Auditoría Interna de la Inteligencia Artificial aplicada a procesos de Negocio

Daniel Tortosa Illana (Telefónica)
Javier Escribano Alarcón (Repsol)

27 de febrero de 2023

AUDITORÍA INTERNA

LA FÁBRICA DE PENSAMIENTO
INSTITUTO DE AUDITORES INTERNOS DE ESPAÑA

Auditoría Interna de la Inteligencia Artificial aplicada a procesos de negocio

The Institute of Internal Auditors

40 años Instituto de Auditores Internos de España

AUDITORÍA INTERNA

LA FÁBRICA DE PENSAMIENTO
INSTITUTO DE AUDITORES INTERNOS DE ESPAÑA

Auditoría Interna de la Inteligencia Artificial aplicada a procesos de negocio

Febrero 2023

MIEMBROS DE LA COMISIÓN TÉCNICA

COORDINACIÓN:

Daniel Tortosa Illana; ICAEW, ROAC. TELEFÓNICA.

Pablo Ausín Sánchez; PMP-PMI. INDITEX.

Luis Enrique Corredera. DELOITTE.

José Ignacio Díaz Aroca; CIA, CISA, CFE, COSO CI, COSO ERM, CESCOM. INDRA.

Javier Echevarría Blanco. BBVA.

Javier Escribano Ajarón; CISA, COBIT, ITIL y PMP. REPSOL.

Alejandro Martínez Morillo; CISA, CDPSE; Lead Auditor 27001. PwC.

Andrés Morales Fernández. KPMG.

Borja Rioja Mata. MAPFRE.

Jaime Sabau Jiménez. EY.

Juan José Villar Roldán. IBERDROLA.

The Institute of Internal Auditors

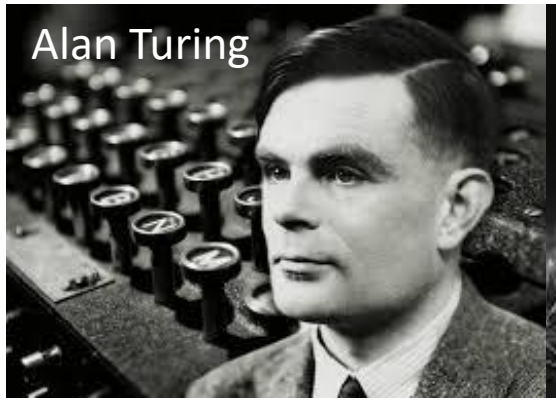
40 años Instituto de Auditores Internos de España

Índice

ÍNDICE

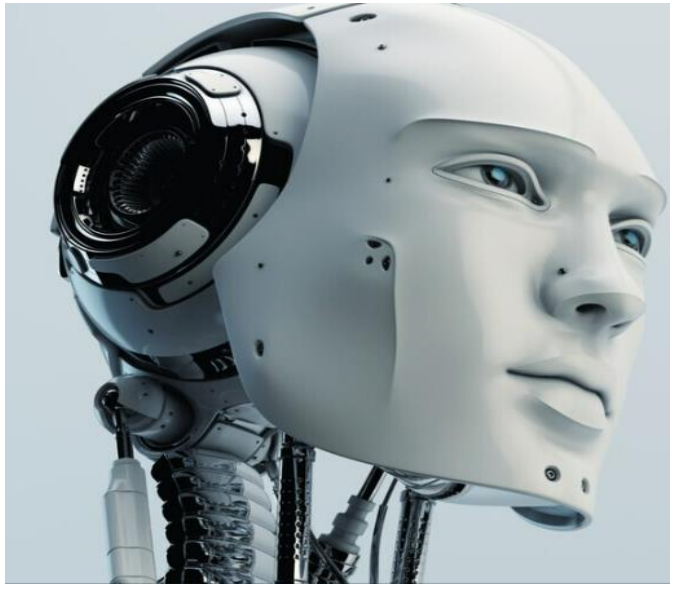
Auditoría Interna de la Inteligencia Artificial aplicada a procesos de negocio

1> ¿QUÉ ES LA INTELIGENCIA ARTIFICIAL (IA)?



Preguntemos a la propia IA

chat.openai.com



BE ¿Qué es la Inteligencia Artificial?

La Inteligencia Artificial es una rama de la informática que se enfoca en el desarrollo de sistemas que imitan la inteligencia humana. Se logra a través de técnicas como aprendizaje automático, procesamiento del lenguaje natural y visión por computadora.

BE ¿para que sirve la inteligencia artificial?

La Inteligencia Artificial se utiliza en una amplia variedad de aplicaciones, tales como:

1. Análisis de datos y toma de decisiones.
2. Automatización de procesos industriales y comerciales.
3. Reconocimiento de voz y lenguaje.
4. Diagnóstico y tratamiento en medicina.
5. Mejora de la eficiencia energética.
6. Investigación y desarrollo en ciencias.
7. Juegos y entretenimiento.

En resumen, la Inteligencia Artificial se utiliza para mejorar la eficiencia, efectividad y calidad en una amplia gama de campos y sectores.

1> LA PENETRACIÓN DE LA INTELIGENCIA ARTIFICIAL EN LAS ORGANIZACIONES.

- **Fuerte inversión y foco en planes de desarrollo de Inteligencia Artificial**, con Europa y Estados unidos a la cabeza.

- Los sectores con alto componente tecnológico (***high tech),*** **empresas de telecomunicaciones y servicios financieros** son las que muestran mayores índices de adopción de modelos de inteligencia artificial, en las áreas de desarrollo de **producto y servicios y área de riesgos en el sector financiero.**

