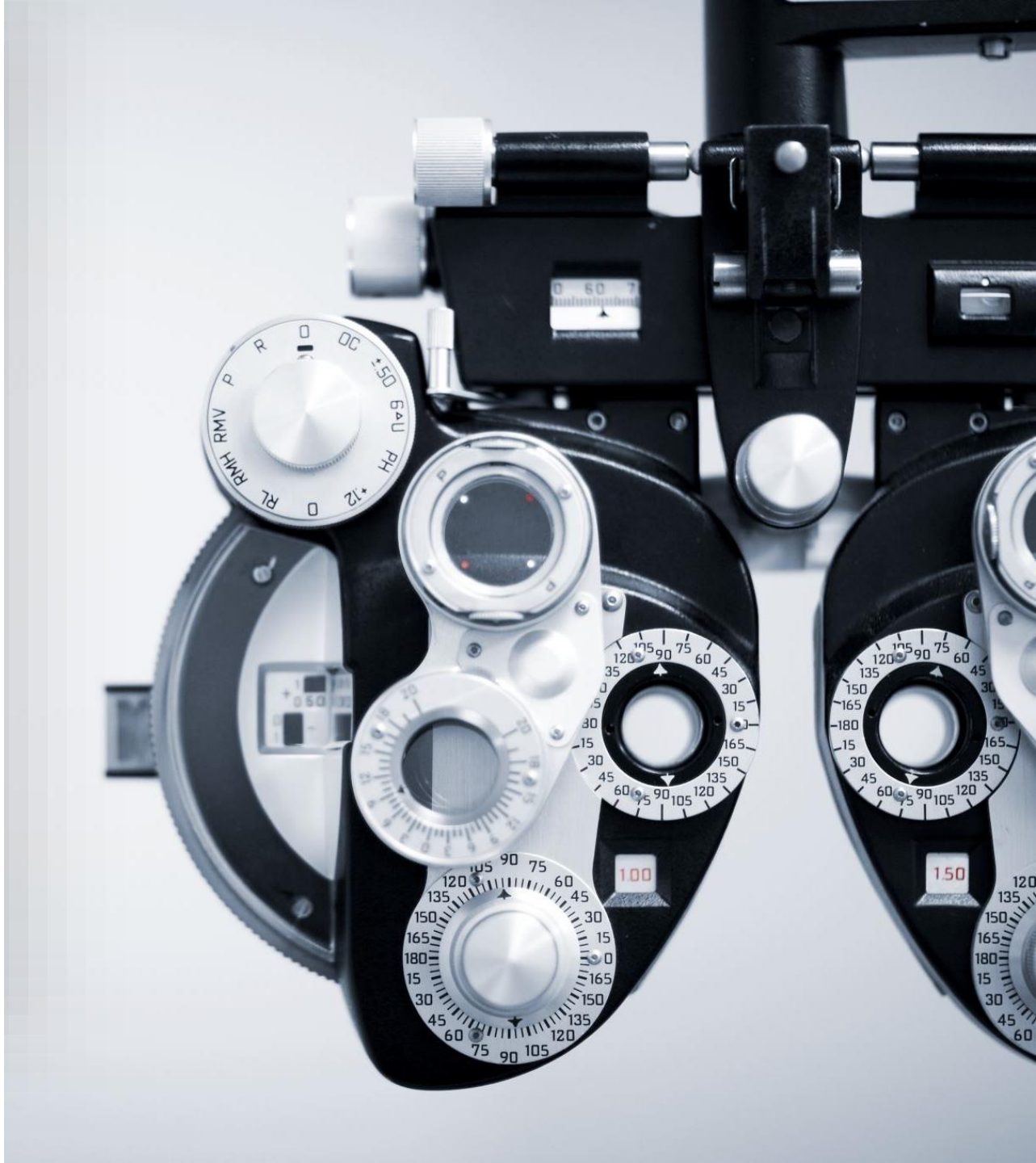
Auditoría Gubernamental

Dr. © Rafael Paredes-Carrasco

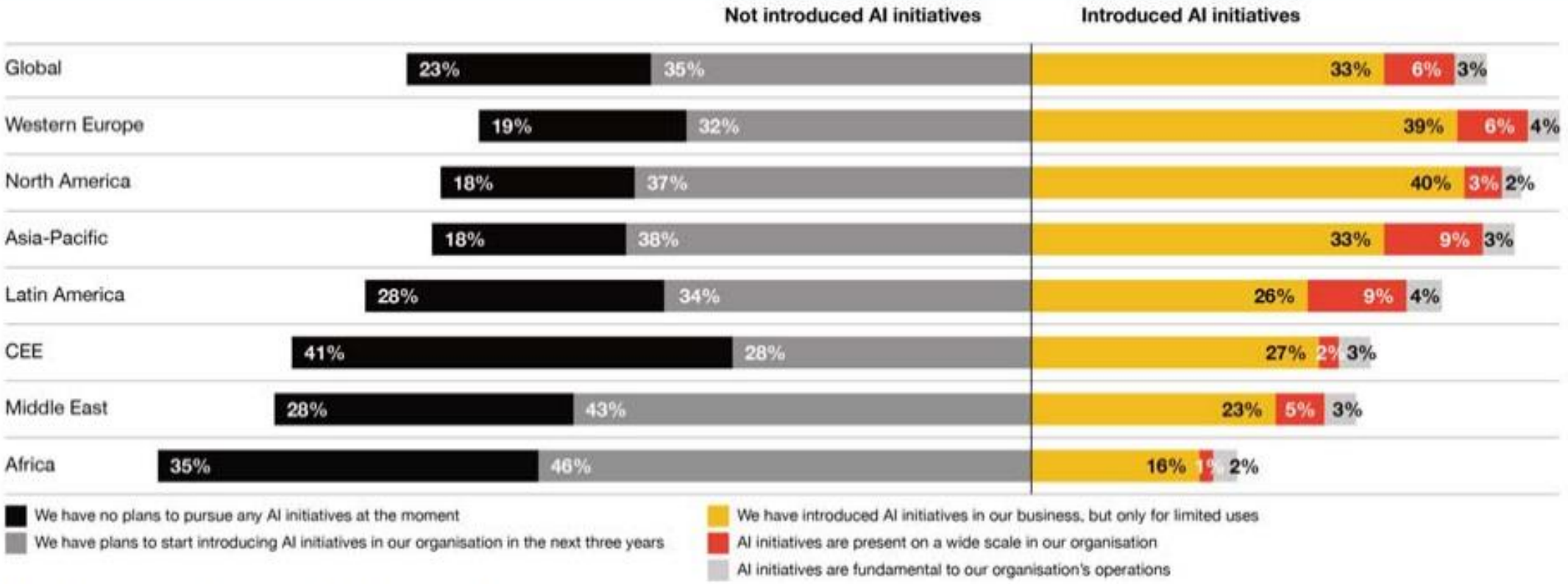
Profesor U. de Chile

Auditoría I de la IA aplicada a procesos

- ¿PUEDEN LAS MÁQUINAS PENSAR?
- El término Inteligencia Artificial fue acuñado en 1956 por el científico de datos John McCarthy.

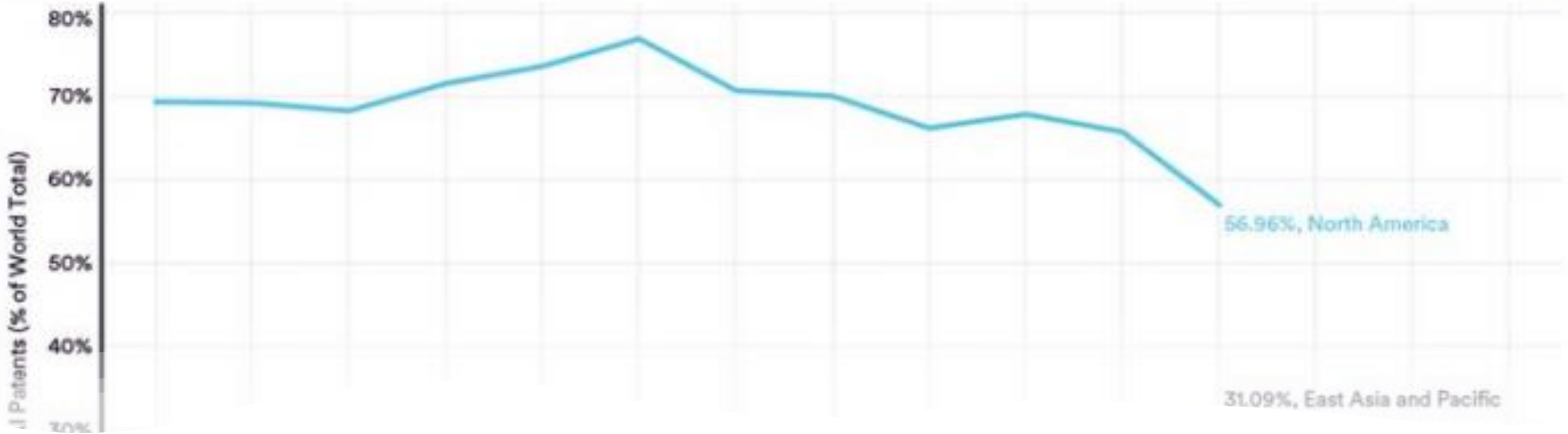


PLANES DE DESARROLLO DE INTELIGENCIA ARTIFICIAL



FUENTE: PwC. 22 nd Anual Global Global CEO Survey[4]

PATENTES DE INTELIGENCIA ARTIFICIAL POR REGIONES



- El uso de herramientas de IA en funciones empresariales ha subido del 50% al 56% en solo un año, según Mckinsey

- Es recomendable que el equipo de Auditoría Interna disponga de una combinación de conocimientos técnicos y humanísticos

LA "INTELIGENCIA ARTIFICIAL ACT": EL CAMINO DE EUROPA HACIA UNA REGULACIÓN COMUNITARIA SOBRE LA INTELIGENCIA ARTIFICIAL

«En cuanto a la Inteligencia Artificial ["IA"], la confianza es una obligación, no un adorno. Mediante estas reglas de referencia, la UE lidera la formulación de nuevas normas mundiales para que garanticen que se pueda confiar en la IA. Al establecer las normas, podremos facilitar el advenimiento de una tecnología ética en todo el mundo y velar por que la UE siga siendo competitiva. Nuestras normas, que son a prueba de futuro y propicias a la innovación, intervendrán cuando sea estrictamente necesario, esto es, cuando estén en juego la seguridad y los derechos fundamentales de los ciudadanos de la UE».

Margrethe Vestager

Vicepresidenta Ejecutiva responsable de la cartera de una Europa Adaptada a la Era Digital.

TIPOS DE SISTEMAS DE INTELIGENCIA ARTIFICIAL RECOGIDOS EN LA "AI ACT"

1 Sistemas de Inteligencia Artificial de Riesgo inaceptable (Art. 5)

Prohibido

- Manipulación del comportamiento, las opiniones y las decisiones humanas.
- Clasificación de personas en función de su comportamiento social.
- Identificación biométrica masiva a distancia y en tiempo real, salvo ciertas excepciones.

