



Universidad Internacional de La Rioja
Escuela Superior de Ingeniería y
Tecnología

Máster Universitario en Inteligencia artificial

**Definición de Controles, Procesos y
Estructuras para el Diseño de un Sistema
de Inteligencia Artificial Responsable**

Trabajo fin de estudio presentado por:	Alvear Aguirre Isabel Gardenia Flores Villamil Edwin Vladimir Ortiz Albán Iván Josué
Tipo de trabajo:	Trabajo de Fin de Máster
Director/a:	PhD Juan Ignacio Rouyet Ruiz
Fecha:	12/feb/2025

Resumen

Con el apogeo de la inteligencia artificial, el hombre utiliza ésta, pero también con cada acción inevitablemente deja información que entrenará algoritmos de IA que tomarán decisiones automáticas.

En este contexto, en occidente han surgido diferentes propuestas para establecer límites éticos y responsables para el uso de IA y la información personal. Este trabajo analiza los riesgos actuales del uso de la IA, las autorregulaciones de las grandes empresas que la desarrollan y la utilizan (Google, Meta, Amazon, Open AI y Microsoft), las reglamentaciones vigentes en la Unión Europea, como las directrices para una IA fiable (2019), el reglamento general de protección de datos (2016), el reglamento de la IA (2024), así como diferentes iniciativas de investigadores y de la academia (Cisco Systems - Universidad de Oxford, Universidad de Stanford) y las combina para ofrecer a las empresas de occidente, una herramienta o marco de trabajo que les permita evaluar su nivel de cumplimiento de IA responsable conforme la normativa vigente para apoyarlos a definir los controles, procesos y estructuras necesarias para ofrecer productos o servicios de IA concebidos, desarrollados e implementados responsablemente, que estén centrados en el ser humano sin arriesgarlo a él o a sus derechos fundamentales.

Con la combinación de las iniciativas de occidente se llegó a 181 preguntas finales (171 relacionadas al producto o servicio de IA y 10 acerca de la empresa), que se aplicaron a una empresa para evaluar su nivel de IA responsable. Como resultado se obtuvo que la empresa alcanzó el nivel global 2 (en una escala de 5) de IA responsable con 422 puntos y un nivel ponderado de IA responsable de 2,4294 que coincide con el nivel global. También se evaluó el nivel de cada área para identificar recomendaciones específicas para afianzar el nivel 2 y llevarla al nivel 4 para ser considerada IA responsable.

Palabras clave:

IA responsable en Occidente, IA ética, IA confiable, Modelo de Madurez en IA responsable, IA responsable, Medición IA responsable.

Abstract

With the rise of artificial intelligence, humans use AI but also with every action they inevitably leave behind information that will train AI algorithms to make automatic decisions.

In this context, different proposals have emerged in the West to establish ethical and responsible boundaries for the use of AI and personal information. This paper analyses the current risks of AI use, the self-regulations of the big companies that develop and use it (Google, Meta, Amazon, Open AI and Microsoft), the regulations in force in the European Union, such as the guidelines for trustworthy AI (2019), the general data protection regulation (2016), the AI regulation (2024), as well as different initiatives from researchers and academia (Cisco Systems - Oxford University, Stanford University) and combines them to provide companies in the West with a tool or framework to assess their level of responsible AI compliance with current regulations to support them in defining the necessary controls, processes and structures to deliver responsibly conceived, developed and implemented AI products or services that are human-centric without putting them or their fundamental rights at risk.

The combination of the Western initiatives resulted in 181 final questions (171 related to the AI product or service and 10 about the company), which were applied to one company to assess its level of responsible AI. As a result, the company reached the overall level 2 (on a scale of 5) of responsible AI with 422 points and a weighted level of responsible AI of 2,4294, which matches the overall level. The level of each area was also assessed to identify specific recommendations to entrench level 2 and take it to level 4 to be considered a responsible AI.

Keywords:

Responsible AI in the West, ethical AI, trustworthy AI, Responsible IA Maturity Model, Responsible IA, Responsible IA measure.

Índice de contenidos

1	Introducción	1
1.1	Motivación	1
1.2	Planteamiento del trabajo	1
1.3	Estructura del trabajo	2
2	Contexto y estado del arte	4
2.1	Riesgos de la IA	4
2.1.1	Preocupaciones de la sociedad en la IA	4
2.1.2	Amenazas típicas de la IA	6
2.2	Gobernanza de la IA -Empresas desarrolladoras del mundo occidental.	8
2.2.1	Google.....	8
2.2.2	Apple.....	10
2.2.3	Amazon	10
2.2.4	Microsoft	12
2.2.5	Meta	15
2.2.6	Open AI	16
2.2.7	Comparación de los principios éticos de las empresas desarrolladoras de IA	18
2.3	Gobernanza de la IA – Grandes empresas occidentales que usan IA	18
2.3.1	Deloitte	19
2.3.2	Repsol	19
2.3.3	Unilever	20
2.3.4	Walmart.....	21
2.3.5	Salesforce.....	22
2.3.6	Telefónica	23

2.4	Regulación en IA en el mundo occidental	23
2.4.1	Directrices éticas para una IA fiable	24
2.4.2	Reglamento General de protección de datos de la Unión Europea.....	29
2.4.3	Reglamento Europeo sobre Inteligencia Artificial.....	31
2.4.4	Orden presidencial de Estados Unidos de América para legislar la IA.....	34
2.4.5	Data Management Association - DAMA.....	37
2.5	Iniciativas de regulación de la IA desde la academia	40
2.5.1	Ethics and Responsible AI Deployment/ Ética y Despliegue Responsable de la IA. 40	
2.5.2	Adding Adding human values on the deepfake: co-designing fact-checking solutions to combat misinformation / Añadir valores humanos al deepfake: codiseñar soluciones de comprobación de hechos para combatir la desinformación	41
2.5.3	Índice de transparencia del modelo fundacional	43
2.5.4	Transparencia de la IA en la era de los LLM: Una hoja de ruta para la investigación centrada en el ser humano	47
2.6	Conclusiones.....	48
3	Objetivos concretos y metodología de trabajo.....	50
3.1	Objetivo general	50
3.2	Objetivos específicos	50
3.3	Metodología del trabajo.....	50
3.3.1	Herramientas	50
3.3.2	Metodología	60
4	Desarrollo específico de la contribución.....	64
4.1	Elaboración del formulario.....	64
4.1.1	Preguntas obtenidas de las Directrices Éticas para una IA fiable	66

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

4.1.2	Preguntas obtenidas del Reglamento General de Protección de Datos	71
4.1.3	Preguntas obtenidas del Reglamento de IA	73
4.1.4	Preguntas obtenidas del Índice de Transparencia Fundacional.....	75
4.1.5	Preguntas obtenidas de los principios éticos comunes de las empresas desarrolladoras de IA.	77
4.2	Aplicación del formulario – caso práctico	83
4.2.1	Recopilación de datos.....	83
4.2.2	Análisis de cumplimiento	84
4.2.3	Sensibilización en la empresa sobre IA responsable	93
4.3	Lecciones aprendidas	94
4.4	Rueda de madurez en IA responsable	94
4.5	Mejoras a realizarse	97
4.5.1	En el formulario	97
4.5.2	Durante el acompañamiento.....	98
4.5.3	Otras sugerencias	98
5	Conclusiones y trabajo futuro	99
5.1	Conclusiones.....	99
5.2	Líneas de trabajo futuro	100
	Referencias bibliográficas.....	103
Anexo A.	Lista de evaluación para una IA fiable.....	111
Anexo B.	Evaluación de los 100 indicadores según el modelo fundacional (FMTI)	124
Anexo C.	Formulario del marco de trabajo.....	129
Anexo D.	Formulario llenado con la empresa.....	140
Anexo E.	Informe para la Empresa	169

Alvear Aguirre Isabel Gardenia

Flores Villamil Edwin Vladimir

Ortiz Albán Iván Josué

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Índice de figuras

Figura 1. Estructura del trabajo a realizarse	2
Figura 2. Patentes de IA concedidas en el CSET.	5
Figura 3. IA fiable y sus niveles	25
Figura 4. Requisitos que deben cumplir un sistema de IA para considerarse una IA fiable. ...	28
Figura 5. El dato y recorrido en el RGPD	30
Figura 6. Niveles de riesgo en las aplicaciones de IA según el Reglamento de IA	31
Figura 7. Gobernanza de la IA en la Unión Europea según el Reglamento de IA	33
Figura 8. Orden presidencial, pasos y principios transversales	36
Figura 9. Ciclo de vida del dato según el DMBOK2	37
Figura 10. Áreas de conocimiento (dimensiones) que propone el DMBOK2	38
Figura 11. Indicadores de la aplicación 2023 del modelo de transparencia fundacional con énfasis en los tres aspectos evaluados.	45
Figura 12. Puntuación global 2024 del Índice de Transparencia del Modelo Fundacional con detalle de los tres aspectos evaluados.	46
Figura 13. Documentación que se utiliza como herramientas para desarrollar el marco de trabajo	51
Figura 14. Áreas de Evaluación de IA responsable.	52
Figura 15. Fases de la metodología planteada para desarrollar el marco de trabajo para una IA responsable.	61
Figura 16. Estado del área 1 evaluada: agentes humanos y supervisión.	84
Figura 17. Estado del área 2 evaluada: solidez técnica y seguridad.	85
Figura 18. Estado del área 3 evaluada Privacidad y gobernanza de datos.	85
Figura 19. Estado del área evaluada 4 – Transparencia.	86
Figura 20. Estado del área evaluada 5 – Diversidad, no discriminación y equidad.	86

Figura 21. Estado del área evaluada 6 – Bienestar social y medioambiental	87
Figura 22. Estado del área evaluada 7 – Rendición de cuentas	88
Figura 23. Nivel ponderado de IA responsable de la empresa evaluada.....	92
Figura 24. Porcentaje relativo de desarrollo en IA responsable por cada área evaluada.	93
Figura 25. Rueda de madurez de IA responsable por subáreas.	96
Figura 26. Rueda de madurez en IA responsable por área evaluada.	97

Índice de tablas

Tabla 1. <i>Organización del trabajo en grupo.</i>	XIII
Tabla 2. <i>Principios éticos para el desarrollo de IA en Google.</i>	8
Tabla 3. <i>Pilares para el uso de la IA en Apple.</i>	10
Tabla 4. <i>Dimensiones para una IA responsable según Amazon.</i>	11
Tabla 5. <i>Herramientas de AWS para una IA responsable.</i>	12
Tabla 6. <i>Dimensiones para una IA responsable según Microsoft.</i>	13
Tabla 7. <i>Acciones que implementa Microsoft durando el ciclo de vida de los datos</i>	14
Tabla 8 <i>Pilares con los cuales se diseñan productos de IA en Meta</i>	15
Tabla 9. <i>Principios éticos de Open IA.</i>	16
Tabla 10. <i>Comparativa de principios éticos de las distintas empresas de IA.</i>	18
Tabla 11. <i>Derechos fundamentales en el contexto de una IA fiable.</i>	26
Tabla 12. <i>Principios éticos en el contexto de una IA fiable.</i>	27
Tabla 13. <i>Obligaciones de proveedores y usuarios de servicios de IA según el nivel de riesgo del sistema</i>	32
Tabla 14. <i>Infracciones y multas máximas por incumplimientos.</i>	34
Tabla 15. <i>Acciones de la Orden presidencial, sus plazos y responsables.</i>	35
Tabla 16. <i>Técnicas logarítmicas sugeridas en el estudio para cumplir con normativas</i>	40
Tabla 17. <i>Resultados del taller multidisciplinario sobre uso ético de las deepfakes</i>	42
Tabla 18. <i>Derechos de los interesados según el RGPD.</i>	53
Tabla 19. <i>Principales obligaciones del responsable de protección de datos según el RGPD.</i>	54
Tabla 20. <i>Aspectos del Reglamento de IA de la UE que se utilizarán en el marco de trabajo.</i>	55
Tabla 21. <i>Los 100 indicadores del FMTI que se revisarán como cuarto ingrediente para el marco de trabajo</i>	57

Tabla 22. <i>Principios éticos comunes de las grandes empresas desarrolladoras de IA que se considerarán para la elaboración del marco de referencia</i>	59
Tabla 23. <i>Áreas y subáreas que proponen las directrices éticas para una IA fiable</i>	65
Tabla 24. <i>Preguntas por área y subárea obtenidas de las directrices éticas para una IA fiable</i>	66
Tabla 25. <i>Preguntas por área y subárea obtenidas del reglamento general de protección de datos de la UE</i>	71
Tabla 26. <i>Preguntas por área y subárea obtenidas del reglamento de IA de la UE</i>	73
Tabla 27. <i>Preguntas por área y subárea obtenidas de los indicadores del índice de transparencia fundacional</i>	75
Tabla 28. <i>Preguntas por área y subárea obtenidas de los principios éticos comunes en las empresas desarrolladoras de IA</i>	77
Tabla 29. <i>Evaluación del formulario para determinar el nivel global de IA responsable</i>	79
Tabla 30. <i>Rangos para determinar el nivel global de IA responsable</i>	81
Tabla 31. <i>Ponderación por cada pregunta del formulario</i>	89
Tabla 32. <i>Puntuación ponderada por subárea evaluada</i>	90

Organización del trabajo en grupo

Considerando que este trabajo requiere investigación de la normativa y esfuerzos de Occidente para regular y establecer una IA responsable, los integrantes nos comprometimos a aportar con al menos una normativa legal, una iniciativa de regulación de investigadores y/o académicos y principios éticos de empresas que desarrollan y usan productos de IA.

Posteriormente se incluyó el aporte individual en el estado del arte y cada uno expondrá a los demás lo investigado para conjuntamente establecer las herramientas que se utilizarán y la metodología que se llevará a cabo para la contribución al conocimiento que se realiza con este trabajo.

Una vez establecidos los documentos base se generará conjuntamente el formulario acordando las preguntas y los niveles de IA responsable. Posteriormente se lo aplicará como piloto con una o dos empresas que acepten ser parte de este estudio.

Los resultados de la evaluación serán analizados en conjunto para realizar el informe a la empresa, registrar los hallazgos y conclusiones en este documento.

Partes que aborda el TFE

Este TFE aborda varias partes:

- Investigación enfocada en el mundo occidental de la normativa legal, regulaciones, iniciativas de investigadores y académicos, así como políticas emitidas por empresas que generan y usan IA para hacerlo de manera responsable.
- Análisis de las políticas, normativas y regulaciones considerando donde se complementan y redundan para utilizar las más relevantes para la elaboración del formulario.
- Elaboración del formulario con los documentos relevantes cuidando de no redundar contenido y en caso de solapamiento entre disposiciones incluir las prioritarias por ser de cumplimiento legal obligatorio.
- Acompañamiento a la(s) empresa(s) que acepten formar parte del piloto de evaluación de IA responsable.
- Análisis de los resultados obtenidos, elaboración del documento para la empresa

- Elaboración del documento que se presentará para el TFE.

Distribución y estructura de la memoria

Se ha planificado organizar el trabajo de escribir la memoria del TFE conforme la tabla a continuación:

Tabla 1. Organización del trabajo en grupo.

Organización del trabajo en grupo - Desarrollo de la memoria	
Apartado de la memoria	Responsables
Introducción	Alvear Aguirre Isabel Gardenia Flores Villamil Edwin Vladimir Ortiz Albán Iván Josué
Contexto y estado del arte	Alvear Aguirre Isabel Gardenia Flores Villamil Edwin Vladimir Ortiz Albán Iván Josué
Objetivos y metodología de trabajo	Alvear Aguirre Isabel Gardenia Flores Villamil Edwin Vladimir Ortiz Albán Iván Josué
Desarrollo específico de la contribución	Alvear Aguirre Isabel Gardenia Flores Villamil Edwin Vladimir
Conclusiones	Alvear Aguirre Isabel Gardenia Flores Villamil Edwin Vladimir

Fuente: Elaboración propia.

Objetivo del TFE desde el punto de vista de la adquisición de conocimientos

Desde el punto de vista de la adquisición de conocimientos, este TFE, nos permitirá profundizar en lo siguiente:

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

- Estado actual de cómo se están desarrollando modelos, productos y servicios de IA de manera responsable ética y fiable.
- Incorporar el conocimiento obtenido en un solo documento (formulario) que fortalecerá nuestra capacidad de análisis deductivo e inductivo.
- Cómo se debería concebir, diseñar, implementar y utilizar la IA de manera responsable con el ser humano y cumpliendo con la normativa legal de occidente.

Mecanismos de coordinación empleados

Para coordinar entre los miembros del grupo se ha creado un grupo de WhatsApp a través del cual se realizará cualquier comunicación. Aprovechando la licencia de Microsoft 365 proporcionada por UNIR se creó un repositorio en ONE DRIVE en el cual se colocará la documentación en la que se trabaja y los respectivos respaldos. Así mismo, se utilizará *Teams* para las reuniones grupales necesarias.

1 Introducción

1.1 Motivación

Con el lanzamiento público de ChatGPT el 30 noviembre de 2022 para interacciones conversacionales naturales con el público general, el desarrollo de la inteligencia artificial (IA) ha sobrepasado cualquier expectativa o pronóstico realizado porque también han ido apareciendo otras inteligencias artificiales generativas que ayudan a crear contenido.

Esta explosión de herramientas de IA ha provocado que el desarrollo y avance de IA haya sido más dinámico que la reglamentación para su diseño y uso. Una de las principales preocupaciones es que el uso de IA que requiere información (inclusive datos personales) para su entrenamiento, ponga en peligro al hombre (propietario de esa información) o a sus derechos fundamentales. En ese contexto han surgido varias iniciativas en el mundo occidental para sugerir mejores prácticas y reglamentaciones, mientras que las empresas no saben que recomendaciones o leyes cumplir para desarrollar IA ética y responsable.

En este escenario, surge la necesidad de poner a disposición de las empresas occidentales una herramienta que consolide todo el cumplimiento de ética de IA de occidente y les permita conocer el nivel de IA responsable se encuentran y reciban recomendaciones necesarias para incrementar su nivel (o madurez en el tema) de responsabilidad en IA.

El objetivo de este TFE es entregar esa herramienta de evaluación a las empresas occidentales para que se aseguren de que están cumpliendo los lineamientos éticos, legales y regulatorios que la normativa actual exige. Es un aporte que se quiere realizar al mundo empresarial que terminará beneficiando a la sociedad porque permitirá que más modelos y productos de IA sean responsables y centrados en atender los derechos humanos.

1.2 Planteamiento del trabajo

Al formar parte de la Maestría Universitaria en Inteligencia Artificial se evidenció que se disponen de muchas herramientas para desarrollo de sistemas de IA, pero ninguna que pueda indicar el nivel de responsabilidad, ética y fiabilidad de un sistema de IA o si está o no cumpliendo con la normativa vigente o si está o no limitando / afectando los derechos fundamentales del hombre.

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

En este contexto surge la idea de proponer una guía para que las empresas de occidente conciban, diseñen, desarrollen e implementen sus sistemas de IA en el marco de la responsabilidad, ética y cumplimiento de las reglamentaciones vigentes. El centro de esa guía será un formulario que señalará el nivel de IA responsable y señalará el camino a recorrer para ir creciendo o madurando en IA responsable.

El formulario contendrá acciones, controles y estructuras que deben tener los sistemas de IA para considerarse responsables y se espera que las empresas lo utilicen para identificar las áreas a mejorar.

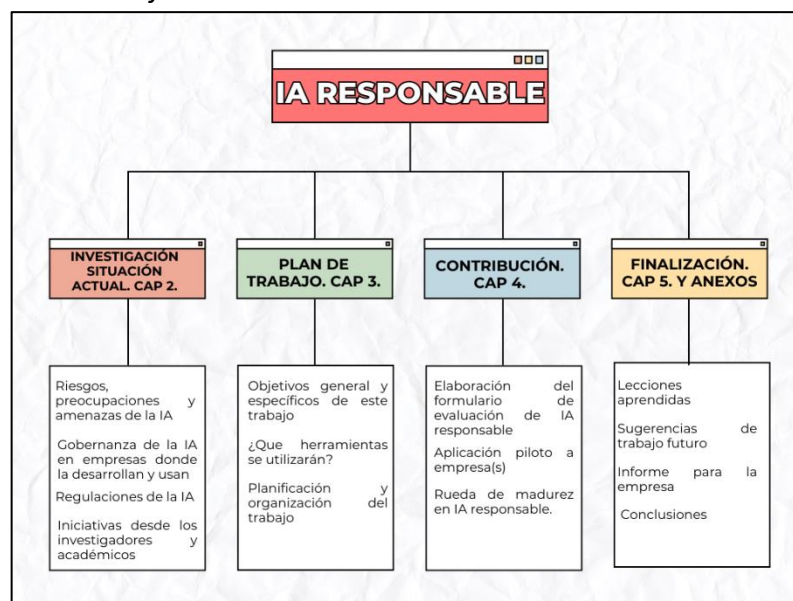
Se planifica realizar un piloto a una o dos empresas de occidente que tengan o usen productos de IA para identificar mejoras que pueden realizarse a la guía.

Finalmente se describirá todo lo aprendido, lo que pudo comprobarse y los trabajos siguientes en el tema.

1.3 Estructura del trabajo

El trabajo a realizarse tiene cuatro etapas principales que pueden visualizarse en la siguiente figura:

Figura 1. Estructura del trabajo a realizarse



Fuente: Elaboración propia

A continuación, una breve descripción del trabajo a realizarse:

- En el capítulo 2 se plasmará la investigación sobre la situación de las IA ética y responsable. Se analizarán los riesgos, preocupaciones y amenazas que representa la IA para el ser humano. También se revisará leyes, reglamentos, iniciativas y mejores prácticas éticas que están surgiendo en occidente.
- En el capítulo 3 se desarrollará el plan de trabajo que regirá la generación de valor que entregará este TFE. Se propondrán los objetivos del trabajo, seleccionando las herramientas a utilizarse y se planificará el trabajo (mediante la definición de la metodología a seguirse).
- En el capítulo 4 se presenta la contribución al conocimiento que aporta este TFE que consiste en la elaboración del formulario de evaluación, la aplicación piloto y la determinación del nivel de madurez en IA responsable según una rueda de madurez propuesta.
- En el capítulo 5 se detallarán los hallazgos, conclusiones de la aplicación del formulario y se propondrá trabajos futuros al respecto.

2 Contexto y estado del arte

El avance de la tecnología ha hecho posible que hoy en día la Inteligencia Artificial sea utilizada en casi todas las áreas en las que se desenvuelve el ser humano, facilitando tareas cotidianas o accediendo rápidamente a la información. Sin embargo, el uso de la Inteligencia Artificial no representa solamente beneficios, sino que también existen preocupaciones crecientes sobre su impacto en la sociedad, la economía y la vida cotidiana, debido al crecimiento exponencial de su uso.

En los últimos años, la Inteligencia Artificial se ha desarrollado exponencialmente debido a la disponibilidad de información histórica de la humanidad, de empresas, gobiernos, organizaciones internacionales inclusive datos personales, redes sociales, internet, además de toda la información generada en tiempo real con el auge del IoT (Internet of Things - Internet de las Cosas). El hecho de que esta tecnología tenga acceso prácticamente a todo el conocimiento escrito, sonoro y gráfico, implica que debe ser manejada de una manera responsable, hecho que ha sido pasado por alto debido al acelerado avance en la adopción de las tecnologías de Inteligencia Artificial que ha progresado a mayor velocidad que las regulaciones, leyes y controles necesarios para encauzar su potencial, procurando respetar el bienestar físico, psicológico, emocional del ser humano y sus derechos.

2.1 Riesgos de la IA

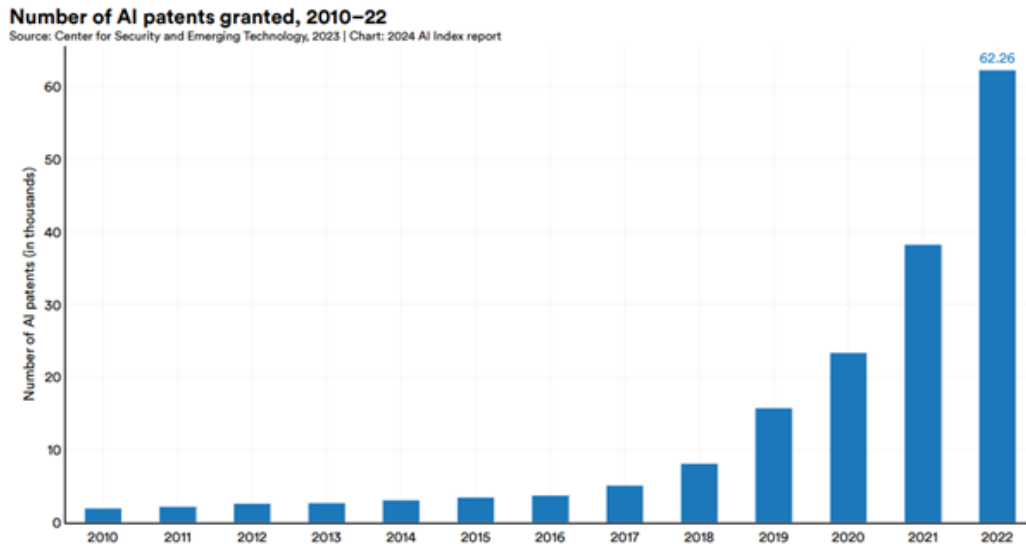
La IA y su avance no solo trae beneficios para la humanidad, sino también varios aspectos que pueden llegar a amenazar a las personas, tales como exposición de los datos personales, posibilidad de ser hackeada o utilizada para fines no pacíficos, entregar información no precisa debido a datos sesgados, con las diferentes aplicaciones de IA, hoy es posible generar contenido real o ficticio para influir en las personas, llegando incluso a manipular su voluntad influyendo para que tome decisiones específicas.

2.1.1 Preocupaciones de la sociedad en la IA

La inteligencia artificial se ha venido integrando en las actividades humanas al punto que el CSET (Center for Security and Emerging Technology – Centro para la seguridad de la tecnología emergente) que se enfoca particularmente en la intersección de la seguridad y la inteligencia

artificial, realiza un registro de patentes relacionadas a la IA y ha visto un crecimiento exponencial desde el año 2019, como señala el AI Index Report 2024 (Maslej et al., 2024).

Figura 2. *Patentes de IA concedidas en el CSET.*



Fuente: Maslej et al. 2024¹.

Esto demuestra un explosivo crecimiento en el uso de IA en los últimos años, ya que día a día se van generando nuevas aplicaciones de IA para diferentes campos de la ciencia, lo que aparte de beneficios está trayendo preocupaciones a la humanidad, específicamente relacionadas con el impacto de la IA en la privacidad, salud, entretenimiento, empleo y la economía a nivel mundial (Maslej et al., 2024).

La principal preocupación es que la IA se desarrolla con datos que son recopilados en todos los ámbitos en los que se desenvuelve el hombre, porque actualmente todas las actividades son registradas ya sea en un computador o en la nube, pudiendo llegar a ser difundidas o adulteradas, afectando moral o físicamente a las personas. Con esta información, es posible determinar inequívocamente su ubicación, sus hábitos, su estado de salud, sus desplazamientos rutinarios, sus preferencias musicales, etc. (Zuboff, 2019), que las empresas

¹ Artificial Intelligence Index Report, Página 38.

desarrolladoras de IA a la fecha explotan para mostrar anuncios publicitarios relacionados a gustos personales y lucran con la información personal.

Existe la preocupación en el imaginario colectivo de que la IA avanzada junto con toda la información que gestiona, pueda “borrar” a la humanidad (Greer, 2022) como en la película “Terminator” (Cameron, 1984), al punto de que el Centro para el Estudio del Riesgo Existencial (Centre for the Study of Existential Risk-CSER) de la Universidad de Cambridge ha incluido a la IA, como una de las seis áreas que pudieran llevar a la extinción humana o al colapso de la civilización y se dedica al estudio y mitigación de sus riesgos (CSER,2024).

Por otra parte, debe considerarse que la IA es programación que aprende y se adapta, lo que no la hace infalible ya que sigue perfeccionándose, esto quiere decir, que también hay el peligro de que las empresas desarrolladoras liberen al mercado tecnología IA que no esté suficientemente probada que generen incidentes que afecten al ser humano, tales como carros autónomos que maten a transeúntes (McGregor & Lam,2018), o que los sistemas de reconocimiento facial sugieran arrestos de personas inocentes (McGregor & Atherton, 2020).

Es necesario también señalar que la IA cada vez es capaz de realizar tareas más desafiantes y complejas, antes solo realizadas por seres humanos, lo que genera interrogantes tales como ¿es la IA un factor de producción diferente?, ¿es la IA un sustituto o complemento del trabajo y/o capital humano? ¿Cuántos empleos eliminará la IA?, ¿Cuántos empleos generará? (Goolsbee, 2019).

Existe la preocupación de que la IA pueda aumentar la desigualdad a nivel mundial, ya que los beneficios de la automatización de tareas y/o trabajos podrían concentrarse en pocas manos dejando rezagados a los países en vías de desarrollo.

2.1.2 Amenazas típicas de la IA

Las amenazas relacionadas con la inteligencia artificial (IA) en 2024 presentan desafíos significativos en diversos ámbitos, desde la desinformación hasta la seguridad. Entre estas amenazas, destacan las alucinaciones, los *deepfakes*, el sesgo, la inyección de mensajes, la manipulación, el desarrollo de armas y el incremento del cibercrimen

Las *alucinaciones* en IA ocurren cuando los modelos generan un resultado que no se corresponde con sus datos de entrenamiento. Esto es especialmente problemático en aplicaciones críticas, como la medicina o el derecho, donde la precisión es fundamental. Por ejemplo, se han documentado casos en los que los sistemas generaron citas o referencias inexistentes, lo que resalta la necesidad de verificar la información producida por la IA y ser conscientes de las limitaciones de los modelos actuales. Estas alucinaciones pueden contribuir a la propagación de desinformación y afectar la credibilidad de los sistemas de IA, limitando su utilidad en situaciones que requieren exactitud y confianza. Además, pueden reflejar sesgos presentes en los datos de entrenamiento, perpetuando prejuicios y potencialmente discriminando a ciertos grupos (Duran, 2024).

