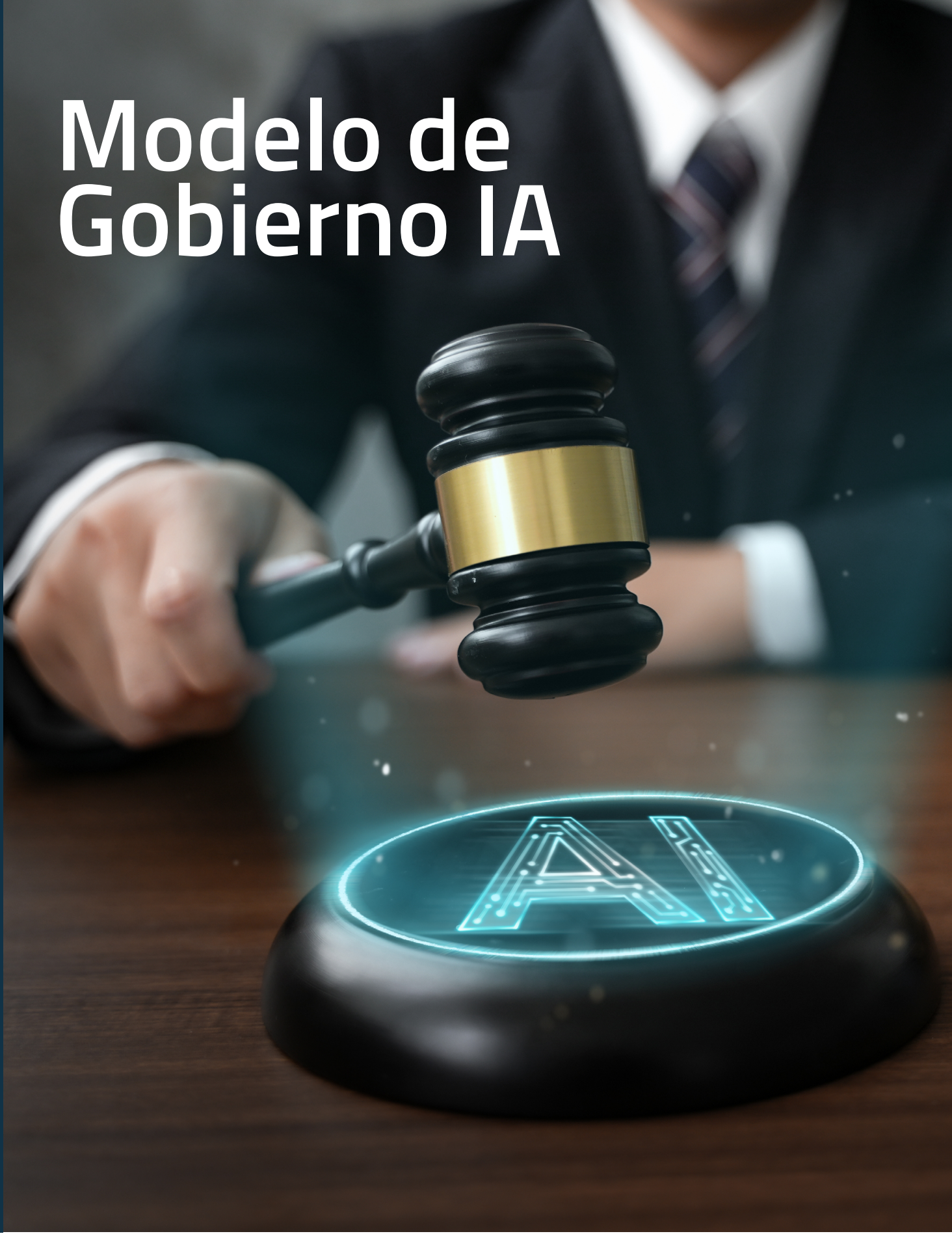
Modelo de Gobierno IA



Una iniciativa de
www.ismsforum.es

isms forum

MODELO DE GOBIERNO IA

AUTORES

COORDINADORES	Ángel Pérez
	Francisco Lázaro
SUBCOORDINADORES	Rubén Cabezas
PARTICIPANTES	Ana de la Higuera
	Berta Balanzategui
	Diego Fernández
	Elena Mora
	Esther García
	Ignacio Hornes
	María Cristina Köhler
	María Cumbreras
	María Pilar Alapont
	Miguel Angel Cabezas
	Pablo Díaz
	Silvia Tortosa
	Xavier Alberdi
GESTOR DE PROYECTOS	Beatriz García
DISEÑO/MAQUETACIÓN	Rim Souri

CONTENIDOS

1 MODELO DE GOBIERNO DE LA INTELIGENCIA ARTIFICIAL

1.1. PRINCIPIOS DIRECTORES

Como hemos visto, la IA se constituirá, previsiblemente, como la piedra angular de todo el proceso de digitalización de las organizaciones, sobre todo en sectores como el financiero, salud y movilidad, entre otros. De acuerdo con los beneficios y riesgos de esta tecnología, resulta primordial establecer en las organizaciones que quieran hacer desarrollos o uso de la IA la configuración de modelos de gobierno que controlen y gestionen el uso de la IA, facilitando el reporte y monitorización del cumplimiento de la diferente normativa que se podría ver afectada y que les permitan aprovechar todas sus fortalezas y mitigar todos los riesgos derivados de las actividades que utilizan IA. Para ello, resultará posible basarse en las diferentes opiniones y guías que hemos recogido en los dos apartados anteriores que describen los principios para una apropiada gobernanza en el desarrollo y uso de sistemas de IA, para así incrementar la seguridad y la confianza en dichas tecnologías.

En este sentido, la aprobación del texto definitivo de la propuesta de Reglamento de IA no debe ser un obstáculo para que las organizaciones puedan avanzar en dichos modelos de gobierno y sus estructuras correspondientes. Entre otros motivos, porque como venimos mencionando en este documento y así se reproduce también en el texto de la nueva regulación europea, muchas otras leyes que se encuentran en vigor inciden también en esta tecnología y, en particular, la normativa de protección de datos personales, resultando de especial relevancia los pronunciamientos de las autoridades de control y supervisión (a través tanto de guías o esquemas o de resoluciones prohibiendo o exigiendo medidas a estos sistemas en virtud del RGPD).

Por ello, las organizaciones deben dotarse de modelos y estructuras de gobierno, desde el diseño y por defecto, para adaptarse a esta nueva realidad que les permita acreditar el cumplimiento de su obligación de "accountability" y responsabilizarse de manera proactiva de los sistemas que utilizan o desarrollan, que debería contemplar, al menos, los siguientes aspectos:

I. Los principios en base a los cuales se desarrollará o usará la inteligencia artificial.

II. El cumplimiento normativo (aspecto en el que se deberá tener en cuenta la normativa de aplicación al sector o producto/servicio, lo que puede resultar en una tarea compleja y que será necesario eficientar y simplificar).

III. La gestión de los riesgos derivados de la IA (desarrollo o uso) y su seguimiento continuo (aspecto en el que habrá que tener en cuenta salud, la seguridad, los derechos fundamentales, la democracia y el Estado de derecho y del medio ambiente si tenemos en la retina la propuesta normativa de Reglamento de IA a lo largo de todo el ciclo de vida del sistema de IA).

IV. Atención a las partes afectadas y a las autoridades competentes.

Junto con ello, las organizaciones deberán establecer las estructuras (comités, órganos, etc.), las normas, procedimientos o protocolos internos, así como dotarse de las herramientas técnicas necesarias y adecuadas para el cumplimiento normativo y programas de formación y concienciación debidamente adaptados a los perfiles involucrados. WW.

Centrándonos, en este primer apartado relativo a los principios rectores de esas Políticas corporativas de las que deberán dotarse las organizaciones y que establecerán el marco general, pendiente todavía de aprobación del Reglamento de IA, existen diferentes enfoques que pueden servir de utilidad a las organizaciones y que, como veremos a continuación, conducen a conclusiones similares.

De este modo, en el cuadro que se muestra a continuación se incluyen los fundamentos establecidos en el documento publicado por los expertos de la Comisión, Directrices para una IA fiable, los principios generales del artículo 5 del RGPD y, finalmente, se muestran los principios que se contemplan en la propuesta de Reglamento de IA del Parlamento.