Ejemplo: Social scoring

2 Sistemas de Inteligencia Artificial de Alto Riesgo (HRAIS, Art. 6)

Permitido si se cumplen los requisitos de la IA de la evaluación ex-ante de conformidad

- Principales aspectos del reglamento (anexo III).
- Regímenes comunes con los que ya están sujetos a una norma armonizada de la UE.
- Lista adicional que debe ser revisada cada año por la EAIB (art. 84).

Ejemplo: Contratación

3 Sistemas de Inteligencia Artificial con obligaciones de transparencia específicas (Art. 52)

Permitido pero sujeto a obligaciones de información/ transparencia

- Interacción con humanos.
- Uso para detectar emociones o determinar categorías basadas en datos biométricos.
- Generación de contenidos manipulados.

Ejemplo: Personificación (bots)

4 Sistemas de Inteligencia Artificial de Riesgo Inexistente o Mínimo

Permitido sin restricciones

Ejemplo: Mantenimiento predictivo

FUENTE:Elaboración Propia. *Tipos de sistemas de Inteligencia Artificial recogidos en la "AI Act".*

SISTEMAS DE INTELIGENCIA ARTIFICIAL CONSIDERADOS DE ALTO RIESGO[9]	
1. Identificación biométrica y categorización de personas físicas.	Sistemas de IA destinados a utilizarse en la identificación biométrica remota en tiempo real, o en diferido, de personas físicas, conforme lo descrito en la letra d) anterior.
2. Gestión y explotación de infraestructuras críticas.	Modelos de IA vinculados a infraestructuras transportes u otros similares que puedan poner en peligro la vida y la salud de los ciudadanos.
3. Educación y formación profesional	Sistemas de IA que puedan determinar el acceso a la educación y la carrera profesional de una persona.
4. Empleo, gestión de trabajadores y acceso al autoempleo	Sistemas de IA destinados a la contratación –por ejemplo, para anunciar las vacantes, seleccionar o filtrar las solicitudes, evaluar a los candidatos en el curso de las entrevistas o pruebas– así como para tomar decisiones sobre la promoción y la terminación de las relaciones contractuales relacionadas con el trabajo, para la asignación de tareas y para el seguimiento y la evaluación del rendimiento y el comportamiento en el trabajo.
5. Acceso y disfrute de los servicios privados esenciales y de los servicios y prestaciones públicas	Sistemas de IA destinados a ser utilizados por las autoridades o en nombre de ellas para evaluar el derecho a prestaciones y servicios de asistencia pública, así como para conceder, revocar o reclamar dichas prestaciones y servicios.

Modelos de Inteligencia Artificial

- El análisis predictivo utiliza métodos y técnicas estadísticas avanzadas para asignar probabilidades de ocurrencia de eventos futuros basándose en datos históricos.

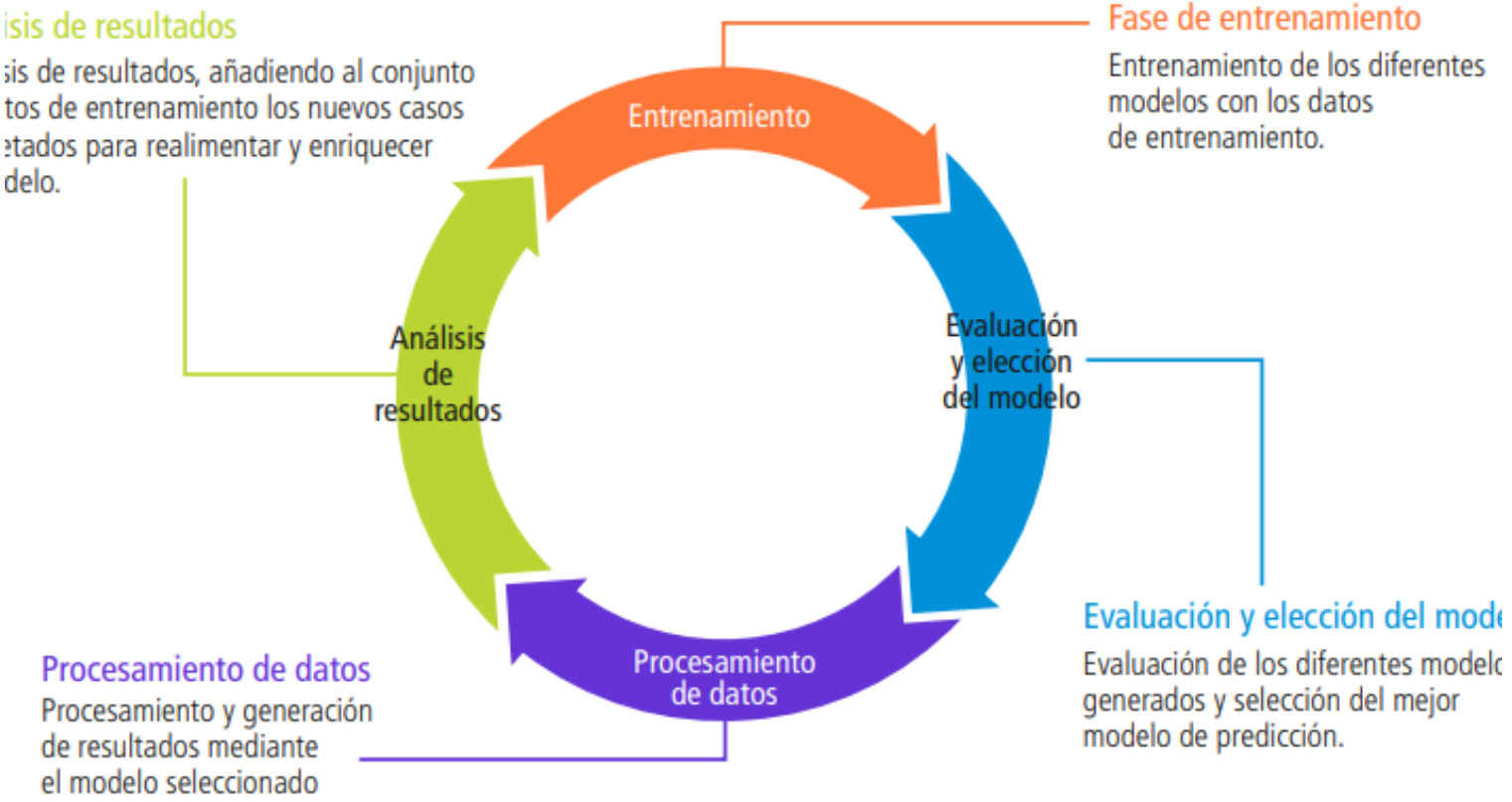
Ejemplos de caso de uso para ilustrar los avances de la IA en distintos campos