Los *deepfakes* representan una amenaza creciente en la era digital. Estas tecnologías avanzadas permiten crear contenido multimedia falso con una precisión impresionante, lo que plantea retos significativos en la verificación de la autenticidad de imágenes y videos que ha llevado inclusive a fraudes, estafas y suplantación de identidad (Peralta, 2024). En sectores como la educación y el entretenimiento, los *deepfakes* también se están explorando para aplicaciones creativas y educativas, aunque su potencial para el mal uso sigue siendo una preocupación global. La habilidad de alterar la realidad digitalmente ha llevado a una reflexión sobre las regulaciones éticas y la necesidad de herramientas de detección más avanzadas (Goggin & Woodward, 2024).

El sesgo en los modelos de inteligencia artificial (IA) es una preocupación significativa. A pesar de los avances, estos sistemas aún exhiben tendencias discriminatorias en sus resultados debido a sesgos inherentes en los datos utilizados para su entrenamiento. Esto puede conducir a decisiones injustas o desiguales en áreas como la contratación laboral, la justicia penal o el acceso a servicios financieros. La falta de diversidad en los conjuntos de datos y en el desarrollo de estos sistemas sigue siendo un obstáculo para la creación de una IA verdaderamente inclusiva y equitativa (López, 2024).

Los modelos de aprendizaje automático se utilizan para mejorar la ciberseguridad, pero también son empleados por los ciberdelincuentes para llevar a cabo ataques sofisticados que van perfeccionándose con cada ataque (Zambrano, 2024).

Si bien es cierto que las redes sociales permiten la conexión entre personas, también se ha comprobado que utilizan IA para priorizar contenido en base a las preferencias de cada usuario, lo que refuerza sus creencias y preferencias a la par que aleja al ser humano de puntos de vista diferentes al suyo (de Paz García ,2021). Los algoritmos están creados para captar la atención del usuario, lo cual en exceso puede ocasionar adicción, dependencia hacia la IA o incluso manipulación con herramientas de IA para influir en elecciones (Cencerrado, 2024).

La irrupción de la IA también ha llegado a la guerra, no solo apoyando la precisión, la estrategia y la eficiencia en los recursos utilizado, llegando inclusive a dejar delegar decisiones letales a máquinas (drones no tripulados), lo que tiene varias consideraciones éticas (Florkin, s.f.) actuales y futuras para evitar que se llegue a presentar una competencia internacional para desarrollar tecnologías militares de IA superiores.

2.2 Gobernanza de la IA -Empresas desarrolladoras del mundo occidental.

Con el objetivo de que la Inteligencia Artificial sirva a la humanidad de una manera ordenada, justa y equitativa, y que siga siendo un apoyo para el avance de la humanidad de una manera responsable, las grandes compañías que la desarrollan han establecido por su cuenta principios y normas que rigen sus respectivas actividades.

2.2.1 Google

El 07 de junio de 2018, mediante su página web, Google publicó sus siete (7) principios con los cuales conciben para que la IA (Pichai, 2018) que se muestran en la Tabla 2:

Tabla 2. Principios éticos para el desarrollo de IA en Google.

Principio	Ejecución
Socialmente beneficiosa	Desarrollar y usar IA en casos donde los beneficios sean mayores a los riesgos o posibles desventajas. Esforzarse por presentar información precisa y de alta calidad respetando las normas culturales, sociales y legales de los países donde operan.

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Principio	Ejecución
Evitar crear o reforzar un sesgo injusto	Considerar que la información puede tener sesgo y tratar de minimizar el sesgo en los datos para evitar que se afecte sensiblemente a las personas en temas como: raza, identidad étnica, género, nacionalidad, ingresos, orientación sexual, capacidad y creencias políticas o religiosas
Estar construida y probada para ser segura	Implementar estrictas prácticas de seguridad para evitar daños no intencionales o accidentales, desarrollando sistemas de IA cautelosos, inclusive probándolos en ambientes controlados y monitoreando su implementación (en algunos casos).
Ser responsables con las personas	Diseñar los sistemas de IA para considerar que la retroalimentación y las tecnologías estarán dirigidas y controladas por humanos
Incorporar principios de diseño de privacidad	Dar oportunidad para que se avise y obtenga consentimiento. Alentar arquitecturas con garantía de privacidad. Planear con transparencia y control sobre el uso de los datos personales
Mantener altos estándares de excelencia científica	Desarrollar con rigor científico, promoviendo la colaboración multidisciplinaria. Compartir el conocimiento de IA para que más personas puedan desarrollar aplicaciones de IA útiles.
Estar disponible para usos que estén de acuerdo con estos principios	Limitar las aplicaciones con potencial de ser dañinas o abusivas. Evaluar propósito y uso general, naturaleza, escala o impacto y como participa Google (integrando o desarrollando aplicaciones a medida) en la tecnología IA. Limitar las aplicaciones con potencial de ser dañinas o abusivas.

Fuente: Elaboración propia.

Además de estos 7 principios, Google no realiza aplicaciones de IA en las siguientes áreas:

- Tecnologías con riesgo potencial de daño.
- Armas o tecnologías con potencial de lesionar directamente a las personas.
- Tecnologías que usan la información para vigilancia que violen los tratados internacionales.
- Tecnologías que infrinjan los derechos humanos.

2.2.2 Apple

Apple fundamenta el uso de la IA en el uso de los datos, combinando la IA generativa con el contexto personal de cada usuario (Cook, 2024), por ello ha cimentado su uso en tres pilares: privacidad, transparencia y seguridad cuya aplicación se muestra en la Tabla 3.

Tabla 3. *Pilares para el uso de la IA en Apple.*

Pilar	Ejecución
Privacidad	Apple prioriza la privacidad del usuario al procesar datos para la IA, utiliza el aprendizaje federado y procesamiento en el propio dispositivo para minimizar la recopilación de datos personales y mantener la información de los usuarios segura.
Transparencia	Apple busca ser transparente sobre cómo se utiliza la IA en sus productos, siempre proporciona información sobre cómo funcionan las funciones de IA que están en sus productos y qué datos del usuario se utilizan.
Seguridad	Apple diseña sus sistemas de IA para que sean seguros y resistentes a los ataques además de implementar medidas de seguridad para proteger los datos de los usuarios y prevenir el uso malicioso de la IA por parte de terceros.

Fuente: Elaboración propia

Además de estas consideraciones, Apple se esfuerza por evitar sesgos en sus algoritmos para garantizar la equidad, la inclusión y que sus productos de IA sean accesibles para todos.

2.2.3 Amazon

En Amazon Web Services (AWS), se promueve el desarrollo responsable de la inteligencia artificial (IA) mediante un enfoque centrado en las personas, priorizando la educación, la ciencia y las necesidades del cliente. Este compromiso se refleja en la integración de la IA

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

responsable en todo el ciclo de vida de los sistemas de inteligencia artificial, basándose en ocho dimensiones definidas por Amazon (Amazon, 2024).

Tabla 4. Dimensiones para una IA responsable según Amazon.

Dimensión	Descripción	Beneficios Clave
Equidad	Garantizar que los sistemas de IA sean justos e inclusivos, identificando y mitigando sesgos en datos y modelos.	Promueve la inclusión, evita desigualdades, fortalece la confianza pública.
Capacidad de explicación	Proporcionar explicaciones claras sobre cómo y por qué un sistema toma decisiones.	Facilita la comprensión, detección de errores y fomenta la transparencia y rendición de cuentas.
Privacidad y seguridad	Asegurar el control de los datos personales y prevenir accesos no autorizados o divulgación indebida.	Protege la información personal, genera confianza y fomenta el uso ético de los datos.
Seguridad	Diseñar sistemas que funcionen de manera ética y segura, evitando resultados perjudiciales o daños involuntarios.	Mitiga riesgos, previene impactos negativos y asegura un funcionamiento seguro.
Capacidad de control	Permitir la supervisión y corrección del comportamiento de los sistemas de IA para alinearlos con valores humanos.	Mitiga riesgos, facilita la resolución de problemas y asegura el beneficio social.
Veracidad y solidez	Garantizar que los sistemas sean confiables y precisos incluso frente a datos inesperados o condiciones cambiantes.	Asegura consistencia, precisión y resistencia a variaciones en el entorno o datos.
Gobernanza	Implementar políticas y procesos responsables en el uso de IA, garantizando cumplimiento legal y ético.	Minimiza riesgos legales y éticos, asegura mejores prácticas y responsabilidad organizacional.
Transparencia	Comunicar información clara sobre el desarrollo, capacidades y limitaciones de los sistemas de IA.	Ayuda a las partes interesadas a tomar decisiones informadas, promueve confianza y ética.

Fuente: Elaboración propia

Dentro de su ecosistema, Amazon ofrece herramientas específicas dentro de Amazon Web Services (AWS) para ayudar a quienes utilicen su tecnología a diseñar, crear y operar sistemas de IA de manera responsable, las cuales se muestran a continuación en la tabla 5.

Tabla 5. *Herramientas de AWS para una IA responsable.*

Herramienta	Utilidad
Amazon SageMaker Clarify	Identifica sesgos en los datos y modelos.
SageMaker Model Monitor	Detecta predicciones inexactas
Amazon Augmented AI	Implementa revisión humana de predicciones de aprendizaje automático
SageMaker Role Manager, Model Cards y Model Dashboard	Fortalecen la gobernanza y el control de los proyectos.

Fuente: Elaboración propia

Amazon recomienda una estrategia de adopción de IA responsable que esté alineada con los objetivos comerciales y un equipo interdisciplinario para mitigar sesgos, mejorar la calidad de los datos y garantizar una gobernanza clara. También señala que la transparencia organizacional debe mantenerse mediante la gobernanza continua y el monitoreo de los sistemas de IA para asegurar que operen de manera justa y segura. Amazon también promueve la formación de líderes en IA a través de iniciativas como el programa de becas AWS AI & ML y colabora con universidades y organizaciones internacionales en la investigación de IA responsable.

2.2.4 Microsoft

Microsoft reconoció que las leyes y regulaciones no habían avanzado al mismo ritmo que los riesgos específicos de la IA ni las necesidades de la sociedad, por lo cual estableció principios, a los que deben regirse los equipos de desarrollo de productos de IA. (Bird, 2024).

Tabla 6. Dimensiones para una IA responsable según Microsoft.

Dimensión	Descripción	Beneficios Clave
Equidad	Garantizar igualdad de acceso y uso, evitando discriminación y resultados estereotipados o degradantes. Promover la inclusión y no perpetuar prejuicios o desigualdades, asignando recursos de manera justa y priorizando a grupos históricamente marginados.	Reducción de desigualdades, mejora en la calidad de vida y mayor aceptación de la IA en diversos sectores.
Confiabilidad y Seguridad	Asegura que los sistemas de IA funcionen de manera fiable bajo las condiciones previstas, registrando errores (falsos positivos y negativos) y adoptando estrategias para reducir su impacto. Implementar monitoreo, retroalimentación y evaluación continuos para optimizar el desempeño de los sistemas.	Mejora en la estabilidad del sistema y confianza del usuario; minimización de riesgos y errores críticos.
Inclusión	Garantizar accesibilidad para todos los usuarios eliminando obstáculos y diseñando tecnologías inclusivas que consideren la diversidad social. Este enfoque busca evitar la exclusión y promover la participación de diferentes comunidades.	Ampliación de la base de usuarios; tecnologías más representativas y efectivas para diferentes comunidades.
Privacidad y Seguridad	Proteger la información personal con mecanismos robustos contra accesos no autorizados, manipulaciones y ataques cibernéticos. Implementar encriptación avanzada, auditorías de seguridad y actualizaciones constantes para adaptarse a nuevas amenazas, garantizando la confidencialidad y la confianza del usuario.	Protección de datos sensibles, fortalecimiento de la confianza del usuario y reducción de riesgos legales.
Transparencia	Asegura que los usuarios entiendan cómo funcionan los sistemas de IA y las decisiones que generan. Diseñar documentación adecuada y evaluaciones continuas para garantizar que el sistema sea	Mejora en la toma de decisiones informadas y fortalecimiento de la confianza entre los desarrolladores y usuarios.

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Dimensión	Descripción	Beneficios Clave
	interpretable y claro para las partes interesadas.	
Responsabilidad	Compromiso con decisiones automatizadas explicables, éticas y alineadas con valores y derechos fundamentales. Los desarrolladores deben evaluar, monitorear y corregir errores o sesgos para evitar impactos negativos en la sociedad, promoviendo la rendición de cuentas.	Fomenta el uso ético de la IA, mejora la aceptación social y fortalece la responsabilidad organizacional.

Fuente: Elaboración propia

Además de estos principios, Microsoft ha establecido un uso ético y responsable del tratamiento de los datos y la IA, lo cual se resume de la siguiente manera:

Tabla 7. *Acciones que implementa Microsoft durante el ciclo de vida de los datos*

Fase del dato	Acciones para manejo ético y responsable
Recopilación	obtención de datos de diversas fuentes para evitar sesgos y garantizar que el sistema sea equitativo
Procesamiento	los datos deben ser limpiados, transformados y anonimizados para proteger la privacidad
Almacenamiento	implementando cifrado y el control de acceso a los datos almacenados.
Uso	los datos se utilizan para entrenar y mejorar los modelos de IA, bajo políticas de privacidad y regulaciones aplicables
Compartición	medidas de seguridad al compartir datos con terceros, asegurando que la información se utilice de manera ética
Retención y eliminación	Los datos no deben almacenarse indefinidamente, y deben eliminarse de manera segura al final de su vida útil.

Fuente: Elaboración propia

2.2.5 Meta

Meta ha desarrollado un enfoque de cinco (5) pilares para garantizar que sus sistemas de IA se diseñen y utilicen de manera responsable, cuidando que siempre se beneficie a las personas y a la sociedad, guiando sus esfuerzos hacia una IA equitativa, justa, responsable, transparente y segura (Meta-Responsible AI 2024). Los pilares son:

Tabla 8 *Pilares con los cuales se diseñan productos de IA en Meta*

Pilar	Descripción
Privacidad y seguridad:	la protección de la información personal es prioridad para Meta durante el desarrollo y uso de IA para garantizar que los datos sean manejados de forma segura y responsable e incluye a todos los empleados de Meta para cumplir con esta responsabilidad.
Equidad e inclusión:	se procura que los productos de Meta traten a todas las personas de manera justa y funcionen igual de bien para todos, sin importar su origen o características demográficas. La equidad implica no solo evitar la discriminación, sino garantizar que las tecnologías beneficien a todos por igual. Meta procura construir conjuntos de datos más inclusivos para evitar los sesgos o prácticas discriminatorias.
Robustez y seguridad	los sistemas de IA se prueban rigurosamente para asegurar que se comporten de manera segura y de acuerdo con su propósito, para mitigar posibles riesgos y garantizar el funcionamiento adecuado de los sistemas que Meta saca al usuario final.
Transparencia y control	Meta promueve la transparencia, brindando a las personas más claridad sobre cómo se recopilan y utilizan sus datos y cómo los sistemas de IA afectan la experiencia de sus personas.
Responsabilidad y gobernanza	Meta construye procesos confiables que responsabilizan a los sistemas de IA y a las decisiones que estos toman. Este pilar implica crear mecanismos para revisar, auditar y asegurar el correcto uso de la tecnología en todos sus productos y servicios.

Fuente: Elaboración propia

2.2.6 Open AI

Open AI busca que sus modelos de IA sean éticos, seguros y útiles, por lo cual ha establecido siete (7) principios que rigen cómo responde la IA a las consultas de los usuarios. El objetivo es mantener una perspectiva justa y objetiva para establecer límites no negociables en el comportamiento de la IA, mantener interacciones éticas, el respeto a la privacidad, el cumplimiento de las leyes y la prevención de respuestas dañinas o inapropiadas (Open AI-Model Specification, 2024).

A continuación, se muestran sus 7 principios:

Tabla 9. Principios éticos de Open IA.

Principios	Explicación
Seguir la Cadena de Mando	La IA debe adherirse a las instrucciones proporcionadas por el desarrollador o usuario, respetando su autoridad al definir el contexto y los objetivos de las interacciones.
Cumplir con las Leyes Aplicables	La IA debe asegurarse de que las respuestas y acciones estén alineadas con las leyes y regulaciones relevantes, evitando consejos o guías que puedan infringir normas legales.
No Proporcionar Riesgos de Información	No se debe divulgar o generar información que pueda causar daño (por ejemplo, instrucciones para actividades ilegales o detalles personales sensibles).
Respetar a los Creadores y Sus Derechos	Se deben respetar los derechos de propiedad intelectual y el trabajo de los creadores de contenido, asegurando que la IA no plagie ni haga un uso indebido de contenido.
Proteger la Privacidad de las Personas	Los datos personales y la información sensible deben ser protegidos. La IA debe evitar compartir o solicitar detalles personales innecesarios.
No Responder con Contenido NSFW (Not Safe)	La IA debe evitar generar o participar en contenido que no sea apto para el trabajo, como temas violentos, adultos o inapropiados.

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Principios	Explicación
For Work – No seguro para el trabajo)	
Excepción: Tareas de Transformación	En casos donde se solicite explícitamente una tarea de transformación, la IA puede participar, pero debe priorizar siempre la seguridad y la adecuación.

Fuente: Elaboración propia

Open AI además ha establecido un conjunto de valores predeterminados con expectativas básicas sobre cómo la IA debe manejar conversaciones y tareas con los usuarios finales. Estos valores deben ser asumidos obligatoriamente por la IA a menos que se especifique lo contrario y son las siguientes:

- Asumir buenas intenciones del usuario o desarrollador, respondiendo de manera constructiva, pero con cautela.
- Hacer preguntas de clarificación cuando sea necesario, si la solicitud del usuario no está clara o podría generar resultados no deseados, la IA debe hacer preguntas para aclarar antes de proceder.
- Ser lo más útil posible sin excederse, para evitar ir más allá de su papel u ofrecer consejos no calificados.
- Apoyar las diferentes necesidades de conversación interactiva y uso programático.
- Asumir un punto de vista objetivo, manteniéndose neutral y basada en hechos en sus respuestas.
- Fomentar la equidad y la amabilidad, y desalentar el odio evitando activamente discursos discriminatorios o de odio.
- No intentar cambiar la opinión de nadie, debe centrarse en proporcionar información sin intentar influir en un punto de vista particular.
- Expresar incertidumbre, si la IA no está segura de una respuesta, debe expresar esa incertidumbre en lugar de generar información incorrecta o engañosa.
- Usar la herramienta adecuada para la tarea solicitada.

- Ser completo pero eficiente, respetando los límites de longitud de las respuestas.

2.2.7 Comparación de los principios éticos de las empresas desarrolladoras de IA

A continuación, se presenta un resumen de los principios éticos por los cuales se rigen las principales compañías del mundo occidental que producen y ofertan servicios de IA, organizados de mayor a menor coincidencia entre las diferentes empresas.

Tabla 10. Comparativa de principios éticos de las distintas empresas de IA.

	Goo gle	Appl e	Ama zon	Micr osof t	Met a	Ope n AI
Privacidad	X	X	X	X	X	X
Seguridad	X	X	X	X	X	X
Transparencia	X	X	X	X	X	X
Inclusión		X	X	X	X	
Evitar Sesgos	X	X		X	X	
Gobernanza			X	X	X	X
Equidad			X	X	X	
Explicabilidad			X			X
Socialmente beneficiosa	X					
Valores predeterminados de interacción						X
Evitar contenido inapropiado						X
Respetar derechos de creadores						X
Cumplir con leyes aplicables						X
Veracidad y solidez			X			
Rendición de cuentas				X		
Controlabilidad			X			
Usos alineados con principios	X					
Excelencia científica	X					
Responsabilidad con las personas	X					

Fuente: Elaboración propia.

2.3 Gobernanza de la IA – Grandes empresas occidentales que usan IA

Las empresas occidentales que usan IA en sus actividades se aseguran de mantener un uso ético, transparente y responsable de la IA. Para ello, es esencial establecer políticas que

protejan los datos, reduzcan sesgos y promuevan beneficios sociales. Varias compañías han desarrollado principios éticos que guían el diseño y aplicación de la IA para alinear sus prácticas con estándares internacionales y garantizar decisiones automatizadas seguras y justas. A continuación, se presentan los pilares éticos definidos por estas empresas.

2.3.1 Deloitte

Deloitte es una empresa que opera en 150 países. Se dedica a entregar servicios profesionales tales como: consultoría, auditoría, impuestos, asesoría financiera, gestión de riesgos y asesoría tecnológica. Trabaja con empresas y organizaciones de casi todas las industrias, debido a su capacidad de resolver problemas complejos, ofrecer soluciones personalizadas a empresas y gobiernos. (Deloitte, s.f.).

Deloitte se ha unido a la iniciativa OdiselA para impulsar el uso responsable y ético de la IA (Deloitte, 2023), lo que asegura que esta compañía está comprometida con el uso responsable de IA dentro y fuera de la organización.

Deloitte también ha emitido sugerencias para sus clientes respecto al uso responsable de IA por parte de sus empleados (Acuña, 2023) que son las siguientes:

- Dado que la IA es de uso libre, si autoriza a sus empleados a utilizarla, debe incluir una política de cómo utilizarla.
- Si la IA será parte del proceso del negocio, se debe evaluar los riesgos de la privacidad.
- Transparentar la IA, especificando las políticas de privacidad y protección de datos, cómo la IA aborda temas éticos (no discriminación, sesgo, equidad) y la forma en la que la IA complementa las tareas.

El compromiso con el uso responsable de IA en Deloitte llega a tal nivel que es una de las empresas que está preocupada de prevenir los riesgos del uso de la IA por parte de sus empleados (Ruiz de Valbuena, 2024).

2.3.2 Repsol

Repsol es una empresa multinacional considerada como una de las principales compañías de energía en el mundo. Opera principalmente con petróleo, gas natural y productos

petroquímicos. Aunque recientemente está haciendo la transición energética hacia las energías renovables. (Repsol, s.f.).

Aunque abiertamente no tiene publicado sus principios éticos en relación con la IA, Repsol ha establecido principios clave para el uso responsable de la inteligencia artificial (IA), incluyendo la creación de un grupo de trabajo multidisciplinar que supervisa su aplicación ética y segura, asegurando el cumplimiento normativo y enfocándose en la seguridad durante su despliegue (Repsol, 2024.).

Primero creó el Centro de Competencia de IA Generativa, para explorar de manera segura y responsable las oportunidades que la IA ofrece a sus negocios y empleados (Repsol, 2023). Como parte de esta exploración, a experimentar con el uso de IA generativa para complementar las tareas internas. Utilizó Copilot Microsoft 365 con 500 empleados (Carrizosa, 2024), para lo cual estableció los pasos obligatorios a seguir en el uso esta herramienta y son los siguientes:

- Solicitar que actúe como una persona con un rol específico.
- Darle un contexto.
- Poner límites (fechas, longitud de la respuesta, estilo de lenguaje).
- Transmitirle pasos detallados para elaborar la tarea.
- Entregarle un formato o modelo previo de lo que se requiere.
- Especificar lo que se desea de resultado con un ejemplo.

Esta iniciativa logró un promedio de optimización de 121 minutos en las tareas, el 84% buscó información, el 83% de los usuarios la utilizaron para resumir documentos y un 71% para encontrar conclusiones.

2.3.3 Unilever

Unilever es una empresa multinacional británica de bienes de consumo, fundada en 1929 tras la fusión de la compañía neerlandesa Margarine Unie y la inglesa Lever Brothers. La empresa se dedica a la producción y comercialización de productos en las categorías de nutrición, higiene y cuidado personal. Opera en más de 190 países y cuenta con una cartera de más de 400 marcas (Unilever, 2023).

Unilever tiene su “Marco de IA responsable” (Hill & Lee, 2024) que rige sus más de 500 sistemas de IA funcionando en todo el mundo (personalización de marketing, optimización de la cadena de suministro y reducción de residuos). Este marco no es público, pero tiene los siguientes principios (Hill & Lee, 2024).:

- Cumplimiento normativo.
- Gestionar los riesgos
- Privacidad de los datos.
- Transparencia.
- Uso responsable, ético y justo de los datos.
- Evitar el sesgo en la IA.

2.3.4 Walmart

Walmart Inc. es una corporación multinacional estadounidense que opera cadenas de grandes almacenes de descuento y clubes de almacenes. Fundada por Sam Walton en 1962, la empresa es la mayor corporación pública del mundo según la lista Fortune Global 500 de 2020. La compañía cuenta con más de 10,600 tiendas en más de 20 países (Statista, 2023).

Walmart usa IA en sus operaciones de la cadena de suministro y en chatbots internos para manuales de trabajo (Walmart, 2024). Es importante señalar que Walmart desarrolló su propio modelo de lenguaje de aprendizaje automático llamado Wallaby que próximamente estará operativo.

Walmart ha establecido seis principios fundamentales para el uso de la IA (Walmart, 2023) que son:

- Transparencia: Ayudar a clientes, miembros y asociados a comprender cómo se utilizan sus datos y la tecnología, incluyendo la IA, y los objetivos detrás de su uso.
- Seguridad: Implementar medidas avanzadas para proteger los datos, revisando continuamente las prácticas de seguridad para mitigar amenazas emergentes.
- Privacidad: Evaluar los sistemas de IA para asegurar que la información sensible o confidencial se utilice de manera que proteja la privacidad.
- Equidad: Analizar las herramientas de IA para identificar y mitigar sesgos que puedan afectar a clientes, miembros y asociados, comprometiéndose a evaluaciones regulares.

- Responsabilidad: Asegurar que la IA esté gestionada por personas, asumiendo la responsabilidad por su impacto.
- Enfoque en el Cliente: Medir la satisfacción de los clientes con las interacciones de IA y considerar sus opiniones, revisando continuamente las herramientas de IA para garantizar su precisión, relevancia y utilidad.

2.3.5 Salesforce

Salesforce, es una empresa líder en soluciones de gestión de relaciones con clientes (CRM) basadas en la nube. La compañía presencia en más de 30 países y oficinas en más de 100 ciudades a nivel mundial (Salesforce Developers, s.f.). Salesforce utiliza inteligencia artificial (IA) en diversos aspectos de su plataforma para optimizar la gestión de relaciones con clientes y mejorar la eficiencia operativa de las empresas.

Salesforce ha establecido un conjunto de Principios de IA de Confianza para guiar el desarrollo y uso ético de la inteligencia artificial en sus productos y servicios (Salesforce, s.f.)

- Precisión: Asegurar que los sistemas de IA proporcionen resultados precisos y verificables, permitiendo a los usuarios validar las respuestas y comprender las fuentes de información.
- Transparencia: Proporcionar claridad sobre cómo funcionan los sistemas de IA, incluyendo la explicación de los procesos y decisiones que toman, para fomentar la confianza de los usuarios.
- Responsabilidad: Asumir la responsabilidad de los impactos de la IA, trabajando para mitigar riesgos y asegurando que su uso esté alineado con los valores éticos y legales.
- Equidad: Desarrollar sistemas de IA que sean justos e imparciales, evitando sesgos y discriminación, y promoviendo la igualdad en su aplicación.
- Privacidad y Seguridad: Proteger los datos personales y garantizar que la información utilizada por los sistemas de IA esté segura, cumpliendo con las normativas de privacidad vigentes.

2.3.6 Telefónica

Telefónica, S.A. es una multinacional del sector de las telecomunicaciones, incluyendo telefonía fija y móvil, banda ancha y televisión por suscripción. Opera en Europa y América Latina (Telefónica, s.f.).

Telefónica utiliza IA para atención al cliente (asistentes virtuales), optimización de redes, ciberseguridad, análisis predictivo del consumo y comportamiento de clientes, automatización de tareas administrativas para lo cual estableció principios para su uso responsable (Telefónica, 2021):

- IA justa: Minimizar el sesgo en los datos de formación para evitar impactos discriminatorios.
- IA transparente y explicable: Asegurar la comprensión del uso de datos y la lógica detrás de las decisiones de IA.
- IA centrada en las personas: Generar beneficios tangibles sin afectar negativamente los derechos humanos.
- Privacidad y seguridad por diseño: Incluir privacidad y seguridad como partes inherentes del ciclo de vida de la IA.
- Colaboración con terceros: Verificar que socios y proveedores cumplan con estos principios.

2.4 Regulación en IA en el mundo occidental

Los esfuerzos para desarrollar la inteligencia artificial se han extendido a nivel global, impulsando su avance de manera exponencial. Este rápido crecimiento ha generado preocupación tanto en organismos internacionales como en gobiernos locales y regionales, lo que ha llevado a la creación de propuestas destinadas a regular su alcance, con el objetivo de proteger la libertad y los derechos del ser humano.

Las iniciativas que se consideran relevantes en este trabajo son las siguientes:

- Directrices éticas para una IA fiable – establecidas por el comité ético de la Unión Europea.

- Reglamento general de protección de datos de la Unión Europea.
- Reglamento de inteligencia artificial de la Unión Europea.
- La orden presidencial de Joe Biden como presidente de Estados Unidos para legislar la IA en dicho país.
- La Asociación de Gestión del Dato (Data Management Association).

Todas estas propuestas se analizan en este capítulo.

2.4.1 Directrices éticas para una IA fiable

En 2018, la Comisión Europea creó un grupo de 52 expertos de alto nivel sobre inteligencia artificial y encargó dos documentos: las directrices éticas sobre la IA y las recomendaciones sobre política e inversión en IA (Smuha, et. al 2019).