DIRECTRICES PARA UNA IA FIABLE (FEBRERO 2019)*	PRINCIPIOS GENERALES DE PROTECCIÓN DE DATOS (ART. 5)	PROPUESTA DEL PARLAMENTO DEL FUTURO REGLAMENTO DE IA
Autonomía humana.	Licitud, lealtad y transparencia.	Intervención y vigilancia humana.
Prevención del daño.	Limitación de la finalidad.	Solidez y seguridad técnicas.
Equidad.	Minimización de datos.	Privacidad y gobernanza de datos.
Explicabilidad.	Exactitud.	Transparencia: trazabilidad y explicabilidad, información a los usuarios e información a personas afectadas.
-	Limitación del plazo de conservación.	Diversidad, no discriminación y equidad.
-	Integridad y confidencialidad.	Bienestar social y medioambiental.
-	Responsabilidad proactiva.	-

**Es necesario recordar en este punto que la propuesta de Directrices concreta los 4 fundamentos expuestos en los siguientes 7 requisitos: i) acción y supervisión humana, ii) solidez técnica y seguridad, iii) gestión de la privacidad de los datos, iv) transparencia, v) diversidad, no discriminación y equidad, vi) bienestar social y ambiental y vii) rendición de cuentas.*

Con estos antecedentes, organizaciones y empresas que pretendan desarrollar o crear sistemas de inteligencia artificial disponen de orientaciones que pueden seguir. A este respecto, los principios generales de la protección de datos resultan en estos momentos de obligado cumplimiento y a luz de las nuevos retos y exigencias que suponen los sistemas de IA constituyen herramientas muy útiles que pueden servir de ayuda a las organizaciones, debido al conocimiento que se tiene de las mismas y que resultan de aplicación desde hace años.

Los principios arriba apuntados se pueden resumir en 2 grandes grupos:

- Establecimiento de un proceso de toma de decisiones explicable, transparente y justo.
- La tecnología debe centrarse en el bienestar y seguridad de las personas.

Estos dos grupos abarcarían todos los principios expuestos en los textos que se han indicado con anterioridad y que pueden concretarse a través de los mismos (se trarta de una cuestión práctica que es, habitualmente, el gran reto al que se enfrentan las organizaciones).

A modo de ejemplo, el proceso de toma de decisiones explicable, transparente y justo que se constituye como un aspecto crítico de la IA, podría alcanzarse a través de la elaboración de políticas informativas para los afectados que informen, de manera veraz, sobre la eficiencia, las capacidades y las limitaciones reales de un determinado sistema de IA (contexto y situación, la posibles existencia de terceras partes, la localización física/virtual de la solución AI, etc.), pudiendo utilizar para ello las fórmulas que se vienen utilizando en la protección de datos personales, de manera que, en cada una de las etapas del ciclo de vida de la IA, se incluyan los requisitos concretos. En este sentido, durante la etapa de entrenamiento se informará claramente al interesado sobre la posibilidad de que resulte identificable a partir de los datos del modelo utilizado (o, en su caso, informarle de que no es posible realizar dicha reidentificación).

De esta manera, por ejemplo, en los procesos de preparación de los data set para los entrenamientos se incluirán los principios de minimización, exactitud o responsabilidad proactiva, de manera que se eliminarían los datos no necesarios, y se preservaría la representatividad de las muestras o su falta de sesgo. Finalmente, a través del principio de responsabilidad proactiva se deberían incluir previsiones en relación con la conservación de los ficheros de registro de los eventos que se ejecutan en el sistema de IA (o "logs") para sustentar los procesos de auditoría y los mecanismos de seguridad, el cumplimiento de obligaciones legales, como el artículo 25 de la de la Ley 10/2010 de prevención de blanqueo de capitales y de la financiación del terrorismo, o la trazabilidad sobre la procedencia de los datos de entrenamiento y validación, así como registros de los análisis que se han realizado sobre la validez de dichos datos y sus resultados.

Finalmente, es necesario indicar que todo ello redundaría, a su vez, en un cumplimiento del segundo principio, el enfoque centrado en el bienestar y seguridad de las personas afectadas por los sistemas de IA, debido a que permitiría a las organizaciones un mejor cumplimiento del principio de transparencia.

1.2. ELEMENTOS PARA LA GOBERNANZA DE UN SISTEMA DE IA

En línea con los principios mencionados en el anterior apartado, las organizaciones deberán crear las estructuras, así como los procedimientos necesarios para dar soporte al cumplimiento de esta nueva normativa. En este sentido, la aceptación y la participación de la alta dirección resultarán esenciales.

En aquellos supuestos en los que la IA utilice información personal, puede resultar de ayuda a las organizaciones aplicar los conocimientos de la normativa de protección de datos, así como las estructuras y los procedimientos internos adoptados con la entrada en vigor del RGPD.

Con carácter previo, las organizaciones deberán realizar un análisis a fin de identificar los objetivos de negocio en el desarrollo y uso de esta tecnología, de manera que una vez que se encuentren definidos se determinará su integración en los diferentes sistemas de AI de la organización que permitirán la consecución de tales objetivos estratégicos.

En ese momento, resultará también necesario que se defina una política interna que regule el uso de sistemas de IA en la organización.

A modo de ejemplo, se podrán determinar una serie de usos aceptables de la IA de conformidad con la cultura y valores de cada Organización, constituyendo un uso inaceptable aquellos que sean ilícitos y no garanticen suficientemente el respeto de la legislación aplicable, o bien aquellos usos que no aseguren el cumplimiento de los principios y valores éticos de la entidad u otros usos de la IA que puedan llegar a provocar daños a terceros, aunque sea de manera accidental.

Una vez se identifiquen los sistemas de IA que serán utilizados para la consecución de los objetivos estratégicos determinados por cada organización, todo ello dentro del marco de cumplimiento normativo definido por la política y los procedimientos internos que la desarrollan, se deberá diseñar un calendario de comunicaciones y un plan de formación y concienciación a finde de garantizar que es conocido por todos los miembros de la organización.

Para lo anterior resultará imprescindible contar con una gerencia comprometida con el respeto de los principios que se incluyen en la política, que es parte activa de las acciones de comunicación y que completa su propia formación puntualmente.

De manera complementaria, resultará necesario incluir los posibles usos de la AI por parte de terceros que prestaran sus servicios a la organización. Los proveedores de servicios y otras terceras partes deberán acreditar el cumplimiento de la normativa aplicable, tanto a nivel legal como de las políticas internas de la organización, con carácter previo a la prestación de los servicios y, de manera especial, antes de incluir cualquier tipo de información pertenecientes a la organización en sus sistemas de AI.

A la vista de todo lo anterior, el gobierno interno de la IA en una organización debe disponer de recursos, experiencia y autoridad para dirigir la implementación de la estrategia de IA mientras se supervisa su desarrollo y uso, que puede incluir los siguientes elementos:

1.2.1. ADAPTACIÓN DE LAS ESTRUCTURAS DE GOBIERNO INTERNAS

El primer gran hito de este marco de gobierno se sitúa en ese texto unificador que concentre, en base a los diferentes roles de la organización respecto de los sistemas de IA (proveedor, usuario/implementador, importador, distribuidor, etc.), los principios rectores que se han identificado en el apartado anterior y el resto de los aspectos exigidos por la propuesta normativa de IA. Estos aspectos deberán ser desarrollados y concretados a través de las correspondientes normas y protocolos internos.

La estructura de gobierno interno debe generalmente comprender tanto una estrategia de IA a nivel organizativo, junto con un Comité de gobierno (o un organismo similar).

La estrategia de IA servirá para demostrar el compromiso de la alta dirección con el desarrollo y uso de la IA en la organización e incorpora instrucciones claras sobre las finalidades para los que se puede utilizar la IA y cómo se debe utilizar.

En este sentido, seguidamente se apunta, con carácter general, los procedimientos/normas o metodologías internas de las que pueden dotarse las organizaciones:

A. Evaluación del cumplimiento del Reglamento de IA.

B. Establecimiento de un sistema de gestión del riesgo que incluya el seguimiento posterior a la puesta en producción de la solución de IA.

C. Realización de las evaluaciones de impacto en derechos fundamentales.

D. Elaboración de un procedimiento de evaluación que incluya la declaración de conformidad CE, consistente en que la tecnología de IA utilizada es conforme con los requisitos de las Directivas de la UE relativas a la comercialización de productos que le son de aplicación.

E. Elaboración del registro de sistemas de IA, asociado o asociable al Registro de actividades de tratamiento exigido por el art. 30 del RGPD y, en el supuesto de que quede incluida en alguno de los ocho ámbitos específicos previstos en la propuesta de Reglamento IA, registrarse en la base de datos de la UE.

F. Notificación de los incidentes y fallos de funcionamiento.

G. Modelos de documentación interna que deberá disponer cada sistema de IA.

H. Normativa interna dirigida a los empleados y proveedores en función de su rol en la creación o uso de la IA.

Adicionalmente, las organizaciones se deberán dotar de las estructuras correspondientes que deberán dar soporte a todo el sistema.

1.2.2. ESTABLECIMIENTO DE UN COMITÉ/FORO DE IA

En determinadas organizaciones, la obligación de disponer de un sistema de gestión de riesgo o la realización de evaluaciones de impacto en derechos fundamentales por los sistemas de IA puede llevar, fácilmente, a la necesidad de disponer de un Comité/Foro que pueda evaluar y hacer seguimiento de estos aspectos, que incluso podría tener entre sus competencias liderar la implementación de la propuesta normativa.

Parece también necesario que este Comité/Foro debería tener una composición multidisciplinar a los efectos de poder abordar de la mano de los diferentes especialistas los riesgos que implican los sistemas de IA (de cumplimiento, de afectación a la salud, derechos fundamentales, de seguridad o asociados a la calidad del dato, diseño de la experiencia de los usuarios, entre otros).