- **Lucha contra el terrorismo.** Actualmente se han desarrollado softwares basados en la aplicación de Machine Learning capaces de identificar patrones ante un posible atentado terrorista, mediante la extracción y combinación de información existente en diferentes medios de internet como redes sociales. También se han desarrollado sistemas capaces de monitorizar y detectar transacciones ligadas a grupos terroristas.

Ejemplos de caso de uso para ilustrar los avances de la IA en distintos campos

- La aplicación de la IA Generativa:
  - **A) Microsoft Copilot** impulsado por el modelo de lenguaje GPT-4, es un chatbot que ofrece funciones avanzadas de generación de código, integrándose en aplicaciones de Microsoft como Visual Studio Code y Word para mejorar la experiencia de desarrollo.
  - **B) Amazon Q** integrado en la infraestructura de AWS, está diseñado para abordar necesidades empresariales específicas, proporcionando soluciones y asistencia en tiempo real dentro de la consola de administración de AWS, con acceso a diversos modelos de IA para respuestas más precisas.
  - **C) IBM WatsonX** es una plataforma integral de datos e IA, ofrece herramientas para el desarrollo de soluciones personalizadas, un almacén de datos eficiente y un kit de herramientas para la gobernanza de la IA.

CONTINUA DEL MODELO BASADO EN APRENDIZAJE SUPERVISADO



laboración propia.

TIPOS DE ALGORITMOS DE MACHINE LEARNING: APRENDIZAJE SUPERVISADO, APRENDIZAJE NO SUPERVISADO Y APRENDIZAJE POR REFUERZO. REDES NEURONALES. INTELIGENCIA ARTIFICIAL GENERATIVA

IA Generativa

- En los últimos años han proliferado los llamados Modelos de Lenguaje Grande (Large Language Models, “LLM”, por sus siglas en inglés).
- Los Modelos de Lenguaje Grande utilizados por la IA Generativa tienen su origen en una arquitectura similar a las redes neuronales profundas.