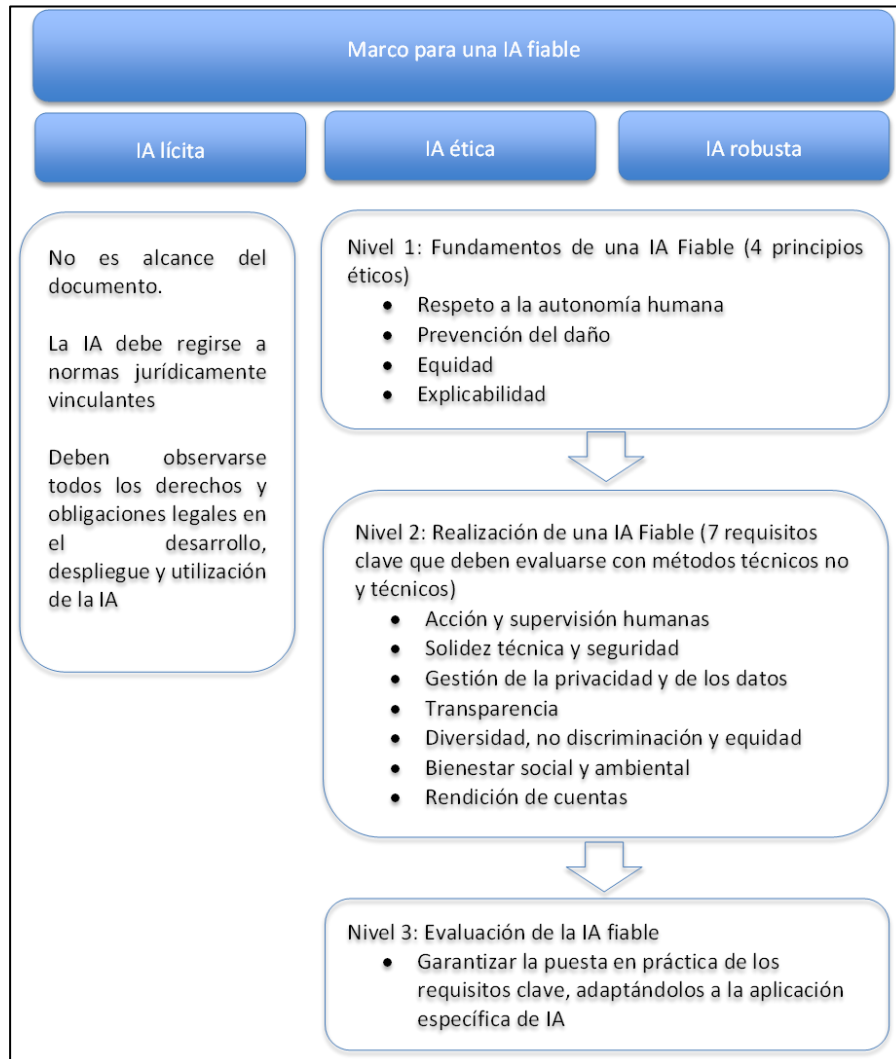
El documento “Directrices éticas para una IA fiable” fue publicado inicialmente como borrador en diciembre 2018, con el cual se realizó una consulta pública y los resultados se publicaron en 2019.

El objetivo de las directrices es “promover una inteligencia artificial fiable” (Smuha, et. al 2019. Pág.7) teniendo como su razón de ser que “los sistemas de IA se centren en las personas y se fundamenten en el compromiso de utilizarlos al servicio de la humanidad y del bien común, con el objetivo de mejorar el bienestar y la libertad de los seres humanos” (Smuha, 2019. et. al Pág 7). El documento propone un marco para una IA fiable apoyado en que la IA sea lícita, ética y robusta.

El aspecto lícito no es alcance de la propuesta de la Unión Europea, pero señala que la IA debe cumplir con todas las leyes y reglamentos locales, regionales e internacionales o tratados internacionales que apliquen o estén vinculados al desarrollo, despliegue y utilización de sistemas de IA, inclusive debe considerar lo que no debe hacerse según las diferentes regulaciones.

Para que la IA sea ética y robusta sugieren que sea vista en 3 niveles: fundamentos, realización y evaluación como se muestra en la figura a continuación:

Figura 3. IA fiable y sus niveles



Fuente: Elaboración propia.

El objetivo del nivel 1 (fundamentos de una IA fiable), es garantizar que cualquier persona pueda prosperar en un mundo basado en inteligencia artificial, respetando sus derechos fundamentales consagrados en leyes y tratados europeos dentro del marco de la democracia y el estado de derecho, así como principios éticos básicos. A continuación, se muestran los derechos fundamentales que debe cumplir una IA fiable.

Tabla 11. *Derechos fundamentales en el contexto de una IA fiable.*

Derecho fundamental	En el contexto de la IA
Respeto a la dignidad humana	Las personas deben ser tratadas como sujetos morales, no como objetos que se pueden filtrar, ordenar, condicionar o manipular, por lo tanto, la IA debe desarrollarse para que esté al servicio de la integridad física y mental del ser humano.
Libertad individual	Exigir la mitigación de la coerción ilegítima directa o indirecta, amenazas a la salud mental, vigilancia injustificada, engaño y manipulación injusta.
Respeto a la democracia, la justicia y el estado de derecho	La IA no debe socavar procesos democráticos o deliberaciones humanas, ni los sistemas de votación.
Igualdad, no discriminación y solidaridad	El resultado de un sistema de IA no debe generar resultados injustamente sesgados
Derechos de los ciudadanos	Deben protegerse los derechos de los ciudadanos para evitar que les afecten negativamente aplicaciones o sistema de IA

Fuente: Elaboración propia.

Para garantizar una IA fiable, es fundamental tener en cuenta principios éticos que, aunque ya están contemplados en diversas normativas y leyes relacionadas con la IA, deben ser observados en todo momento por los profesionales que la desarrollan o utilizan. Estos principios se presentan a continuación en la Tabla 12.

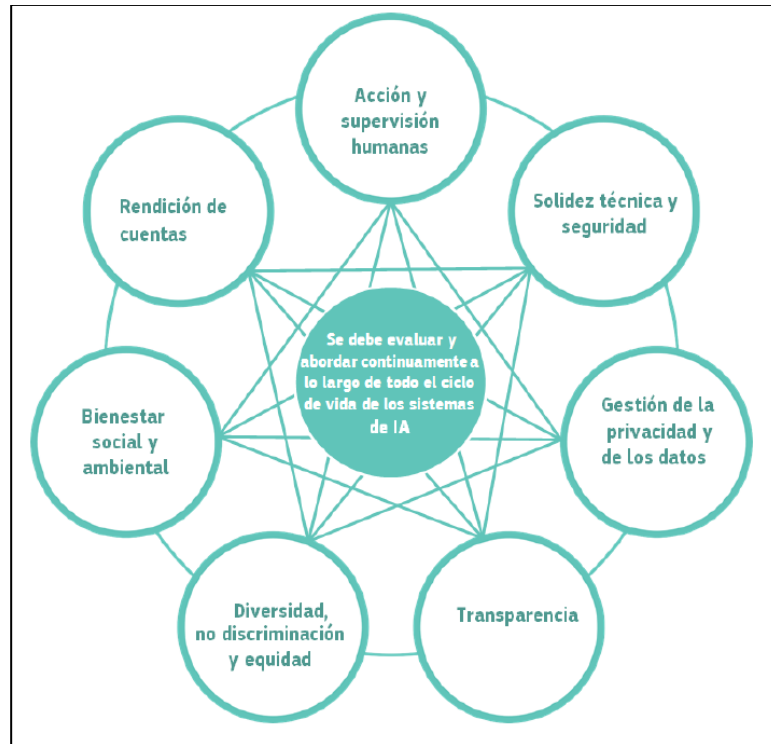
Tabla 12. Principios éticos en el contexto de una IA fiable.

Principios éticos	En el contexto de la IA
Respeto a la autonomía humana	<p>La IA no deben subordinar, coaccionar, engañar, manipular, condicionar o dirigir a los seres humanos injustificadamente.</p> <p>La IA debe aumentar, complementar y potenciar las aptitudes cognitivas, sociales y culturales de las personas.</p>
Prevención del daño	<p>La IA no debe provocar daños, agravarlos o perjudicar a los seres humanos física o mentalmente o afectar al entorno natural y/o seres vivos.</p> <p>Los sistemas de IA deben ser seguros para evitar que puedan utilizarse con fines malintencionados.</p>
Equidad	<p>Evitar los sesgos injustos, discriminación o estigmatización garantizando una distribución igualitaria de beneficios y costos de IA.</p> <p>Capacidad de oponerse a las decisiones tomadas por sistemas de IA o por las personas que los manejan.</p> <p>Respetar la proporción entre los objetivos y los medios para conseguirlos.</p>
Explicabilidad	<p>Comunicar abiertamente las capacidades y finalidades de los sistemas de IA.</p> <p>Todos los procesos deben ser transparentes inclusive en los algoritmos “caja negra”.</p>

Fuente: Elaboración propia.

En el nivel 2 (realización de una IA fiable), las directrices éticas establecen siete (7) requisitos que deben ser evaluados con métodos técnicos y no técnicos a lo largo del ciclo de vida de un sistema de IA, los cuales se muestran a continuación:

Figura 4. Requisitos que deben cumplir un sistema de IA para considerarse una IA fiable.



Fuente: Smuha et al., 2020²

Para el nivel 3 (evaluación de una IA fiable), las directrices éticas proponen una lista de evaluación para los sistemas de IA que interactúan directamente con los usuarios. Su objetivo es ser utilizada como guía por los desarrolladores y responsables concientizando sobre la implementación de controles, procesos y estructuras que aseguren el despliegue de una IA fiable

La evaluación debe indicar que tan fiable es el sistema de IA tomando en cuenta si cumple con todas las leyes y reglamentos aplicables (IA lícita), si garantiza el respeto de los principios y valores éticos (IA ética) y si tanto técnica como socialmente sea robusta (IA robusta) para que no se generen daños incidentales a pesar de tener buenas intenciones.

La lista completa para evaluar una IA fiable se encuentra en el Anexo A de este documento.

² Directrices éticas para una IA fiable. Página 20.

2.4.2 Reglamento General de protección de datos de la Unión Europea

Con el uso masivo de internet y la aparición de nuevas tecnologías, en la Unión Europea se incrementó la preocupación por el tratamiento masivo de los datos, el perfilado y la vigilancia de las personas físicas. Por este motivo, el Parlamento Europeo y del Consejo promulgaron la Directiva 95/46/CE, una legislación destinada a establecer derechos de los interesados (titulares de los datos personales), reglas para la transferencia internacional de datos personales y disposiciones para los responsables de la protección de datos. Sin embargo, esta directiva quedó obsoleta debido a la falta de uniformidad en su adopción por parte de los Estados miembros.

En 2016, se promulgó el documento denominado *“Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos.”* , conocido ampliamente como Reglamento General de Protección de Datos de la Unión Europea – RGPD (Parlamento Europeo y Consejo de la Unión Europea (2016).

El RGPD se creó con el objetivo de establecer normas para el tratamiento de datos personales y su libre circulación dentro o fuera del territorio de la Unión Europea siempre que dicha información pertenezca a personas dentro de dicho territorio. Para el RGPD, dato personal es toda información sobre una persona física que permita identificarla como su nombre, su número de identificación, o *“varios elementos propios de la identidad física, fisiológica, genética, psíquica, económica, cultural o social de dicha persona”* (Parlamento Europeo y Consejo de la Unión Europea (2016); Pág., 3).

Para el tratamiento de datos, el RGPD establece dos roles adicionales: el responsable (quien decide cómo y por qué se tratan los datos) y el encargado (quien ejecuta lo señalado por el responsable). Estos roles pueden ser desempeñados por personas naturales o jurídicas.

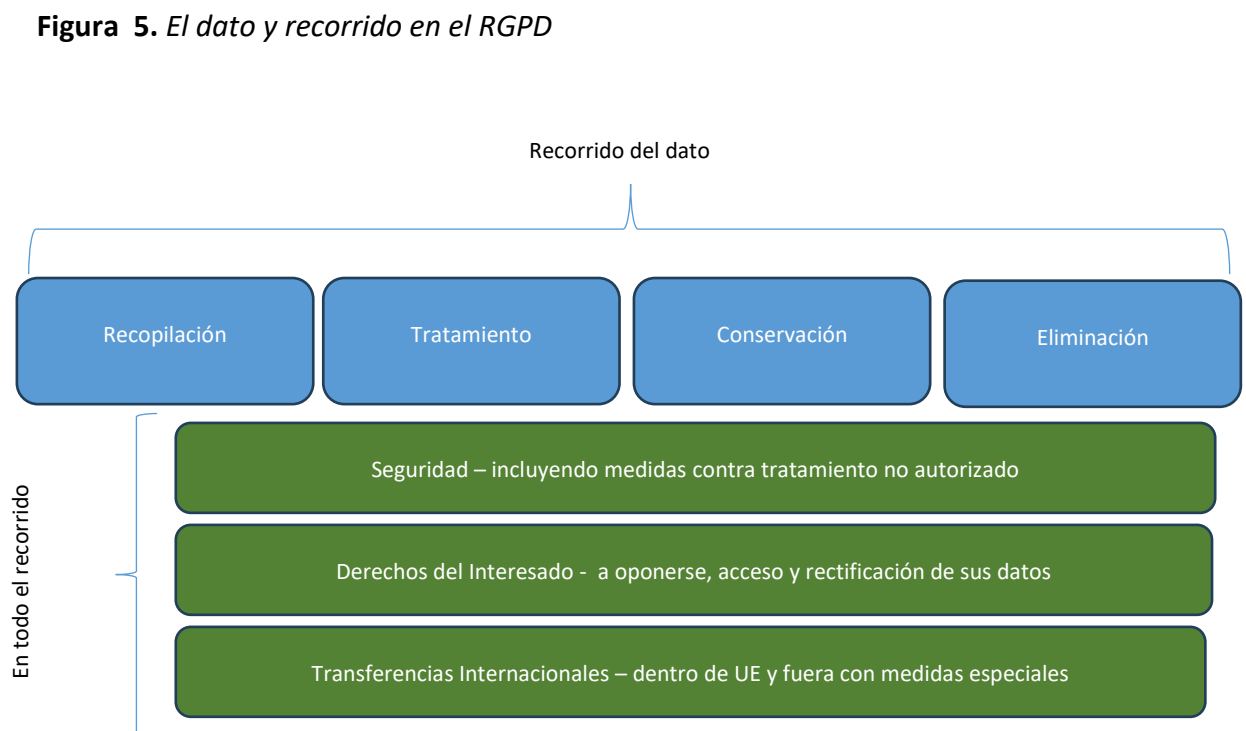
El RGPD establece directrices detalladas para la gestión de los datos desde su recopilación hasta su eliminación. Los datos deben:

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

- Obtenerse (Recopilación) de forma lícita, legal y transparente, informando su propósito y recopilar solo lo estrictamente necesario.
- Tratar (Tratamiento) solo lo estrictamente necesario, implementando medidas de seguridad como encriptaciones.
- Almacenarse (Conservación), corregirse y mantenerse actualizados solo el tiempo necesario considerando los fines para los que se recopilaron y trataron.
- Olvidarse (Eliminación) , cuando ya no son necesarios o cuando lo solicita el interesado

Durante todo el ciclo de vida del dato, el RGPD obliga a que transversalmente se implemente seguridad (para mantenerse íntegros y confidenciales), se observen los derechos de los interesados y se regulan las transferencias internacionales (dentro y fuera de la Unión Europea).

Una explicación gráfica del recorrido del dato según el RGPD puede visualizarse en la figura a continuación:



Fuente: Elaboración propia.

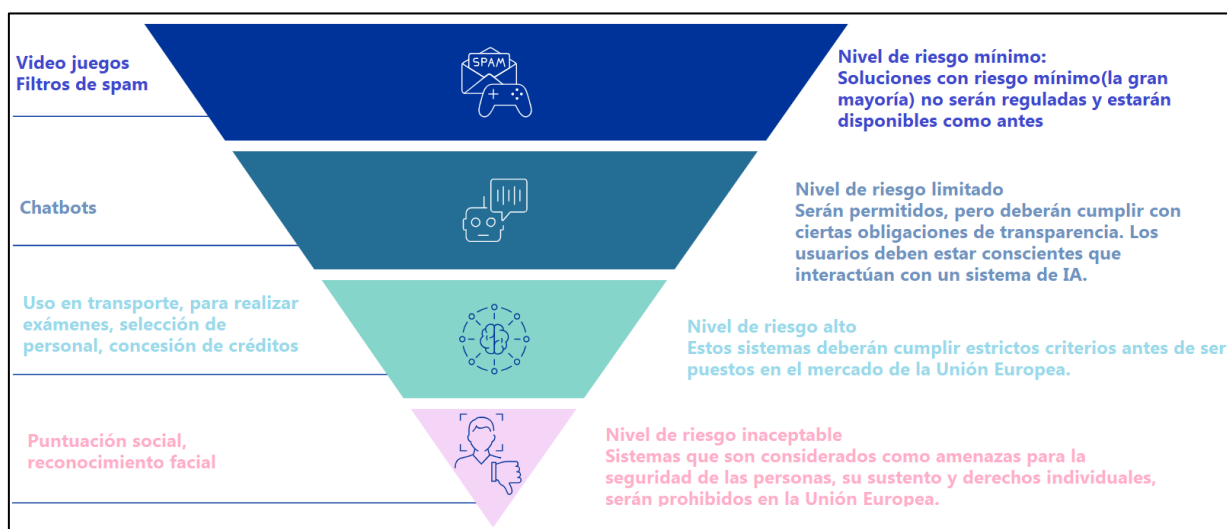
2.4.3 Reglamento Europeo sobre Inteligencia Artificial

El documento “Reglamento (UE) 2023/2088 del Parlamento Europeo y del Consejo relativo a la Inteligencia Artificial” es conocido comúnmente como la Ley de Inteligencia Artificial de la Unión Europea, el “Act AI” o el Reglamento Europeo sobre Inteligencia Artificial, fue aprobada finalmente el 21 de mayo de 2024, entró en vigor el 01 de agosto de 2024 y será de aplicación obligatoria para todas las empresas a partir del 02 de agosto de 2026 (Consilium, 2024).

El objetivo del Reglamento es garantizar que los sistemas de IA se desarrollen y utilicen de manera responsable dentro de la Unión Europea, por lo cual se imponen obligaciones a los proveedores de IA y a los responsables del despliegue de tecnologías de IA (Consilium, 2024).

El documento clasifica a los sistemas de IA en cuatro categorías según su riesgo: mínimo, limitado, alto e inaceptable, como se visualiza en la figura a continuación:

Figura 6. Niveles de riesgo en las aplicaciones de IA según el Reglamento de IA



Fuente: Consilium, 2024³ y elaboración propia.

Los sistemas que este reglamento considera como de riesgo inaceptable son los que pueden influir de manera considerada en los derechos fundamentales de las personas como manipulación subliminal, vigilancia masiva indiscriminada y están prohibidos de desarrollarse

³ Inteligencia artificial: política de la UE. Página 2.

o implementarse en ninguna circunstancia en territorio de la Unión Europea (Unión Europea, 2024).

Los sistemas de alto riesgo que influyen en sectores estratégicos como salud, seguridad pública, seguridad social, empleo deben cumplir con estrictas obligaciones para los desarrolladores y proveedores de este tipo de sistemas. Las obligaciones previstas tienen que cumplirse durante todo el ciclo de vida de estos sistemas de IA y están relacionadas con gestión de riesgos, gobernanza de datos, documentación técnica detallada, registro de estos sistemas, supervisión humana, seguridad, robustez, precisión y explicabilidad.

Las obligaciones para proveedores y usuarios de sistemas de IA según este reglamento se resumen en la tabla a continuación:

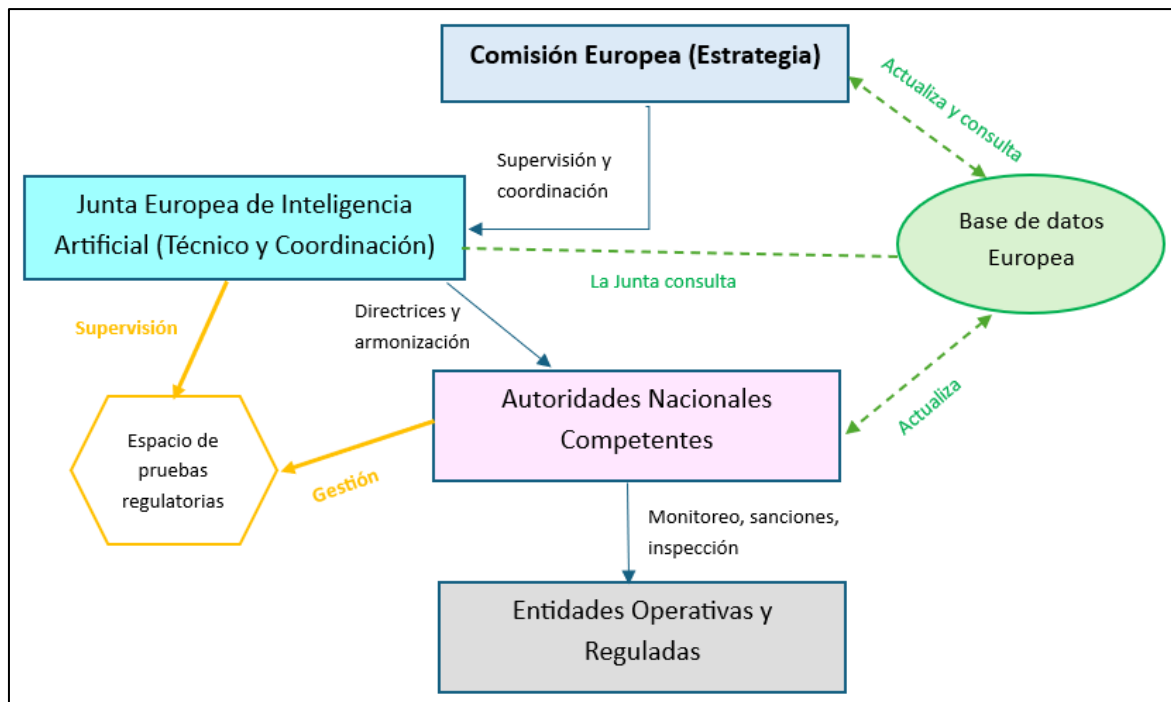
Tabla 13. *Obligaciones de proveedores y usuarios de servicios de IA según el nivel de riesgo del sistema*

Nivel de riesgo	Obligaciones de proveedores	Obligaciones de Usuarios
Bajo	No se aplican requisitos adicionales	No se aplican requisitos adicionales
Limitado	Garantizar transparencia mínima. Informar al usuario que interactúa con IA. Especificar que el contenido fue creado por IA. Minimizar que los usuarios sean manipulados o engañados. Diseño simple de fácil comprensión.	No utilizar para manipular o engañar a otras personas No evadir los principios de transparencia del proveedor Usar estos sistemas de manera informada.
Alto	Realizar evaluación de riesgos. Documentar el sistema. Garantizar gobernanza de datos. Proporcionar supervisión humana. Registrar en la base pública de IA y Garantizar seguridad robusta.	Cumplir las instrucciones del proveedor, supervisión humana, mantener registros del uso, informar de incidentes a las autoridades.
Inaceptable	Prohibido desarrollar o implementar este tipo de sistemas.	Prohibido utilizar este tipo de sistemas.

Fuente: Elaboración propia.

El reglamento de IA también establece una estructura jerárquica de organismos que se encargan de regular y supervisar el cumplimiento de lo especificado en el mencionado documento. En el gráfico a continuación se visualizan los organismos y sus responsabilidades:

Figura 7. Gobernanza de la IA en la Unión Europea según el Reglamento de IA



Fuente: Elaboración propia.

- La Comisión Europea asume el rol de supervisor estratégico de la aplicación del reglamento ya que debe garantizar su implementación uniforme en todo el territorio. Coordina con la Junta Europea emisión de directrices. Actualiza y consulta la base de datos europea de sistemas de alto riesgo.
- La Junta Europea de Inteligencia Artificial, es un organismo nuevo creado con este reglamento, se encarga de coordinar entre los estados miembros. Emite directrices y recomendaciones para armonizar la aplicación del reglamento en todos los estados.
- Autoridades Nacionales Competentes, son organismos en cada estado miembro que garantiza el cumplimiento del reglamento a nivel local. Se encargan del monitoreo, sanciones e inspección a las entidades operativas y reguladas.

- Entidades operativas y reguladas son los proveedores y usuarios de los sistemas de IA en la Unión Europea.

El reglamento señala que se crea una “Base de Datos Europea” donde se registra la información detallada de los sistemas de IA de alto riesgo, incluyendo registros, actualizaciones y resultados de auditorías. Esta base es alimentada por la Comisión Europea y por las Autoridades Nacionales Competentes para que pueda ser consultada por la Junta Europea.

El reglamento también implementa los “Espacios de pruebas regulatorias” o Sandboxes que son entornos controlados para probar sistemas de IA en condiciones reales, para fomentar la innovación responsable garantizando el cumplimiento del reglamento. Son gestionados por las Autoridades Nacionales Competentes y supervisados por la Junta Europea.

Para asegurar la adopción y cumplimiento de este reglamento se han establecido multas como se detalla a continuación:

Tabla 14. *Infracciones y multas máximas por incumplimientos*

Tipo de Infracción	Ejemplo de Infracción	Multa Máxima Prevista
Incumplimiento de prohibiciones específicas (sistemas de riesgo inaceptable)	Uso de sistemas de manipulación subliminal o vigilancia masiva prohibida	Hasta 30 millones de euros o el 6% de la facturación global anual, lo que sea mayor
Incumplimiento de requisitos aplicables a sistemas de alto riesgo	Falta de gobernanza de datos o supervisión humana en sistemas de alto riesgo	Hasta 20 millones de euros o el 4% de la facturación global anual, lo que sea mayor
Falta de cooperación con las autoridades nacionales competentes o la Comisión Europea	Negativa a permitir auditorías o inspecciones de las autoridades competentes	Hasta 10 millones de euros o el 2% de la facturación global anual, lo que sea mayor
Suministro de información incorrecta, incompleta o engañosa	Proveer documentación técnica incompleta sobre sistemas de IA	Hasta 10 millones de euros o el 2% de la facturación global anual, lo que sea mayor

Fuente: Elaboración propia.

2.4.4 Orden presidencial de Estados Unidos de América para legislar la IA.

El nombre original del documento es “Order on Safe, Secure, and Trustworthy Artificial Intelligence” cuya traducción sería Orden Ejecutiva sobre el Desarrollo y Uso Seguro, Confiable y Responsable de la Inteligencia Artificial, conocida como orden presidencial sobre IA, fue emitida el 30 de octubre de 2023, por Joe Biden, presidente de los Estados Unidos de América. El propósito de este decreto es tratar los peligros que podrían presentarse con los progresos

acelerados en IA, poniendo en marcha acciones preventivas y de control que aseguren la seguridad nacional (Casa Blanca, 2023).

El decreto pretende garantizar la seguridad y confiabilidad en los sistemas de IA en territorio de Estados Unidos, proteger los derechos civiles, evitar sesgos o discriminación, fortalecer la seguridad nacional, mitigar riesgos en biotecnología, ciberseguridad o infraestructuras críticas y promover la innovación responsable de la IA en los Estados Unidos.

Según la Casa Blanca (2023), esta orden será implementada a través de acciones coordinadas entre varios organismos gubernamentales (agencias y departamentos federales) principalmente por:

- Departamento de comercio (NIST-National Institute of Standards and Technology) para establecer estándares y directrices.
- Departamento de seguridad nacional (DHS- Department of Homeland Security): para identificar y abordar amenazas de seguridad relacionadas con IA y colaborar con el sector privado para proteger infraestructuras críticas.
- Oficina de gestión y presupuesto (OMB- Office of Management and Budget) emitirá guías y políticas para la gobernanza y uso responsable de la IA en las agencias federales.

Tabla 15. *Acciones de la Orden presidencial, sus plazos y responsables.*

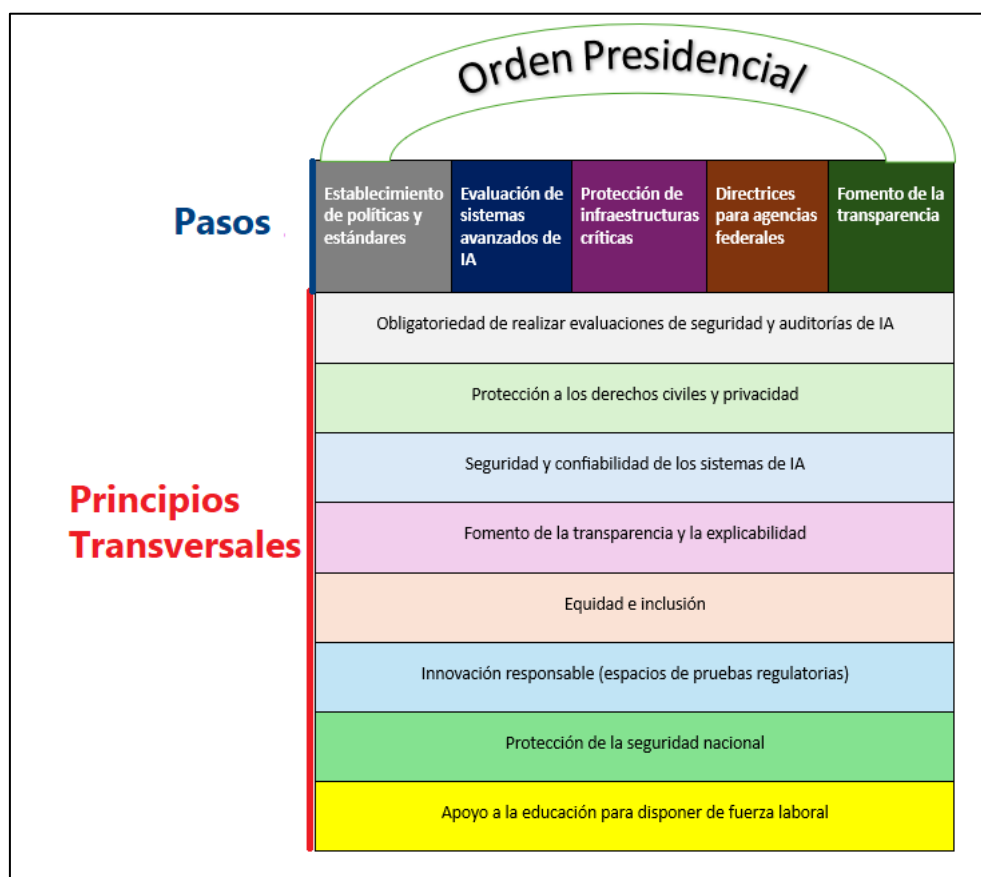
Paso de la Orden	Plazo para completar	Agencia Responsable
Establecimiento de políticas y estándares	240 días	NIST (Departamento de Comercio)
Evaluación de sistemas avanzados de IA	90 días	Proveedores de IA, con supervisión del gobierno
Protección de infraestructuras críticas	1 año	Departamento de Seguridad Nacional (DHS)
Directrices para agencias federales	9 meses	Oficina de Gestión y Presupuesto (OMB)
Fomento de la transparencia	1 año	Varias agencias federales

Fuente: Elaboración propia.