En este sentido, resulta particularmente familiar en la gestión de los riesgos de la privacidad y en las evaluaciones de impacto en la protección de datos, en particular, la formación de estos equipos multidisciplinares que abordan los riesgos jurídicos y los riesgos de seguridad que se derivan de los tratamientos de información de personal. De hecho, recientemente, a la luz de las ultimas guías publicadas por la AEPD, el análisis de los algoritmos y de los sistemas de IA se ha introducido como un aspecto añadido a evaluaciones de impacto, con el objetivo de garantizar tanto el cumplimiento del RGPD como la mitigación de los riesgos concretos que podrían derivarse del hecho de que un tratamiento de datos se realice a través de un sistema de IA.

Finalmente, tanto si la organización ha determinado estos roles como si se encuentra en proceso de hacerlo, parece que la propuesta de Reglamento de IA obligará a ajustar aún más los esquemas para determinar quiénes y cómo se gestionarán las obligaciones concretas que se aplicarán a los sistemas de IA. De este modo, resultará necesario concretar en las entidades quiénes son los encargados de la gestión y seguimiento del riesgo, la calidad y registro de los sistemas de IA, así como de las evaluaciones de impacto en materia de derechos fundamentales y del resto de requisitos que establece la propuesta normativa (obligaciones en registro en la base de la UE, la comunicación de fallos u otros aspectos internos, como por ejemplo los relacionados con los datos y su gobernanza, la redacción y conservación de la documentación técnica, los registros, la vigilancia humana o la ciberseguridad).

Asimismo, resultará necesario abordar las posibles coincidencias o solapamientos con los roles y competencias derivados del cumplimiento de otras normativas (por ejemplo, con protección de datos personales donde ya existen obligaciones muy similares en la realización de evaluaciones de impacto en la privacidad o la comunicación de brechas de seguridad o, en el caso de sectores particularmente regulados como las entidades de crédito, la normativa sectorial aplicable a estos modelos, donde también existen Comités/Foros que supervisan los riesgos de los mismos y disponen de metodologías concretas para su desarrollo y aprobación por el supervisor sectorial).

Algunas de las competencias que se pueden atribuir a este Comité/Foro o bien, asignar estas tareas específicas a algún otro órgano ya existente e integrar el gobierno de la IA como parte más de sus responsabilidades, son las siguientes:

- **SUPERVISIÓN:** Esta una función puede ser definida de la manera más precisa posible con el objetivo de evitar cargas de trabajo excesivas, o que se pierda visibilidad sobre aquellos sistemas de AI que requieran una mayor atención, debido a los riesgos que pudiera implicar para los individuos, la sociedad o el medioambiente y el ecosistema.

  La definición de la responsabilidad de este Comité/Forr de gobierno de la IA deberá incluir la revisión y aprobación de los sistemas de AI cuyo uso implique un riesgo alto para los individuos, la sociedad o el medioambiente, así como para la organización misma.

  Ha de ser capaz, por otro lado, de aceptar o no a los nuevos miembros que sean propuestos por las distintas funciones o negocios de la organización, verificando su aptitud, experiencia, etc. a fin de que la supervisión se produzca de manera efectiva y no se convierta en un mero cumplimiento formal.

- **REPORTE A LA ALTA DIRECCIÓN:** Al igual que ocurre en materia de protección de datos o de ciberseguridad, resultará necesaria la comunicación de aquellos incidentes que se hayan producido durante la utilización de los sistemas de IA que pueda suponer riesgos legales, de cumplimiento normativo u operativos que presenten un riesgo alto o relevante para la organización, así como la posibilidad de causarle un daño reputacional.

  Además, será necesario identificar aquellos indicadores (o métricas) que sean relevantes y que hayan de ser informados regularmente a los gerentes de la organización, quienes podrán verificar el éxito o no de la estrategia definida para AI o para los objetivos de negocio que utilizan IA.

  A modo de resumen, este Comité u órgano de la entidad deberá, en su conjunto, ser responsable de que se evalúen beneficios derivados del uso de sistemas concretos de AI, así como los posibles inconvenientes y riesgos; y que esa evaluación incluya el funcionamiento de los algoritmos utilizados para la toma de decisiones, los datos utilizados y quiénes son los actores que participan en los proyectos.

1.2.3. ROLES Y RESPONSABILIDADES

Se identifican a continuación las distintas figuras que las organizaciones deberían integrar en los equipos multidisciplinares, sin perjuicio de que puedan incorporarse otras áreas que también participen en el desarrollo o utilización de los sistemas de inteligencia artificial:

- **CAIO:** El Chief Artificial Intelligence Officer es una figura que sólo existe en organizaciones más maduras en la adopción de IA y tiene como misión principal definir la estrategia de IA de la organización, velar por que se logre su implementación exitosa, así como identificar los riesgos asociados e impulsar acciones para su gestión.
- **CDO** (Chief Data Officer): Con relación a todos los temas relativos al tratamiento de datos y su gobernanza, así como las herramientas que deben permitir el cumplimiento del Reglamento de IA en general.
- **DPO** (Data Protection Officer): en relación con el asesoramiento y cumplimiento de la normativa de protección de datos.
- **CISO** (Chief Information Security Officer): en relación con los aspectos relativos a la seguridad de la información.
- Analistas de sistemas, arquitectos de sistemas y científicos de datos que participarán en el diseño, desarrollo, seguimiento y mantenimiento del sistema de IA.
- Los profesionales legales y de cumplimiento garantizarán el cumplimiento de la legislación aplicables (propiedad intelectual, riesgo reputacional, consumidores y usuarios, etc.), así como otra normativa interna y políticas en el desarrollo y uso de la IA.

- Los equipos de negocio y operativos que deberán utilizar la IA de acuerdo con las políticas y procedimientos de las organizaciones.
- Atención al cliente y Comunicación para interactuar con todas las partes interesadas, incluidos los clientes reguladores y el público en general, y abordar sus preocupaciones respecto al uso de esta tecnología.
- Recursos Humanos, por las posibles implicaciones de los sistemas de IA en materia laboral.

Las competencias en relación con la determinación de los riesgos, en particular sobre la afectación a derechos fundamentales o la evaluación sobre el bienestar medioambiental o los requerimientos en relación con el consumo de energía de estos modelos, son aspectos relevantes que las entidades deberán configurar y establecer dentro de los equipos de las organizaciones.

Con relación a la determinación de la afectación de los riesgos para los derechos fundamentales, la última guía de la AEPD sobre gestión de riesgos y evaluaciones de impacto puede ser un buen referente para seguir por los equipos involucrados para que, junto con la afectación al derecho fundamental a la protección de datos, puedan, a su vez, concluir la afectación en materia de derechos fundamentales. Será necesario evaluar las necesidades de formación adicional que requerirán los mismos para abordar esta tarea con éxito.

Para el segundo de los aspectos, el bienestar medioambiental y el consumo de energía, los equipos del CDO y de sistemas informáticos entendemos que pueden estar en una buena disposición para ayudar en el cumplimiento de los requerimientos técnicos y de información que parece requerir la propuesta de norma.

1.2.4. ELABORACIÓN DE POLÍTICAS Y PROCEDIMIENTOS INTERNOS

En la elaboración de políticas internas con respecto al uso de sistemas de IA se debe diferenciar aquellas destinadas a proteger la propiedad intelectual y otros bienes de la organización, de aquellas otras en las que se establecen las condiciones, los usos y las garantías necesarias para poder utilizar sistemas de IA para la consecución de los objetivos estratégicos de la organización.

En el primer supuesto, se deben definir los diferentes sistemas de IA, la enumeración de los actores involucrados, así como los usos aceptables y aquellos que se encuentran prohibidos, todo ello centrado en la protección de los intereses legítimos de la organización.

En el segundo caso, se deben desarrollar políticas y procedimientos con un enfoque responsable, sostenible y centrado en el ser humano, desarrollados e implementados con los intereses y el bienestar de las personas y la sociedad en mente de manera que amplifiquen y aumenten, en lugar de desplazar, las capacidades humanas y preserve el control de las personas. En el presente apartado nos centraremos en el segundo supuesto centrándonos en la relación de la IA con la ética y los valores que a ella se asocian.

En este sentido, una política destinada a interactuar con sistemas de IA éticos, así como los procedimientos destinados a su implementación, deberán considerar su ciclo de vida productivo al completo, no olvidando aquellas etapas cuando el sistema de IA deje de estar en uso y resulta necesario retirarlo o sustituirlo por otro.

Por este motivo, se enumeran a continuación una serie principios éticos que podrían ser tenidos en cuenta por las Organizaciones a la hora de redactar su política para el uso ético de la IA:

- Respeto, protección y promoción de los derechos humanos, de las libertades fundamentales y de la dignidad humana.
- Respeto, protección y promoción del medioambiente y del ecosistema.
- Respeto, protección y promoción de la diversidad y la inclusión, favoreciendo la participación de los individuos y sin que las circunstancias individuales les hagan ser objeto de exclusión (raza, edad, género, nacionalidad, etc.).