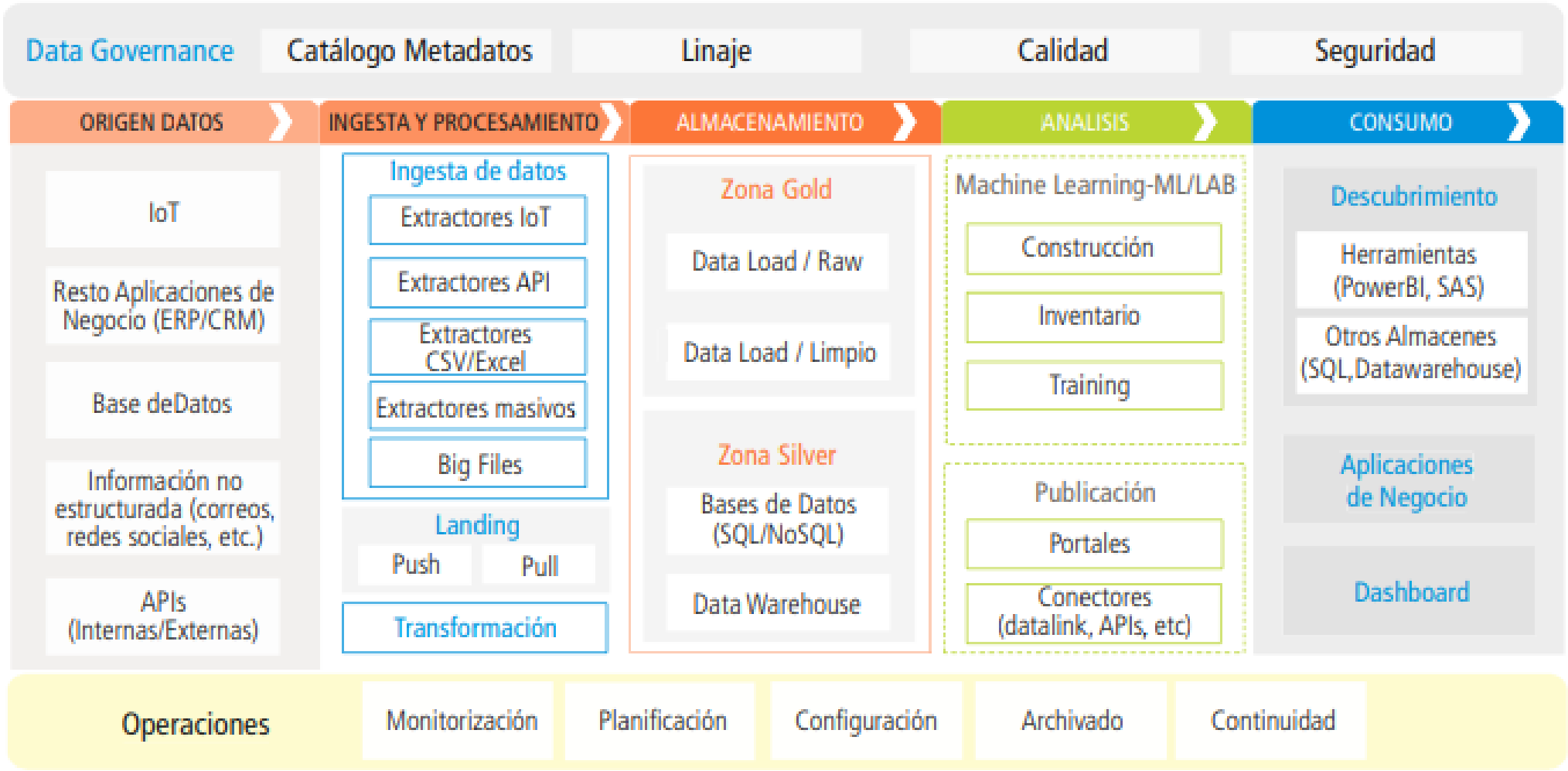
IA Generativa

- Algunos de los casos de uso más relevantes serían:
  - **Análisis de texto y datos.** Los modelos pueden analizar grandes volúmenes de texto, como correos electrónicos, documentos legales, o feedback de clientes, para extraer insights, tendencias y patrones significativos. Esto apoya la toma de decisiones basada en datos y ayuda a identificar áreas de mejora en productos y servicios.
  - **Entrenamiento y desarrollo.** Los LLM se emplean para crear simulaciones y escenarios de entrenamiento personalizados que ayudan en el desarrollo de habilidades del personal. Pueden facilitar escenarios interactivos para la capacitación en servicio al cliente, ventas, y más, adaptándose a las respuestas de los usuarios para ofrecer una experiencia de aprendizaje más efectiva.

ARQUITECTURA DE DATOS Y TI

- Los estándares de la industria aconsejan que toda solución basada en Inteligencia Artificial venga acompañada de una arquitectura que permita su automatización y facilite su uso y explotación. Desde un punto de vista de control interno, es muy relevante conocer la arquitectura de datos y TI de los sistemas de IA objeto de auditoría.