Para asegurarse que las agencias cumplan con las directrices establecidas en la orden presidencial se creó el rol especializado de Chief AI Officer (CAIO)- Oficial jefe de IA cuya misión es supervisar y coordinar en cada agencia federal las actividades relacionadas con IA.

Dentro de estos pasos de implementación de la orden presidencial se incluyen principios transversales que deben considerarse durante la implementación (Guarneras, 2024) y se muestra todo el conjunto a continuación:

Figura 8. Orden presidencial, pasos y principios transversales



Fuente: Elaboración propia.

Tomando en cuenta que ya ha pasado más 1 año desde la expedición de esta orden, la fecha máxima para cumplimiento era el 29 de octubre de 2024, sin embargo, no se pudo conseguir un reporte público de que una agencia ya haya terminado con su labor en el cumplimiento de esta orden presidencial.

Es importante señalar que, el 23 de enero de 2025 el presidente Donald Trump firmó una Orden Ejecutiva derogando las políticas de IA establecidas por Joe Biden en la Orden Presidencial del 30 de octubre de 2023 por considerarla que limita el desarrollo y despliegue de IA. La nueva orden ejecutiva se enfocará en promover y consolidar el liderazgo de Estados

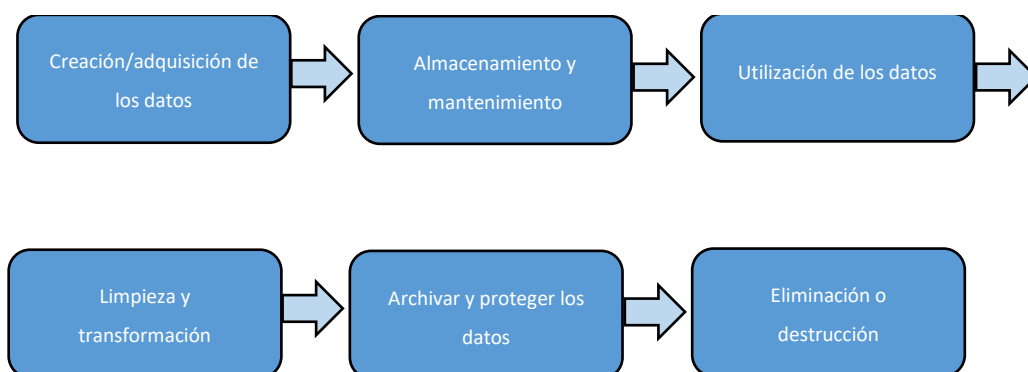
Unidos en IA a nivel mundial sin dejar a un lado su seguridad nacional. A la fecha de finalización de este trabajo, la Casa Blanca aún no define las siguientes acciones (Casa Blanca, 2025)

2.4.5 Data Management Association - DAMA

DAMA (Data Management Association International) es una organización internacional, cuya sede principal se encuentra en Estados Unidos, que promueve las mejores prácticas en la gestión de datos. Su objetivo es ayudar a las empresas y profesionales a gestionar sus datos de manera efectiva y estratégica. DAMA propone considerar a los datos como activos estratégicos de la organización, que su gestión debe estar alineada con los objetivos del negocio y que la gobernanza de los datos es necesaria para garantizar su calidad.

DAMA propone un marco de referencia estándar para la gestión de los datos que lo recoge en el DAMA-DMBOK (Data Management Body of Knowledge) o *DM Handbook*, cuya versión vigente es la 2ª edición. En este documento se recoge el ciclo de vida del dato que, según DAMA, que va desde la adquisición de los datos, pasando por el almacenamiento, utilización, limpieza, archivo hasta su eliminación, como se muestra la figura a continuación:

Figura 9. *Ciclo de vida del dato según el DMBOK2*



Fuente: Elaboración propia.

DAMA en su DMBOK2 propone 11 áreas de conocimiento (dimensiones) para la gestión de los datos, las cuales se muestran en la figura a continuación:

Figura 10. Áreas de conocimiento (dimensiones) que propone el DMBOK2



Fuente: Elaboración propia.

1. Gobernanza de datos: es transversal a la organización, estableciendo políticas y procesos para que los datos sean utilizados de manera eficiente y alineados con los objetivos estratégicos del negocio. Todas las áreas están obligadas a seguir estas directrices. Garantiza la calidad, seguridad, integridad y disponibilidad de los datos, además del cumplimiento de regulaciones y normativas legales relacionadas con su gestión. También define roles y responsabilidades dentro de una estructura organizada, como el Chief Data Officer (CDO), el consejo de gobernanza de datos y los administradores de datos.
2. Gestión de arquitectura de datos: Se enfoca en el diseño, estructuración y documentación de la arquitectura de datos para garantizar que los datos estén alineados con las necesidades del negocio y los sistemas tecnológicos de la organización.
3. Gestión de modelado e información de datos: Incluye la creación, documentación y mantenimiento de modelos de datos conceptuales, lógicos y físicos para estructurar y organizar los datos de manera efectiva.

4. Gestión de almacenamiento y operaciones de datos: gestiona las tecnologías y los procesos necesarios para almacenar, procesar y administrar los datos de manera segura y eficiente en toda su vida útil.
5. Gestión de seguridad de datos: asegura la protección de los datos frente a accesos no autorizados, pérdida o alteración, ya que establece controles de acceso, políticas de encriptación y medidas de seguridad.
6. Gestión de calidad de datos: garantiza que los datos sean precisos, completos, consistentes y confiables, a través de procesos de validación, limpieza y monitoreo continuo.
7. Gestión de metadatos: se ocupa del manejo de metadatos, proporcionando contexto, significado y trazabilidad a los datos para mejorar su gestión y comprensión
8. Gestión de datos maestros y de referencia: garantiza la estandarización y consistencia de los datos clave en la organización, como datos maestros (clientes, productos) y datos de referencia (códigos, categorías).
9. Gestión de datos transaccionales e históricos: maneja a los datos operacionales que se generan en tiempo real y los datos históricos que se archivan para análisis y cumplimiento de normativas.
10. Gestión de integración e interoperabilidad de datos: diseña y coordina los procesos para integrar y compartir datos entre sistemas, asegurando la interoperabilidad y consistencia en el flujo de información.
11. Gestión de documentos y contenido: consiste en organizar, almacenar y gestionar información no estructurada, como documentos, imágenes, videos y otros tipos de contenido digital.

Las prácticas señaladas en el DMBOK2 orientan a las empresas en la alineación de sus estrategias de datos con los objetivos del negocio. Además, permiten garantizar el cumplimiento de las normativas relacionadas con la gestión de datos, transformándolos en un activo estratégico clave para la organización.

2.5 Iniciativas de regulación de la IA desde la academia

Las universidades y centros de investigación han aportados lineamientos para gestionar la IA de una manera responsable y ética, sin limitar su desarrollo y respetando los derechos de las personas, a continuación, se detallan algunas de ellas.

2.5.1 Ethics and Responsible AI Deployment/ Ética y Despliegue Responsable de la IA.

El Centro de Investigación de Cisco Systems y el Departamento de Ciencias de la Computación y Ciencias de la Ingeniería de la Universidad de Oxford, realizaron un estudio para explorar el equilibrio entre las capacidades de la IA y la protección de la privacidad personal mediante el uso de técnicas algorítmicas procurando cumplir el reglamento general de protección de datos de la Unión Europea (Radanliev & Santos, 2023).

El objetivo principal del estudio fue identificar soluciones técnicas que permitan desarrollar sistemas de IA que estén alineados con los principios éticos y las regulaciones. En este contexto, los autores proponen tres técnicas algorítmicas que ofrecen soluciones viables para asegurar la privacidad y la ética en el desarrollo de sistemas de IA, las que a continuación se detallan:

Tabla 16. *Técnicas logarítmicas sugeridas en el estudio para cumplir con normativas*

Técnica Algorítmica	Aplicación para cumplir desarrollo de sistemas de IA responsable
Privacidad Diferencial	Añade ruido estadístico a los datos o resultados de análisis para proteger la identidad de los individuos en un conjunto de datos. Esto permite realizar análisis estadísticos o aprendizaje automático sobre los datos sin revelar información sensible.
Encriptación Homomórfica	Permite realizar cálculos directamente sobre datos cifrados, sin necesidad de descifrarlos. Los resultados obtenidos permanecen encriptados y solo el receptor autorizado puede descifrarlos para acceder a la información procesada.
Aprendizaje Federado	Permite entrenar modelos de inteligencia artificial utilizando datos que permanecen en dispositivos locales, sin transferirlos a un servidor central. En lugar de enviar los datos, los dispositivos envían actualizaciones del modelo que se combinan para mejorar el modelo general.

Fuente: Elaboración propia.

Según Radanliev & Santos (2023), el uso de estas técnicas permitirá proteger la privacidad sin comprometer el despliegue y desarrollo de la inteligencia artificial (IA) en ámbitos sensibles como salud, finanzas, justicia penal, la defensa y los recursos humanos que presentan retos únicos y dilemas morales que requieren una atención muy cuidadosa.

Aunque según los autores, se han dado pasos importantes hacia la regulación de la IA, aún existe una falta de consenso global al respecto, lo que constituye un obstáculo significativo para equilibrar la evolución de la IA frente a la privacidad. Observan que debido a la divergencia en las normativas entre regiones se dificulta establecer estándares universales, dejando brechas en la protección de derechos fundamentales del ser humano mientras la IA sigue desarrollándose.

Para garantizar un despliegue ético y responsable de la IA, señalan que es esencial combinar avances tecnológicos con un marco normativo sólido y estrategias éticas claras para regular de sistemas de IA de tal manera que desde su diseño hasta su uso diario respeten la privacidad, la equidad y los valores humanos.

2.5.2 Adding Adding human values on the deepfake: co-designing fact-checking solutions to combat misinformation / Añadir valores humanos al deepfake: codiseñar soluciones de comprobación de hechos para combatir la desinformación

Los medios sintéticos son contenidos virtuales realistas, como videos, imágenes, sonidos y textos, creados utilizando inteligencia artificial y técnicas de aprendizaje automático (Maia et al., 2024). Dentro de este grupo, las deepfakes representan un subconjunto específico que se enfoca en generar contenidos altamente personalizados y manipulados.

Investigadores de la Universidad de Oporto y la Universidad Católica de Río de Janeiro realizaron un estudio para explorar las tecnologías de generación de medios sintéticos con el objetivo de mejorar la rapidez y eficiencia en los procesos de verificación de hechos, asegurando el respeto a los derechos fundamentales de privacidad, transparencia y autenticidad. Esta investigación surge como respuesta a la proliferación de desinformación generada mediante medios sintéticos, cuya verificación manual resulta lenta e insuficiente para lograr una verificación automatizada más eficiente y confiable.

Utilizaron la metodología VSD (Value Sensitive Design – Diseño Sensible a los Valores) que busca incorporar los valores humanos de privacidad, transparencia, autenticidad, inclusividad y confianza desde el diseño hasta la implementación de tecnologías de deepfake para verificación de hechos. Esta metodología se aplicó en un taller multidisciplinario realizado durante la conferencia C+J-DataJ 2023,⁴ donde se les solicitó a los participantes analizar escenarios de uso de deepfakes para verificar hechos, evaluar herramientas tecnológicas de medios sintéticos y proponer soluciones adaptadas a los mencionados valores humanos.

Las herramientas evaluadas en el taller fueron AIDA, GaB, Synthesia y D-ID. AIDA (Algoritmo de Interpretación de Datos) es un sistema que automatiza la creación de reportes de noticias basados en datos llegando a integrar voces sintéticas y personajes 3D. GaB es un personaje virtual para conectarse interactivamente con audiencias jóvenes. Synthesia es una plataforma de IA que crea videos utilizando avatares generados por deepfake. D-ID es una API que transforma imágenes estáticas en videos avanzados.

El taller arrojó resultados sorprendentes que se detallan a continuación:

Tabla 17. Resultados del taller multidisciplinario sobre uso ético de las deepfakes

Resultados	Detalles
Beneficios del uso de las deepfakes en verificación de hechos	<ul style="list-style-type: none"> • Rapidez en la verificación de hechos (realizado con AIDA) • Personalización para adaptar los mensajes a diferentes audiencias (realizado con GaB) • Accesibilidad generando contenido para múltiples plataformas (realizado con Synthesia).
Desafíos éticos	<ul style="list-style-type: none"> • Generar confianza ya que las deepfakes están asociadas con desinformación. • Los usuarios pueden percibir que los resultados verificados con estos métodos no son auténticos. • Garantizar privacidad cuando se usan datos personales para generar contenido.

⁴ C+J-DataJ 2023, (Computation + Journalism). Se llevó a cabo en Zúrich, Suiza, del 22 al 24 de junio de 2023

Resultados	Detalles
Requisitos tecnológicos	<ul style="list-style-type: none">• Incorporación de mecanismos que permitan rastrear el proceso de generación de contenido para asegurar la transparencia.• Diseño de herramientas que señalen el nivel de confiabilidad de los resultados.• Implementar protecciones para salvaguardar los datos que se utilizan en la generación de medios sintéticos.

Fuente: Elaboración propia.

La investigación recomienda desarrollar las tecnologías de deepfake de manera ética para combatir la desinformación actual. Estos sistemas deben diseñarse incorporando mecanismos de rastreo, protección de datos personales y principios de inclusión. Además, se enfatiza la importancia de realizar auditorías independientes y fomentar la colaboración interdisciplinaria, promoviendo así una implementación ética en ámbitos como la educación y el periodismo.

2.5.3 Índice de transparencia del modelo fundacional

“Un modelo fundacional es un segmento de la inteligencia artificial en el que el algoritmo central se crea ingiriendo cantidades masivas de datos, con un modelo resultante que puede aplicarse a una amplia variedad de tareas” (Martinez-Polo, 2024, Pág.1)

Los modelos fundacionales pueden ser de lenguaje (GPT, BERT, LaMDA), de código (CODEX) o de imágenes (DALL-E2).

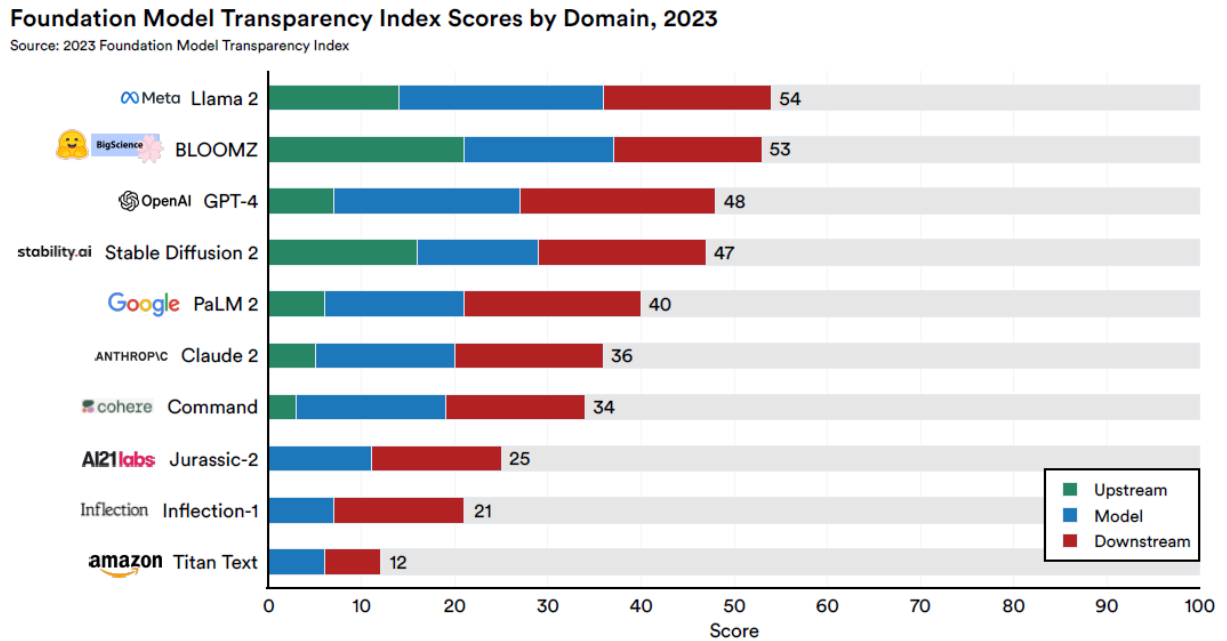
La Universidad de Stanford a través de su Instituto para la Inteligencia Artificial Centrada en Humanos (Stanford Institute for Human-Centered Artificial Intelligence – HAI), creó una iniciativa interdisciplinaria para investigar los modelos fundacionales (Center for Research on Foundation Models -CRFM).

Considerando que la transparencia permite la rendición de cuentas, la equidad, la innovación responsable y la confianza en la IA, el CFRM decidió evaluar y promover la transparencia en el desarrollo y uso de los modelos fundacionales, creando el Índice de Transparencia del Modelo de Fundación (Foundation Model Transparency Index – FMTI).

“El Índice de Transparencia de Modelos Fundacionales 2023 especifica 100 indicadores detallados que codifican exhaustivamente la transparencia de los modelos de fundaciones, abarcando los recursos previos utilizados para construir un modelo de fundación (por ejemplo, datos, mano de obra, computación), detalles sobre el propio modelo (por ejemplo, tamaño, capacidades, riesgos) y el uso posterior (por ejemplo, canales de distribución, políticas de uso, geografías afectadas)” (Bommasani et. al. 2023.Pág.1).

El FMTI y sus 100 indicadores se aplicaron a 10 empresas desarrolladoras de modelos fundacionales o de cimentación: AI21 LABS, Amazon, Anthropic, Cohere, Google, Hugging Face, Inflection, Meta, OpenAI y Stability AI con los siguientes resultados:

Figura 11. Indicadores de la aplicación 2023 del modelo de transparencia fundacional con énfasis en los tres aspectos evaluados.



Fuente: Bommasani, et. al. 2023⁵.

Como se puede evidenciar todas las empresas que generan este tipo de IA tienen mucho por mejorar, la mayor puntuación es 54%, la media es 37%, señalando una falta de transparencia generalizada, por lo que es necesario y urgente que implementen cambios.

El estudio concluyó que los “resultados muestran que el statu quo se caracteriza por una falta de transparencia generalizada entre los desarrolladores.... Consideramos que se trata de una grave acusación contra el ecosistema en su conjunto... los modelos de fundación se están desarrollando, desplegando y adoptando a un ritmo frenético: para que esta tecnología promueva el interés público, es necesario un cambio real que rectifique la falta fundamental de transparencia del ecosistema” (Bommasani et. al. 2023.Pág.63).

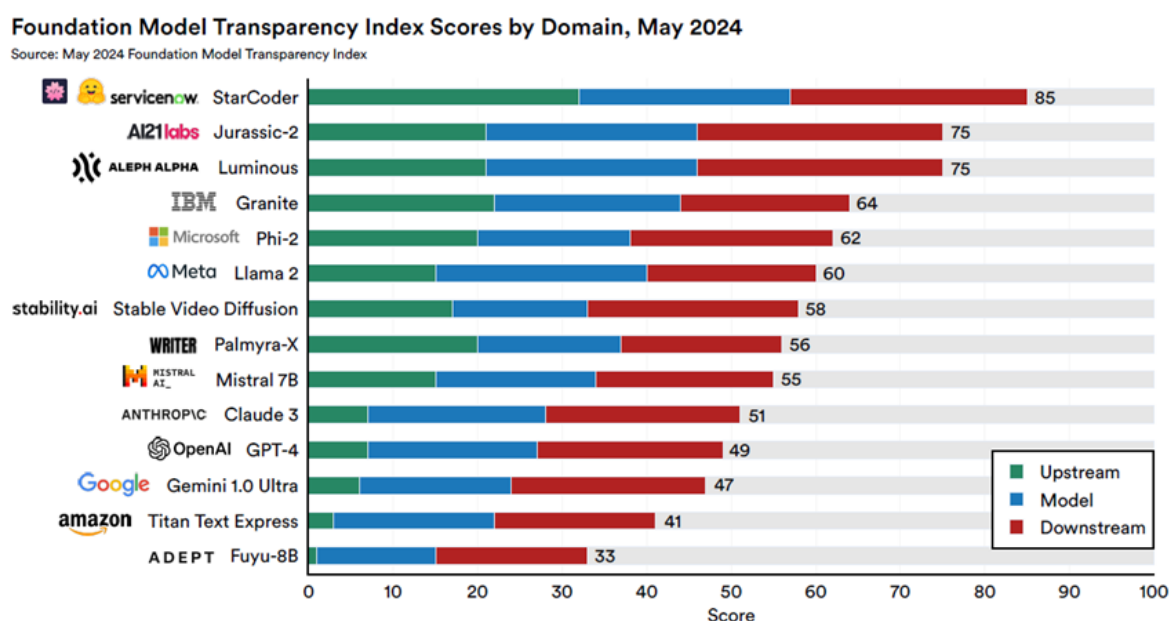
En mayo 2024, se publicó la versión 1.1 del FMTI (Bommasani, et. al. 2024), no solo con información pública (como el estudio del 2023) sino también se permitió que las empresas presenten información que antes no era de dominio público, lo que mejoró significativamente

⁵The Foundation Model Transparency Index. Página 31

los resultados. En esta versión, se incluyeron a 6 empresas: Adept, Aleph Alpha, IBM, Microsoft, Mistral y Writer.

En términos generales, en la versión 1.1 del índice de transparencia, las empresas obtuvieron mejores puntuaciones como se evidencia en la Figura No. 12 a continuación:

Figura 12. Puntuación global 2024 del Índice de Transparencia del Modelo Fundacional con detalle de los tres aspectos evaluados.



Fuente: Bommasani, et. al. 2024⁶.

En esta oportunidad, la puntuación media subió de 37% a 58%, pero se sigue evidenciando que en la fase previa (upstream), el manejo de los datos aún está lejos la transparencia deseada. “Los desarrolladores no revelan los datos, la mano de obra y los cálculos utilizados para construir sus modelos, a menudo no publican evaluaciones reproducibles de riesgos o mitigaciones, y no comparten información sobre el impacto que sus modelos están teniendo en los usuarios o en los sectores del mercado. La transparencia en el ecosistema de los modelos fundacionales lleva varios años disminuyendo, y es poco probable que esta tendencia se invierta a corto plazo. (Bommasani, et. al. 2024.Pág.14)

⁶ The Foundation Model Transparency Index v1.1. Página 7

2.5.4 Transparencia de la IA en la era de los LLM: Una hoja de ruta para la investigación centrada en el ser humano

Un LLM (Large Language Model) o Modelo de Lenguaje de Gran Tamaño, es un modelo de aprendizaje automático que se entrena con grandes cantidades de datos de texto. Este tipo de modelos tienen la capacidad de comprender y generar lenguaje humano (Amazon Web Services, s.f.).

Al momento, están en apogeo los LLM por su versatilidad y capacidad para generar contenido, sin embargo, su gran debilidad es la falta de transparencia, porque transparencia no es sólo visibilizar el código o los datos, es también entregar información relevante a las partes interesadas sobre las capacidades, limitaciones, funcionamiento y uso adecuado de estas poderosas herramientas (Holstein et al, 2023).

Para asegurar que los LLM se desarrollen transparentemente centrados en el ser humano, las autoras proponen los siguientes pasos:

- Definir el concepto de transparencia para usuarios finales, desarrolladores y responsables de la elaboración de políticas.
- Establecer principios para mejorar la transparencia durante todo el desarrollo del LLC, documentando todo el modelo (arquitectura, limitaciones, riesgos), divulgar las dudas (donde los LLC no son confiables o presentan resultados poco claros) y evaluar el impacto real (en la confiabilidad de los LLM) de las decisiones tomadas durante el diseño.
- Formalizar estrategias para la comprensión de los LLM dando explicaciones claras, creando interfaces dinámicas para permitir la exploración de los modelos e implementando sistemas que demuestren los resultados de los LLM en diferentes contextos para demostrar la confiabilidad de los modelos.
- Involucrar a las partes interesadas empezando con el co-diseño con los usuarios finales, haciendo auditorías externas e independientes de la transparencia y confiabilidad de los modelos y proporcionando información a los usuarios finales para que su interacción con los LLM sea debidamente informada.

- Estudiar cómo todos los esfuerzos de transparencia influyen en la confianza de los usuarios finales.
- Proporcionar herramientas y recursos para que se puedan evaluar y regular los LLM.

La propuesta hace hincapié que la transparencia en los LLM no es un esfuerzo individual sino el resultado de un trabajo colaborativo e interdisciplinario siempre centrado en las necesidades del ser humano para lograr implementar estrategias universales que hagan que los LLM sean confiables y utilizados ética y responsablemente por la sociedad.

2.6 Conclusiones

- Ante el crecimiento exponencial sin limitaciones del uso de la IA, han aparecido iniciativas de las propias empresas desarrolladoras de IA, de las empresas que usan, desde los gobiernos y desde la academia para su uso y desarrollo ético. Sin embargo, el verdadero reto, es lograr un equilibrio entre el desarrollo de la IA con las leyes y reglamentos que la regulan para mantenerla centrada en el ser humano.
- No existe una normativa global estandarizada y aceptada que rija el desarrollo ético de la IA de una manera obligatoria para todos los actores relacionados con esta actividad, por lo que urge la establecer una normativa para el uso estandarizado de controles procesos y estructuras que permitan el desarrollo, control y auditoría de la IA responsable a través de la aplicación de políticas de regulación, monitoreo y auditorías que garanticen el cumplimiento de estándares técnicos, legales y éticos.
- Se ve la necesidad de consolidar un marco de trabajo que consolide estos esfuerzos en un modelo que pueda ser aplicable a productos y servicios de IA tanto existentes como en desarrollo, por lo cual este trabajo se orienta a proponer dicho marco.
- Dentro del desarrollo de una IA responsable, se debe incluir como un punto de control que todo proceso de IA debe tener una opción de detención forzosa sin condiciones, dado que una característica de la inteligencia es la interpretación de los estados de proceso, y una IA por si sola pudiera asumir que su estado de proceso es correcto a pesar de que no esté cumpliendo con principios de una IA ética y responsable. Por lo que la intervención humana será necesaria para prevenir daños a personas a animales o a infraestructuras físicas o naturales o al medio ambiente.

Alvear Aguirre Isabel Gardenia

Flores Villamil Edwin Vladimir

Ortiz Albán Iván Josué

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

3 Objetivos concretos y metodología de trabajo

3.1 Objetivo general

Desarrollar y validar un marco de trabajo considerando controles, procesos, estructuras, licitud y ética en el mundo occidental, para guiar a las organizaciones durante el diseño e implementación de la IA responsable, a fin de que identifiquen el nivel en el que se encuentran y obtengan directrices para mejorar.

3.2 Objetivos específicos

- Identificar los parámetros mínimos que un sistema de IA debe cumplir con respecto a las regulaciones y leyes de occidente.
- Identificar los controles efectivos para asegurar que los sistemas de IA cumplan con estándares técnicos y éticos, minimizando el riesgo de sesgos y errores.
- Establecer niveles de IA responsable.
- Diseñar un listado de confirmación para la IA responsable que permita controlar que los procesos sean sistemáticos y detallados, que existan estructuras organizacionales, con políticas y procedimientos de auditoría durante el desarrollo e implementación de la IA.
- Disponer de una metodología para elaborar un plan de acción en productos y/o servicios existentes para incrementar el nivel de IA responsable.
- Aplicar la metodología realizando un piloto en una empresa para evaluar resultados y proponer mejoras del marco de trabajo.

3.3 Metodología del trabajo

3.3.1 Herramientas

Este apartado presenta los fundamentos teóricos, normativos y éticos que sustentan la elaboración del marco de trabajo para evaluar la fiabilidad y ética en sistemas de inteligencia artificial (IA). Se detallan las directrices, normativas legales, marcos de transparencia y principios éticos utilizados, estableciendo las bases conceptuales para la metodología propuesta.

Una vez estudiado exhaustivamente el marco legal en occidente, las iniciativas académicas y los principios adoptados por las grandes empresas que generan y/o usan IA, se utilizará la información especificada a continuación para desarrollar el marco de trabajo.

Figura 13. Documentación que se utiliza como herramientas para desarrollar el marco de trabajo



Fuente: Elaboración propia

Directrices éticas para una IA Fiable

Las directrices éticas para una IA fiable establecen siete requisitos o áreas fundamentales para asegurar que los sistemas de IA sean éticos, robustos y centrados en el ser humano, las cuales son las siguientes:

- Agentes humanos y supervisión
- Solidez técnica y seguridad
- Privacidad y gobernanza de los datos
- Transparencia
- Diversidad, no discriminación y equidad
- Bienestar social y medio ambiental

- Rendición de cuentas

Figura 14. Áreas de Evaluación de IA responsable.



Fuente: Elaboración propia.

Dado que las directrices éticas tienen un enfoque holístico respecto a la IA responsable, éstas serán la base y la herramienta primordial sobre la cual se construirá el marco de trabajo.

Reglamento General de Protección de Datos de la Unión Europea-RGPD

El RGPD se utilizará como segundo nivel para construir el marco de trabajo, ya que proporciona normativas relacionadas a la protección de datos y los derechos de los interesados (propietarios de los datos), lo que asegurará que el marco que se propone garantizará que se evaluará el producto o servicio de IA en lo relacionado al cumplimiento normativo de protección de datos personales.

A continuación, se muestran los artículos del RGPD que serán incluidos en el marco en relación con los derechos de los interesados:

Tabla 18. *Derechos de los interesados según el RGPD.*

Derecho	Faculta a los interesados a:
Al acceso	Acceder a sus datos y obtener información sobre su uso (Art.15)
De rectificación	Corregir datos incompletos o inexactos (Art.16)
De supresión	Solicitar la eliminación de sus datos (derecho al olvido Art. 17)
A limitar el tratamiento	Restringir el procesamiento de sus datos (Art.18)
A la portabilidad de datos	Recibir los datos en un formato y transferirlos a otro responsable (Art.20)
De oposición	Oponerse al tratamiento de sus datos (Art.21)
No ser objeto de decisiones individuales automatizadas	Autorizar acogerse a estas decisiones individuales o perfilamiento mediante su consentimiento explícito (Art.22)
Información	Ser informados sobre la recopilación y uso de sus datos personales (Art. 13)
A retirar el consentimiento	Retirar el consentimiento si el tratamiento se basa en él (Art.7)
A presentar un reclamo ante autoridades de control	Presentar un reclamo si consideran que el tratamiento de sus datos infringe el RGPD (Art.77)
A tutela judicial efectiva	Recurrir a los tribunales si consideran que sus derechos han sido vulnerados (Art.78-79)

Fuente: Elaboración propia.