- Norteamérica y Asia a la cabeza en **patentes de inteligencia artificial.**

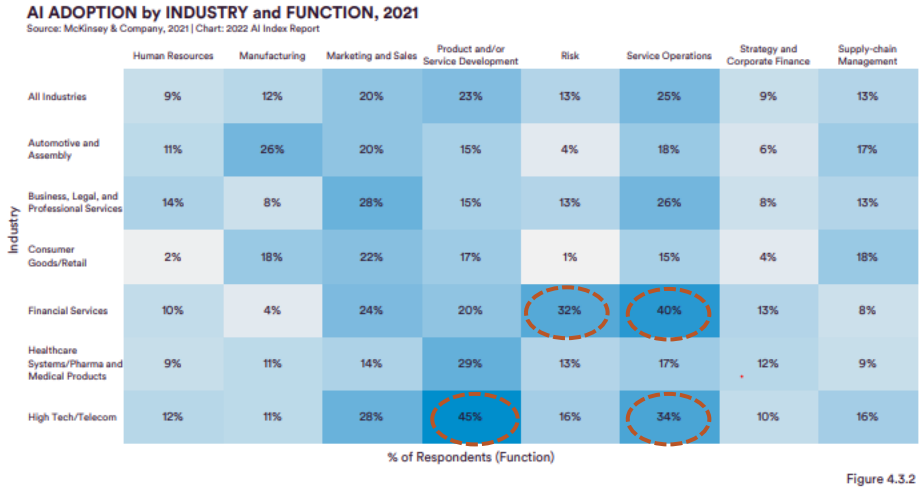
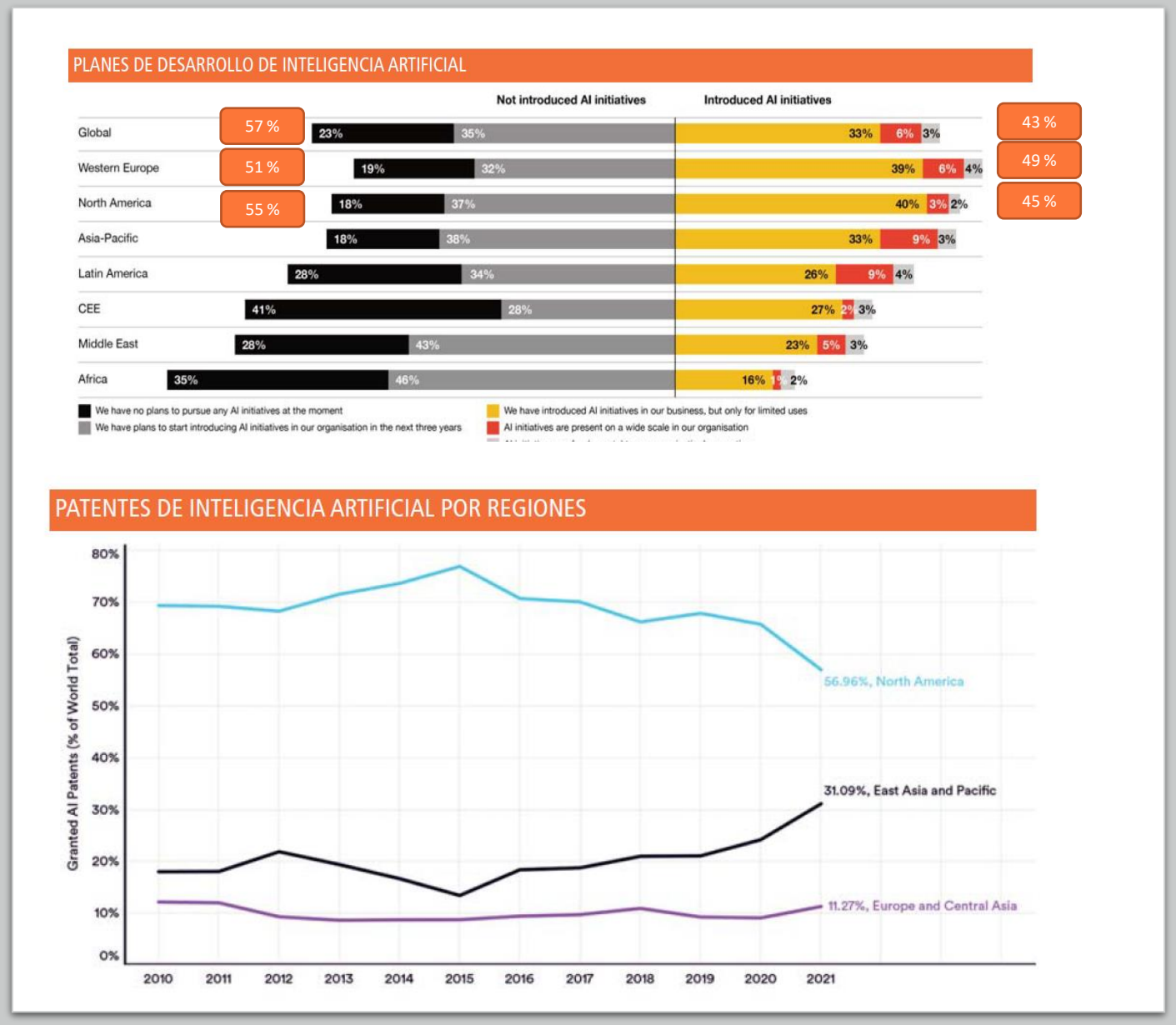


Figure 4.3.2



1> La “Inteligencia Artificial Act”.

En abril 2021, la Comisión Europa aprobó una propuesta de Reglamento sobre normas armonizadas en materia de IA.
Se establece una clasificación de los sistemas de IA en función del uso, con una serie de requisitos y limitaciones para cada uno.

Sistemas de Inteligencia Artificial			
Riesgo Inaceptable (Art. 5)	**Alto Riesgo (Art.6 & Anexo-III)**	**Riesgo Medio (Art. 52)**	**Riesgo Inexistente o Mínimo**
PROHIBIDOS	**PERMITIDOS con evaluación de conformidad ex-ante,** sobre la evaluación y mitigación de riesgos, calidad de datos, supervisión humana, entre otros ítems evaluables.	**PERMITIDOS** pero sujetos a obligaciones de información/transparencia específicos.	**PERMITIDOS** sin restricciones
• **Trato perjudicial o desfavorable** individual o colectivo derivado de: a) Manipulación del comportamiento, opiniones o decisiones humanas o b) Clasificación de personas en función de su comportamientos social. • **Identificación biométrica masiva** a distancia y en tiempo real, salvo en ciertas excepciones por orden judicial y razones de seguridad o prevención de delitos.	• Identificación biométrica y categorización de personas físicas. • Gestión y explotación de infraestructuras críticas. • Educación y formación profesional. • Empleo, gestión de trabajadores y acceso al autoempleo • Acceso y disfrute de los servicios privados esenciales y de los servicios y prestaciones públicas • Asuntos relacionados con la aplicación de la ley • Gestión de la migración, el asilo y el control de fronteras • Administración de justicia y procesos democráticos	• Interacción con humanos. • Uso para detectar emociones o determinar categorías basadas en datos biométricos. • Generación y/o recomendación de contenidos.	No incluidos en los anteriores, permitidos sin restricciones.
Ejemplo: Social Scoring	**Ejemplo: Contratación**	**Ejemplo: Atención al cliente (Bots)**	**Ejemplo: Análisis Predictivo**

1> La “Inteligencia Artificial Act”.

El Régimen sancionador de la Ley IA, establece sanciones niveles de sanción en función de la gravedad del incumplimiento.
En términos comparativos se prevé más gravosa que el Régimen General de Protección de Datos.

“Inteligencia Artificial Act”

Régimen General de Protección de Datos.