Esta parte de una futura política para el uso de sistema de IA tiene un contenido más institucional, pero debemos recordar otros principios que condicionarán de una manera más material y práctica usos presentes o futuros de la IA:

- Respeto de la autonomía de los individuos.
- Prevención de daños, presentes o futuros, materiales o inmateriales.
- Criterios de justicia durante todo el proceso en el que participa la IA.
- Transparencia (explicación del qué, cómo y cuándo de un sistema de IA).

En la propia política o en alguno de los diferentes procedimientos resultará conveniente considerar el cumplimiento de los requisitos establecidos por la Comisión Europea:

- Intervención humana.
- Robustez técnica y seguridad.
- Privacidad y gobernanza del dato, ya sea este personal o de otro tipo.
- Diversidad, no discriminación y justicia.
- Bienestar para la sociedad y el medioambiente.
- Responsabilidad proactiva.

Para hacer efectivo todo lo que se regula en las diferentes políticas y los procedimientos se hace también necesario establecer un programa de monitorización o, incluso, de auditoría con el fin de verificar que la implementación ha sido efectiva y confirmar de manera fehaciente que únicamente sistemas de IA que cumplen los estándares, internos y legales, son utilizados en nuestra organización

1.2.5. FORMACIÓN (ASPECTOS LEGALES Y ÉTICOS)

A la hora de realizar las tareas de formación y concienciación de los diferentes actores en el diseño y utilización de los sistemas de IA, es necesario atender a aquellas obligaciones legales (por ejemplo, como realizar una evaluación de conformidad y la documentación necesaria para acreditar su cumplimiento), así como a los principios éticos adoptados por la organización y que deber configurarse como una parte integral de las acciones formativas.

Es aconsejable realizar una valoración de los canales formativos existentes en la organización, así como los recursos disponibles a fin de integrar las nuevas materias de la forma más efectiva posible.

1.2.6. DETERMINACIÓN DE LOS OBJETIVOS DE NEGOCIO

Una vez definida a nivel general la estrategia de la IA, la entidad podrá plantearse diferentes cuestiones con un alcance más amplio relacionadas con el cumplimiento de los objetivos de negocio. En este sentido, se deberá analizar diferentes aspectos como, por ejemplo, la manera de incorporar el talento necesario para el desarrollo de esta tecnología, la definición de la estructura y el entorno operativos requeridos para la consecución de las finalidades acordadas, dedicar recursos a la investigación y el desarrollo de plataformas y algoritmos en sistemas de AI para la organización, entre otros.

También se deberá valorar la posible adquisición de soluciones disponibles en el mercado, lo que implicará la necesidad de establecer un procedimiento de revisión de sistemas de AI proporcionados por terceros, por ejemplo, en el sistema de homologación de proveedores que incorpore todas las garantías a las que obliga la IA.

1.2.7. APOYO DE LA GERENCIA

La inclusión de los sistemas de IA en el plan estratégico de la organización podría constituirse como un primer indicativo del compromiso de la gerencia con el uso éticos de esta tecnología, verificar que la estructura de gobernanza de la organización es la adecuada y poder monitorizar el uso adecuado de la IA.

Para ello, resultará necesario valorar quién dentro de la gerencia asumirá la responsabilidad de la implantación y supervisión de los sistemas de IA. Una buena práctica podría consistir en la designación de un miembro del Comité de Dirección para liderar el equipo multidisciplinar y presidir el Comité o Foro de la IA dentro de la Organización.

Aunque resulta innegable la necesidad del apoyo de la Gerencia de la Organización, no se debe olvidar la asignación de los diferentes responsables del diseño y uso ético de la IA a fin de conseguir su involucración más allá de un mero apoyo formal.

1.3. GESTIÓN DEL RIESGO PARA LOS DERECHOS Y LIBERTADES

La propuesta de Reglamento europeo de la IA establece una serie de reglas para proveedores y usuarios en función del nivel de riesgo. Aunque muchos sistemas de IA plantean un riesgo mínimo, resultará necesario evaluarlos todos. En este sentido, la futura norma propone que los sistemas de IA que puedan utilizarse en distintas aplicaciones se analicen y clasifiquen según el riesgo que supongan para los usuarios, de manera que los distintos niveles de riesgo implicarán una mayor o menor regulación.

De este modo, la legislación de la UE pretende implementar normas basadas en el riesgo para la IA lo que implica que dicha gestión del riesgo se constituirá como una actividad continua para las organizaciones.

Es importante destacar que no se establece de forma específica qué medidas deberán implementar las organizaciones de forma obligatoria o la forma de implementarlos, limitándose a proporcionar un amplio grado de flexibilidad para que distintos sistemas de IA se adecúen a la norma. Sin embargo, la nueva normativa sí establece que dichas medidas se han de seleccionar siguiendo un análisis basado en el riesgo, y específicamente diferentes categorías de riesgo. De conformidad con este enfoque, algunas de las medidas que la propuesta de Reglamento establece se aplicarán sólo cuando los sistemas de IA que debido a que afecten negativamente a la seguridad o a los derechos fundamentales se considerarán de alto riesgo, mientras que otras deberán modularse en función del nivel y tipo de riesgo (limitado).

Por este motivo, los sistemas de IA de riesgo inaceptable se considerarán una amenaza para las personas y serán prohibidos.

A la vista de lo anterior, las organizaciones deberán adaptar la aplicación de las medidas previstas en la propuesta de Reglamento a las características de los sistemas de IA que utilicen. Lo que puede ser adecuado para una organización que utiliza sistemas de riesgo alto o IA generativa, no resultará obligatorio para una entidad que utilice unos sistemas de IA de riesgo limitado, que únicamente deberán cumplir unos requisitos mínimos de transparencia que permitan a los usuarios tomar decisiones con conocimiento de causa.

La gestión de los riesgos de la IA en las organizaciones supondrá, sin duda, un reto para las organizaciones, para ello resultará necesario establecer un modelo de gobierno que definirá políticas y procedimientos adecuados para garantizar una aplicación efectivo del enfoque de riesgos. Además, la adopción de un marco de gobierno de la IA permitirá la colaboración de todos niveles de la organización, nivel estratégico, táctico y operativo, y proporcionará la capacidad para alinear el uso de la tecnología de IA con los objetivos corporativos.

Por todo ello, el modelo de gobernanza de la IA diseñado por la Organización debe contemplar distintas fases y determinar las responsabilidades legales, éticas y normativas que asumirá la Organización.

La evaluación de riesgos debe ser realizada por un equipo multifuncional compuesto por personal de los diferentes departamentos afectados antes del desarrollo y uso de un nuevo sistema de IA, o cuando se produzcan importantes actualizaciones de un sistema de IA existente. Todas las evaluaciones de riesgos deben ser documentadas de manera adecuada y los resultados de la evaluación de riesgos deben ser revisados y respaldados por el Comité o Foro de Gobierno de IA (o un Comité similar).

1.3.1. DEFINICIÓN DEL PROCESO PARA ANÁLISIS DE RIESGOS

Es importante considerar que el proceso para el análisis de los riesgos derivados del uso de sistemas de IA es la forma efectiva en la que se produce la implementación material de aquella política de uso de los sistemas de IA definida por la organización. Es únicamente a través de la operacionalización de esa política cuándo podremos estar seguros de que se respetan los principios necesarios para asegurar el uso de sistemas de IA sólo cuando cumplen con las obligaciones legales de aplicación.

Durante ese proceso se han de identificar los beneficios derivados del uso del sistema de AI en concreto, así como los posibles inconvenientes y riesgos que puedan producir a individuos, a la sociedad en su conjunto o al medioambiente y el ecosistema. En el capítulo de los riesgos, el proceso deberá hacer posible la prevención de los riesgos a través de la implementación de los controles más adecuados o su mitigación, cuando la prevención no sea posible. Es además necesario considerar la monitorización de todo lo anterior a través de los mecanismos adecuados.

En aquellos supuestos en los que la Organización dispone de una metodología de análisis de riesgos al uso y un procedimiento de cumplimiento normativo, es recomendable considerarlos e integrar el uso de los sistemas de AI como una nueva categoría aplicable a los mismos antes de crear nuevos procesos que puedan hacer más compleja la gestión de la materia.

Este mismo criterio resultaría aplicable a la constitución del Comité de Gobierno de la IA y, debido a este motivo, se han definido distintos perfiles con el objetivo de conseguir una composición más ágil y flexible que pueda tener un mejor encaje en distintos tipos de organización.

De manera complementaria, para que sea eficiente, la gestión del riesgo para los derechos y libertades ha de estar integrada en los procesos de gobernanza y las políticas internas.

I **INVENTARIADO DE IAS EN LA ORGANIZACIÓN**: Identificar los sistemas de IA que complementan o sustituyen la toma de ciertas decisiones en la Organización y determinar los actores involucrados y los responsables de su uso.

La Organización deberá disponer de un repositorio centralizado de los sistemas de IA y, en el caso de aquellos que traten información personal, deberán vincularlo con el Registro de Actividades del Tratamiento (RAT).