ARQUITECTURA DE DATOS Y TI

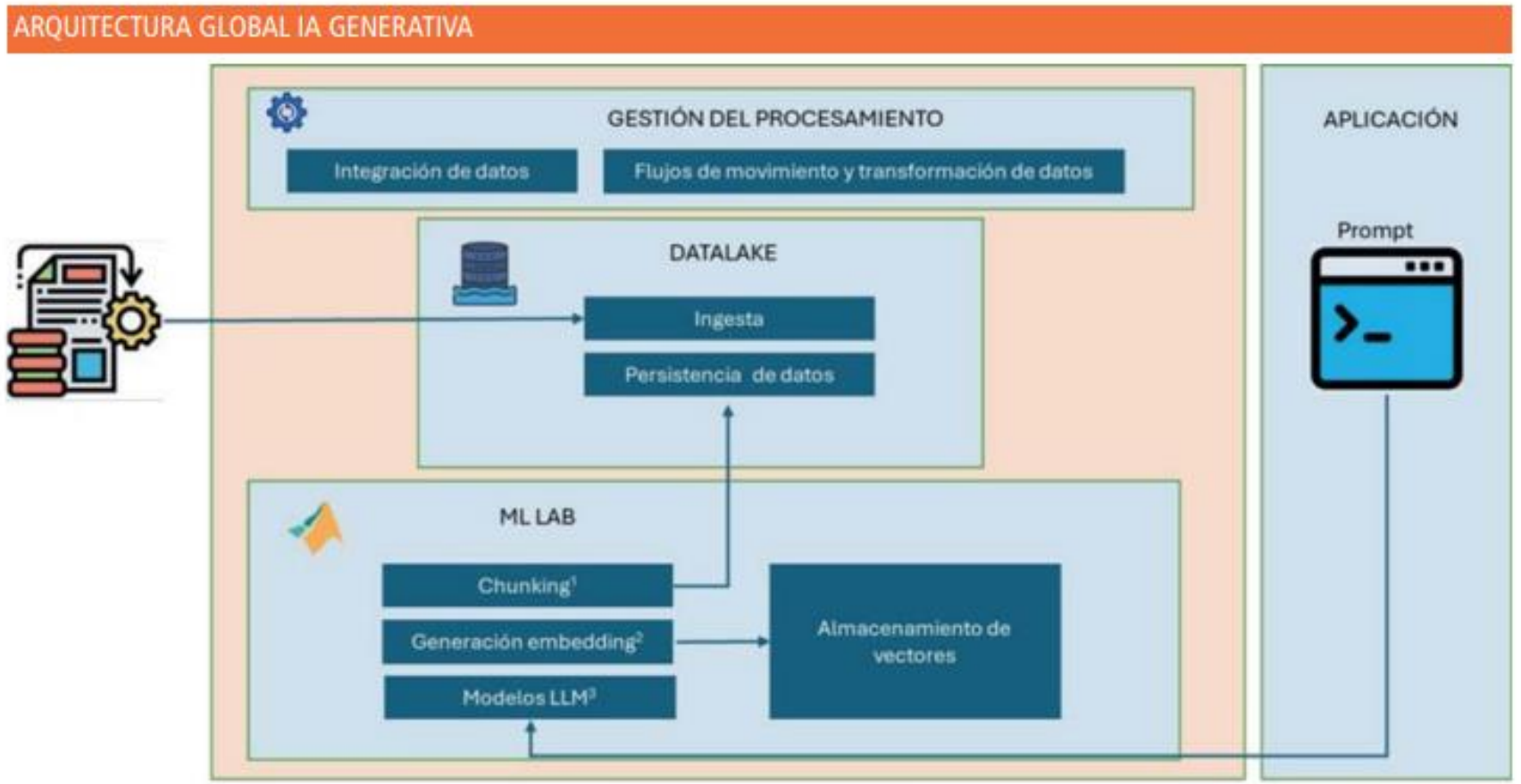


FUENTE: Elaboración Propia

4 flujos principales en cualquier solución de IA Generativa

- **Chunking o fragmentado de datos**: es una técnica que permite combinar o agrupar múltiples unidades de información en un número limitado de fragmentos, de forma que sea más fácil procesar y recordar la información.
- **Embedding**: se trata de una técnica de procesamiento de lenguaje natural que convierte el lenguaje humano en vectores matemáticos, lo que permite que las computadoras procesen el lenguaje de manera más efectiva al tratar las palabras como datos.

4 flujos principales en cualquier solución de IA Generativa



Marco de CI y riesgos de los procesos con IA

MONITORIZACIÓN CONTINUA DE RIESGOS	SUPERVISIÓN DE LA ARQUITECTURA DE DATOS Y TI	REVISIÓN DE LOS MODELOS DE IA	SUPERVISIÓN DE LA IMPLEMENTACIÓN	MONITORIZACIÓN TRAS LA PUESTA EN PRODUCCIÓN
Identificación de riesgos generales por el despliegue de sistemas de IA, y *risk-assessment* específico en función de la complejidad de los algoritmos y objetivos perseguidos con cada modelo de IA	• Actividades definidas para garantizar el acceso, tratamiento, privacidad, protección y destrucción de los datos, especialmente en aquellos en modelos predictivos sobre comportamiento humano. • Proporcionar seguridad razonable sobre los sistemas de TI y prevención de ciber-ataques.	• Asegurar un entendimiento adecuado de la operatividad de los algoritmos y *output* esperado. • Definición de métricas e indicadores de desempeño para la monitorización continua. • Revisión del comportamiento de los modelos en fases tempranas del despliegue de los motores.	• Desarrollar las pruebas de implementación suficientes para garantizar que el despliegue de los motores atiende los objetivos esperados. • Aprobaciones pertinentes de los Comités de Proyecto establecidos previo al *go-live*.	• Revisión periódica de las métricas e indicadores de desempeño. • Mecanismos de control para la identificación de comportamientos anómalos de desempeño, incluyendo las acciones correctivas necesarias a los algoritmos.

ENTORNO DE CONTROL (GOBIERNO Y CULTURA): MEJORES PRÁCTICAS Y RECOMENDACIONES

Existencia de una estrategia definiendo los principales órganos de gobierno para la implementación de sistemas de IA, iniciándose en aquellos modelos de mayor retorno e incrementando la inversión a medida que el *know-how* y *expertise* interno generado asegure el atendimiento de los objetivos establecidos.

La estrategia en la implementación de modelos de IA debe contar con las premisas generales de medición de desempeño de los modelos y, por extensión, de atendimiento de los objetivos perseguidos con el despliegue de los modelos de IA.

Plan estratégico de supervisión y evaluación del modelo de Gobierno para la mejora continuada del entorno de control y su reporte los órganos de administración correspondientes.

RIESGOS	COMENTARIOS
Riesgos de Gobierno	Relacionados con las estructuras internas de las organizaciones, con las políticas, metodologías y la toma de decisiones en los procesos, incluyendo la supervisión a alto nivel. Se debe enfatizar en los riesgos que puedan impactar en la gestión y liderazgo, en la independencia en la toma de decisiones, en el impulso a la transparencia y en la rendición de cuentas.
Riesgos Regulatorios	Vinculados con las áreas legales y de cumplimiento de las regulaciones. Cabe destacar los riesgos de cumplimiento asociados con las regulaciones externas (p.ej. RGPD) o internas (p.ej. Código Ético). Estos riesgos están relacionados con las actividades de los modelos IA, así como las decisiones y acciones relacionados con ella, sean coherentes con los valores y las responsabilidades éticas, sociales y legales de la Compañía.
Riesgo Reputacional	Relacionado con aquellos riesgos derivados por la presencia de sesgos en los modelos, o sanciones impuestas por incumplimiento normativo. También por la exposición a riesgos externos generados por terceros (ver debajo).
Riesgo de Sostenibilidad	El consumo de energía que permite la operatividad y funcionamiento de los sistemas de Inteligencia Artificial (por ejemplo, los modelos de IA Generativa) puede tener un impacto en los modelos y políticas de sostenibilidad de las empresas, por ejemplo, en lo relacionado con los compromisos adquiridos de reducción de gases de efecto invernadero, eficiencia energética o huella de carbono.