Incumplimiento	Sanción
Incumplimiento relativo a **prácticas prohibidas**	hasta **€30 millones** o el 6% del volumen de negocio (*).
Incumplimiento de cualquier otro **requisito u obligación**	hasta **€20 millones** o el 4% del volumen de negocio (*).
Suministro de **información incorrecta, incompleta o engañosa**	hasta **€10 millones** o el 2% del volumen de negocio (*).

Sanciones **más graves: €20 millones** o 4% volumen de negocio.

Sanciones **menos graves: €10 millones** o 2% volumen de negocio.

(*) Volumen de negocio anual considerado a escala mundial del ejercicio financiero anterior.

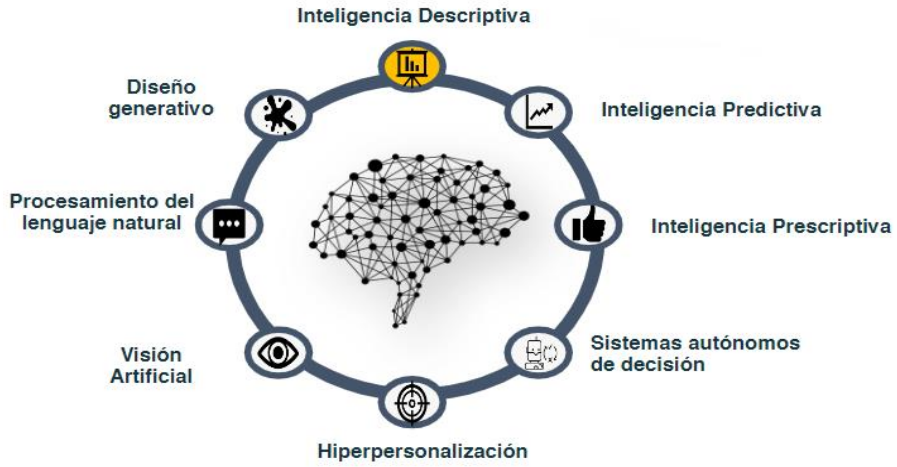
2> TIPOS DE ALGORÍTMOS DE INTELIGENCIA ARTIFICIAL (IA)

DEFINICIONES

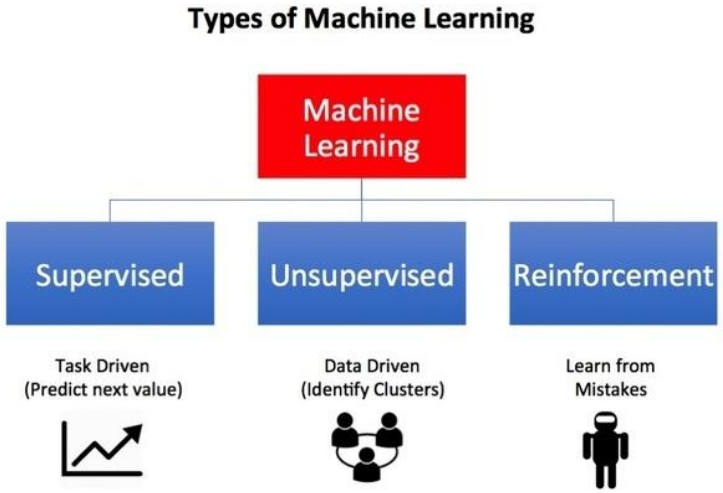
- "Sistemas que muestran un **comportamiento inteligente analizando su entorno y actuando**, con cierto grado de autonomía, **para alcanzar objetivos específicos**. Pueden basarse en interacciones con el mundo virtual a través de software (asistentes de voz, análisis de imágenes, motores de búsqueda, reconocimiento de voz, etc.) o estar integrados en hardware (robots avanzados, coches autónomos, drones, aplicaciones en IoT –internet de las cosas-)"

- "Software y hardware **diseñados por humanos** que actúan en dimensiones físicas o digitales percibiendo su entorno mediante el uso de datos, interpretándolos y **decidiendo la mejor acción a tomar para lograr un objetivo complejo**. Para ello, utilizan tecnologías que combinan datos, algoritmos y potencia de cálculo"

NIVELES DE INTELIGENCIA



TIPO DE APRENDIZAJE



2> ARQUITECTURA Y CASOS DE USO

VENTAS Y MARKETING

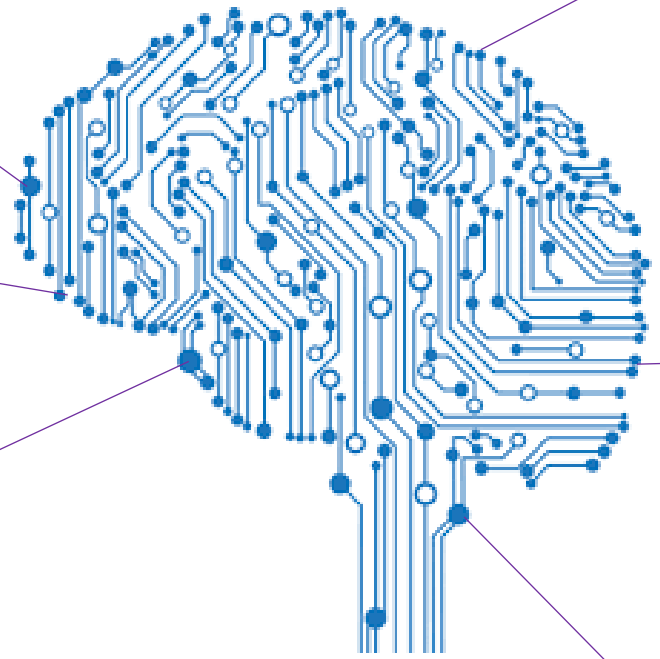
Uso de datos en tiempo real extraídos de múltiples canales y punto de contacto, identificando las preferencia de los clientes para ofrecerles contenidos, productos y servicios a medida

MEDICINA

Chatsbots que nos pregunta por síntomas y realizan diagnósticos. Agilizar la investigación y análisis de secuencias de ADN y ARN para acelerar el desarrollo de vacunas, o aplicando técnicas de *deep learning* en la detección de enfermedades a través de imágenes

FRAUDE

Identificación de actividades fraudulentas y controlar el blanqueo de capitales, gracias a la capacidad de analizar millones de transacciones por segundo y detección de anomalías.