II **IDENTIFICACIÓN Y EVALUACIÓN DE RIESGOS:** La identificación y evaluación del nivel de riesgo de un sistema de IA es una operación necesaria para determinar el modo en que se van a aplicar las medidas tendentes a suprimirlo o minimizarlo, en particular cuando el riesgo detectado es alto según las condiciones establecidas en la nueva normativa.

Para ello, resultará necesario analizar el potencial impacto que la utilización de IA puede tener sobre los derechos de las personas. Asimismo, garantizar que los procesos soportados por esta IA respetan los derechos y libertades de los afectados, que son justos y no discriminatorios.

La evaluación de los riesgos deberá ponderarse con los requisitos y limitaciones que impone la normativa de protección de datos y otra normativa sectorial aplicable a la Organización.

La complejidad del proceso de evaluación de riesgos se ajustará, no al tamaño de la Organización, la disponibilidad de recursos o el sector de actividad, sino al posible impacto del uso de IA sobre los interesados.

Los riesgos del uso de IA podrán afectar a ámbitos como la sostenibilidad y el impacto medioambiental. Un uso sin garantías de la IA puede afectar al acceso a determinadas certificaciones de carácter medioambiental que obtienen ciertas empresas.

En este sentido, existen una serie de factores que se deberán considerar en la realización del análisis de riesgos:

- Requerimientos de protección de datos.
- Volumen, sensibilidad y calidad de los datos.
- Seguridad de los datos.
- Impacto potencial en los individuos y en la sociedad en su conjunto.
- Probabilidad, severidad y duración del impacto.
- Medidas mitigadoras.

III **GESTIÓN DE LOS RIESGOS DETECTADOS:** Tanto para determinar el riesgo, como para establecer las medidas apropiadas, resultará necesario analizar todo el ciclo de vida de la IA, dividiendo el mismo en sus distintas fases con el objetivo de gestionar las posibles peculiaridades de cada una de ellas.

A modo de ejemplo, pueden encontrarse distintas etapas comunes a todos los sistemas de IA, en especial cuando impliquen un desarrollo tecnológico:

» Requisitos de negocio y de diseño (funcionales y no funcionales).

» Obtención de los datos y preparación; (iii) Entrenamiento y pruebas (desarrollo).

» Implantación y seguimiento (explotación).

En este sentido, la gestión de los riesgos comportará reducir el potencial perjuicio disminuyendo la probabilidad de que se materialicen o el impacto que representan.

La Organización debe diseñar una estrategia de gestión de riesgos con la finalidad de:

- Garantizar la disponibilidad, integridad y confidencialidad de los datos personales.
- Garantizar el ejercicio de derechos de los interesados.
- Garantizar los principios relativos al tratamiento de información personal: legitimidad, minimización y transparencia (revelar cuando un contenido haya sido generado por la IA).

Los riesgos asociados al uso de componentes IA deberán estar sujetos a planificación y acciones de mitigación proporcionales a la gravedad de los daños que puedan conllevar.

En especial, la Organización deberá tener en cuenta los riesgos inherentes a las operaciones realizadas por los algoritmos: sesgos humanos, errores en la implementación, vulnerabilidades de seguridad, etc.

El riesgo más característico lo constituirá aquel que se deriva del sesgo en los sistemas de toma de decisiones sobre las personas o su discriminación trasladando a la implementación del componente IA los prejuicios existentes en la sociedad (algorithmic discrimination).

Una correcta gestión de riesgos llevará a la Organización a poder identificarlos, analizar su nivel de impacto y finalmente, implementar medidas mitigadoras y evaluar el riesgo residual remanente y mantenerlo controlado. La ISO 31000 establece la siguiente definición de riesgo residual: “aquel riesgo que subsiste, después de haber implementado los controles”.

Hay que recordar que la gestión de riesgos debe tener un enfoque tanto preventivo (implementar políticas que regulen el uso ético de las herramientas de IA en la Organización) como un enfoque reactivo (notificación de brechas y gestión de las consecuencias).

Un aspecto esencial para conseguir una gestión de riesgos adecuada es la formación y concienciación de todos los recursos humanos, implicando a la Organización desde la Dirección.

Según se ha indicado en anteriores apartados de este documento, el nuevo marco regulador define niveles diferentes de riesgo en la IA con obligaciones diferentes para cada uno de estos niveles.

IV **ANÁLISIS DE LOS RIESGOS REGULATORIOS:** La Organización deberá elaborar un plan de cumplimiento normativo que identifique el contexto normativo que afecta a la actividad de la Organización y crear un órgano que supervise el modelo de cumplimiento respaldado por la Dirección.

La transparencia, la gestión de la calidad y el buen gobierno son claves para alcanzar este cumplimiento.

Identificadas las obligaciones legales, normativas y éticas, la Organización deberá establecer un plan de acción para la adecuación al marco regulatorio vigente.

Al margen de la normativa específica de IA, se deberán analizar otros posibles impactos regulatorios, como por ejemplo la protección del secreto profesional y fuga de información confidencial, riesgos específicos para la privacidad, propiedad intelectual de los contenidos y publicidad ilícita, protección del consumidor, derecho de la competencia, entre otros.

A modo de ejemplo, la normativa de protección de datos determina la obligación de realizar una Evaluación de Impacto en la Privacidad cuando los niveles de riesgo asociados al tratamiento de información personal son elevados, circunstancia que obligará a las entidades a ir un paso más de la mera gestión del riesgo, puesto que exige una formalidad adicional que consiste en la necesaria realización de la Evaluación de Impacto en la Privacidad, tal y como establece el artículo 35.3.a del RGPD, cuando se realice la elaboración de perfiles basados en tratamientos automatizados.

V **ELABORACIÓN DEL INFORME, IMPLANTACIÓN DE LAS RECOMENDACIONES Y MEJORA CONTINUA:** Los resultados del análisis realizado previamente deberán concretarse en la adopción de una serie de medidas específicas y concretas para la gestión del riesgo, algunas de ellas orientadas a reforzar las obligaciones de cumplimiento en función de dicho riesgo. Para ello, será necesario establecer procesos de revisión de los resultados generados por la IA para verificar su adecuación y respeto con los compromisos éticos, legales y normativos de la Organización.

Hay que diferenciar el proceso de gestión del riesgo de los informes que se generan en el proceso. La gestión del riesgo es un proceso que se acredita documentalmente. Documentar el proceso da respuesta a la obligación de 'accountability'.

Disponer de la documentación dará soporte a la ejecución eficaz y eficiente de la gestión del riesgo por el uso de IA para los derechos y libertades y permitirá demostrar que así se ha realizado.

Todas las medidas que emanen de las recomendaciones deberán ser objeto de revisión, evaluación y mejora permanentes.

1.3.2. CICLO DE VIDA DE UNA SOLUCIÓN DE IA

INTRODUCCIÓN

Una vez definidos los conceptos anteriores, se puede comenzar con el análisis de los aspectos regulatorios de aquellas actividades que utilizan un componente de IA. De conformidad con lo que se ha indicado en apartados previos, un componente de IA puede estar integrado junto con otras actividades que realice la organización, incluso, de distintas entidades, como por ejemplo los servicios de algunos proveedores tecnológicos que ofrecen módulos de IA para incorporar en las actividades de las organizaciones.

Con carácter previo a abordar el ciclo de vida de la IA, es importante destacar algunas de las principales razones que provocan el fracaso de la gran mayoría de proyectos de IA en las organizaciones:

- **CARENCIAS DE CAPACIDADES QUE PUEDEN SURGIR A VARIOS NIVELES:** en un primer momento, pueden estar relacionadas con las capacidades de la ciencia de datos dentro de la organización o a disposición de ésta, pero es probable que también incluyan la ingeniería de datos y la gestión operativa. Es posible que los procesos clave no estén adecuadamente definidos o implementados.

- **DATOS NO DISPONIBLES O DE MALA CALIDAD:** Los datos son el combustible principal de la IA y pueden ser volátiles o, atendiendo a su intensidad regulatoria, resultar necesario una comprensión clara del linaje, la calidad, la representatividad estadística y los casos de uso de los datos. La accesibilidad de los datos es a menudo un reto para los proyectos de IA cuando los datos se almacenan en repositorios de datos heredados múltiples y dispares, con escasa vinculación entre los conjuntos de datos. No es infrecuente que el 80% del tiempo necesario para crear un sistema de IA se dedique a la depuración y manipulación de los datos.

- **FALTA DE BASE TECNOLÓGICA:** Los sistemas de IA dependen de las bases técnicas necesarias y aunque muchos proveedores tecnológicos están incorporando capacidades de IA en sus productos existentes, las empresas también están digitalizando y diseñando sus plataformas tecnológicas, procesos y herramientas para permitir el rápido despliegue de casos de uso de IA. Las bases tecnológicas necesarias suelen incorporar el almacenamiento de datos, el acceso y las capacidades de depuración, como los procesos y canalizaciones de datos, la selección de marcos algorítmicos de Machine Learning (ML) e IA y el acceso a las mejores soluciones de IA como servicio (AIaaS), y en la actualidad no todas las organizaciones cuentan con estas bases.