El RGPD también señala obligaciones para el responsable del tratamiento de datos personales, por lo cual el marco de trabajo incluirá preguntas acerca de cómo se están cumpliendo dichas obligaciones en el producto/servicio de IA evaluado.

Los artículos que se tomarán en cuenta son los siguientes:

Tabla 19. *Principales obligaciones del responsable de protección de datos según el RGPD.*

Principales obligaciones del responsable de protección de datos según el RGPD	Artículo
Cumplir con los principios relativos al tratamiento de datos personales	5
Definir las bases legales para el tratamiento de los datos	6
Establecer las condiciones para que el consentimiento del interesado sea válido para el tratamiento de los datos	7
Regular el tratamiento de datos sensibles del interesado	9
Establecer como se van a tratar los datos personales relativos a infracciones criminales de los interesados	10
Proporcionar información clara y accesible a los interesados	12
Tipo de información debe proporcionarse cuando se obtenga la información desde el interesado y desde otra fuente	13-14
Facilitar el ejercicio de los derechos de los interesados	15-22
Cumplir con lo señalado en el RGPD	24
Integrar la protección de datos desde el diseño del sistema y de los procesos	25
Cumplir sus obligaciones con el encargado del tratamiento	28
Llevar un registro de todo el tratamiento de los datos personales	30
Cooperar con la autoridad de control	31
Implementar medidas de seguridad para el tratamiento de los datos	32
Notificar violaciones al RGPD a la autoridad de control y a los interesados	33-34
Evaluación de impacto relativa a la protección de datos.	35
Consulta previa a la autoridad de control.	36
Designación del delegado de protección de datos.	37

Principales obligaciones del responsable de protección de datos según el RGPD	Artículo
Cumplir con las normas sobre transferencias internacionales de datos.	44-50
Cumplir con las obligaciones específicas para situaciones como el tratamiento en el ámbito laboral, con fines de archivo en interés público	85-91

Fuente: Elaboración propia.

Ley de IA de la Unión Europea – Reglamento de IA

La Ley de Inteligencia Artificial de la UE o Reglamento de la IA que entró en vigor el 01 de agosto de 2024 será tercer nivel ingrediente del marco de trabajo a desarrollarse, específicamente se incluyen los aspectos indicados en la siguiente tabla.

Tabla 20. Aspectos del Reglamento de IA de la UE que se utilizarán en el marco de trabajo.

Aspectos	Explicación
Disposiciones Generales	<p>Se enfoca en garantizar seguridad, claridad de roles y responsabilidades, capacitación técnica y ética.</p> <p>Fomenta el desarrollo de sistemas de IA seguros y centrados en el ser humano</p>
Prácticas de IA Prohibidas	<p>Prohíbe sistemas que manipulen comportamientos, exploten vulnerabilidades de grupos específicos o realicen clasificaciones discriminatorias.</p> <p>Limita la recolección y uso de datos biométricos sin consentimiento y el uso de IA para inferir emociones fuera de contextos específicos como la medicina.</p> <p>Busca minimizar riesgos y proteger derechos fundamentales al restringir prácticas perjudiciales y garantizar un uso ético.</p>
Sistemas de IA de Alto Riesgo	<p>Regula sistemas de IA clasificados como de alto riesgo, con criterios para su diseño ético y responsable.</p> <p>Impone requisitos técnicos para seguridad, gestión de riesgos, transparencia, trazabilidad y ciberseguridad.</p> <p>Exige que los proveedores implementen medidas de conformidad y documentación técnica, con controles de calidad y notificaciones de incidentes para asegurar la protección del usuario.</p>

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Aspectos	Explicación
	Permite flexibilidad regulatoria adaptándose a cambios tecnológicos para mantener la protección y el desarrollo ético.
Obligaciones de Transparencia	<p>Establece que proveedores e implantadores deben informar a usuarios sobre la naturaleza y funcionamiento de los sistemas de IA.</p> <p>Exige etiquetado claro de contenidos generados por IA y transparencia en sistemas de reconocimiento emocional y biométrico.</p> <p>Fomenta códigos de práctica para fortalecer la confianza pública y la relación ética con la tecnología.</p>
Modelos de IA de Propósito General	<p>Clasifica modelos de IA por su impacto potencial y establece requisitos para gestión de riesgos y transparencia.</p> <p>Los proveedores deben colaborar en evaluaciones de conformidad y gestionar riesgos sistémicos.</p> <p>Fomenta la creación de códigos de buenas prácticas para promover el desarrollo ético y la cooperación entre actores</p>
Medidas de Apoyo a la Innovación	<p>Crea espacios de prueba regulada para experimentar con IA bajo supervisión, asegurando seguridad y ética en el desarrollo.</p> <p>Permite el uso controlado de datos personales para innovación en sistemas de interés público.</p> <p>Ofrece apoyo a PYMEs y startups para integrarse en el ecosistema de IA, asegurando cumplimiento de normas éticas.</p>
Base de Datos de Sistemas de IA de Alto Riesgo	<p>Implementa una base de datos central para mejorar la transparencia y supervisión de sistemas de IA de alto riesgo.</p> <p>Busca proteger la seguridad y derechos de los ciudadanos, asegurando cumplimiento con las normativas.</p>
Control Posterior a la Comercialización	<p>Requiere que proveedores realicen monitoreo continuo y gestionen incidentes para asegurar la seguridad de los sistemas de IA.</p> <p>Las autoridades tienen facultades para evaluar conformidad, emitir alertas y tomar medidas correctivas.</p>

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Aspectos	Explicación
	Proveedores deben cooperar en auditorías y seguir procesos establecidos para resolver problemas de cumplimiento.

Fuente: Elaboración propia.

Índice de transparencia del modelo fundacional -FMTI

Como cuarto ingrediente del modelo se incluyó al índice de transparencia del modelo fundacional propuesto en el Instituto para la Inteligencia Artificial Centrada en el Humano de la Universidad de Stanford, ya que con estos indicadores se evalúo a los grandes modelos fundacionales y algunos indicadores también son aplicables para el marco de trabajo que se propone

Se revisarán los 100 indicadores que están divididos en tres aspectos: previos al modelo, relacionados con el modelo y posteriores al modelo.

Tabla 21. *Los 100 indicadores del FMTI que se revisarán como cuarto ingrediente para el marco de trabajo*

Previo al modelo (Upstream)	Modelo	Posterior al modelo (Downstream)
Tamaño de los datos	Modalidad de entrada	Toma de decisiones
Fuentes de datos	Modalidad de salida	Proceso de liberación
Creadores de datos	Componentes del modelo	Canales de distribución
Selección de fuentes de datos	Tamaño del modelo	Productos y servicios
Conservación de datos	Arquitectura modelo	Detección de contenidos generados por máquinas
Aumento de datos	Documentación centralizada de los modelos	Modelo de licencia
Filtración de datos nocivos	Protocolo de acceso a modelos externos	Condiciones de servicio
Datos protegidos por derechos de autor	Acceso al modelo externo Blackbox	Usuarios permitidos y prohibidos
Licencia de datos	Acceso completo al modelo externo	Usos permitidos, restringidos y prohibidos
Información personal en los datos	Descripción de capacidades	Aplicación de la política de uso
Utilización de mano de obra humana	Demostración de capacidades	Justificación de la medida de ejecución
Empleo de trabajadores de datos	Evaluación de las capacidades	Mecanismo de recurso en caso de infracción de la política de uso
Distribución geográfica de los trabajadores de datos	Reproducibilidad externa de la evaluación de capacidades	Comportamientos permitidos, restringidos y

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Previo al modelo (Upstream)	Modelo	Posterior al modelo (Downstream)
		prohibidos de los modelos
Salarios	Evaluación de las capacidades de terceros	Aplicación de una política de comportamiento ejemplar
Instrucciones para crear datos	Descripción de las limitaciones	Interoperabilidad de las políticas de uso y comportamiento de los modelos
Protección laboral	Demostración de limitaciones	Interacción del usuario con el sistema de IA
Terceros colaboradores	Evaluación de las limitaciones por terceros	Limitaciones de uso
Acceso a datos externos consultables	Descripción de riesgos	Política de protección de datos de los usuarios
Acceso directo a datos externos	Demostración de riesgos	Uso permitido y prohibido de los datos de los usuarios
Uso del ordenador	Evaluación de daños no intencionados	Protocolo de acceso a los datos de uso
Duración del desarrollo	Reproducibilidad externa de la evaluación de daños no intencionados	Protocolo de control de versiones
Hardware informático	Evaluación del daño intencionado	Registro de cambios
Propietario de hardware	Reproducibilidad externa de la evaluación del daño intencionado	Política de amortización
Consumo de energía	Evaluación de riesgos de terceros	Mecanismo de retroalimentación
Emisiones de carbono	Descripción de las medidas paliativas	Resumen de los comentarios
Impacto medioambiental más amplio	Demostración de mitigación	Consultas gubernamentales
Etapas modelo	Evaluación de las medidas paliativas	Mecanismo de control
Objetivos del modelo	Reproducibilidad externa de la evaluación de las medidas paliativas	Aplicaciones posteriores
Marcos básicos	Evaluación de las medidas paliativas de terceros	Sectores del mercado afectados
Dependencias adicionales	Evaluación de la fiabilidad	Personas afectadas
Medidas de protección de la intimidad	Reproducibilidad externa de la evaluación de la fiabilidad	Informes de uso
Mitigación de los derechos de autor	Evaluación de la duración de la inferencia	Estadísticas geográficas
	Inferencia cálculo evaluación	Mecanismo de recurso
		Documentación centralizada para uso posterior
		Documentación para un uso posterior responsable

Fuente: Bommasani et al.,2023⁷ y Elaboración propia.

El detalle de cada indicador se encuentra en el Anexo B de este documento.

Principios éticos comunes en las empresas desarrolladoras de IA

Finalmente, la quinta herramienta que se utilizará para realizar el marco de trabajo, son los principios éticos que son comunes en las mayores empresas desarrolladoras de IA que se encontraron son 10 y se muestran a continuación:

Tabla 22. *Principios éticos comunes de las grandes empresas desarrolladoras de IA que se considerarán para la elaboración del marco de referencia.*

Principio	Detalle
Privacidad	Asegurar que los datos personales de los usuarios se manejen de manera segura y confidencial, respetando la privacidad y control que las personas deben tener sobre su información personal.
Seguridad	Implementar medidas para proteger los sistemas de IA y los datos de usuarios contra accesos no autorizados, ataques y otras vulnerabilidades que puedan comprometer la integridad de la información.
Transparencia	Proporcionar claridad sobre cómo se desarrollan, implementan y operan los sistemas de IA, permitiendo a los usuarios comprender cómo se toman las decisiones y qué datos se utilizan para ello
Inclusión	Garantizar que los sistemas de IA sean accesibles y útiles para todas las personas, independientemente de sus antecedentes, capacidades o condiciones, promoviendo la diversidad
Evitar Sesgos	Desarrollar modelos de IA que minimicen los prejuicios inherentes, asegurando resultados justos y equitativos para todos los usuarios, evitando la discriminación basada en características personales.
Equidad	Diseñar sistemas que garanticen el trato justo e imparcial para todos los usuarios, evitando ventajas injustas para cualquier grupo o individuo.

⁷ The Foundation Model Transparency Index, Páginas 17-20.

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Principio	Detalle
Explicabilidad	Crear sistemas de IA cuyas decisiones y procesos sean comprensibles para los usuarios, permitiendo que se expliquen de manera clara y razonable.
Evitar contenido inapropiado	Asegurar que los sistemas de IA no generen, promuevan o distribuyan contenido dañino, ofensivo o inapropiado.
Rendición de cuentas:	Establecer mecanismos claros para identificar responsabilidades en el diseño y uso de la IA, permitiendo que las entidades sean responsables por sus acciones y decisiones.
Responsabilidad con las personas	Centrar el desarrollo de la IA en el bienestar humano, asegurando que la tecnología mejore la vida de las personas y no cause daño.

Fuente: Elaboración propia.

No se consideró la orden presidencial de Estados Unidos de América para legislar la IA porque a la fecha de publicación de este trabajo, aún no se han reportado públicamente sus resultados y lo propuesto en dicho documento coincide en gran medida con el Reglamento Europeo sobre IA que si consideró para el marco de trabajo.

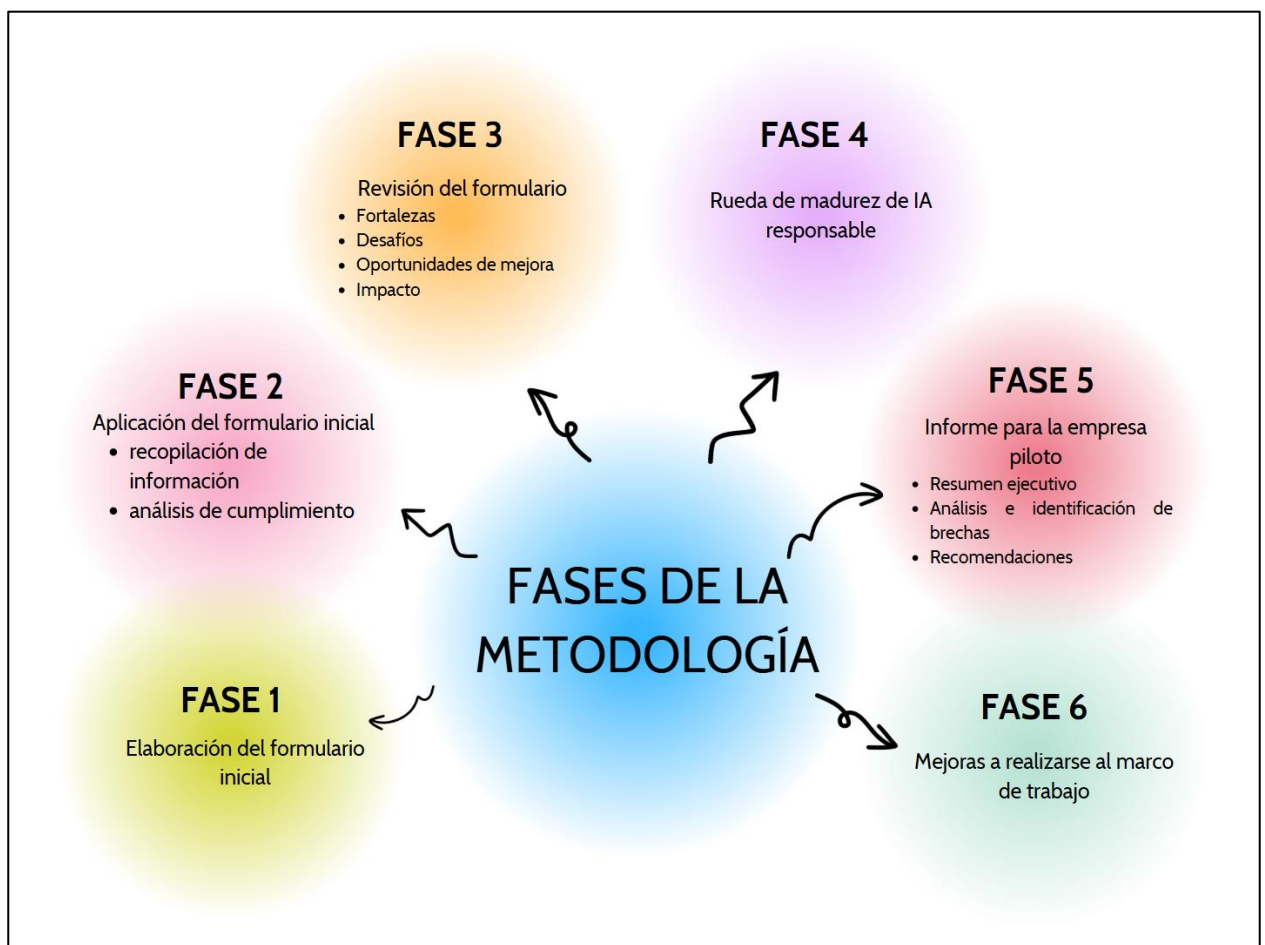
No se utilizó como herramienta el DMBOK2 de DAMA debido a que el reglamento general de protección de datos de la Unión Europea es de cumplimiento obligatorio mientras que el DMBOK2 son prácticas sugeridas.

3.3.2 Metodología

La metodología descrita a continuación establece el proceso sistemático utilizado para desarrollar y validar el marco de trabajo para la evaluación de IA responsable en 6 fases y varias subfases. Esta metodología se fundamenta en las herramientas descritas en el numeral anterior que están vigentes en el mundo occidental.

A continuación, se presentan las fases y sus respectivos pasos.

Figura 15. Fases de la metodología planteada para desarrollar el marco de trabajo para una IA responsable.



Fuente: Elaboración propia.

Fase 1: Elaboración del formulario inicial

El primer paso en el desarrollo del marco de trabajo consiste en la elaboración de un formulario inicial que sirva como herramienta para recopilar información clave sobre los sistemas de inteligencia artificial en evaluación. Este formulario se diseña considerando las herramientas descritas en el numeral 3.3.1

El formulario inicial se estructura en secciones claras tomando como base las 7 áreas de evaluación de las directrices éticas para una IA fiable y se va complementando con los demás documentos descritos en la sección herramientas.

Adicionalmente se establecen indicadores cuantitativos para facilitar una evaluación objetiva.

Fase 2: Aplicación del formulario inicial

Una vez desarrollado, el formulario inicial se aplicará a una empresa que disponga de sistemas de IA en desarrollo o implementación y que esté de acuerdo en ser parte de este estudio.

En esta fase se busca recopilar datos empíricos con los que la empresa se encuentra operando, los cuales permitirán medir el nivel de cumplimiento de IA responsable considerando las normativas regulatorias y prácticas de las grandes proveedoras de IA.

Dentro de esta fase se realizarán dos pasos fundamentales: la recopilación de datos, es decir, aplicar el formulario conjuntamente con personal designado de la empresa, aclarando de ser necesario las preguntas y orientando en caso de ser necesario. El segundo paso es realizar el análisis de cumplimiento, que consiste en revisar las respuestas obtenidas para identificar áreas donde el sistema, producto o servicio de IA es responsable según los estándares establecidos, así como identificar brechas que necesiten ser abordadas.

La aplicación del formulario no solo realiza una evaluación diagnóstica del nivel de responsabilidad en los sistemas, productos o servicios de IA, sino que también permite sensibilizar a los responsables en la empresa sobre la importancia de cumplir con principios éticos y normativos.

Los resultados obtenidos se utilizarán de base para la elaboración de las recomendaciones específicas que se entregarán a la empresa para que mejoren su nivel de IA responsable en los sistemas, productos y/o servicios evaluados.

Fase 3: Revisión del formulario

Una vez completada la aplicación del formulario inicial, se procede a analizar lo que podría mejorarse durante el proceso de evaluación. Este análisis permite identificar fortalezas,

desafíos y oportunidades de mejora en el marco de trabajo considerando tanto el diseño del formulario como su aplicación. Los pasos en esta etapa son:

- Identificar fortalezas: identificando las áreas y/o preguntas del formulario que resultaron útiles para evaluar con precisión el nivel de IA responsable, así como los aspectos del marco de trabajo que fueron mejor recibidos por la empresa evaluada.
- Desafíos encontrados: documentando dificultades enfrentadas durante la aplicación del formulario, como falta de datos o resistencia del personal a responder ciertas preguntas y analizar sus causas.
- Oportunidades de mejora del formulario: si es necesario realizar ajustes al formulario inicial, como la mejora en la redacción o la simplificación de ciertas preguntas, para optimizar la efectividad y aplicabilidad del marco de trabajo.
- Impacto en la empresa: si es posible evaluar cómo durante la aplicación del formulario se sensibilizó el personal de la empresa sobre la importancia de la IA responsable y su cumplimiento.

Esta revisión realizada luego de la aplicación del formulario permitirá perfeccionar el marco de trabajo.

Fase 4: Rueda de madurez de IA responsable

Con base en los resultados obtenidos durante la aplicación del formulario y el análisis de las lecciones aprendidas, se diseñará una rueda de madurez de IA responsable que es única para cada empresa. Esta herramienta tiene como objetivo proporcionar una representación visual y comprensible del nivel de madurez de la empresa en términos de implementación de IA responsable.

La rueda de madurez incluirá 5 niveles, lo cual se utiliza como una herramienta estratégica para identificar las áreas que requieren atención prioritaria y para diseñar planes de acción que permitan a las empresas avanzar hacia niveles más altos de madurez en IA responsable.

Fase 5. Informe para la empresa piloto

Tras completar la evaluación de los sistemas, productos o servicios de IA mediante el formulario y los análisis posteriores, se elaborará un informe detallado para la empresa piloto

participante. Este informe presentará un resumen ejecutivo para los responsables de las decisiones de la empresa con la información concisa de los resultados principales. También se realizará un análisis detallado de los resultados obtenidos en cada una de las 7 áreas de evaluación, detallando las normativas que no están cumpliendo. El informe finalizará con recomendaciones específicas para abordar las brechas identificadas, priorizando aquellas que tendrán mayor impacto en la empresa y la apoyarán para alcanzar el mínimo nivel de IA responsable en el menor tiempo posible.

El informe buscará no solo evaluar el estado actual del sistema de IA, sino también servir como guía práctica para que la organización avance hacia un nivel superior de madurez en IA responsable.

Fase 6. Mejoras a realizarse al marco de trabajo

La última fase de esta metodología es sugerir ajustes para perfeccionar el marco de trabajo en función de los resultados obtenidos durante su aplicación piloto. Este paso garantiza que el marco sea robusto, práctico y adaptable a distintos contextos empresariales. Las mejoras que podrán sugerirse son incorporación o modificación de preguntas existentes para aumentar su relevancia y claridad, ajustes en los niveles de madurez de IA responsable para adaptarla mejor a las diferentes realidades que se pueden presentar en las empresas e incluir nuevas regulaciones, si es que ya está disponible públicamente los resultados de la orden presidencial de Estados Unidos para legislar la IA. Inclusive se puede sugerir mejorar el formulario mediante el desarrollo de como guías específicas o herramientas digitales, que faciliten la aplicación del marco.

Esta fase asegurará que el marco de trabajo propuesto evolucione, manteniendo su relevancia y efectividad en el entorno tecnológico y regulatorio de la IA responsable.

4 Desarrollo específico de la contribución

4.1 Elaboración del formulario.

Las directrices éticas para una IA fiable proporcionan 7 áreas y subáreas en los cuales se debería evaluar una IA y son los siguientes:

Tabla 23. Áreas y subáreas que proponen las directrices éticas para una IA fiable.

Área	Subárea
1. Agentes humanos y supervisión	1.1 Derechos fundamentales
	1.2 Agentes humanos
	1.3 Supervisión humana
2. Solidez técnica y seguridad	2.1 Resistencia a los ataques y seguridad
	2.2 Plan de emergencia y seguridad general
	2.3 Precisión
	2.4 Fiabilidad y reproducibilidad
3. Privacidad y gobernanza del dato	3.1 Respeto de la intimidad y protección de datos
	3.2 Calidad e integridad de los datos
	3.3 Acceso a los datos
4. Transparencia	4.1 Trazabilidad
	4.2 Explicabilidad
	4.3 Comunicación
5. Diversidad, no discriminación y equidad	5.1 Evitar el sesgo injusto
	5.2 Accesibilidad y diseño universal
	5.3 Participación de las partes interesadas
6. Bienestar social y medioambiental	6.1 IA sostenible y respetuosa con el medio ambiente
	6.2 Impacto social
7. Rendición de cuentas	7.1 Auditoría
	7.2 Minimizar y notificar el impacto negativo
	7.3 Documentar los compromisos y compensaciones
	7.4 Capacidad de reparación

Fuente: Elaboración propia.

Como ya se indicó en el capítulo 3, el marco de trabajo considera como base las directrices éticas para una IA fiable, las cuales señalan que la evaluación de fiabilidad de una IA debe realizarse en 7 áreas y 22 subáreas que cubren todo el espectro de la responsabilidad en IA.

Considerando los documentos descritos en el numeral 3.3.1, se elaboraron preguntas para cada una de las áreas y subáreas mencionadas, con el objetivo de cubrir toda la normativa y prácticas reales en IA responsable, de la siguiente manera:

- Directrices éticas para una IA fiable (70 preguntas).
- Reglamento General de Protección de Datos (44 preguntas).
- Reglamento de IA de la Unión Europea (19 preguntas).
- Modelo de Transparencia Fundacional (18 preguntas).
- Prácticas éticas comunes de las grandes empresas de IA (16 preguntas).

A continuación se listan las preguntas obtenidas de cada documento, identificando cada una de ellas con un color específico.

4.1.1 Preguntas obtenidas de las Directrices Éticas para una IA fiable

Tabla 24. Preguntas por área y subárea obtenidas de las directrices éticas para una IA fiable.

DIRECTRICES ETICAS PARA UNA IA FIABLE	
Área de evaluación	Sub-área y preguntas propuestas
1. Agentes humanos y supervisión	1.1 Derechos fundamentales
	¿Ha llevado a cabo una evaluación de impacto sobre los derechos humanos fundamentales para identificar y documentar las posibles compensaciones antes del desarrollo del sistema y ha analizado las posibles contramedidas para reducir o justificar los riesgos identificados?
	¿Ha establecido mecanismos para recibir información sobre los sistemas de IA que pueden vulnerar los derechos humanos fundamentales?
	1.2 Agentes humanos
	¿Ha evaluado cuál es el valor añadido específico del sistema de AI para ayudar al usuario final humano en sus tareas y objetivos?
	¿Ha evaluado la asignación de tareas entre el sistema de AI y las personas para lograr una interacción significativa?
	¿Comunica el sistema de IA a los usuarios (finales) que está interactuando con un agente no humano y que una decisión, contenido, consejo o resultado es fruto de una decisión algorítmica?
	¿Se han proporcionado a los usuarios los conocimientos y herramientas necesarios para comprender e interactuar con los sistemas de IA de forma satisfactoria?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

DIRECTRICES ETICAS PARA UNA IA FIABLE	
Área de evaluación	Sub-área y preguntas propuestas
	¿Se ha permitido a los usuarios cuestionar razonablemente el sistema y evitar una dependencia excesiva del mismo?
	¿Se permite a los usuarios modificar una decisión basada únicamente en un tratamiento automatizado cuando ésta produzca efectos jurídicos en los usuarios o les afecte de forma significativa?
2. Solidez técnica y seguridad	2.1 Resistencia a los ataques y seguridad
	¿Ha tenido en cuenta distintos tipos y naturalezas de vulnerabilidades, como la contaminación de datos, la fuga de modelos, la infraestructura física o los ciberataques, y ha adoptado las contramedidas respectivas?
	¿Ha comprobado cómo se comporta su sistema, producto o servicio de IA en situaciones y entornos inesperados?
	¿Ha tomado medidas preventivas adecuadas contra el posible uso indebido del modelo, producto o servicio de IA (como bots maliciosos, desinformación, creación de virus, etc.)?
	2.2 Plan de emergencia y seguridad general
	¿Ha calculado el impacto probable de un fallo de su sistema de IA cuando proporciona resultados erróneos, deja de estar disponible o proporciona resultados socialmente inaceptables (por ejemplo, discriminación)?
	¿Existe un plan para mitigar o gestionar los riesgos de resultados erróneos o socialmente inaceptables?
	¿Ha definido umbrales y establecidos procedimientos de gobernanza para activar planes alternativos/de emergencia?
	¿Ha comprobado y considerado el nivel de riesgo que tendría el sistema de IA funcionando con el plan alternativo?
	¿Dispone de una póliza de seguros para hacer frente a posibles daños (a los usuarios, al medioambiente, etc) causados por el sistema, producto o servicio de IA?
	¿Ha considerado y tenido en cuenta las normas de responsabilidad y protección de los consumidores en el desarrollo de producto, o servicio de IA?
	2.3 Precisión
	¿Ha evaluado qué nivel y definición de precisión se requerirían en el contexto del sistema de IA y del caso de uso?
	¿Ha comprobado qué perjuicios se causarían si el sistema de IA realiza predicciones inexactas?
	¿Ha establecido métodos para medir si su sistema está realizando una cantidad inaceptable de predicciones inexactas?
	2.4 Fiabilidad y reproducibilidad
	¿Ha establecido y documentado una estrategia para supervisar y comprobar si el sistema de IA cumple los objetivos, propósitos y aplicaciones previstos en el diseño e implementación?
	¿Ha documentado contextos específicos o condiciones particulares necesarias para garantizar la reproductibilidad del sistema de IA?
	3.1 Respeto de la intimidad y protección de datos

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

DIRECTRICES ETICAS PARA UNA IA FIABLE	
Área de evaluación	Sub-área y preguntas propuestas
3. Privacidad y gobernanza de datos	En función del caso de uso, ¿ha establecido un mecanismo que permita a otras personas señalar problemas relacionados con la privacidad o la protección de datos en la recopilación y procesamiento de datos del sistema, producto o servicio de IA?
	¿Consideró formas de desarrollar el sistema de IA (o de entrenar el modelo, producto o servicio), con y sin un uso mínimo de datos potencialmente sensibles personales?
	¿Incorporó en su sistema, producto o servicio de IA, mecanismos de notificación y control de los datos personales en función del caso de uso (como el consentimiento válido y la posibilidad de revocarlo, cuando proceda)?
	¿Ha tomado medidas para mejorar la privacidad, como el cifrado, la anonimización y la agregación?
	En caso de que exista un responsable de la protección de datos (RPD), ¿se ha implicado a esta persona en una fase temprana del proceso de diseño e implantación?
	3.2 Calidad e integridad de los datos
	¿Ha establecido mecanismos de supervisión para la recogida, el almacenamiento, el tratamiento y el uso de los datos?
	¿Ha evaluado en qué medida controla la calidad de las fuentes de datos externas utilizadas para su modelo, producto o servicio de IA?
	3.3 Acceso a los datos
	¿ Utiliza protocolos, procesos y procedimientos para gestionar y garantizar una gobernanza adecuada de los datos?
	¿Ha evaluado quién puede acceder a los datos de los usuarios y en qué circunstancias?
	¿Se ha asegurado de que estas personas están cualificadas con las competencias necesarias para comprender los detalles de la política de protección de datos?
	¿Ha garantizado un mecanismo de supervisión para registrar cuándo, dónde, cómo, por quién y con qué fin específico se ha accedido a los datos?
4. Transparencia	4.1 Trazabilidad
	¿Ha establecido medidas que garanticen la trazabilidad de los métodos de diseño, desarrollo y validación del sistema algorítmico?
	¿Ha establecido medidas que garanticen la trazabilidad de los resultados del sistema algorítmico?
	4.2 Explicabilidad
	¿Ha evaluado hasta qué punto las decisiones y, por tanto, los resultados tomados por el sistema de IA pueden ser comprendidos por sus propios constructores?
	¿Ha garantizado una explicación de por qué el sistema tomó una determinada decisión que dio lugar a un determinado resultado que todos los usuarios puedan entender?
	¿Puede cambiar y actualizar la forma de entrenar y probar los datos a lo largo del tiempo?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