CIBERSEGURIDAD

Mejorar la anticipación y neutralización de amenazas, reaccionando con mayor rapidez gracias a la capacidad análisis de grandes volúmenes de información y el aprendizaje continuo sobre casos reales

INDUSTRIA 4.0

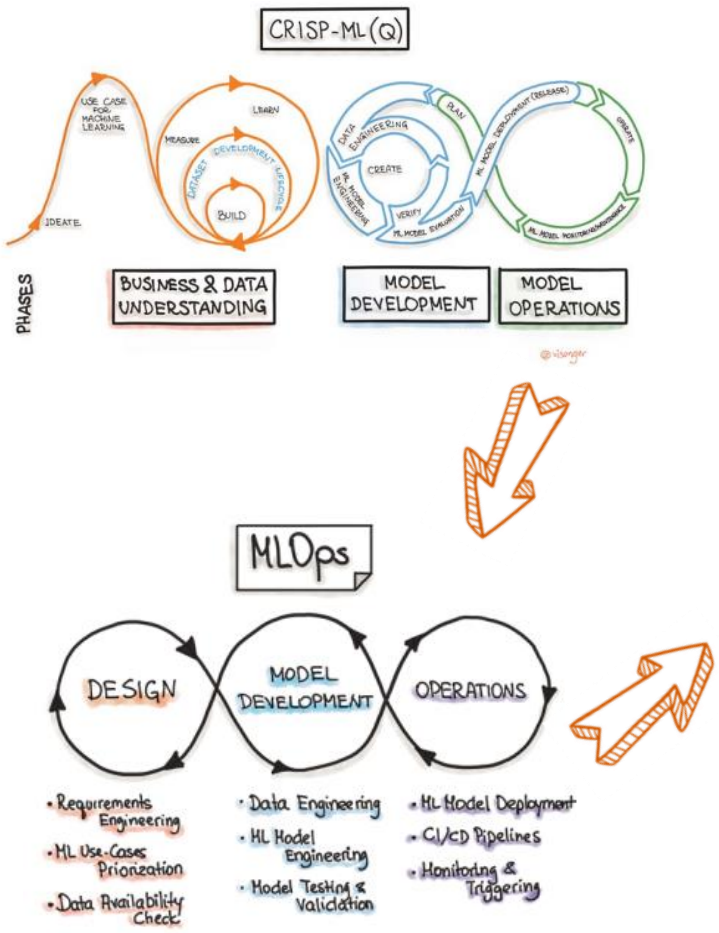
Combinado con los beneficios de IoT permite optimizar los procesos de mantenimiento gracias a los análisis predictivos que reducen los periodos de inactividad y costes

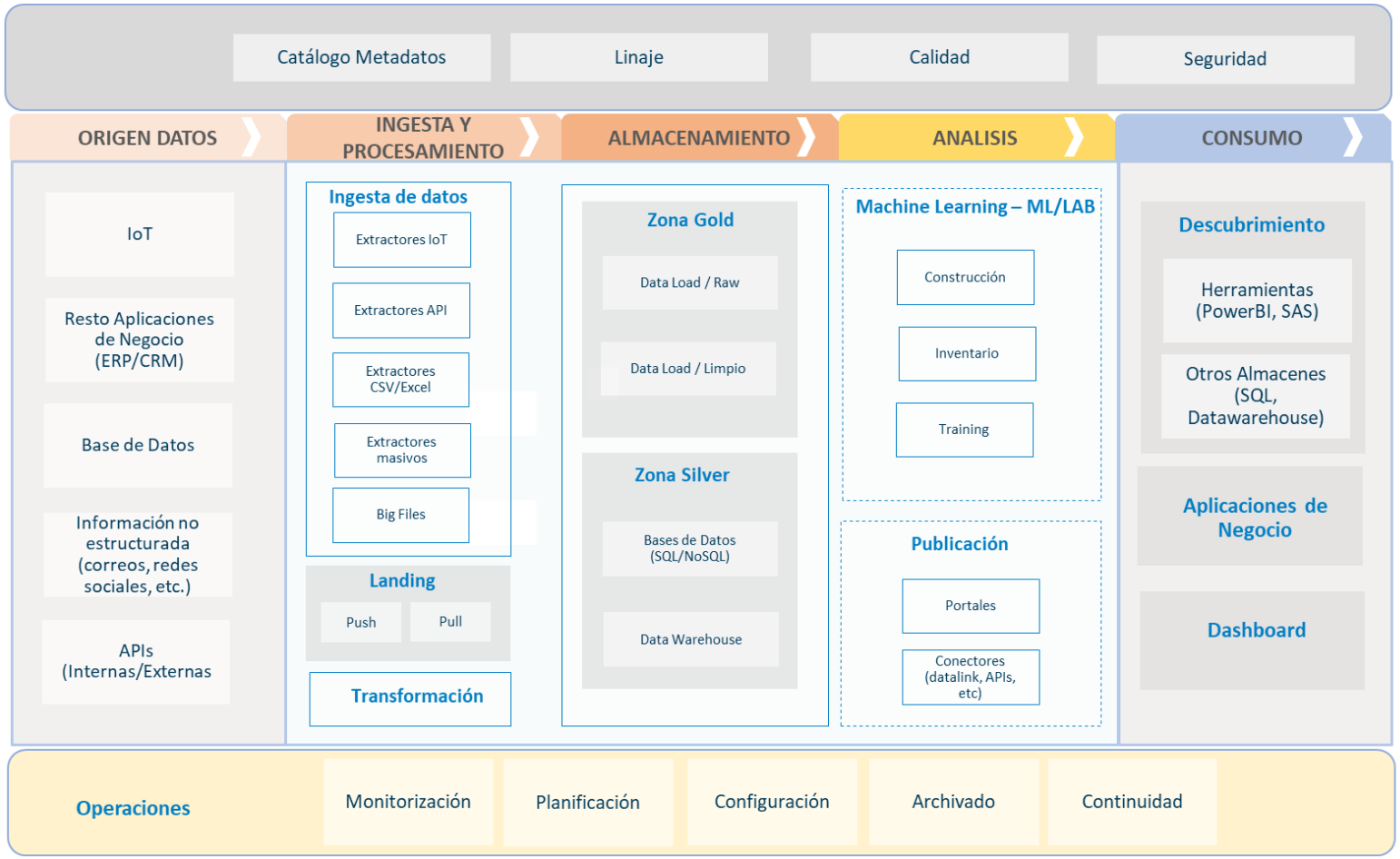
EDUCACIÓN

Evaluar las habilidades y áreas de mejora de los estudiantes, creando materiales de aprendizaje personalizados (creación de libros de texto personalizados). Laboratorios virtuales en el campo de la medicina o apoyo al proceso creativo de artistas y diseñadores con la IA Generativa

2> ARQUITECTURA Y CASOS DE USO

Cross-Industry Standard Process for development of Machine Learning applications with Quality Assurance





CAPACIDADES

- Capacidad de procesamiento de alto volumen
- Calidad y registro de datos
- Solidez, seguridad y precisión
- Monitorización y supervisión humanada

3> Marco Integrado de Control Interno y Riesgos de los procesos de negocio con IA.

COSO The Committee of Sponsoring Organizations of the Treadway Commission

Role de Auditoria Interna

Gobierno y Cultura | Evaluación de Riesgos | Actividades de Control | Información y Comunicación | Supervisión y Evaluación.