- **ESTRUCTURA DEFICIENTE DE GOBIERNO:** Garantizar desde el principio un proceso de gobernanza claro con controles adecuados es fundamental, pero a menudo se pasa por alto. Este proceso debe abarcar tanto las funciones tradicionales de supervisión de proyectos (presupuesto, control de fases, apoyo en el despliegue y entrega, gestión de las partes interesadas internas y revisión), como un conjunto más amplio de controles éticos y normativos que, en última instancia, requerirán la supervisión de la Alta Dirección o incluso de un Comité ético específico para ello.

- **FALTA DE CONOCIMIENTOS SOBRE LA FINANCIACIÓN Y/O GESTIÓN DE PROYECTOS DE IA:** Los proyectos de IA requieren a menudo competencias en gestión de proyectos que pueden no existir en la organización. Puede que sea necesario mejorar la capacidad interna de gestión de proyectos para llevar a cabo proyectos de IA. En este tipo de proyecto puede resultar diferenciador prestar especial atención a los datos, la ética y las comunicaciones con las partes interesadas externas. El uso de metodologías ágiles y gestión de proyectos es adecuado, y la experimentación iterativa es fundamental, así como la gestión de riesgos durante el proceso, ya que el riesgo de "deriva" del modelo o un entorno cambiante pueden socavar la eficacia del algoritmo.

ETAPAS

El ciclo de vida de la IA es el proceso iterativo que siguen las soluciones de IA, desde la idea inicial de la herramienta en la mente de las partes interesadas de la organización, hasta el uso y mantenimiento de la herramienta en un entorno operativo. En algunos casos, también puede incluir el desmantelamiento de la herramienta, si la empresa determina que ya no es necesaria.

Conviene recordar que, de conformidad con los requerimientos del borrador del Reglamento de la IA, resultará necesario realizar un análisis de toda la actividad en la que participe un sistema de la IA, de manera que el componente IA y el resto de los elementos que conforman toda la actividad deben ser estudiados de manera conjunta y analizando sus posibles requerimientos de manera global.

Es importante matizar que, al pasar un sistema de IA a través de diferentes etapas en su desarrollo, resultará necesario adaptar los requerimientos regulatorios o los diferentes matices o particularidades de cada una de estas etapas. Desde un punto de vista organizacional, se puede dividir el ciclo de vida de la IA en cuatro fases:

1.3.2.1 DISEÑO

Un modelo sólo tiene éxito si resuelve los problemas adecuados para la organización, en el orden de prioridad correcto, con las partes interesadas adecuadas, reuniendo las habilidades adecuadas, alineadas bajo un equipo de liderazgo que comprenda tanto el potencial como las limitaciones y los riesgos de estas tecnologías. Es por ello, que antes de abordar la implantación de una solución de IA, es preciso evaluar si se han realizado las siguientes acciones (WEF: Empowering AI Leadership):

- Definir la estrategia y visión empresarial como base para ayudar a definir problemas concretos que desea que la IA ayude a resolver.
- Definir una hoja de ruta priorizada de posibles soluciones de IA, incluido un análisis de impacto y costes, y establecer métricas de éxito.
- Mayor concienciación sobre los riesgos empresariales inherentes al desarrollo de una solución de IA (razones por las que las soluciones IA fracasan).
- Contratar a las personas adecuadas para ejecutar el caso de uso.
- Confeccionar un equipo de proyecto adecuado y multidisciplinar.
- Alinearse con las partes interesadas, principalmente, entre negocio y la ingeniería de datos que permita una comunicación fluida e integrar su trabajo.
- Evaluar el impacto de la solución inteligente identificando el propósito de la aplicación, quienes van a ser los responsables del despliegue y cuáles serán los beneficios, daños o tensiones que este sistema inteligente puede provocar.
- Analizar y revisar la información recolectada en la evaluación de impacto y compartir las conclusiones con el gobierno de IA de la organización quienes revisan si el enfoque del proyecto se ajusta al marco ético requerido por el caso de uso y en caso de presentar desajustes, serán escalados para ofrecer una respuesta adecuada para la realización del proyecto.
- Asegurar que las obligaciones legales y los principios éticos que se han definido en los apartados anteriores se implementan en el proceso de desarrollo de IA.

En conjunto, estas acciones pueden considerarse la fase de diseño de una aplicación de IA.

1.3.2.2. DESARROLLO

Como hemos indicado, los datos son un requisito fundamental para entrenar y utilizar cualquier modelo de IA, por lo que las organizaciones deben adoptar prácticas sólidas que les permitan recopilar y tratar grandes cantidades de datos de alta calidad.

A menudo, la recopilación, el almacenamiento y la depuración de los datos clave para crear un modelo constituyen la mayor parte del trabajo para que dicho modelo funcione. Dependiendo del tipo de solución inteligente que se desarrolle el modelo requerirá tipos específicos de datos sobre el tema en cuestión. No disponer de datos suficientes o de la suficiente calidad suele dar lugar a resultados imprecisos y a una fiabilidad limitada. En estos casos, se puede recurrir a recursos internos y externos de la organización para recopilar datos. En cualquier caso, es clave comprender la "cadena de suministro de datos" para garantizar la gestión de la calidad y la responsabilidad.

Los datos de origen son los generados directamente por la propia organización (por ejemplo, transacciones, como compras de clientes). Las organizaciones tienden a depender más de los datos origen, lo que los hace fiables y de libre uso, una vez considerados todos los aspectos normativos y éticos. Cuando los datos de origen son insuficientes, las organizaciones recurren a datos externos. Los datos de terceros suelen ser más sólidos y completos, ya que combinan varias fuentes, pero se puede asumir un alto riesgo al usarlos, ya que a menudo es difícil confirmar los procesos de recopilación de datos.

Además del abastecimiento de datos, hay que tener en cuenta toda la cadena de suministro de datos. A veces denominada DataOps, la cadena de suministro de datos es una combinación de prácticas, políticas y tecnologías necesarias para manejar, analizar y aprovechar los datos de la organización. La optimización de esta cadena de suministro mejora la agilidad, reduce el déficit de datos y da confianza a la empresa en los datos que se utilizan. Es importante destacar que además de disponer de datos suficientes, también hay que asegurarse de que se cumplen normas de alta calidad para ello hay que tomar de referencia las seis dimensiones de calidad siguientes:

- **EXHAUSTIVIDAD:** Entendiéndose como la capacidad de garantizar que los datos son completos. Lo que suele expresarse como el porcentaje de campos que tienen un valor (campo con un valor dividido por el total de campos esperados).
- **SINGULARIDAD**: Permite garantizar que los datos sólo se registran una vez, con un mínimo de valores duplicados, lo que se representa como un porcentaje (entidades únicas esperadas divididas por entidades únicas reales).
- **DISPONIBILIDAD:** Representa la medida de la demora entre el momento en que los datos se generan y el momento en que están listos para ser utilizados. Se trata de una dimensión importante para los datos, que son sensibles al tiempo y cambian con rapidez.
- **VALIDEZ:** Permite garantizar que los datos se capturan en el formato o sintaxis correctos.

- **PRECISIÓN:** Es la capacidad de medir si los datos representan al "sujeto del mundo real" que se está estudiando. Aunque esto puede ser difícil de cuantificar, las técnicas de recopilación de datos y las reglas de entrada deben estar en su lugar para reducir el sesgo y los riesgos de recopilación inexacta.

- **COHERENCIA:** Posibilita garantizar la alineación entre los datos almacenados en múltiples lugares sobre el mismo tema. Suele medirse en porcentaje (número de coincidencias/número total de campos duplicados).

Esta lista de dimensiones no es exhaustiva y no todos los campos de datos requerirán que se cumplan todas las dimensiones, pero estas dimensiones proporcionan una visión de las características de "datos de calidad". Además de la calidad, la organización debe asegurarse de que existen prácticas adecuadas para el cumplimiento legal, la gestión de riesgos, el acceso y la seguridad, el control de calidad y la usabilidad de los datos. A menudo, esto se consigue aplicando prácticas formales de gobernanza de datos, incluidas funciones y herramientas/sistemas especializados. El despliegue de un buen modelo de gobierno de datos requerirá además de las políticas y procedimientos establecidos por los responsables de la gestión de datos de la organización, implantar las herramientas técnicas adecuadas, como diccionarios de datos, catálogos de datos y cuadros de mando de calidad de datos.

Es importante en esta fase tener en cuenta las siguientes acciones:

I. Identificar requisitos. Sigue un modelo similar al modelo de desarrollo tradicional, se priorizan los requerimientos principales y se identifican los principales modelos de IA que van a formar parte de la solución inteligente. En este momento es clave definir el modelo de gobierno de los conjuntos de datos seleccionados para la solución de conformidad con los requisitos éticos identificados previamente. También será necesario definir las estrategias para la gestión de los requisitos éticos, con el objetivo de mitigar riesgos y tensiones previamente identificados.

II. Definir el uso de datos. Consiste en la realización de un análisis de la procedencia de los datos, identificando a los responsables de la recolección de estos, se procede al etiquetado de estos y se limitan los casos de uso que pueden utilizar dichos datos en caso de limitaciones de protección de datos o cualquier otro aspecto bloqueante.