DIRECTRICES ETICAS PARA UNA IA FIABLE	
Área de evaluación	Sub-área y preguntas propuestas
	4.3 Comunicación
	¿Ha establecido mecanismos para informar de forma clara a los destinatarios o usuarios (finales) sobre las razones y los criterios en los que se basan los resultados del sistema de IA?
	¿Ha establecido procesos que tengan en cuenta los comentarios de los usuarios y los utilicen para adaptar /mejorar el sistema?
	¿Se comunicaron a los posibles usuarios, los riesgos potenciales o percibidos, como la parcialidad, la fiabilidad o las posibles deficiencias?
	Según el caso de uso, ¿ha considerado la comunicación de los riesgos potenciales o percibidos, hacia otras audiencias, terceros o el público en general?
5. Diversidad, no discriminación y equidad	5.1 Evitar los prejuicios injustos
	¿Ha evaluado las posibles limitaciones de sesgo derivadas de la composición de los conjuntos de datos utilizados (por ejemplo, diversidad, representatividad)?
	¿Ha realizado pruebas para poblaciones específicas o casos de uso problemáticos?
	¿Cómo se comprueba y controla el posible sesgo durante la fase de desarrollo, despliegue y uso del sistema?
	¿Ha tenido en cuenta, además de a los usuarios (finales), a otras personas potencialmente afectadas de forma indirecta por el sistema, producto o servicio de IA?
	¿Se ha asegurado de contar con una definición operativa adecuada de "sesgo" que aplique en el diseño de sistemas de IA?
	5.2 Accesibilidad y diseño universal
	¿Ha evaluado si el sistema de IA es utilizable por personas con necesidades especiales o con discapacidad o en riesgo de exclusión?
	¿Ha considerado el reglamento general de protección de datos y el reglamento de IA, durante la fase de desarrollo del sistema de IA?
	¿Ha tenido en cuenta el impacto de su sistema de IA en el público potencial de usuarios específicos?
	¿Ha evaluado si el equipo que participa en la creación del sistema de IA es representativo de la población destinataria en general, teniendo en cuenta también otros grupos que podrían verse afectados tangencialmente?
	¿Ha evaluado si podría haber personas o grupos que se vieran desproporcionadamente afectados por las consecuencias negativas por la implementación y uso del sistema IA?
	5.3 Participación de las partes interesadas
	¿Ha previsto algún mecanismo para incluir la participación de partes interesadas con distintas necesidades en el desarrollo y el uso del sistema de IA?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

DIRECTRICES ETICAS PARA UNA IA FIABLE	
Área de evaluación	Sub-área y preguntas propuestas
	¿Recibió información de diferentes grupos multi-disciplinarios, que representaran diferentes contextos y experiencias para desarrollar el sistema, producto o servicio de IA?
	¿Recibió comentarios sobre el uso del sistema por parte de otros equipos o grupos que representaron diferentes contextos y experiencias?
6. Bienestar social y medioambiental	6.1 IA sostenible y respetuosa con el medio ambiente
	¿Ha establecido mecanismos para medir el impacto medioambiental del desarrollo, despliegue y uso del sistema de IA (por ejemplo, el tipo de energía, cantidad, climatización que están siendo utilizadas por los centros de datos)?
	¿Ha tomado medidas para reducir el impacto ambiental del ciclo de vida de su sistema de IA?
	6.2 Impacto social
	¿Ha evaluado si su sistema de IA anima a los humanos a desarrollar apego y empatía hacia el sistema?
	¿Se ha asegurado de que el sistema de IA señala claramente que su interacción social es simulada y que no tiene capacidades de "entender" y "sentir"?
	¿Se ha asegurado de que se comprenden bien por parte de los usuarios, las repercusiones sociales positivas y negativas del sistema de IA?
7. Rendición de cuentas	7.1 Auditabilidad
	¿Se ha asegurado de que el sistema de IA pueda ser auditado por un organismo interno autónomo?
	7.2 Minimizar y notificar el impacto negativo
	¿Ha evaluado qué trabajadores o ramas del equipo participan en todas las fases del sistema de IA (desde el diseño hasta el uso)?
	¿Se brindó capacitación y educación a los diferentes grupos interesados para fomentar el desarrollo de prácticas relacionadas con la rendición de cuentas?
	¿Se capacita también en el posible marco jurídico aplicable al sistema de IA?
	¿Ha previsto algún tipo de orientación externa para supervisar la ética y la responsabilidad del sistema de IA, además de las iniciativas internas?
	7.3 Documentación de los compromisos
	¿Se ha implementado un mecanismo para identificar los intereses y valores relevantes asociados al sistema de IA, así como las posibles compensaciones entre ellos?
	7.4 Capacidad de recurso
	¿Ha establecido un conjunto adecuado de mecanismos que permitan la reparación en caso de que se produzca algún daño o impacto adverso?
	¿Cuenta con mecanismos para comunicar a los usuarios finales y a terceros las opciones disponibles para ejercer sus derechos o presentar sus quejas?

Fuente: Elaboración propia.

4.1.2 Preguntas obtenidas del Reglamento General de Protección de Datos

Tabla 25. Preguntas por área y subárea obtenidas del reglamento general de protección de datos de la UE.

REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS	
Área de evaluación	Sub-área y preguntas propuestas
1. Agentes humanos y supervisión	1.2 Agentes humanos
	¿Se han designado claramente los responsables y encargados del tratamiento de datos, y conocen sus obligaciones relacionadas con el control y la supervisión humana (RGPD-Art. 24, 28)?
	¿Existe un Delegado de Protección de Datos (DPD) designado (RGPD-Art. 37, 38)?
	¿Se realizan evaluaciones de impacto en la protección de datos para identificar riesgos asociados con procesos automatizados que afectan a los usuarios (RGPD-Art.35)?
	1.3 Supervisión humana
	¿Se aplican medidas técnicas y organizativas para garantizar que los datos personales se procesen de manera lícita (conforme a la ley) y transparente respetando los principios de protección de datos (RGPD-Art.25)?
	¿Se obtiene el consentimiento explícito del interesado para decisiones automatizadas (RGPD-Art-6, 9)?
2. Solidez técnica y seguridad	2.2 Plan de Fallback y seguridad general
	¿Tiene implementados planes de contingencia para proteger la infraestructura de hardware, software y datos de su sistema de IA (RGPD-Art.32)?
	¿Tiene implementados mecanismos de reversión a versiones previas o plan de fallback para su sistema de IA(RGPD-Art.32)?
	2.3 Precisión
	¿Tiene implementados mecanismos para garantizar la exactitud de los datos procesados (RGPD-Art.5)?
	2.4 Fiabilidad y reproducibilidad
	¿Tiene mecanismos y procesos para garantizar la fiabilidad del sistema de IA (RGPD-Art.32)?
3. Privacidad y gobernanza del dato	3.1 Respeto de privacidad y protección de datos
	¿El personal de su empresa está capacitado para garantizar la transparencia en el tratamiento de datos personales? (RGPD-Art.12)
	3.2 Calidad e integridad de los datos
	¿Existen en su organización procesos para rectificar datos personales inexactos? (RGPD-Art.16)
	3.3 Acceso a los datos
	¿Los interesados tienen garantías de que se ejecutará su derecho de acceso de sus datos personales y como son tratados? (RGPD-Art.15)
4. Transparencia	4.1 Trazabilidad
	¿Su organización verifica que los procesos de trazabilidad se cumplen? (RGPD-Art.30)
	4.2 Explicabilidad

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS	
Área de evaluación	Sub-área y preguntas propuestas
	¿Su empresa tiene políticas para garantizar la explicabilidad de las decisiones automatizadas? (RGPD-Art.22)
	¿Existen procedimientos para proporcionar explicaciones claras sobre el tratamiento automatizado de datos? (RGPD-Art.22)
	4.3 Comunicación
	¿Existen métodos para comunicar a los interesados sobre el uso de sus datos personales? (RGPD-Art.12)
	¿Existen procesos para verificar la claridad y comprensión de las notificaciones enviadas a los interesados, en relación con el uso de sus datos (RGPD-Art.12)?
	¿Existen estrategias para mejorar la comunicación al usuario final, sobre las políticas de tratamiento de datos? (RGPD-Art.14)
	¿Existen procedimientos para gestionar las solicitudes de información realizadas por los interesados? (RGPD-Art.15)
5. Diversidad, no discriminación y equidad	5.1 Evitar el sesgo injusto
	¿Existen medidas para identificar y eliminar posibles sesgos en los procesos automatizados? (RGPD-Art.22)
	¿Existen auditorías para garantizar la equidad en las decisiones automatizadas? (RGPD-Art.35)
	¿Existen procesos internos para mitigar los riesgos de sesgos injustos en los sistemas de IA? (RGPD-Art.25)
	¿Se realizan capacitaciones para los equipos responsables (personal a cargo) para evitar sesgos en el diseño de sistemas? (RGPD-Art.24)
	5.2 Accesibilidad y diseño universal
	¿La empresa incorporó los principios de diseño universal en los procesos tecnológicos relacionados al sistema, producto o servicio de IA? (RGPD-Art.25)
	¿Existen herramientas implementadas en el modelo, producto o servicio de IA, para asegurar la accesibilidad de los datos personales a los interesados? (RGPD-Art.15)
	5.3 Participación de los interesados
	¿Existen mecanismos para involucrar a los interesados en las decisiones relacionadas con el tratamiento de datos? (RGPD-Art.12)
6. Bienestar social y medioambiental	¿Toma en cuenta las opiniones de los interesados en los procesos de toma de decisiones? (RGPD-Art.12)
	¿Dispone de métodos de retroalimentación para los interesados en relación a las políticas de tratamiento de datos? (RGPD-Art.13)
	6.1 IA sostenible y respetuosa con el medio ambiente
	¿Se han implementado políticas para garantizar que su modelo, producto o servicio de IA sea sostenible y respetuoso con el medio ambiente? (RGPD-Art.5)
	¿Dispone de estrategias para minimizar el consumo energético de su modelo, producto o servicio de IA? (RGPD-Art.25)
	¿Se evalúan los impactos medioambientales asociados con el uso del modelo, producto o servicio de IA? (RGPD-Art.35)
	¿Existen sistemas para el monitoreo de la sostenibilidad de las operaciones del modelo, producto o servicio IA? (RGPD-Art.32)
	6.2 Impacto social

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

REGLAMENTO GENERAL DE PROTECCIÓN DE DATOS	
Área de evaluación	Sub-área y preguntas propuestas
	¿Adopta medidas para garantizar un impacto social positivo del tratamiento de datos personales en su sistema de IA? (RGPD-Art.35)
	¿Su organización ha implementado medidas proactivas, para evaluar y mejorar los efectos sociales de usar su modelo, producto o sistema de IA? (RGPD-Art.24)
	¿ Su organización participa activamente, en la evaluación de impacto de sus actividades con el uso de datos personales? (RGPD-Art.13)
7. Rendición de cuentas	7.1 Auditoría
	¿Mantiene registros de auditoría para evaluar el cumplimiento normativo? (RGPD-Art.30)
	7.2 Minimizar y notificar el impacto negativo
	¿Dispone de procedimientos para detectar y notificar impactos negativos relacionados con el tratamiento de datos? (RGPD-Art.33)
	¿Posee mecanismos para mitigar riesgos derivados del tratamiento de datos? (RGPD-Art.32)
	¿Cuenta con procedimientos de comunicación establecidos para notificar sobre incidentes de seguridad? (RGPD- Art.34)
	7.3 Documentar los compromisos y compensaciones
	¿ Mantiene registros para documentar compromisos relacionados con la protección de datos? (RGPD-Art.30)
	¿Dispone y almacena documentación que acredita su cumplimiento con las obligaciones del RGPD? (RGPD-Art.24)
	7.4 Capacidad de reparación
	¿Dispone de mecanismos para garantizar la capacidad de reparación en caso de incumplimientos? (RGPD-Art.32)
	¿Cuenta con estrategias para indemnizar o reparar a los afectados en caso de incidentes de brechas de seguridad que impliquen riesgos significativos para sus derechos y libertades ? (RGPD-Art.34)
	¿Existen procedimientos para gestionar las reclamaciones de los interesados en caso de incidentes de seguridad? (RGPD-Art.24)

Fuente: Elaboración propia.

4.1.3 Preguntas obtenidas del Reglamento de IA

Tabla 26. Preguntas por área y subárea obtenidas del reglamento de IA de la UE.

REGLAMENTO DE IA	
Área de evaluación	Sub-área y preguntas propuestas
1. Agentes humanos y supervisión	1.2 Agentes humanos
	¿Los agentes humanos tienen el control sobre el sistema producto o servicio de IA?
	¿Los agentes humanos pueden intervenir en caso de fallo del sistema de IA?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

REGLAMENTO DE IA	
Área de evaluación	Sub-área y preguntas propuestas
	1.3 Supervisión humana
	¿La supervisión humana permite detener el sistema de IA en cualquier momento de operación?
2. Solidez técnica y seguridad	2.1 Resistencia ante ataques y seguridad
	¿Se realizan auditorías periódicas para garantizar la seguridad del sistema, producto o servicio de IA?
	2.2 Plan de Fallback y seguridad general
	¿El sistema de IA puede recuperar su funcionalidad tras un fallo crítico?
	2.3 Precisión
	¿El sistema de IA ha sido probado en múltiples escenarios para garantizar su precisión?
3. Privacidad y gobernanza del dato	¿Los resultados del sistema han sido validados con datos reales?
	3.1 Respeto de privacidad y protección de datos
	¿Se protege la confidencialidad de los datos personales utilizados por el sistema?
	3.2 Calidad e integridad de los datos
	¿Existen procesos para verificar la integridad de los datos utilizados por el sistema?
	3.3 Acceso a los datos
4. Transparencia	¿Existen auditorías para controlar el acceso a los datos del sistema?
	4.1 Trazabilidad
	¿Es posible rastrear las decisiones tomadas por el sistema de IA?
	4.2 Explicabilidad
	¿Las decisiones del sistema de IA son explicables a los usuarios?
	4.3 Comunicación
5. Diversidad, no discriminación y equidad	¿El sistema informa claramente sobre sus limitaciones y capacidades?
	¿Los usuarios son informados si están interactuando con un sistema de IA?
	5.2 Accesibilidad y diseño universal
	¿El sistema está diseñado para ser accesible a personas con discapacidades (usabilidad real)?
	5.3 Participación de los interesados

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

REGLAMENTO DE IA	
Área de evaluación	Sub-área y preguntas propuestas
	¿Se consultó a las partes interesadas durante el diseño del sistema?
6. Bienestar social y medioambiental	6.2 Impacto social
	¿El sistema de IA contribuye positivamente al bienestar social en general?
7. Rendición de cuentas	7.1 Auditoría
	¿Existen registros claros para facilitar las auditorías del sistema?
	7.3 Documentar los compromisos y compensaciones
	¿El sistema incluye documentación sobre compromisos éticos?

Fuente: Elaboración propia.

4.1.4 Preguntas obtenidas del Índice de Transparencia Fundacional

Tabla 27. Preguntas por área y subárea obtenidas de los indicadores del índice de transparencia fundacional.

ÍNDICE DE TRANSPARENCIA FUNDACIONAL	
Área de evaluación	Sub-área y preguntas propuestas
1. Agentes humanos y supervisión	1.1 Derechos Fundamentales
	¿Existen indicadores específicos para verificar el cumplimiento de los derechos fundamentales de la UE en el modelo, servicio o producto?
	1.2 Agentes Humanos
	¿Se verifica la eficacia de los agentes humanos en la supervisión del modelo, producto o servicio?
	1.3 Supervisión Humana
	¿Existen protocolos claros para que humanos supervisen el modelo, producto o servicio de IA durante su operación?
2. Solidez técnica y seguridad	2.1 Resistencia ante ataques y seguridad
	¿El modelo, producto o servicio de IA cuenta con estrategias para resistir ataques externos?
	2.3 Precisión
	¿El modelo garantiza niveles consistentes de precisión durante su operación?
	2.4 Fiabilidad y reproducibilidad
	¿Los resultados del modelo, producto o servicio de IA, son reproducibles en diferentes entornos?
3. Privacidad y gobernanza del dato	3.1 Respeto de privacidad y protección de datos

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

ÍNDICE DE TRANSPARENCIA FUNDACIONAL	
Área de evaluación	Sub-área y preguntas propuestas
	¿Se cumplen normativas de privacidad (RGPD), en los datos utilizados por el modelo, producto o servicio de IA?
	3.3 Acceso a los datos
	¿Existen restricciones claras para limitar el acceso a los datos sensibles al modelo, producto o servicio de IA?
4. Transparencia	4.1 Trazabilidad
	¿Es posible rastrear el origen de los datos utilizados en el modelo?
	4.2 Explicabilidad
	¿Se proporciona una explicación clara sobre cómo el modelo toma decisiones?
	4.3 Comunicación
	¿El modelo comunica riesgos y beneficios de manera transparente al público?
5. Diversidad, no discriminación y equidad	5.1 Evitar el sesgo injusto
	¿Se verifica la equidad del modelo para beneficiar a todos los grupos de usuarios?
	5.2 Accesibilidad y diseño universal
	¿El modelo cumple con estándares internacionales de accesibilidad (cumplimiento técnico)?
6. Bienestar social y medioambiental	6.1 IA sostenible y respetuosa con el medio ambiente
	¿El modelo incorpora estrategias para minimizar el impacto ambiental del modelo, producto o servicio de IA?
	6.2 Impacto social
	¿Se utilizan indicadores sociales para medir el impacto del modelo, producto o servicio?
7. Rendición de cuentas	7.1 Auditoría
	¿El modelo es auditado de manera regular para garantizar la transparencia?
	7.2 Minimizar y notificar el impacto negativo
	¿Los usuarios son notificados sobre posibles impactos negativos identificados en el modelo, producto o servicio de IA?
	7.3 Documentar los compromisos y compensaciones
	¿Se documentan los compromisos éticos relacionados con el uso del modelo?

Fuente: Elaboración propia.

4.1.5 Preguntas obtenidas de los principios éticos comunes de las empresas desarrolladoras de IA.

Tabla 28. Preguntas por área y subárea obtenidas de los principios éticos comunes en las empresas desarrolladoras de IA.

PRINCIPIOS ÉTICOS COMUNES EN LAS EMPRESAS DESARROLLADORAS DE IA	
Área de evaluación	Sub-área y preguntas propuestas
2. Solidez técnica y seguridad	2.1 Resistencia ante ataques y seguridad
	¿Las auditorías programadas garantizan la resistencia a ataques en seguridad?
	¿La documentación de resistencia a ataques respalda su efectividad en el marco de seguridad?
3. Privacidad y gobernanza del dato	3.1 Respeto de privacidad y protección de datos
	¿Las auditorías programadas garantizan la práctica adecuada de procesamiento de datos en lo relacionado a privacidad?
	¿La documentación de procesamiento de datos respalda su efectividad en lo relacionado a privacidad?
4. Transparencia	4.2 Explicabilidad
	¿Las auditorías programadas garantizan la práctica adecuada de explicabilidad con relación a la transparencia?
	¿La documentación de explicabilidad respalda su efectividad en el marco de transparencia?
5. Diversidad, no discriminación y equidad	5.1 Evitar el sesgo injusto
	¿Las auditorías programadas garantizan la práctica adecuada de mitigación de sesgos?
	¿La documentación de mitigación de sesgos respalda su efectividad en lo relacionado a evitar sesgos?
	5.2 Accesibilidad y diseño universal
	¿Las auditorías programadas garantizan la práctica adecuada de diversidad cultural en inclusión?
	¿La documentación de diversidad cultural respalda su efectividad en el marco de inclusión?
	¿Las auditorías programadas garantizan la práctica adecuada de tratamiento igualitario en equidad?
	¿La documentación de tratamiento igualitario respalda su efectividad en el marco de equidad?
6. Bienestar social y medioambiental	6.2 Impacto social
	¿Las auditorías programadas garantizan la práctica adecuada de respeto a derechos humanos que permitan cumplir la normativa legal?
	¿La documentación de respeto a derechos humanos respalda su efectividad en lo relacionado a cumplir con leyes?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

PRINCIPIOS ÉTICOS COMUNES EN LAS EMPRESAS DESARROLLADORAS DE IA	
Área de evaluación	Sub-área y preguntas propuestas
7. Rendición de cuentas	7.1 Auditoría
	¿Las auditorías programadas aseguran la correcta implementación de prácticas relacionadas con la rendición de cuentas?
	¿La documentación generada en las auditorías demuestra su eficacia en el contexto de la rendición de cuentas?
	7.2 Minimizar y notificar el impacto negativo
	¿Las auditorías programadas garantizan la práctica adecuada de filtrado de temas sensibles en evitar contenido inapropiado?
	¿La documentación sobre el filtrado de temas sensibles garantiza su eficacia en la prevención de contenido inapropiado?
	7.3 Documentar los compromisos y compensaciones
	¿Las auditorías programadas garantizan la práctica adecuada de resultados fiables en veracidad?
	¿La documentación sobre resultados fiables garantiza su efectividad en términos de veracidad?

Fuente: Elaboración propia.

Con esta documentación se llegó a un formulario de 171 preguntas cuyo detalle se encuentra en el Anexo C.

Las preguntas se elaboraron en Microsoft Excel y se establecieron respuestas cerradas entre las siguientes cinco (5) posibilidades:

1. La acción nunca se ha realizado.
2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.
3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.
4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.
5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.

Al seleccionar cada respuesta automáticamente se cuantifica el valor respectivo entre 1 y 5 que servirá para conocer el nivel global de IA responsable de la empresa, producto o servicio de IA.

En base a estas respuestas, se estableció que el marco de trabajo tiene 5 niveles y que son:

- **Nivel 1 – Iniciando:** La empresa no cuentan con políticas ni controles claros para el desarrollo de IA responsable. La concienciación sobre el tema es limitada o nula.
- **Nivel 2 -Básico:** La empresa ha comenzado a implementar controles básicos y a desarrollar políticas iniciales relacionadas con la IA responsable. El nivel de conocimiento de IA responsable es básico
- **Nivel 3- Intermedio:** La empresa tiene conocimientos intermedios de IA responsable y ejecuta regularmente ciertas acciones.
- **Nivel 4- Avanzado:** La empresa tiene un alto conocimiento de IA responsable y procura ejecutar todas las acciones que le es posible implementar, por lo cual este nivel se considera el ideal para declarar que se ha llegado a un mínimo en IA responsable.
- **Nivel 5 – Desarrollado:** Conoce y aplica plenamente la IA responsable en el modelo, producto o servicio de IA.

Considerando que diseñó el formulario en una hoja en Excel, es posible realizar una evaluación cualitativa global casi automáticamente, obteniéndose puntuaciones entre 171 (mínima) y 855 puntos (máxima). La sumatoria de los valores asignados a cada pregunta señalará el nivel de la IA responsable en la empresa, producto o servicio de IA.

Tabla 29. Evaluación del formulario para determinar el nivel global de IA responsable.

Áreas	Preguntas	Subáreas	Preguntas por subárea	Puntajes				
				Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5
1. Agentes humanos y supervisión	21	1.1 Derechos fundamentales	3	3	6	9	12	15
		1.2 Agentes humanos	12	12	24	36	48	60
		1.3 Supervisión humana	6	6	12	18	24	30
2. Solidez técnica y seguridad	27	2.1 Resistencia ante ataques y seguridad	7	7	14	21	28	35

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Áreas	Preguntas	Subáreas	Preguntas por subárea	Puntajes				
				Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5
		2.2 Plan de emergencia y seguridad general	9	9	18	27	36	45
		2.3 Precisión	7	7	14	21	28	35
		2.4 Fiabilidad y reproducibilidad	4	4	8	12	16	20
3. Privacidad y gobernanza de datos	21	3.1 Respeto de la intimidad y protección de datos	10	10	20	30	40	50
		3.2 Calidad e integridad de los datos	4	4	8	12	16	20
		3.3 Acceso a los datos	7	7	14	21	28	35
4. Transparencia	25	4.1 Trazabilidad	5	5	10	15	20	25
		4.2 Explicabilidad	9	9	18	27	36	45
		4.3 Comunicación	11	11	22	33	44	55
5. Diversidad, no discriminación y equidad	32	5.1 Evitar el sesgo injusto	12	12	24	36	48	60
		5.2 Accesibilidad y diseño universal	13	13	26	39	52	65
		5.3 Participación de los interesados	7	7	14	21	28	35
6. Bienestar social y medioambiental	17	6.1 IA sostenible y respetuosa con el medio ambiente	7	7	14	21	28	35
		6.2 Impacto social	10	10	20	30	40	50
7. Rendición de cuentas	28	7.1 Auditoría	6	6	12	18	24	30
		7.2 Minimizar y notificar el impacto negativo	10	10	20	30	40	50
		7.3 Documentar los compromisos y	7	7	14	21	28	35

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Áreas	Preguntas	Subáreas	Preguntas por subárea	Puntajes				
				Nivel 1	Nivel 2	Nivel 3	Nivel 4	Nivel 5
		compensaciones						
		7.4 Capacidad de reparación	5	5	10	15	20	25
	171	TOTAL, PUNTAJES	171	171	342	513	684	855

Fuente: Elaboración propia.

El nivel global de IA responsable en el cual se encuentra el modelo, producto o servicio de IA será determinado por la evaluación cuantitativa de las respuestas a cada una de las interrogantes planteadas en el cuestionario del Anexo C considerando la siguiente tabla:

Tabla 30. Rangos para determinar el nivel global de IA responsable.

Nivel	Puntaje mínimo	Puntaje máximo
1.- Iniciando	171	341
2.- Básico	342	512
3.- Intermedio	513	649
4.- Avanzado	650	799
5.- Desarrollado	800	855

Fuente: Elaboración propia.

A pesar de que según los cálculos de la tabla 29, al nivel 4 se llegaría con 684 puntos, este marco de trabajo establece que alcanzando los 650 puntos ya constituye el nivel 4, mínimo puntaje para considerar que un producto o servicio de IA es responsable según la normativa occidental plasmada en este trabajo.

Para conocer la realidad particular de la empresa, su organización interna, su nivel de compromiso con IA, se propone realizar el siguiente cuestionario de 10 preguntas que deberán ser solventadas previo a cualquier reunión con la empresa a la cual se realizará la evaluación de IA responsable.

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

1. Tipo de empresa: pública (perteneciente al estado) o privada.
2. Número de empleados: si no es posible responder el número exacto indicar al menos un rango.
3. Número de aplicaciones, productos o servicios de IA que se encuentran en
 - a. Desarrollo:
 - b. Implementación o despliegue:
 - c. Uso:
4. Especificar el giro del negocio de la empresa.
5. ¿El giro del negocio de la empresa es IA?
 - a. Si su respuesta es NO, ¿el producto o servicio de IA es para consumo propio o para clientes?
 - b. Si su respuesta es SI, ¿el producto o servicio de IA lo utilizan clientes finales o un tercero?
6. ¿Su producto o servicio de IA tiene acceso a información delicada de los usuarios finales (nombre, ID, datos biométricos, dirección física, número telefónico)?
7. ¿Su empresa y su personal tiene conocimiento de la ley de IA y del Reglamento General de Protección de Datos de la Unión Europea?
8. Por favor compartir un organigrama de su empresa.
9. Breve historia de su empresa.
10. Su empresa dispone de los siguientes roles, por favor especificar número de personas en cada uno:
 - a. Científico de datos,
 - b. Legal (relacionado a cumplimiento de la ley de IA o del reglamento de protección de datos),
 - c. Diseño de sistemas (analista funcional de IA),
 - d. Desarrollo de sistemas (técnico desarrollador de IA),
 - e. Chief Technology Officer (director de tecnología),
 - f. Experiencia de usuario,
 - g. CEO,
 - h. Director de información (Chief Information Officer).

Es necesario aclarar que el marco de trabajo propuesto utiliza el formulario para conocer el nivel de IA de una empresa de occidente en tres aspectos:

- Global: con la puntuación total que la ubica en niveles 1 a 5. (Tabla 30).
- Por cada una de las 7 áreas evaluadas.
- Integral: ponderando cada una de las áreas.

Para determinar el nivel de IA responsable por cada una de las 7 áreas e integral se utilizará la rueda de madurez descrita en el numeral 4.4

4.2 Aplicación del formulario – caso práctico

4.2.1 Recopilación de datos

Para la recopilación de datos, dos empresas privadas de Ecuador aceptaron realizar el formulario para conocer el nivel de IA responsable.

La primera es una empresa que se dedica a medir el nivel de popularidad de personajes políticos en las redes sociales, con sede en Quito - Ecuador. La segunda es una entidad bancaria con sede en Guayaquil – Ecuador.

La planificación original era aplicar el formulario a la empresa pequeña con giro del negocio en IA, obtener lecciones aprendidas, mejorar el formulario y aplicarlo en la empresa grande. Sin embargo, en la empresa pequeña no se logró terminar la aplicación del formulario por temas personales de su directora ejecutiva que le impidieron continuar con las reuniones virtuales. Por tal razón, se aplicó el formulario únicamente a la entidad bancaria, que, aunque tomó más tiempo del esperado por ser una empresa tan grande, se obtuvieron resultados inesperados.

La entidad bancaria con varias décadas de existencia, opera solo en Ecuador y según señalan sus empleados, está a la vanguardia en innovación y tecnología, razón por la cual tiene aproximadamente 192 modelos, productos y/o servicios de IA en diseño, desarrollo, implementación y uso, una gran cantidad considerando que su giro del negocio no es IA. Tiene acceso a información delicada de sus clientes finales como son: nombre completo, número de identificación nacional y/o pasaporte, dirección física y número telefónico.