- ✓ **Órganos de Gobierno.**
  - Comités Ejecutivos.
  - Efectividad Control Interno.
  - Evaluación Riesgos.
- ✓ Plan estratégico.
- ✓ Cultura empresarial y valores éticos.
- ✓ Capacitación y conocimiento.

- ✓ **Publicación interna** sobre mejores prácticas, políticas internas y valores éticos / morales en la utilización de sistemas de IA.
- ✓ **Publicación externa** Compartir con la opinión pública los principios sobre IA.
- ✓ **Top Management, accionistas y Consejo Administración** son informados de los aspectos relevantes del avance, desempeño real, y de las iniciativas sobre sistemas de IA alcanzados.
- ✓ Existencia de Planes de Emergencia para abordar imprevistos.

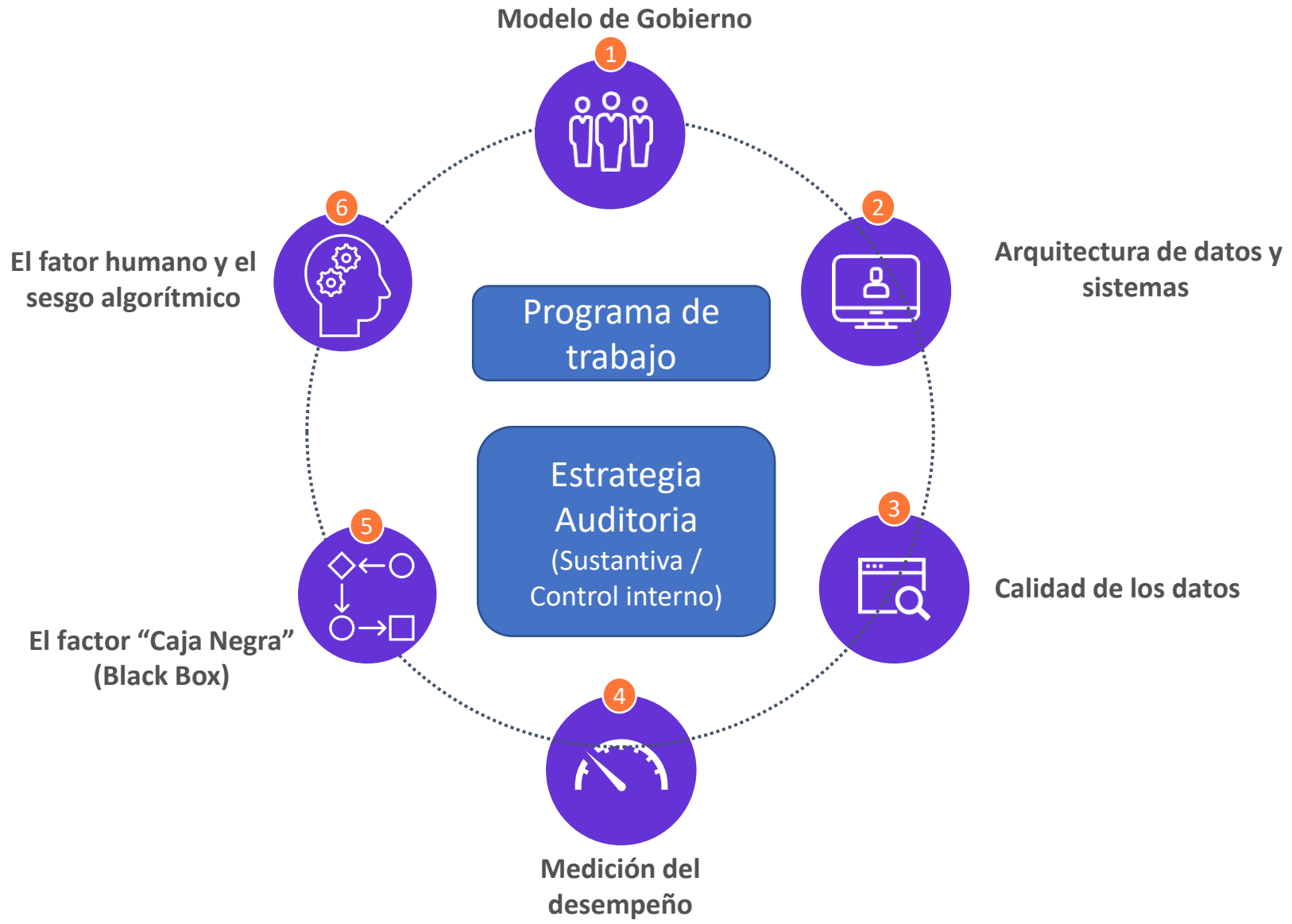
- ✓ **Risk Assessment** continuo end-to-end de proyectos de implementación y desarrollo de sistemas de IA.
- ✓ **Definición de un plan a largo plazo (o estratégico)** de auditoría de los modelos de IA, que acompañe la estrategia de la compañía a este respecto.
- ✓ **Pruebas sustantivas o de auditoría** del control interno sobre la integridad, precisión y confiabilidad de los modelos de IA:
- ✓ **Comunicación,** por parte de Auditoría Interna a la Comisión de Auditoria y otros stakeholders.

3> Marco de Control Interno y Riesgos de los procesos de negocio con Inteligencia Artificial.

Evaluación de Riesgos					
Riesgos de Gobierno	**Riesgos Operacionales y/o de Negocio**	**Riesgos Financieros**	**Riesgos Regulatorios o *Compliance***	**Riesgos Tecnológicos y de Ciberseguridad**	**Riesgo Reputacional**
Estructuras internas, políticas, metodologías y procesos de toma de decisiones en los procesos con sistemas de IA **Top Management continuos risk assessment** permitan identificar y evaluar cambios significativos.	Riesgos intrínsecos durante **ciclo de vida del desarrollo e implementación** de un sistema de IA.	Sistemas de IA que participan en procesos **con potenciales impactos en la información financiera**. .	Riesgos de cumplimiento de **regulaciones externas** (RGPD) o **internas** (Código Ético).	Asociados a los **sistemas y la ciberseguridad** de los modelos desarrollados de IA.	Relacionados con la **presencia de sesgos** en los modelos, o **sanciones impuestas** por incumplimiento normativo.
[Ejemplos: Ausencia *governance* o riskassessment continuo]	[Errores de procesamiento, integridad y completitud de los datos, desviaciones o sesgos en resultados].	[Ausencia de controles o supervisión "humana" de registros contables estimados por sistemas de IA].	[Compliance riskS, ausencia de actividades detectivas, preventivas y reactivas]	[Riesgos derivados a controles generales de TI, cyber risks].	[Influencia política, cultural y/o social de los desarrolladores]