III. Elaborar Planes de pruebas prestando especial atención a los casos de prueba que pueden colisionar con los principios éticos.

IV. Evaluación del prototipo. Con esta acción se pretende identificar y analizar los potenciales escenarios de fallo del modelo que sustenta el sistema inteligente, buscando vulnerabilidades con técnicas de muestra adversaria y definiendo el plan para mitigarlos.

V. Criterios de despliegue. En esta actividad se evalúan las condiciones de despliegue, se tienen en cuenta los datos que ya se pueden utilizar en un escenario productivo, y realizar una definición del testeo holístico de todas las funcionalidades.

VI. Revisión y pruebas. Durante esta actividad se ejecutan los planes de revisión y pruebas, teniendo en consideración los principios éticos definidos y finalmente se escalan los potenciales gaps encontrados.

1.3.2.3. DESPLIEGUE

Después de definir, seleccionar y entrenar el modelo pasamos a la fase de despliegue, esto es cuando la solución inteligente pasa a producción. Uno de los pasos más complicados e importantes del ciclo de vida de la IA, es la implantación. La implantación es el método por el que se integra un modelo de aprendizaje automático en un entorno de producción existente para tomar decisiones empresariales prácticas basadas en datos. Esta integración suele requerir la coordinación entre los científicos de datos, los equipos informáticos, los desarrolladores de software y negocio. En esta fase, es donde encuentra la mayoría de las organizaciones, si los sistemas informáticos que corren en máquinas con software antiguo, debido a claras incompatibilidades con lenguajes tradicionales de creación de modelos IA como Python. El éxito del despliegue es esencial para obtener valor de negocio de un modelo de IA, pero también es uno de los aspectos más complicados de la implantación de la IA. Es fundamental que la integración y el despliegue se consideren al principio del ciclo de vida de la IA y no en el momento del despliegue para tener en cuenta las posibles incompatibilidades.

1.3.2.4. SEGUIMIENTO Y CONTROL

Una vez implantado un modelo de IA, hay que supervisarlo continuamente y ajustarlo para garantizar que la información obtenida siga siendo fiable y acorde con los valores y principios éticos de la empresa. Es importante asegurar que los principios éticos se encuentran en los niveles aceptados durante la vida de la solución. Las etapas finales del ciclo de vida garantizan el mantenimiento del modelo hasta su retirada.

Como es sabido, los conocimientos de muchos modelos de IA, en concreto de ML, se obtienen a partir de una afluencia constante de nuevos datos. Es por ello por lo que, estos modelos necesitarán una cuidadosa supervisión, ajuste y reentrenamiento para garantizar que siguen produciendo las percepciones correctas. La función de supervisión de la organización comprobará que los requisitos éticos y normativos sean adecuados, así como aspectos más concretos como la representatividad estadística en curso, los posibles sesgos, los bucles de retroalimentación y otras características más técnicas que dependen del modelo.

Será preciso realizar auditorías periódicas de los supuestos subyacentes realizados en las primeras fases de selección y formación del modelo a la luz de los resultados actuales del modelo, ya que la introducción de nuevos datos exigirá un reentrenamiento del modelo a medida que se produzcan cambios. Cada vez que se vuelva a entrenar el modelo, habrá que tener en cuenta la explicabilidad, la interpretabilidad y la privacidad, entre otros aspectos definidos en el marco del desarrollo de soluciones IA responsables.

Antes de concluir, es preciso hablar de la última etapa, la retirada del modelo. Aunque no suele incluirse en el ciclo de vida, una pregunta clave que deben plantearse la organización es cuándo retirar un algoritmo o solución inteligente. Una de las causas podría ser que el sistema no cumple algunos de los principios éticos definidos por la organización

o alguna de sus funcionalidades no se ejecuta de forma adecuada, lo que podría suponer un riesgo organizativo importante.

Es importante que la organización incorpore, como parte de sus procesos de gobernanza y principios éticos, criterios para evaluar el funcionamiento continuo de la solución, así como la gestión del riesgo, incorporando un conjunto de estrategias de mitigación en caso de que una solución deje de funcionar según lo previsto o impacte de forma significativa en la organización o las personas.

1.3.2.5. RETIRADA FINAL DEL COMPONENTE DE IA

Cuando una organización decide, por diferentes motivos, la retirada de una solución de IA es necesario retirar el sistema actual y determinar si pudiera resultar necesario su utilización en un fututo, determinando para ello la compatibilidad de los datos, el modelo o de sistema utilizado previamente.

Por este motivo, se deberán identificar los supuestos en los que los datos van a ser explotados por un nuevo sistema o bien, si será precisa la utilización de más datos o una transformación de los actuales con el objetivo de que el nuevo sistema pueda ejecutarse en producción. Esta cuestión podría tener relevancia al dificultar una posible reutilización del sistema de IA, así como el mantenimiento de un entorno aislado que se deba retirar por motivos de compatibilidad. Las conclusiones obtenidas por un sistema de IA que se retire, sus datos, tanto de entrada como de entrenamiento, pueden necesitar de requisitos adicionales de carácter normativo para poder ser utilizados en un nuevo sistema que deberán ser analizados por las organizaciones.

2 MARCOS DE REFERENCIA PARA LA VALIDACIÓN DE ALGORITMOS

En un plano más técnico, identificamos 4 marcos de referencia en relación con la evaluación y auditoría de los sistemas de IA:

1. Requisitos para Auditorías de Tratamientos que incluyan IA, de la Agencia Española de Protección de Datos (AEPD), con un enfoque en el tratamiento de los datos personales.
2. Auditoría de algoritmos gubernamentales, de la corte de auditoría de los Países Bajos, donde se evalúan 4 perspectivas diferentes con la ética como denominador común en todas ellas.
3. Auditoría en tres niveles a los grandes modelos lingüísticos (LLM) de IA, propuesto por la Universidad de Oxford, dividiendo entre los niveles de gobierno, modelo y aplicación, pero resaltando su interrelación.
4. El modelo del BSI alemán que busca lograr una IA confiable y obtener marcos de referencia de auditoría para los sistemas de IA.

2.1. REQUISITOS PARA AUDITORÍAS SEGÚN AEPD

La AEPD ha publicado bastante documentación sobre la inteligencia artificial y la protección de datos personales, destacando el trabajo "Requisitos para Auditorías de Tratamientos que incluyan IA", que proporciona las orientaciones metodológicas particulares y un listado de objetivos de control y controles específicos que podrían ser seleccionados para incluir en el proceso de auditoría de protección de datos de un tratamiento que incorpore componentes o soluciones de IA. Dichos objetivos de control y controles se agrupan en cinco grandes bloques:

- Identificación y Transparencia del Componente
- Propósito del Componente de IA
- Fundamentos del Componente de IA
- Gestión de los Datos
- Verificación y Validación

2.2. ENFOQUE DE LA CORTE DE AUDITORÍA DE LOS PAÍSES BAJOS

Por otra parte, la Corte de Auditoría de los Países Bajos ha definido un marco de referencia orientado a auditar los algoritmos gubernamentales, ya que, debido a la su utilización, se están automatizando muchos procesos en los que se apoya el gobierno de cara a conceder ayudas, o a iniciar un proceso sancionador.

I GOBERNANZA Y RESPONSABILIDAD

Determina si se ha definido una finalidad clara del algoritmo utilizado, se hace una evaluación periódica de los riegos asociados a la utilización del algoritmo, incluso aunque se haya externalizado.

II MODELO Y DATOS

Aborda los controles para garantizar que el algoritmo funciona correctamente. La exhaustividad es un requisito para todo tratamiento de datos, donde ningún ciudadano o ente singular debiera ser excluido involuntariamente del tratamiento de datos, ya que puede dar lugar a resultados incorrectos. La comprobación de la exactitud de los datos utilizados por el modelo es el siguiente paso lógico tras de la exhaustividad.

Se debe comprobar si contiene un sesgo indeseable en relación con determinados individuos o grupos. Para ello, estudiando las causas de los posibles sesgos y cómo tratarlos en caso de que se produzcan.

III PRIVACIDAD

Se deben elaborar registros de tratamiento que describan el tipo de datos personales que recogen y tratan. Realizar evaluaciones de impacto de la protección de datos personales. Minimizar los datos personales recogidos, es decir, no recopilar más datos de los necesarios para alcanzar el objetivo previsto. Transparencia en cuanto a la privacidad y al tratamiento de datos personales, incluyendo que algoritmos se emplean en su política de privacidad.

IV CONTROLES GENERALES DE TI

Deben proporcionarse controles de acceso a los algoritmos para que tanto los datos utilizados como el propio algoritmo estén protegidos. Aunque el desarrollo o la gestión de un algoritmo se haya externalizado, la responsabilidad de que el algoritmo funcione correctamente recae en el organismo que usa el algoritmo.