EVALUACIÓN DE RIESGOS: MEJORES PRÁCTICAS Y RECOMENDACIONES

Existencia de políticas internas de gestión de riesgos, específicas para los sistemas de IA; con la finalidad de que toda la compañía conozca y se implique en la identificación de riesgos relacionados con modelos de IA, de una manera continua. Se deben incluir directrices generales para el diseño e implementación de estrategias de mitigación de riesgos.

Existencia de un Inventario o Mapa de Riesgos con la naturaleza de cada uno de los riesgos identificados, incluyendo su criticidad, y cuando posible, cuantificación de la probabilidad de ocurrencia o potencial impacto financiero, en los sistemas de TI participantes en el proceso, así como otras categorías de riesgos.

ACTIVIDADES DE CONTROL: MEJORES PRÁCTICAS Y RECOMENDACIONES

Existencia de procedimientos internos y/o matrices de riesgos / controles, identificando el diseño de las actividades de control, y sus atributos clave; incluyendo la documentación soporte que evidencie la efectividad de los controles, así como los responsables de su ejecución y revisión para el abordaje *end-to-end* de los riesgos en procesos de negocio con sistemas de IA implementados.

El diseño de las actividades de control considera tanto la implementación como los controles recurrentes en sistemas de IA; siendo estos últimos evaluados de forma periódica para identificar riesgos no abordados por las correspondientes actividades de control interno, así como cambios en el diseño de los controles necesarios durante el ciclo de vida de los sistemas de IA.

Diseño y establecimiento de actividades oportunas de "revisión humana" de los comportamientos y resultados de los algoritmos de IA, para asegurar que éstos reflejan el objetivo original y, además, se utilizan de manera legal, ética y responsable.

INFORMACIÓN Y COMUNICACIÓN: MEJORES PRÁCTICAS Y RECOMENDACIONES

La organización publica sus mejores prácticas sobre valores éticos y morales en la utilización de sistemas de IA.

Compartir con la opinión pública los principios sobre Inteligencia Artificial, limitando aquellos aspectos que preocupan más a organismos públicos y privados.

Los máximos responsables de la compañía, accionistas y Consejo de Administración son informados de los aspectos relevantes del avance, desempeño real, y de las iniciativas sobre sistemas de IA alcanzados.

SUPERVISIÓN Y EVALUACIÓN: MEJORES PRÁCTICAS Y RECOMENDACIONES

Planificación y ejecución, por parte de Auditoría Interna, de las actividades de supervisión y evaluación de los modelos, así como validación de la idoneidad de los procesos de identificación de riesgos.

Definición de un plan a largo plazo (o estratégico) de auditoría de los modelos de IA, que acompañe la estrategia de la compañía a este respecto.

Definición de pruebas sustantivas o de auditoría del control interno sobre la integridad, precisión y confiabilidad de los datos sobre los que se construyen los algoritmos de IA.

ROL DE AI

- AI debe ayudar a evaluar, comprender y comunicar el efecto que los algoritmos tendrían sobre la capacidad de crear valor.
- Entre otros, AI aporta aseguramiento sobre la gestión de los riesgos relacionados con la confiabilidad de los datos y los algoritmos.

Programa de trabajo ilustrativo para la auditoría del CI de la IA aplicada en procesos

ESTRATEGIA

1. Modelo de Gobierno de sistemas de Inteligencia Artificial.
2. Arquitectura de datos y sistemas de TI.
3. Calidad de los datos.
4. Medición del desempeño.
5. El factor “Caja Negra” (Black Box) en los sistemas de IA.
6. El factor humano y el sesgo algorítmico.

MODELO DE GOBIERNO DE SISTEMAS DE IA

OBJETIVOS DE CONTROL O ACTIVIDADES DE CONTROL INTERNO GENERALES	PROCEDIMIENTOS DE AUDITORÍA INTERNA[25]
Implementación de un modelo de gobierno adecuado a la complejidad y riesgos de los sistemas de IA utilizados, desde un punto de vista de diseño y de operatividad del modelo de gobierno, incluyendo análisis ROI durante la fase de diseño de modelos de IA.	• **Revisión de la estructura organizativa y modelo de gobierno.** Determinar si el diseño del modelo de gobierno es adecuado y/o suficiente, y si opera tal y como fue diseñado, así como coincidente con los valores éticos y otras políticas internas de la organización. • **Revisión de los análisis ROI realizados,** con las metodologías de cálculo y justificativa de conclusión sobre la implementación de modelos de IA.
Existencia de una adecuada **segregación de funciones** sobre el sistema de IA.	• **Revisión de una matriz de segregación de funciones** que manifieste las responsabilidades y funciones conflictivas sobre el uso y gestión de la IA

ARQUITECTURA DE DATOS Y SISTEMAS DE TI

OBJETIVOS DE CONTROL O ACTIVIDADES DE CONTROL INTERNO GENERALES	PROCEDIMIENTOS DE AUDITORÍA INTERNA
Existencia de procesos y procedimientos formalizados sobre los **perfiles y roles de accesos** de los usuarios de los sistemas de IA. **Disponibilidad de controles de accesos** de usuarios a los sistemas de IA de acuerdo con los perfiles y roles predefinidos en los procedimientos de gestión de sistemas de IA, incluyendo: a) **Protocolo de autenticación de acceso** a los sistemas de IA (longitud mínima, caducidad predefinida contraseñas, entre otros aspectos de autenticación). b) **Gestión de cuentas de acceso y permisos de acceso** a los sistemas de IA.	• **Evidenciar la existencia y verificar la idoneidad de mecanismos de control** para la autenticación de los usuarios a los sistemas de IA. • **Revisión de los procedimientos de gestión altas/bajas** de usuarios y del listado de personas autorizadas a acceder a los sistemas de IA. • **Evidenciar una adecuada segregación de funciones** ente los diferentes roles (operadores, desarrolladores de algoritmos, propietarios de los datos, entre otros). • **Evidenciar la revisión periódica** de los permisos de acceso y gestión de accesos de usuarios con privilegios.