La empresa tiene alrededor de 3.000 empleados a nivel nacional. Aceptó realizar varias sesiones de máximo 2 horas de duración con líderes de sus departamentos de: innovación, data y analítica y desarrollo e innovación. Los entrevistados señalaron que el personal si tiene conocimientos del reglamento general de protección de datos de la UE y del reglamento de IA de la UE.

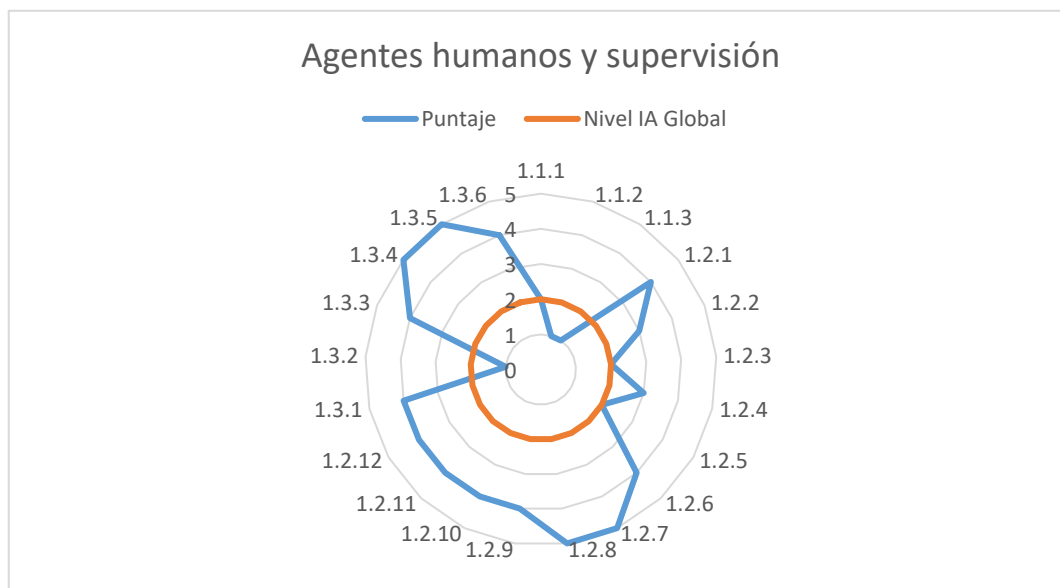
4.2.2 Análisis de cumplimiento

El formulario que se completó con las respuestas de la entidad financiera se encuentra en el Anexo D, y se obtuvo una puntuación global de 422 puntos. De acuerdo con la tabla de rangos de nivel global de IA responsable, ese puntaje corresponde al nivel 2 – básico.

Este nivel nos servirá de referencia para emitir las recomendaciones a fin de que cada una de las áreas se afiance en el nivel 2-básico y de ahí implemente paulatinamente las recomendaciones hasta alcanzar al menos el nivel 4 – Avanzado para que pueda considerarse que su productos y modelos de IA son responsables.

A continuación, se visualizan los gráficos de cómo se encuentra cada área evaluada con el formulario del marco de trabajo.

Figura 16. Estado del área 1 evaluada: agentes humanos y supervisión.

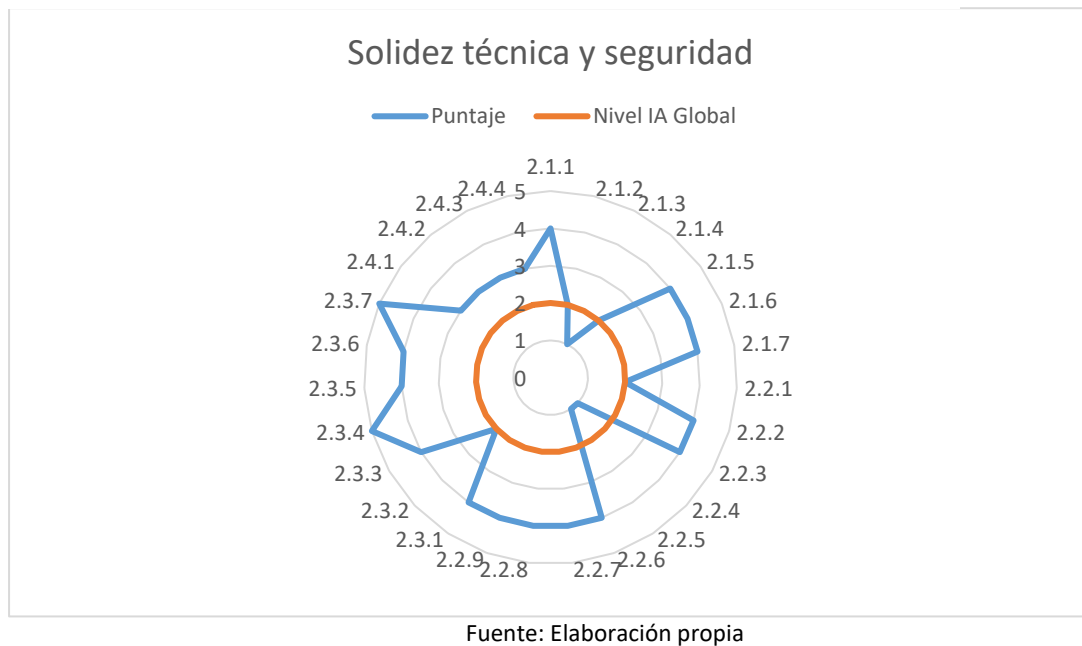


Fuente: Elaboración propia.

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

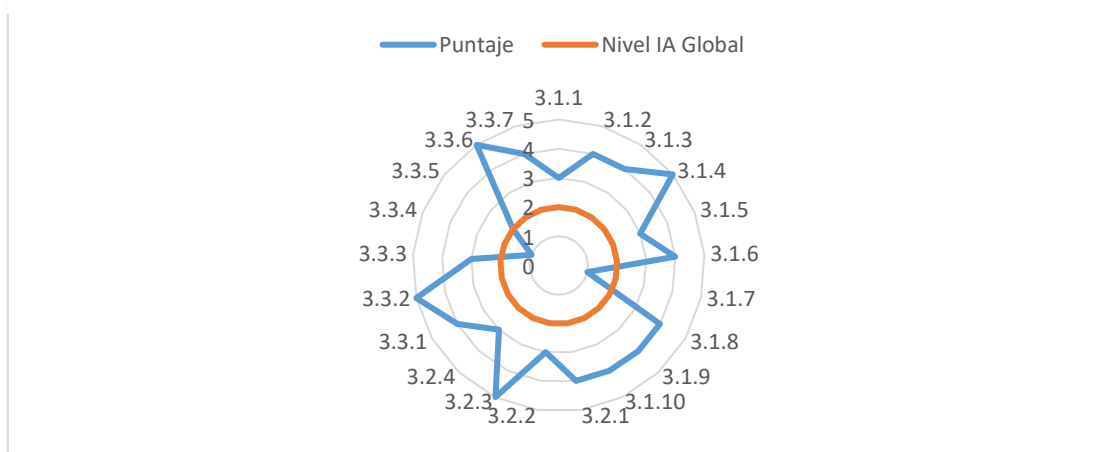
En el área 1, correspondiente a agentes humanos y supervisión la empresa evaluada tiene menos de 2 puntos (o nivel 2) en las preguntas 1.1.2, 1.2.3 y 1.3.2 que corresponden a derechos fundamentales y supervisión humana.

Figura 17. Estado del área 2 evaluada: solidez técnica y seguridad.



En el área 2, correspondiente a solidez técnica y seguridad la empresa evaluada tiene menos de 2 puntos (o es inferior a nivel 2) en las preguntas 2.1.3, 2.2.4 y 2.2.5 que corresponden a resistencia a los ataques, seguridad y plan de emergencia.

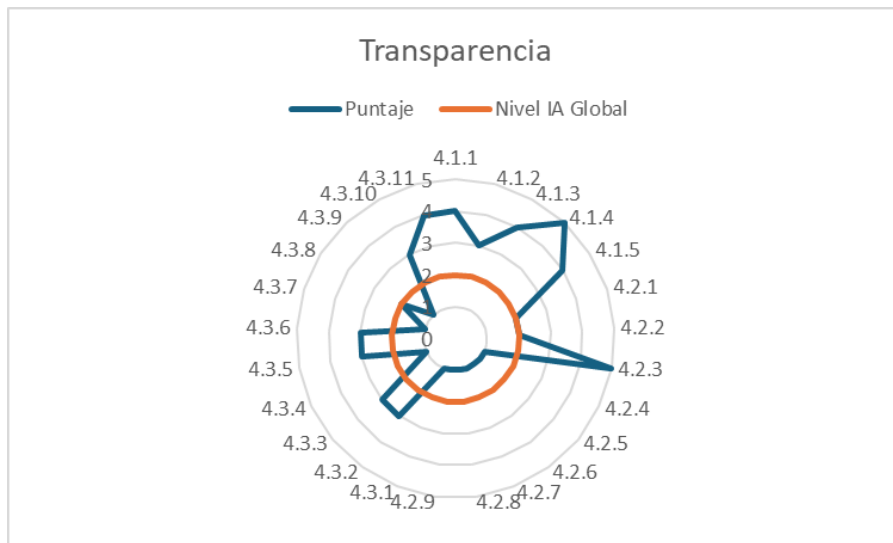
Figura 18. Estado del área 3 evaluada Privacidad y gobernanza de datos.



Fuente: Elaboración propia

En el área 3, correspondiente a privacidad y gobernanza del dato la empresa evaluada tiene menos de 2 puntos (o es inferior a nivel 2) en las preguntas 3.1.7 y 3.3.4 que corresponden a respeto a la intimidad, protección de datos y acceso a los datos.

Figura 19. Estado del área evaluada 4 – Transparencia.

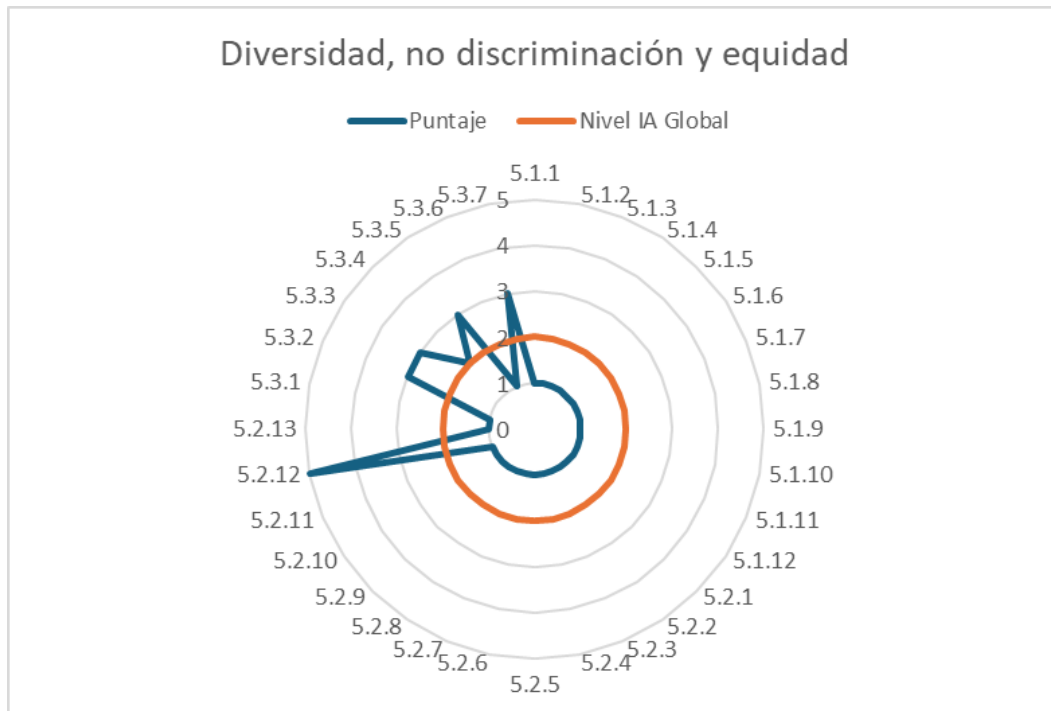


Fuente: Elaboración propia

En el área 4, correspondiente a privacidad, la empresa evaluada tiene menos de 2 puntos (o es inferior a nivel 2) en las preguntas 4.2.4, 4.2.5, 4.2.6, 4.2.7, 4.2.8, 4.2.9, 4.3.1, 4.3.4, 4.3.7, y 4.3.9 que corresponden a respeto a explicabilidad y comunicación.

Figura 20. Estado del área evaluada 5 – Diversidad, no discriminación y equidad.

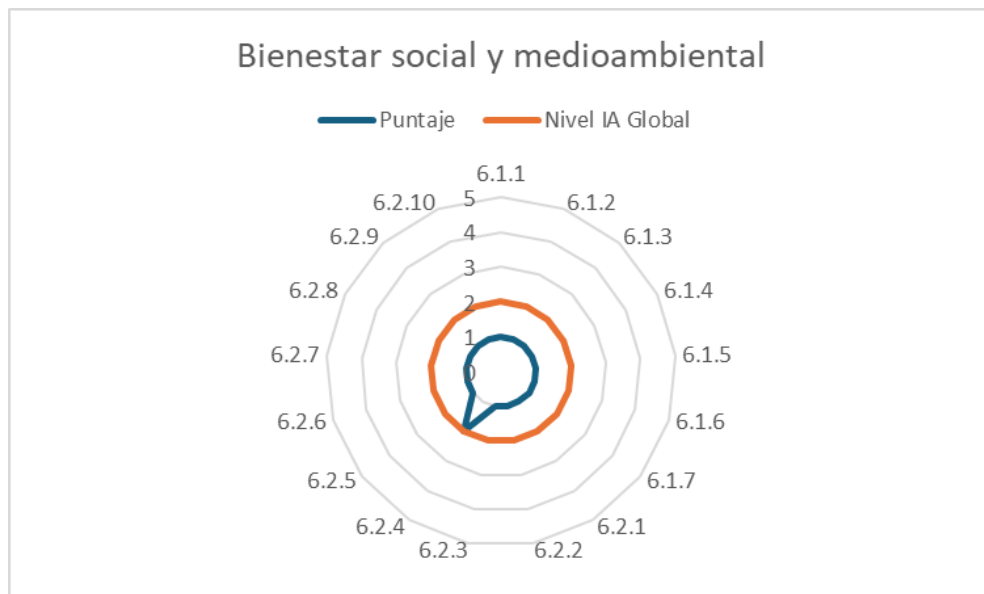
Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable



Fuente: Elaboración propia

En el área 5, correspondiente a diversidad, no discriminación y equidad, la empresa evaluada tiene menos de 2 puntos (o es inferior a nivel 2) en las preguntas 5.1.1, 5.1.2, 5.1.3, 5.1.4, 5.1.5, 5.1.6, 5.1.7, 5.1.8, 5.1.9, 5.1.10, 5.1.11, 5.1.12, 5.2.1, 5.2.2, 5.2.3, 5.2.4, 5.2.5, 5.2.6, 5.2.7, 5.2.8, 5.2.9, 5.2.10, 5.2.11, 5.2.13, 5.3.1 y 5.3.6 que corresponden a evitar el sesgo injusto, accesibilidad, diseño universal y participación de las partes interesadas. Es importante señalar que solo en 6 preguntas es igual o superior al nivel 2.

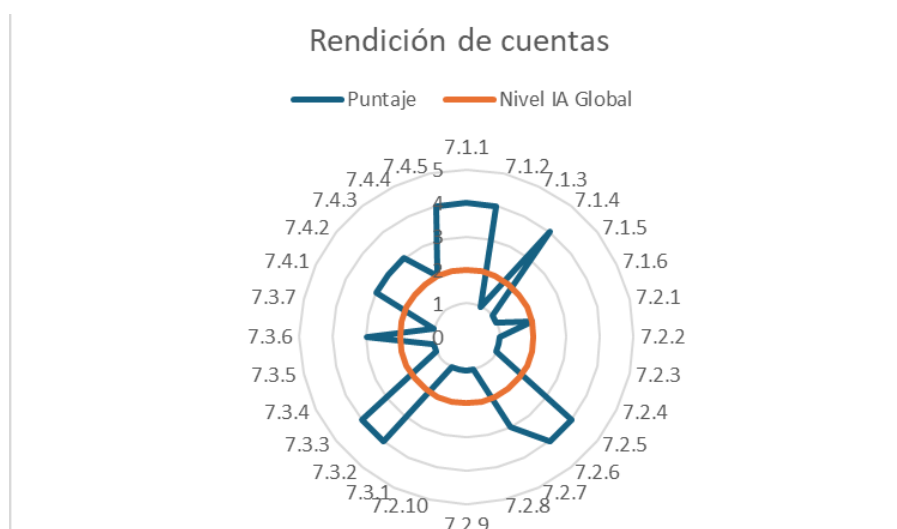
Figura 21. Estado del área evaluada 6 – Bienestar social y medioambiental



Fuente: Elaboración propia

En el área 6, correspondiente a bienestar social y medioambiental, la empresa evaluada tiene menos de 2 puntos (o es inferior a nivel 2) en las preguntas 6.1.1, 6.1.2, 6.1.3, 6.1.4, 6.1.5, 6.1.6, 6.1.7, 6.2.1, 6.2.2, 6.2.3, 6.2.5, 6.2.6, 6.2.7, 6.2.8, 6.2.9 y 6.2.10 que corresponden a IA sostenible y respetuosa con el medio ambiente e impacto social de la IA. Es importante señalar que solo en 1 preguntas es igual al nivel 2 (pregunta 6.2.4).

Figura 22. Estado del área evaluada 7 – Rendición de cuentas



Fuente: Elaboración propia

En el área 7, correspondiente a rendición de cuentas, la empresa evaluada tiene menos de 2 puntos (o es inferior a nivel 2) en las preguntas 7.1.3, 7.1.5, 7.1.6, 7.2.2, 7.2.3, 7.2.4, 7.2.8, 7.2.9, 7.2.10, 7.3.1, 7.3.4, 7.3.5 y 7.3.7 que corresponden a auditoría, minimizar y notificar el impacto negativo y documentar compromisos y compensaciones.

Como es evidente en las figuras previas, en todas las áreas hay subáreas que obtuvieron evaluación inferior a 1 y es ahí donde se enfocarán las recomendaciones prioritarias hacia la empresa.

Dado que cada subárea tiene un número distinto de preguntas, se asignó un factor de ponderación a cada subárea, con el objetivo de que cada una de ellas aportara un mismo peso a la evaluación, conforme a la siguiente tabla:

Tabla 31. Ponderación por cada pregunta del formulario

Áreas	Porcentaje en la rueda	Preguntas	Subáreas	Porcentaje de cada subárea	Preguntas por subárea	Porcentaje por pregunta	Factor de ponderación
1. Agentes humanos y supervisión	14,29	21	1.1 Derechos fundamentales	4,77	3	1,59	0,0159
			1.2 Agentes humanos	4,76	12	0,4	0,004
			1.3 Supervisión humana	4,76	6	0,79	0,0079
2. Solidez técnica y seguridad	14,28	27	2.1 Resistencia ante ataques y seguridad	3,57	7	0,51	0,0051
			2.2 Plan de emergencia y seguridad general	3,57	9	0,4	0,004
			2.3 Precisión	3,57	7	0,51	0,0051
			2.4 Fiabilidad y reproducibilidad	3,57	4	0,89	0,0089
3. Privacidad y gobernanza de datos	14,29	21	3.1 Respeto de la intimidad y protección de datos	4,77	10	0,48	0,0048
			3.2 Calidad e integridad de los datos	4,76	4	1,19	0,0119
			3.3 Acceso a los datos	4,76	7	0,68	0,0068
4. Transparencia	14,29	25	4.1 Trazabilidad	4,77	5	0,95	0,0095

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Áreas	Porcentaje en la rueda	Preguntas	Subáreas	Porcentaje de cada subárea	Preguntas por subárea	Porcentaje por pregunta	Factor de ponderación
			4.2 Explicabilidad	4,76	9	0,53	0,0053
			4.3 Comunicación	4,76	11	0,43	0,0043
5. Diversidad, no discriminación y equidad	14,29	32	5.1 Evitar el sesgo injusto	4,77	12	0,4	0,004
			5.2 Accesibilidad y diseño universal	4,76	13	0,37	0,0037
			5.3 Participación de los interesados	4,76	7	0,68	0,0068
6. Bienestar social y medioambiental	14,28	17	6.1 IA sostenible y respetuosa con el medio ambiente	7,14	7	1,02	0,0102
			6.2 Impacto social	7,14	10	0,71	0,0071
7. Rendición de cuentas	14,28	28	7.1 Auditoría	3,57	6	0,6	0,006
			7.2 Minimizar y notificar el impacto negativo	3,57	10	0,36	0,0036
			7.3 Documentar los compromisos y compensaciones	3,57	7	0,51	0,0051
			7.4 Capacidad de reparación	3,57	5	0,71	0,0071
TOTAL	100%	171		100%			

Fuente: Elaboración propia

Este factor de ponderación de cada pregunta se aplica a la puntuación obtenida en cada respuesta a la pregunta para obtener la puntuación ponderada. La puntuación ponderada de todas las preguntas de cada subárea se suma y se obtiene la puntuación ponderada de la subárea, como se muestra en la tabla a continuación:

Tabla 32. Puntuación ponderada por subárea evaluada.

Área	Subárea	Puntuación ponderada de la subárea en función de las preguntas de ella
1. Agentes humanos y supervisión	1.1 Derechos fundamentales	0,1113
	1.2 Agentes humanos	0,176
	1.3 Supervisión Humana	0,1817
2. Solidez técnica y seguridad	2.1 Resistencia a los ataques y seguridad	0,1071
	2.2 Plan de emergencia y seguridad general	0,112
	2.3 Precisión	0,1428

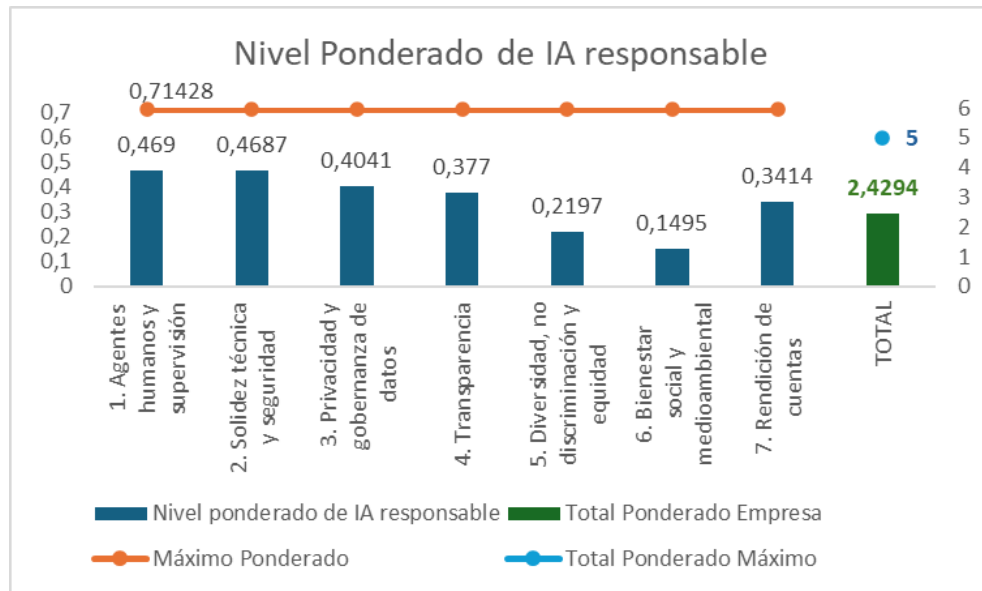
Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Área	Subárea	Puntuación ponderada de la subárea en función de las preguntas de ella
	2.4 Fiabilidad y reproducibilidad	0,1068
3. Privacidad y gobernanza de datos	3.1 Respeto de la intimidad y protección de datos	0,0624
	3.2 Calidad e integridad de los datos	0,1785
	3.3 Acceso a los datos	0,1632
4. Transparencia	4.1 Trazabilidad	0,19
	4.2 Explicabilidad	0,0795
	4.3 Comunicación	0,1075
5. Diversidad, no discriminación y equidad	5.1 Evitar el sesgo injusto	0,048
	5.2 Accesibilidad y diseño universal	0,0629
	5.3 Participación de las partes interesadas	0,1088
6. Bienestar social y medioambiental	6.1 IA sostenible y respetuosa con el medio ambiente	0,0714
	6.2 Impacto social	0,0781
7. Rendición de cuentas	7.1 Auditoría	0,09
	7.2 Minimizar y notificar el impacto negativo	0,0684
	7.3 Documentar los compromisos y compensaciones	0,0765
	7.4 Capacidad de reparación	0,1065

Fuente: Elaboración propia

Con la puntuación por subárea, es posible conocer el nivel ponderado de IA responsable por área evaluada, el cual se muestra en la siguiente figura:

Figura 23. Nivel ponderado de IA responsable de la empresa evaluada.

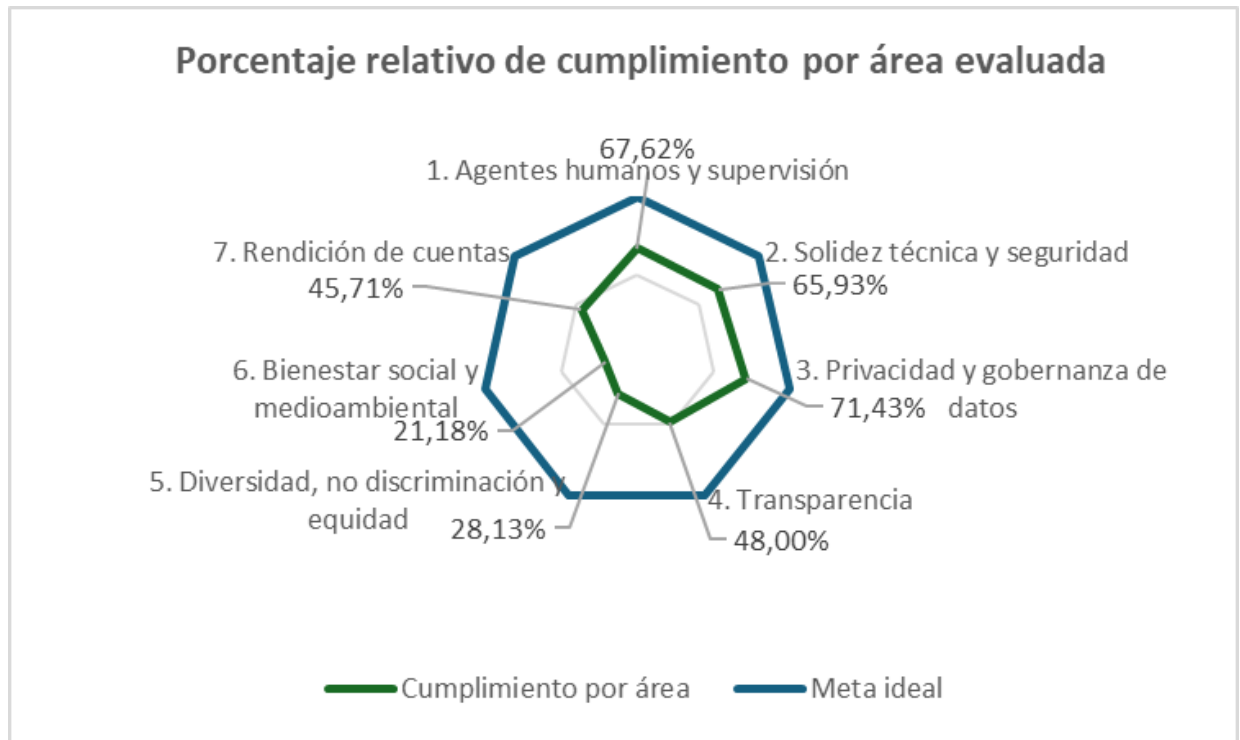


Fuente: Elaboración propia

El nivel ponderado de IA responsable de la entidad financiera evaluado es 2,4294 de un máximo de 5, lo cual confirma que la empresa se encuentra en nivel 2 y es necesario trabajar en todas las áreas evaluadas ya que ninguna llega a valores cercanos a 0,71426 que es valor máximo ponderado por área.

Considerando el puntaje obtenido en cada área, se obtiene el porcentaje relativo de desarrollo en IA responsable de la empresa evaluada en cada área, como se evidencia en la figura 24.

Figura 24. Porcentaje relativo de desarrollo en IA responsable por cada área evaluada.



Fuente: Elaboración propia

Es necesario aclarar que el porcentaje es relativo, el nivel de IA responsable se asigna por el área con menor porcentaje de desarrollo, en este caso con 21,18% se confirma el nivel 2- básico.

Adicionalmente se identificaron (figuras 16-22) subáreas que tienen cumplimiento nivel 1- iniciando que requieren atención prioritaria de la empresa para que se pueda afirmar que ha alcanzado el nivel 2- básico por completo y de ahí se oriente a llegar a nivel 4 - Avanzado.

4.2.3 Sensibilización en la empresa sobre IA responsable

Se identificó que en algunas áreas y sub-áreas de IA responsable, los delegados de la entidad durante la ejecución del cuestionario tomaban conciencia que algunos aspectos ni siquiera habían sido visualizados antes de las sesiones virtuales, lo cual es sumamente importante ya que muestra que, desde la aplicación del formulario, el marco de trabajo está ayudando a la concienciación de la IA responsable en modelos, productos o servicios de IA.

Los delegados de la empresa notaron durante la aplicación del formulario que sus productos de IA no tomaron en consideración el sesgo injusto, el medio ambiente, la rendición de cuentas, los interesados y que tampoco consideraron el bienestar social.