Riesgos intrínsecos de los modelos de IA		
Datos (modelos *data driven)*	**Algoritmos programados (o mantenidos) de forma inadecuada**	**Incapacidad de interpretación o interpretación incorrecta de los outputs de los modelos de IA**
Cualquiera utilización de **bases de datos errónea o calidad inadecuada** puede provocar resultados de los algoritmos inestables y/o incorrectos. **Governance de los datos** utilizados durante las distintas iteraciones de los modelos. La **selección incorrecta de datos** puede tener repercusiones éticas, sociales o de negocio, podría estar desconsiderando información de un grupo socia o datos relevantes necesarios.	Errores en la **programación y desarrollo de código** de los algoritmos de los modelos de IA. La **fase de desarrollo de implementación** de los algoritmos de modelos IA resulta ser la más crítica, especialmente en aquellos sistemas de IA con una mayor sofisticación, por lo que cualquier error de código o programación puede llevar a resultados inapropiados. **Retroalimentación, mantenimiento y mejora del código** en base a los resultados de los modelos en sus distintas iteraciones	Incapacidad de interpretar o entender de forma correcta los resultados que se obtienen una vez las bases de datos son escaneados por los algoritmos de IA, **tomando a cabo decisiones (o no) con resultados no deseados o inesperados**.
[Ejemplos: Integridad y exactitud de los datos utilizados como inputs para los modelos de IA]	[Teste de código incompleto o inexistencia, puesta en producción sin los pruebas de calidad pertinentes, ausencia de mejoras del código]	[Algoritmos de IA basados en redes neuronales puede contener con mayor profundidad esta tipología de riesgos intrínsecos]

4> Programa de trabajo ilustrativo



4> PROGRAMA DE TRABAJO - MODELO DE GOBIERNO (1/6)

Disponer de un modelo de Gobierno formado por las políticas, normas y procedimientos que faciliten la dirección, gestión y monitorización acorde a los riesgos derivados de su utilización

	OBJETIVOS DE CONTROL	PROCEDIMIENTOS DE AUDITORÍA INTERNA
1 GOBIERNO	**Implementación de un modelo de gobierno, incluyendo la definición de políticas y procedimientos internos** suficientes y adecuados destinados al buen gobierno de los sistemas de IA	**Revisión de las políticas internas y procedimientos.** Valorar si abordan los aspectos mínimos y riesgos intrínsecos de los modelos de IA, incluyendo (lista no exhaustiva); identificación roles y responsabilidades, arquitectura de TI y datos, estrategia y objetivos de los modelos de IA, medición de desempeño y métricas de los sistemas de IA.
2 NORMAS Y REGULACIÓN	Análisis del impacto de las **normativas y regulaciones externas** aplicables (p.ej. Normativas de regulación europea y/o local, incluyendo RGPD u otra regulación) e implementación de un adecuado sistema de cumplimiento regulatorio.	• **Revisión de *checklist* de verificación** de la normativa de aplicación al Sistema de IA (Reglamento de Protección de Datos, Normativa medioambiental, etc.). • **Evaluar si los sistemas de cumplimiento regulatorio** son suficientes y adecuados para atender los requerimientos de las normativas externas aplicables.
3 INVENTARIO	**Definición de las características de los algoritmos** sujetos a un análisis de riesgos y modelo de gobierno, garantizando un inventario actualizado de los mismos	**Revisión del inventario** actualizado de modelos y de su documentación establecida en los procedimientos definidos.
4 RIESGOS	**Implementación de un modelo de evaluación y estrategias de mitigación de riesgos**, considerando un *risk-assessment* continuo y revisable en el tiempo.	• Revisión del proceso para la identificación y construcción de un **inventario y/o mapa de riesgos**, incluyendo la idoneidad de las estrategias planteadas para su mitigación. • Valorar una adecuada **segregación de funciones** para asegurar las métricas de **impacto de los riesgos** (métricas y/o factores cuantitativos y cualitativos) identificados de forma continua en el tiempo.
5 MODELOS DE CONTROL	Establecimiento de mecanismos de control para identificar aquellos modelos de IA que, de forma directa o indirecta, **pudieran tener impacto en la información financiera y no financiera** cuentan con los pertinentes controles de revisión, previos a su contabilización	• **Revisión del inventario de sistemas de IA** para garantizar el control de aquellos con impactos en la información financiera vs no financiera. • Análisis de los **datos utilizados por el control owner para el registro contable,** así como verificación de la documentación soporte y evidencias de revisión previa a la contabilización.

4> PROGRAMA DE TRABAJO - ARQUITECTURA DE DATOS Y SISTEMAS TI (2/6)

Garantizar que la arquitectura tecnológica que gestiona los modelos analíticos y sus datos, dispone de las medidas y controles adecuados que aseguren la confidencialidad, integridad y disponibilidad de la información

	OBJETIVOS DE CONTROL	PROCEDIMIENTOS DE AUDITORÍA INTERNA
1 GESTIÓN DE ACCESO	Políticas de **autenticación de acceso** (longitud mínima, caducidad predefinida contraseñas, entre otros aspectos de autenticación) y procesos de gestión de cuentas de acceso y permisos asociados	• Existencia y verificación de la idoneidad de mecanismos de **control de autentificación de usuarios**. • Revisión de los **procedimientos de gestión de ABM** de usuarios y el inventario de autorizantes. • Evidenciar una adecuada **segregación de funciones**. • **Revisión periódica de permisos** de acceso, tanto para usuarios normales como privilegiados.
2 GESTIÓN DEL CAMBIO	Proceso de Gestión de **Cambios en Infraestructuras y Datos** (actualización de software, migración de entornos, etc.) y la existencia de diferentes **entornos** para ello (desarrollo, test y producción)	• Existencia de proceso de **aprobación de cambios en infraestructura y desarrollos**. • Evidencias que la gestión del cambio incluye **pruebas de aceptación** previo a la puesta en producción. • Existencia de mecanismos de **monitorización** periódico de los parámetros clave de la arquitectura. • Comprobar la existencia de un proceso de identificación, evaluación, priorización e **implantación de parches y nuevas versiones de software** en los sistemas de IA.
3 CONTINUIDAD	Definición de las p**olíticas de respaldo** para los sistemas implicados en el modelo, y la existencia de un **plan de continuidad** ante incidencias (Plan de Recuperación de Desastres).	• Evidencias la **existencia de copias de respaldo** de datos según las n**ecesidades de los modelos y sus requisitos de retención**. • **Revisión de los Planes de Continuidad y Recuperación**, y la evidencia de eventos ocurridos o simulados.
4 CIBERSEGURIDAD	Asegurar que los Sistemas de IA se encuentran d**ebidamente protegidos ante ciber-incidentes** y se encuentran dentro de as políticas de ciberseguridad de la organización.	• Evaluar si los sistemas IA se encuentran **integrados dentro de la estrategia de ciberseguridad** de la compañía, **adecuadamente bastionados y sujetos a evaluaciones periódicas de seguridad.**
5 PRIVACIDAD	Los **datos** que utilizados se **encuentran protegidos con los estándares necesarios** para atender los requerimientos de la normativa aplicable sobre protección de datos (Reglamento General de Protección de Datos).	• Revisión de la a**decuada implantación de las políticas de protección** sobre el universo de datos utilizados por los sistemas de IA, especialmente aquellos datos susceptibles y/o sensibles de acuerdo con las políticas internas y regulación (p.e. GDPR, PCI-DSS, etc.).