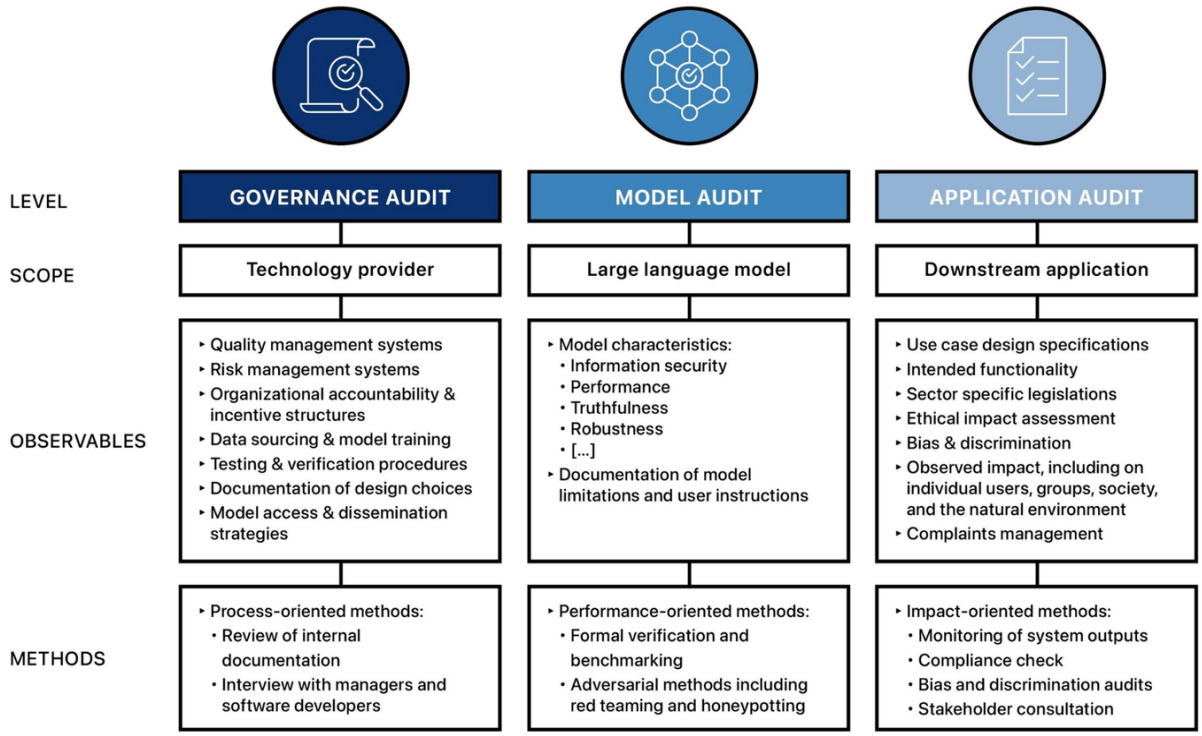
La ética se integra de las cuatro perspectivas descritas, a través de cuatro principios éticos:

- Respeto de la autonomía humana
- Prevención de daños
- Equidad
- Explicabilidad y transparencia

2.3. ENFOQUE DE AUDITORÍA SEGÚN ESTUDIO DE UNIDAD DE OXFORD

Un estudio de investigación, liderado por la Universidad de Oxford, plantea un enfoque de auditoría de tres niveles para abordar los retos de gobernanza de los grandes modelos lingüísticos (LLM) en la IA:

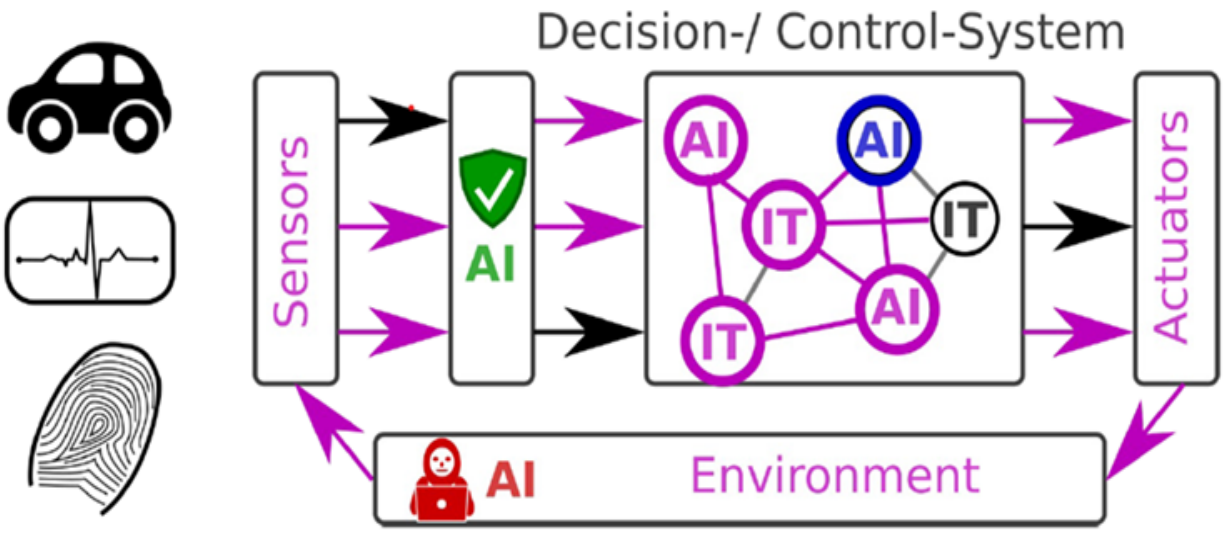
- Auditorías de gobernanza (de los proveedores de tecnología que diseñan y difunden los LLM).
- Auditorías de modelos (de los LLM después de la formación previa pero antes de su publicación).
- Auditorías de aplicaciones (de aplicaciones basadas en LLMs, en un dominio específico).



Se plantean tres ámbitos de auditoría independientes pero que están interrelacionados ya que la salida de una auditoría sirve como entrada para las auditorías del resto de niveles. Es un modelo que podría extrapolarse a otros modelos de IA.

2.4. ENFOQUE DEL BSI ALEMÁN

Por otra parte, la aproximación del BSI alemán consiste en considerar un sistema de IA como la suma de diversos módulos o componentes: sensores, actuadores, módulos de TI, módulos de IA y un entorno de utilización.



Los aspectos relevantes del sistema de IA para el BSI son:

- Seguridad.
- Protección.
- Rendimiento.
- Interpretabilidad/Explicabilidad.
- Trazabilidad.
- Gestión de riesgos.

Los aspectos como la privacidad, el sesgo y la ética no los considera del ámbito de actuación del BSI, sino del usuario. Considera los módulos de IA tanto en su ciclo de vida (filas agrupadas como AI module life cycle) como en su utilización en una aplicación en particular (filas agrupadas como Embedding) y de este modo se evalúan los distintos aspectos del sistema utilizando una matriz como la de la siguiente figura, dónde el número indica la posibilidad de auditoría en una escala de 0 (nada auditable) a 10 (completamente auditable).

	Lifecycle Phase / Aspect	Security	Safety	Performance	Robustness	Interpret-/ Explainability	Tracability	Risk Management	
Embedding	organization	3	2	5	3	4	6	6	
	use case specific requirements & risks	5	5	5	5	4	4	6	
	Embodiment & situatedness of AI module	5	5	5	5	6	2	5	
AI module life cycle	planning phase	4	4	5	4	4	6	6	Out of scope: user focused criteria ("Ethics": Bias, Data Privacy, Human oversight, ...)
	data acquisition and QA phase	4	5	6	6	4	6	6	
	training phase	5	5	5	5	6	6	6	
	evaluation phase	5	5	5	5	6	6	6	
	deployment and scaling phase	4	2	5	3	4	6	6	
	operational (& maintenance) phase	5	2	5	3	4	6	6	

Los aspectos que considerar son:

- **Seguridad:** propia de los sistemas de TI, robustez pasiva y activa del sistema de IA frente a ataques, especialmente los específicos de la IA y con respecto a los tres objetivos de seguridad integridad, confidencialidad y disponibilidad
- **Protección (safety):** Protección de personas, organizaciones y activos frente a daños (físicos).
- **Rendimiento:** Rendimiento del sistema de IA con respecto a las métricas de rendimiento.
- **Robustez:** Robustez pasiva y activa frente a variaciones naturales de las entradas (situaciones), incluidas las que no se han tenido en cuenta durante el entrenamiento.
- **Interpretabilidad y Explicabilidad**: Capacidad de los seres humanos para comprender el proceso de proceso de decisión del sistema de IA.
- **Trazabilidad:** Trazabilidad del sistema de IA a lo largo del ciclo de vida.
- **Gestión de riesgos:** Identificación, análisis y priorización de riesgos y aplicación coordinada de recursos para minimizar el impacto del riesgo.

Este modelo permite una ampliación hacia los aspectos que tratan otros marcos de referencia como la privacidad o la ética.

MODELO DE GOBIERNO IA

www.ismsforum.es
info@ismsforum.es
(+34) 915 63 50 62

@ISMSForum
ISMS Forum

Una iniciativa de

isms forum