CALIDAD DE LOS DATOS

OBJETIVOS DE CONTROL O ACTIVIDADES DE CONTROL INTERNO GENERALES	PROCEDIMIENTOS DE AUDITORÍA INTERNA
Definición y documentación de un **proceso de lectura de los datos de entrada**, asegurando su integridad y exactitud.	• **Revisión** de los procedimientos del proceso de lectura de datos utilizado. • **Obtener** una muestra de los datos de entrada y verificar que la organización ha incorporado protocolos de lectura adecuados, garantizando la integridad y exactitud de los datos añadidos al modelo. • **Revisión de *log* de errores de entrada de datos** y validar que son revisados y resueltos antes de la ejecución de los modelos de IA.
Existencia de un proceso de **testeo de la calidad de los *inputs*** empleados por el modelo de Inteligencia Artificial, así como de las transformaciones realizadas (por ejemplo, normalización).	• **Obtener las evidencias** de que se han definido valores máximos y mínimos sobre variables cuantitativas, y que existen controles que detectan la presencia de valores anómalos. • **Revisión de los controles** sobre variables con valores nulos, o sobre variables de control *checksum* o similar.

EL FACTOR BLACK BOX EN LOS SISTEMAS DE IA

- Se refiere a aquellos algoritmos de IA que, por su complejidad y/o sofisticación, los mecanismos internos de ejecución ente los datos de entrada y los de salida son difícilmente entendibles o explicables.

EL FACTOR BLACK BOX EN LOS SISTEMAS DE IA

OBJETIVOS DE CONTROL O ACTIVIDADES DE CONTROL INTERNO GENERALES	PROCEDIMIENTOS DE AUDITORÍA INTERNA
Existencia un procedimiento documentado de análisis de sensibilidad del modelo ante fluctuaciones en los datos de entrada al modelo, y la interpretación de los resultados, con la finalidad de reducir el riesgo del factor *black box*, considerando la definición de: a) Métricas de precisión, exactitud y rendimiento de los sistemas de IA. b) Valores predefinidos de falsos positivos y falsos negativos. c) Mecanismos de supervisión para la adaptabilidad de sistemas de IA no supervisados (o de aprendizaje continuo) a nuevos datos y supervisión de la idoneidad de las conclusiones sostenibles en el tiempo con el aprendizaje continuo.	• **Evaluar el establecimiento de las métricas** o conjunto de métricas agregadas para determinar en los sistemas de IA su precisión, exactitud, sensibilidad u otro parámetro de rendimiento relativo a la aplicación del principio de exactitud de los datos. • **Evidenciar** los análisis realizados e interpretados sobre los valores de las tasas de falsos positivos y falsos negativos que arroja el componente IA de cara a determinar la precisión, la especificidad y la sensibilidad del comportamiento de los sistemas de IA. • **Evidenciar** la supervisión realizada del grado de adaptabilidad a nuevos datos o tipos de datos de entrada para el caso de modelos de IA no supervisados. • **Validar** los mecanismos de supervisión continua de modelos de aprendizaje continuo, con el objetivo de **verificar** que las conclusiones extraídas siguen siendo válidas, el componente es capaz de adquirir nuevo conocimiento y no se está produciendo una pérdida de las asociaciones previamente aprendidas durante el aprendizaje inicial.

EL FACTOR HUMANO Y EL SESGO ALGORÍTMICO

- Valores éticos y morales.
- **El sesgo algorítmico** (algorithm bias). En los sistemas de IA, el sesgo algorítmico ocurre cuando los valores de los humanos que lo diseñan y desarrollan terminan, de alguna forma intencionada o no, en los algoritmos de IA que desarrollan.

OBJETIVOS DE CONTROL O ACTIVIDADES DE CONTROL INTERNO GENERALES	PROCEDIMIENTOS DE AUDITORÍA INTERNA
Actividades de control diseñadas con el objetivo de impedir que los **resultados de los sistemas de IA sean utilizados de forma ilegal o delictiva, o incumpliendo cualquier regulación externa o política empresarial interna.**	• **Revisión** de los objetivos o estrategia de implantación de los sistemas de IA, **e identificar** cualquier brecha legal, o en la regulación externa o políticas internas. • **Revisión** de los resultados de los sistemas de IA, **y asegurar** que los mismos son utilizados sin intenciones ilícitas o legales, o en contra de la regulación externa o políticas internas de la compañía.

Referencias

Agencia Española de Protección de Datos. Requisitos para auditorías de tratamientos de datos personales que incluyan Inteligencia Artificial. 2021.

Instituto de Auditores Internos de España (2024). Auditoría Interna de la Inteligencia Artificial aplicada a procesos de negocio. La fábrica de pensamiento.

Recuero, P. Los 2 tipos de aprendizaje en Machine Learning: supervisado y no supervisado. 2017 [Blog] http://data-speaks.luca-d3.com/2017/11/que-algoritmo-elegir-en-ml-aprendizaje.htm