4> PROGRAMA DE TRABAJO - CALIDAD DE LOS DATOS (3/6)

Garantizar la integridad, exactitud y confiabilidad de los datos que alimentan los algoritmos de IA, asegurando que la organización está preparada para la gestión de grandes volúmenes de información

	OBJETIVOS DE CONTROL	PROCEDIMIENTOS DE AUDITORÍA INTERNA
1 INGESTA DE INFORMACIÓN	Definición y documentación de un **proceso de extracción de los datos de fuentes origen, asegurando su integridad y exactitud**.	• Revisión de los **procedimientos del proceso de extracción** de información. • **Verificar** que la organización ha incorporado **protocolos** de extracción de información, garantizando la **integridad y exactitud** de los datos añadidos al modelo analítico. • Monitorización del ***log*** **de errores** y validación de que son revisados y resueltos antes de la ejecución del modelo de IA.
2 CALIDAD DEL DATO	Existencia de un **proceso de testeo** de la calidad de los de la información cargada para consumo de los modelos de IA, así como de las **transformaciones** realizadas	• Comprobar la existencia de **valores máximos** y **mínimos** sobre las variables, y que existen controles que detectan la presencia de valores anómalos. • Revisión de controles sobre las variables con valores nulos, o sobre variables de control ***checksum*** o similar
3 MONITORIZACIÓN	Las **fuentes y repositorios** de datos están **supervisados** y monitorizados de forma continua	• Evidenciar la **supervisión y documentación de las fuentes y repositorios de datos** que alimentas los modelos de IA. • Evidenciar que existen **aprobaciones** sobre los **cambios de las fuentes y repositorios de datos**, incluyendo una **evaluación de calidad** de los datos derivada de posible cambios.
4 CONSISTENCIA DEL DATOS	Los modelos de IA cuentan con actividades de control para la **medición de la integridad, exactitud y confiabilidad de los datos**, siendo estos monitorizados con **métricas o informes de excepción** para su **análisis y resolución** por los usuarios propietarios de los sistemas IA	• **Revisión de informes de excepciones y métricas** sobre la calidad del dato. • Evidencias las **actividades llevadas a cabo por los propietarios** de los sistemas IA sobre la **resolución de excepciones** y de las métricas de calidad de datos.

4> PROGRAMA DE TRABAJO – MEDICIÓN DEL DESEMPEÑO (4/6)

Actividades que definen métricas sobre el desempeño de los algoritmos con el objetivo de asegurar que el comportamiento de los modelos de IA atienden a las necesidades del negocio y las medidas de supervisión humana que minimizan los riegos intrínsecos de la tecnología

	OBJETIVOS DE CONTROL	PROCEDIMIENTOS DE AUDITORÍA INTERNA
1 MEDICIÓN	Procedimientos destinados a la **medición continua del desempeño** en términos de a**ctividad, parametrización, reporting y métricas** que ayuden a monitorizar y asegurar que el modelo cumple con los objetivos de negocio	• Revisión de los **procedimientos y protocolos que miden el desempeño** garantizando que existe una **supervisión humana** adecuada. • Evaluar la **idoneidad y suficiencia de las métricas** definidas e implantadas. • Revisión de las actividades de **resolución de desviaciones o excepciones** (incl. Desarrollos correctivos). • Evaluar la **frecuencia de reevaluación, reajuste o reinicio de componentes** en los datos de entrada o cambios de criterios de toma de decisiones.
2 INTERPRETACIÓN RESULTADOS	Procedimiento documentado de **interpretación de los resultados de los algoritmos**, con márgenes de tolerancia, umbrales y cualquier otro tipo de control. Así, como las acciones a llevar a cabo en los casos en los que los resultados no sean los esperados	• Revisión del **procedimiento de interpretación de resultados**, así como la idoneidad de los criterios de interpretación. • Asegurar que el **propietario de los modelos cuenta con la experiencia y conocimiento suficiente.** • Evidencias las **acciones** ejecutadas para los casos de **desviaciones significativas**. • Existen **controles detectivos para datos de entrada erróneos** o fuera de rango. • Se ha evaluado el **comportamiento** de modelo ante casos de uso o **entorno imprevistos**
3 BACKTESTING	Proceso que permite medir la **precisión del modelo y su rendimiento**, de forma que puedan **replicarse resultados del pasado con datos históricos** de un periodo concreto.	• Revisar los **procedimientos de backtesting de la compañía** y seleccionar aleatoriamente una ejecución anterior para el ***reperformance*** independiente con el objetivo de comprobar los datos de salida del modelo
4 STRESS-TESTING	Pruebas destinadas a medir los **resultados esperados** de los datos de salida del algoritmo	• Revisar los **procedimientos de stress-testing realizados** por los propietarios de los modelos durante la implantación y mantenimiento de los resultados de salida de los algoritmos. • **Reperformance** por parte del equipo de auditoría de una ejecución aplicando stress-testing en los datos de entrada del modelo

4> PROGRAMA DE TRABAJO – EL FACTOR " NEGRA" – "BLACK BOX" TESTING (5/6)

En términos de la ciencia de datos, el factor caja negra (o black box testing) se refiere a aquellos algoritmos de IA que por su complejidad y/o sofisticación, los mecanismos internos de ejecución entre los datos de entrada y salida son difícilmente entendible o explicables.

	OBJETIVOS DE CONTROL	PROCEDIMIENTOS DE AUDITORÍA INTERNA
1 ANÁLISIS DE SENSIBILIDAD	**Procedimientos black box testing** implementados: • Métricas de precisión, exactitud y rendimiento. • Valores predefinidos de falsos positivos o outliers. • Mecanismos de adaptabilidad de los modelos a nuevos data sets. • Performance análisis de sensibilidad y expectativas.	• Revisión de los **resultados y acciones correctivas adoptadas** de los procedimientos black box testing realizados. • Evidenciar las investigaciones de los **resultados realizados sobre los falsos positivos y outliers.** • Revisión de actividades implementadas para la adaptación de los modelos a **data sets cambiantes o nuevos**. • Validar los mecanismos de supervisión de los **modelos de aprendizaje continuo.**
2 MONITORIZACIÓN CONTINUA	Controles implementados de **monitorización continua** destinados a: a) Supervisión de los datos de entrada y resultados de modelos de supervisados y/o b) Alarmas / KPIs de performace implementados para identificar inestabilidades de los modelos de IA.	• Evidenciar la ejecución de las pruebas de monitorización implementadas, con especial foco a las acciones correctivas (o no) realizadas y documentadas. • Revisión y entendimiento de los procesos de supervisión y monitorización continua implementados, concluyendo sobre la suficiencia (o no) de los mismos.
3 PROCESOS INEFICACES	Implementación de mecanismos de monitorización continua para la identificación de **procesos ineficaces de los sistemas de** IA (es decir, ocurre un incidente importante o la solución ha evolucionado/aprendido de manera inapropiada).	• Revisión y valoración de la idoneidad de los procesos para la identificación de sistemas de IA ineficaces. • Establecimiento de expectativas y análisis de variaciones del aprendizaje en el tiempo.
4 SCANNING DATOS	Ante ineficiencias de los sistemas de IA, **existen mecanismos de reversión para la corrección de algoritmos y acceso disponible a datos "limpios",** con la finalidad de alcanzar la eficacia de los modelos de IA de forma oportuna en el tiempo.	• **Evidenciar** como las actividades de reversión implementadas históricas, consiguieron abordar ejecuciones de los sistemas de IA ineficaces.

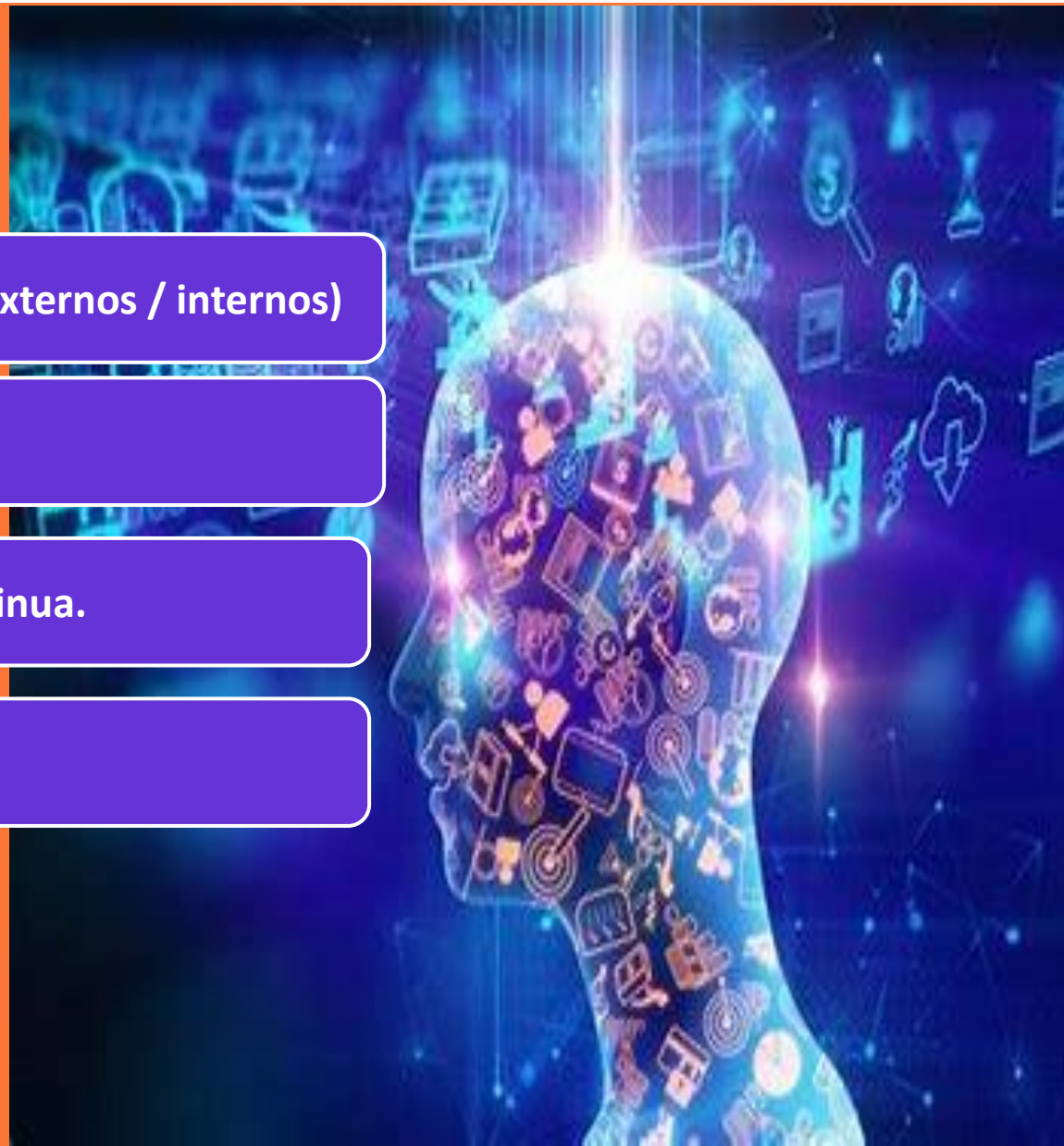
4> PROGRAMA DE TRABAJO – EL FACTOR HUMANO Y EL SESGO ALGORÍTMICO (6/6)

El factor humano en el diseño, implementación y mantenimiento de sistemas de IA es uno de los aspectos a considerar más relevantes, especialmente ante modelos de IA de autoaprendizaje no supervisado, con potencial impacto adverso o no deseado en la sociedad y en los procesos de negocio de las organizaciones.

	OBJETIVOS DE CONTROL	PROCEDIMIENTOS DE AUDITORÍA INTERNA
1 REGULACIÓN INTERNA / EXTERNA	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.
2 SESGOS ALGORÍTMICOS	Asegurar que los resultados de los modelos de IA están libres de **sesgos algoritmos,** intencionados o no.	• Revisión de los objetivos de los sistemas de IA para descartar cualquier tipo de sesgo (intencionado o no) en la fase de diseño de los sistemas de IA. • Revisión de los resultados perseguidos por los sistemas de IA, y compararlos con los objetivos para identificar cualquier desviación y determinar si la causa fue un sesgo algorítmico. • Revisión de la existencia de procedimientos y/o protocolos para identificar sesgos algoritmos motivados por datos históricos con sesgo.

Recomendaciones & Takeaways

- Atentos a la evolución del mercado y requerimientos normativos (externos / internos)
- *Early conversations* con los *process owners.*
- Inversión en formación y actualización continua.
- Cultura *data driven.*



¡GRACIAS!

@Auditorinterno

Instituto de Auditores Internos de España

Síguenos en www.auditoresinternos.es