4.3 Lecciones aprendidas

Durante la aplicación del formulario en la empresa, se tuvo que resolver algunas situaciones que dejaron las siguientes lecciones aprendidas:

- Es necesario incluir una descripción o en cada pregunta o subárea ya que en muchas ocasiones las preguntas no resultaron de fácil entendimiento para el personal encuestado.
- Durante la aplicación del formulario, realizar una introducción general y pequeñas explicaciones al iniciar las preguntas de cada área evaluada.
- Cuando se realice el acompañamiento para responder al formulario, si hay más de una persona como delegado desde la empresa, dar el tiempo necesario para que lleguen a un consenso entre ellos antes de que puedan responder.
- Para la aplicación del formulario es necesario tener a disposición la información que se utilizó como herramientas para la elaboración del formulario ya que los encuestados solicitan explicaciones detalladas de por qué deben cumplir ciertas normativas y es necesario responder prontamente señalando la normativa, regulación o mejor práctica de IA responsable a la que hace referencia la pregunta.
- Dentro del tiempo estimado de aplicación del formulario considerar tiempo de contingencia para enfrentar imprevistos como explicaciones detalladas de las preguntas, de los procesos dentro de la empresa, de creencias respecto al cumplimiento de las diferentes normativas, etc.
- Se pudo evidenciar que la redacción de algunas preguntas puede ser revisada para que sea de fácil entendimiento para personas que no están familiarizadas completamente con la terminología y/o la normativa en IA responsable.

4.4 Rueda de madurez en IA responsable

Para elaborar la rueda de madurez en IA responsable se consideran las áreas y subáreas evaluadas. Si una subárea tiene un valor mínimo de 1, toda el área obtiene el valor de 1, por

lo tanto la empresa debe implementar acciones para reflejar su nivel 2 en cada área y subárea. A pesar de estas acciones aun le quedará mucho trabajo por hacer para llegar al nivel 4 que constituye el nivel en el cual ya es IA responsable. El plan de acción ideal es mostrar todos los pasos para alcanzar al menos el nivel 4, pero considerando la estructura organizacional actual y el nivel de conocimiento en ética de IA del personal de la empresa se le sugirió que primero escale todas sus áreas (con sus respectivas subáreas) a nivel 2 y luego continuar ejecutando y adoptando las acciones necesarias hasta llegar al nivel mínimo que se le consideraría IA responsable (nivel 4).

Es importante señalar que en algunas subáreas se tienen valores superiores a 2, pero también existen preguntas donde la respuesta refleja nivel 1, por lo tanto, se sugiere dar prioridad a la visualización de valores mínimos ya que en esas subáreas es donde se realizarán las sugerencias a la empresa para subir todas sus áreas a nivel 2.

A continuación, se muestran las 22 subáreas con sus respectivos valores mínimos, promedios e ideales:

Figura 25. Rueda de madurez de IA responsable por subáreas.



Fuente: Elaboración propia

Requieren especial atención las siguientes 16 subáreas que tienen valor mínimo 1 en alguna pregunta del formulario:

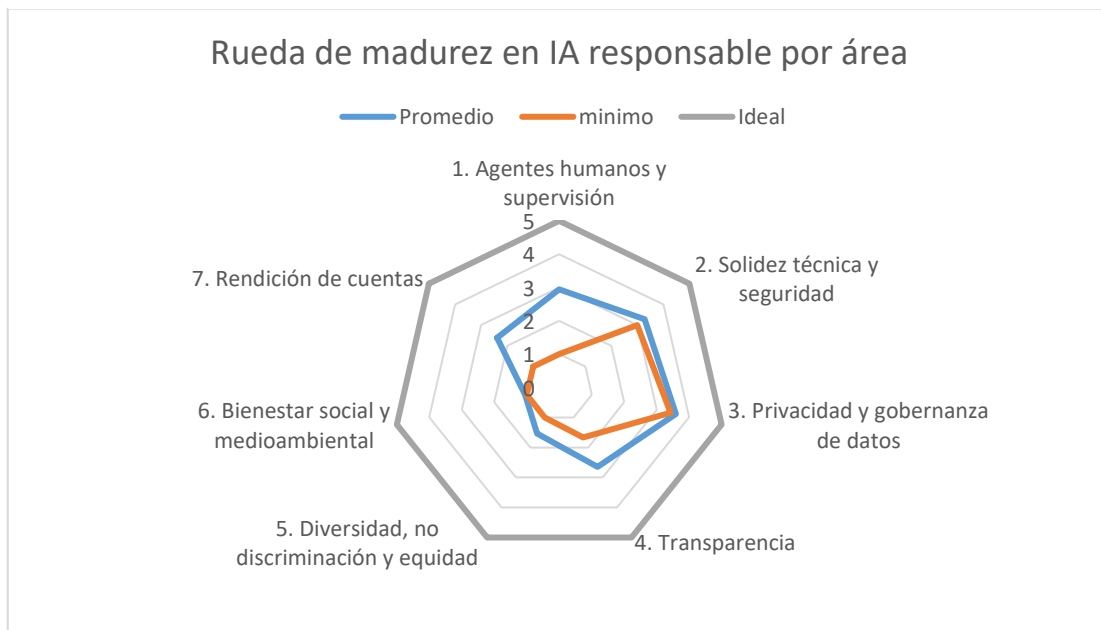
- 1.1 Derechos fundamentales
- 1.3 Supervisión humana
- 2.1 Resistencia a los ataques y seguridad
- 2.2 Plan de emergencia y seguridad general
- 3.1 Respeto de la intimidad y protección de datos
- 3.3 Acceso a los datos
- 4.2 Explicabilidad
- 4.3 Comunicación
- 5.1 Evitar el sesgo injusto
- 5.2 Accesibilidad y diseño universal
- 5.3 Participación de las partes interesadas

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

- 6.1 IA sostenible y respetuosa con el medio ambiente
- 6.2 Impacto social
- 7.1 Auditoría
- 7.2 Minimizar y notificar el impacto negativo
- 7.3 Documentación de los compromisos

A continuación, se muestra la rueda de madurez de IA responsable considerando las áreas evaluadas:

Figura 26. Rueda de madurez en IA responsable por área evaluada.



Fuente: Elaboración propia

Se puede visualizar que el área de bienestar social y medioambiental requiere intervención urgente porque tiene valores mínimos en todas sus subáreas. Estas serán las recomendaciones más importantes para la empresa.

El informe para la empresa se encuentra en el Anexo E.

4.5 Mejoras a realizarse

4.5.1 En el formulario

- Mejorar la redacción de las preguntas para hacerlas comprensibles para todas las personas inclusive para aquellas que no están familiarizadas con ética de IA.

- Añadir explicaciones y ejemplos de situaciones para cada pregunta a fin de que el entrevistador pueda agilizar la aplicación del formulario.
- Optimizar la cantidad de preguntas para reducir el tiempo de entrevistas.
- En lo posible hacer un mismo número de preguntas en todas las áreas y subáreas para aclarar la ponderación final, debido a que incluir ponderación en las preguntas para que todas las áreas tengan el mismo peso en la calificación (0,7142), complica bastante el modelo.
- Automatizar la aplicación del formulario para que el entrevistador tenga toda la información disponible y el informe sea más sencillo de obtener.

4.5.2 Durante el acompañamiento

- Acordar una charla inicial para despejar dudas y que el entrevistador tenga idea del nivel de conocimiento en normativas y regulaciones de IA responsable por parte de los delegados de las empresas.
- Explicar que el formulario evalúa únicamente la normativa y regulación occidental, no se relaciona con leyes de China, India y Japón.
- Detallar durante la primera reunión que se obtendrá un indicador que es parte de tres indicadores todo el marco de IA responsable.
- Hacer hincapié en la última reunión de acompañamiento durante la aplicación del formulario que, si en alguna pregunta se obtuvo un nivel de 1, toda la subárea recibió un puntaje menor.
- Si en alguna pregunta la empresa obtiene el nivel 1 por desconocimiento del tema, no es necesario realizar las preguntas sobre la auditoría de ese cumplimiento para optimizar tiempo de entrevistas.

4.5.3 Otras sugerencias

- Tener a la mano la reglamentación y normativas que se utilizaron para elaborar el formulario para solventar cualquier duda adicional.
- Llevar un resumen ejecutivo de los reglamentos de IA y de protección de datos para entregarlos a los entrevistados.
- Automatizar la obtención de gráficas de IA responsable.

5 Conclusiones y trabajo futuro

5.1 Conclusiones

La regulación de la inteligencia artificial (IA) responsable está siendo abordada por diversos actores, incluidos gobiernos, instituciones académicas, investigadores, organizaciones internacionales y empresas que desarrollan y usan IA. En este contexto, se identificó la necesidad de consolidar la normativa occidental más relevante para asistir a las empresas en el desarrollo de modelos, productos y servicios de IA de manera ética y centrada en el ser humano.

El objetivo principal de este trabajo fue evaluar el estado actual de las empresas en términos de IA responsable según la normativa de occidente y proporcionar recomendaciones que les permita avanzar hacia niveles más altos de cumplimiento y madurez en este tema. Para ello, se diseñó un formulario compuesto por 171 preguntas, que abordan 7 áreas y 22 subáreas de IA responsable. Este formulario, basado en controles, procesos y estructuras definidos por la normativa occidental, garantiza una implementación ética de la IA.

El formulario se aplicó en un proyecto piloto a una entidad bancaria con operaciones en Ecuador. Los resultados del piloto revelaron que la entidad alcanzó un nivel de IA responsable de 2 – Básico, el cual es un buen punto de partida, pero aún se considera insuficiente ya que el nivel mínimo para indicar que una IA es responsable es el nivel 4 en todos los indicadores. Posteriormente, se entregó un informe detallado con recomendaciones para afianzar el nivel 2 y avanzar progresivamente al nivel 4 y 5.

Como parte del trabajo, se estableció un marco de trabajo que define 5 niveles de madurez en IA responsable. Este marco se centra en tres indicadores clave derivados del formulario diseñado:

- Indicador global que se obtiene como la sumatoria de todas las respuestas del formulario.
- Indicador por áreas, representado gráficamente cada una de ellas para evaluar el cumplimiento actual.

- Indicador ponderado por áreas y subáreas, que considera la puntuación máxima de 5 (conforme los 5 niveles de IA responsable establecidos) que refleja directamente el nivel de madurez de la empresa en cada área y subárea evaluada.

Estos resultados y herramientas validaron el formulario y el marco de trabajo, confirmando que esta propuesta brinda a las empresas un camino claro y estructurado para implementar una IA responsable que sea ética, transparente y centrada en el bienestar humano.

En conclusión, también se cumplieron los objetivos específicos planteados para este trabajo de la siguiente manera:

- Identificación de los parámetros mínimos que debe cumplir un producto o servicio de IA para ser considerado responsable, estableciendo que el nivel 4 del modelo es el estándar mínimo.
- Diseño de un listado de confirmación que permita verificar la conformidad de los procesos de IA con estándares técnicos y éticos, minimizando sesgos y errores, y asegurando la existencia de políticas y procedimientos adecuados durante todo el ciclo de vida de la IA.
- Proporcionar una metodología para generar informes con acciones específicas que permitan a las empresas incrementar su nivel de IA responsable.
- Aplicar este enfoque en una empresa, evaluando su nivel actual de IA responsable y proponiendo acciones concretas para afianzar su nivel 2 actual y avanzar hacia el nivel 4.

5.2 Líneas de trabajo futuro

- Validar este marco de trabajo con la aplicación a empresas que se rijan con la normativa europea para verificar el nivel de IA responsable en un entorno donde cumplir esa legislación es obligatorio.
- Aplicar este marco de trabajo a empresas cuyo giro del negocio sea el desarrollo de productos o servicios de IA, ya que tendrán mejor desarrollada su estructura alrededor de IA.

- Incluir las políticas de IA que se generarán por la ejecución de la Orden Ejecutiva de Donald Trump del 23 de enero de 2025, ya que a la fecha de finalización de este trabajo aún no son de conocimiento público, con esta información mejorará la evaluación del estado de IA responsable.
- Añadir al marco de trabajo las siguientes normativas que ya se encuentran publicadas:
 - Ley de Responsabilidad Algorítmica que rige en algunos estados de Estados Unidos de América.
 - Marco de trabajo de la gestión del riesgo de IA (*NIST Framework for AI Risk Management*), publicado por el Instituto Nacional de Estándares y Tecnología de Estados Unidos de América.
 - Estrategia Nacional en IA (*National AI Strategy*) publicado por el Gobierno de Reino Unido para fomentar la innovación y regular la IA.
 - Ley de Implementación de la Carta Digital (*Bill C-27*) emitida por el Gobierno de Canadá para regular la IA de alto impacto.
 - Marco Regulatorio de IA que se encuentra vigente en Brasil.
 - Marco Ético para la IA emitido por la UNESCO a la cual algunas empresas desarrolladoras de IA se han suscrito.
 - Principios para una IA confiable emitido por Organización para la Cooperación y el Desarrollo Económicos (OCDE).
- Además de las regulaciones anteriores, se puede hacer un marco de trabajo global si se llega a incluir las siguientes normativas:
 - Guía Ética de la IA, emitida por el Gobierno de China
 - Regulación sobre algoritmos de recomendación vigente en China.
 - Marco Regulatorio de la IA que se encuentra en desarrollo en Australia.
 - Legislación de IA y Derechos Humanos que se encuentra en desarrollo en Chile.
- Desarrollar una herramienta que permita la actualización del formulario con nuevas normativas y que entregue automáticamente los reportes para las empresas evaluadas.

Alvear Aguirre Isabel Gardenia

Flores Villamil Edwin Vladimir

Ortiz Albán Iván Josué

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Referencias bibliográficas

- Acemoglu, D., Restrepo, P. (2019). *Robots and jobs: Evidence from US labor markets*. Journal of Political Economy. 2188-2244. <https://shapingwork.mit.edu/wp-content/uploads/2023/10/Robots-and-Jobs-Evidence-from-US-Labor-Markets.p.pdf>
- Acuña, M.(2023). *Inteligencia Artificial Responsable*. Deloitte. <https://www2.deloitte.com/content/dam/Deloitte/cl/Documents/risk/cyber/cl-Inteligencia-Artificial-Responsable.pdf>
- Ahmed, S., & Hussain, M. (2022). *Optimizing Resource Management in Agriculture through AI Applications*. International Journal of Agricultural Innovations, 10(4), 200-220. www.agriinnovationsjournal.com/vol10-issue4-resource-management-ai
- Amazon. (2024). *Desarrollo responsable de inteligencia artificial en Amazon Web Services*. <https://aws.amazon.com/es/ai/responsible-ai/>
- Amazon Web Services. (s.f.). *What is a Large Language Model?* AWS. <https://aws.amazon.com/es/what-is/large-language-model/>
- Bird, S.(2024). Announcing new tools in Azure AI to help you build more secure and trustworthy generative AI applications. <https://azure.microsoft.com/en-us/blog/announcing-new-tools-in-azure-ai-to-help-you-build-more-secure-and-trustworthy-generative-ai-applications/>
- Bommasani R., Klymann K., Longpre S., Kapoor S., Maslej N. Xion B., Zhang D., Liang P. (2023). *The Foundation Model Transparency Index*. Stanford Centre for Research on Foundation Models, Stanford Institute for Human-Centered Artificial Intelligence, Stanford University. Octubre 2023. <https://arxiv.org/pdf/2310.12941>
- Bommasani R., Klymann K., Longpre S., Kapoor S., Maslej N. Xion B., Zhang D., Liang P. (2024). *The Foundation Model Transparency Index v1.1*. Stanford Centre for Research on Foundation Models, Stanford Institute for Human-Centered Artificial Intelligence, Stanford University. May 2024. <https://crfm.stanford.edu/fmti/paper.pdf>

- Camaréna, S. (2020). *Predictive Analytics in Crop Yield Using Artificial Intelligence*. *Agricultural Systems Research*, 8(3), 95-110. www.agrisystemsresearch.com/predictive-analytics-ai-2020
- Cameron, J. (Director). (1984). *Terminator* [Película]. Hemdale Film Corporation; Pacific Western Productions; Euro Film Funding.
- Carrizosa, S. (2024). *Un catecismo de seis mandamientos: así enseñan las empresas del Ibex a sus empleados cómo hay que hablar con los robots*. *El País*. (2024-12-08). <https://elpais.com/economia/negocios/2024-12-09/un-catecismo-de-seis-mandamientos-asi-ensenan-las-empresas-del-ibex-a-sus-empleados-como-hay-que-hablar-con-los-robots.html>
- Casa Blanca. (2023). *Executive order on safe, secure, and trustworthy artificial intelligence*. *White House*. (2023, 30 de octubre). <https://www.whitehouse.gov/briefing-room/statements-releases/2023/10/30/fact-sheet-president-biden-issues-executive-order-on-safe-secure-and-trustworthy-artificial-intelligence/>
- Casa Blanca. (2025). *Fact Sheet: President Donald J. Trump Takes Action to Enhance America's AI Leadership*. *White House*. (2025, 23 de enero). <https://www.whitehouse.gov/fact-sheets/2025/01/fact-sheet-president-donald-j-trump-takes-action-to-enhance-americas-ai-leadership/>
- Cook, T. (2024). *Presentamos Apple Intelligence, el sistema de inteligencia personal que coloca poderosos modelos generativos en el corazón mismo del iPhone, el iPad y la Mac*. <https://www.apple.com/la/newsroom/2024/06/introducing-apple-intelligence-for-iphone-ipad-and-mac/>
- Consilium. (2024). *Inteligencia artificial: política de la UE*. (2024-12-13) <https://www.consilium.europa.eu/es/policies/artificial-intelligence/>
- CSER (2024). *Risks from Artificial Intelligence*. University of Cambridge. <https://www.cser.ac.uk/research/risks-from-artificial-intelligence/>
- Cencerrado, A. (2024). *La crisis existencial de nuestro siglo: ¿estamos perdiendo el control de nuestras vidas por culpa de las redes sociales?* (*El País* – 2014-09-17).

<https://elpais.com/tecnologia/2024-09-18/la-crisis-existencial-de-nuestro-siglo-estamos-perdiendo-el-control-de-nuestras-vidas-por-culpa-de-las-redes-sociales.html?utm>

DAMA International (2017). *DAMA-DMBOK: Data Management Body of Knowledge* (2nd ed.). Technics Publications. <https://www.dama.org/cpages/body-of-knowledge>

De Paz García, A. (2021). *El impacto de los sistemas de recomendación en la propagación de la desinformación en redes sociales*. Universidad Autónoma de Madrid. https://repositorio.uam.es/bitstream/handle/10486/701292/paz_garcia_alfonso_de_tfm.pdf?sequence=1&isAllowed=y

Deloitte. (s.f.). *Acerca de Deloitte*. Deloitte. <https://www.deloitte.com/es/es/about.html>

Deloitte (2023). *Deloitte se incorpora a OdiselA para impulsar el uso responsable y ético de la inteligencia artificial*. Deloitte. <https://www.deloitte.com/es/es/about/press-room/deloitte-se-incorpora-a-odiseia-para-impulsar-el-uso-responsable-y-etico-de-la-inteligencia-artificial.html>

Durán, I. (2024) *MIT revela más de 700 riesgos de la IA: Ciberataques y alucinaciones son algunos de los más frecuentes*, Infobae, <https://www.infobae.com/tecno/2024/08/14/mit-revela-mas-de-700-riesgos-de-la-ia-ciberataques-y-alucinaciones-son-algunos-de-los-mas-frecuentes/>

Florkin, J. (s.f.) *La IA en la guerra moderna: mejora de la precisión, la eficiencia y la estrategia*. <https://julienflorkin.com/es/militar/guerra-moderna/La-IA-en-la-guerra-moderna/>

Guarneros, F. (2023, 31 de octubre). *Guía para entender la orden ejecutiva de Joe Biden sobre la regulación de la IA*. Revista Expansión. <https://expansion.mx/tecnologia/2023/10/31/guia-orden-ejecutiva-joe-biden-regulacion-ia>

Goggin, B., Woodward, A. (2024). *Los usos creativos y educativos de los deepfakes: Balance entre ética y tecnología*. *Journal of Digital Ethics*, 12(3), 45-62. <https://doi.org/10.1016/j.jde.2024.11.005>

- Goolsbee, A. (2019). *Public Policy in an AI Economy*. The Economics of Artificial Intelligence: An Agenda". The University of Chicago Press. Pág. 309-316.
- Greer, J. (2022). *The Apocalypse is Coming or Why I've Been Existentially Depressed*.
<https://www.johncgreer.com/the-apocalypse-is-coming/>
- Hill, A., & Lee, C. (2024). *Llega la Ley de IA de la UE: cómo se prepara Unilever*. Unilever España.
<https://www.unilever.es/news/2024/llega-la-ley-de-ia-de-la-ue-como-se-prepara-unilever/>
- Holstein, K., Krafft, P. M., Veale, M., Binns, R., & Andrus, M. (2023). *AI Transparency in the Age of LLMs: A Human-Centered Research Roadmap*. Harvard Data Science Review.
<https://doi.org/10.1162/99608f92.aelq19qy>
- McGregor, S., Atherton, D. (2020). *Incident 74: Detroit Police Wrongfully Arrested Black Man Due To Faulty FRT*. Artificial Intelligence Incident Database. Responsible AI Collaborative. <https://incidentdatabase.ai/cite/74/>
- Lam, K., McGregor, S. (2018). *Incident 4: Uber AV Killed Pedestrian in Arizona*. Artificial Intelligence Incident Database. Responsible AI Collaborative.
<https://incidentdatabase.ai/cite/4/>
- López, A. (2024). *Sesgo en la inteligencia artificial: Retos y oportunidades para una IA inclusiva*.
<https://www.arturolopezvalerio.com/2024/05/20/sesgo-de-la-inteligencia-artificial/>
- Maia, C. H., Ariel, P., & Nunes, S. (2024). *Adding human values on the deepfake: Co-designing fact-checking solutions to combat misinformation*. AI and Ethics.
<https://doi.org/10.1007/s43681-024-00619-y>
- Martinez-Polo, A. (2024). *¿Que son los modelos fundacionales?*. Tech Translated: Foundation models, (PWC), Pág.1. <https://ideas.pwc.es/archivos/20240301/que-son-los-modelos-fundacionales/>
- Maslej, N., Fattorini, L., Perrault, R., Parli, V., Reuel, A., Brynjolfsson, E., Etchemendy, J., Ligett, K., Lyons, T., Manyika, J., Niebles, J., Shoham, Y., Wald, R., Clark, J. (2024). *"The AI Index 2024 Annual Report"*. Stanford University. <https://aiindex.stanford.edu/report/>

- McGregor, S., Lam, K. (2018) *Artificial Intelligence Incident Database*. Responsible AI Collaborative. <https://incidentdatabase.ai/cite/4/>
- Meta (2024). *Responsible AI: Resources and best practices for responsible development of products built with large language models*. <https://ai.meta.com/responsible-ai/>
- OpenAI (2022). *Introducing ChatGPT*. OpenAI. (2022-11-30). <https://openai.com/index/chatgpt/>
- OpenAI (2024). *Model Specification: May 8, 2024*. <https://cdn.openai.com/spec/model-spec-2024-05-08.html>
- Parlamento Europeo y Consejo de la Unión Europea (1995). *Directiva 95/46/CE del Parlamento Europeo y del Consejo, de 24 de octubre de 1995, relativa a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos*. Diario Oficial de las Comunidades Europeas, L 281, 31-50. <https://eur-lex.europa.eu/legal-content/ES/ALL/?uri=CELEX%3A31995L0046>
- Parlamento Europeo y Consejo de la Unión Europea (2016). *Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE (Reglamento general de protección de datos)*. Diario Oficial de la Unión Europea, L119, de 4 de mayo de 2016. <http://data.europa.eu/eli/reg/2016/679/oj>
- Peralta, L. (2024). *Los usos criminales de los 'deepfakes' se disparan: estafas, pornografía y suplantación de identidad*. El País. <https://elpais.com/proyecto-tendencias/2024-10-02/los-usos-criminales-de-los-deepfakes-se-disparan-estafas-pornografia-y-suplantacion-de-identidad.html>
- Pichai, S. (2018). *IA en Google: Nuestros principios*. <https://latam.googleblog.com/2018/06/ia-en-google-nuestros-principios.html>
- Radanliev, P., & Santos, O. (2023). *Ethics and Responsible AI Deployment*. <https://arxiv.org/abs/2311.14705>

Repsol. (s.f.) *Historia de Repsol.*

<https://www.repsol.com/es/conocenos/historia/index.cshtml>

Repsol. (2023). *Repsol crea el primer centro de competencias de inteligencia artificial generativa del sector energético europeo.* Repsol. <https://www.repsol.com/es/sala-prensa/notas-prensa/2023/repsol-crea-primer-centro-competencias-inteligencia-artificial-generativa-sector-energetico-europeo/index.cshtml>

Repsol. (2024). *Descubriendo la inteligencia artificial generativa en Repsol.* Revista Accionistas Repsol. Julio 2024. <https://revistaaccionistas.repsol.com/numero-44/descubriendo-la-inteligencia-artificial-generativa-en-repsol>

Ruiz de Valbuena, I., (2024) *Las empresas se ponen las pilas para prevenir los riesgos del uso de la IA por sus empleados.* Cinco Días. <https://cincodias.elpais.com/legal/2024-08-23/las-empresas-se-ponen-las-pilas-para-prevenir-los-riesgos-del-uso-de-la-ia-por-sus-empleados.html?>

Santander. (2023). *Cinco cosas que tenés que saber para un uso responsable de la inteligencia artificial.* Santander Post. <https://santanderpost.com.ar/articulo/cinco-cosas-que-tenes-que-saber-para-un-uso-responsable-de-la-inteligencia-artificial/>

Salesforce. (s.f.). *Descubre los Principios de IA de confianza de Salesforce.* <https://www.salesforce.com/es/blog/descubre-los-principios-de-ia-de-confianza-de-salesforce/>

Salesforce Developers. (s.f.). *¿Dónde queda Salesforce?* <https://salesforce-developers.com/donde-queda-salesforce/>

Schlarman, S., & Warriar, A. (2019). *DevSecOps: A Practical Guide for Security Teams and DevOps Professionals.* Packt Publishing. www.packtpub.com/en-us/product/implementing-devsecops-practices-9781803231495

Smuha, N., Smuha, A., Bauer, W., Bergmann, U., Bielikova, M., Bonefeld-Dahl, C., Bonnet, Y., Boufara, L., Brunessaux, S., Chatila, R., Coeckelbergh, M., Dignum – Coponente, V., Floridi, L., Gagné – Coponente, J., Giovannini, C., Goodey, J., Haddadin, S., Hasselbalch, G., Heintz, F., Hidvegi, F., Hilgendorf, E., Höckner, K., Jégo-Laveissière, M., Kärkkäinen,

- L., Köszegi, S., Kroplewski, R., Ling, E., Lucas, P., Martinkenaite, I., Metzinger – Coponente, T., Muller, C., Noga, M., O’Sullivan, B., Pachi, U., Petit – Coponente, N., Peylo, C., (2019), *Directrices éticas para una IA fiable*. Comisión Europea. <https://digital-strategy.ec.europa.eu/es/library/ethics-guidelines-trustworthy-ai>
- Soler, A. (2024). *Apple Intelligence: Un modelo propio con 3 mil millones de parámetros*. Paréntesis Media. <https://www.parentesis.media/apple-intelligence-un-modelo-propio-con-3-mil-millones-de-parametros/>
- Spence, M. (2024). *La promesa de la inteligencia artificial para la economía mundial*. Fondo Monetario Internacional. <https://www.imf.org/es/Publications/fandd/issues/2024/09/AIs-promise-for-the-global-economy-Michael-Spence>
- Statista. (2023). *Walmart – Datos estadísticos*. <https://es.statista.com/temas/8019/walmart/>
- Telefónica. (s.f.). *Telefónica: compañía líder en telecomunicaciones*. <https://www.telefonica.com/es/>
- Telefónica. (2021). *IA: Uso Responsable*. Telefónica. <https://www.telefonica.com/es/wp-content/uploads/sites/4/2021/08/ia-uso-responsable.pdf>
- Unilever. (2023). *Our company*. <https://www.unilever.com/our-company/>
- Unión Europea. (2024). *Marco legislativo sobre inteligencia artificial: Normas y principios para un desarrollo responsable*. Diario Oficial de la Unión Europea, 13 de junio de 2024. <https://eur-lex.europa.eu>
- Vehent, J. (2018). *Securing DevOps: Security in the Cloud*. Manning Publications. www.manning.com/books/securing-devops
- Walmart Inc. (2023, 17 de octubre). *Our Responsible AI Pledge: Setting the Bar for Ethical AI*. <https://corporate.walmart.com/news/2023/10/17/our-responsible-ai-pledge-setting-the-bar-for-ethical-ai>
- Walmart Inc. (2024). *Walmart se suma al mundo de la inteligencia artificial: así la usa*. <https://laopinion.com/2024/10/17/walmart-se-suma-al-mundo-de-la-inteligencia-artificial-asi-la-usa/>

- Wang, B., Chen W., Pei, H., Xie, C, Kang, M., Zhang, C., Xu, C., Xion, Z., Dutta, R., Schaeffer, R., Troung, S., Arora, S., Mazekia, M., Hendrycs, D., Lin, Z., Chen, Y., Koyejo, S., Song, D., Li, B. (2024) *DecodingTrust: A Comprehensive Assessment of Trustworthiness in GPT Models*. University of Illinois. <https://arxiv.org/pdf/2306.11698>
- Zambrano, A. (2024). *Impacto de la inteligencia artificial en los ciberataques*. Revista Sinapsis. <https://www.itsup.edu.ec/myjournal/index.php/sinapsis/article/download/895/2078>
- Zuboff, S. (2019). *The age of surveillance capitalism: The fight for a human future at the new frontier of power*. (PublicAffair). Pág. 210-214

Anexo A. - Lista de evaluación para una IA fiable.

1. Acción y supervisión humanas

Derechos fundamentales:

- En aquellos casos de usos en los que puedan producirse efectos potencialmente negativos para los derechos fundamentales, ¿ha llevado usted a cabo una evaluación del impacto sobre los derechos fundamentales? ¿Ha identificado y documentado los posibles equilibrios entre los diferentes principios y derechos?
- ¿Interactúa el sistema de IA con el proceso de adopción de decisiones por parte de usuarios finales humanos (por ejemplo, con las acciones recomendadas, las decisiones que es preciso adoptar o la presentación de opciones)?
 - ¿Existe en esos casos el riesgo de que el sistema de IA afecte a la autonomía humana al interferir con el proceso de adopción de decisiones del usuario final de forma imprevista?
 - ¿Ha considerado usted si el sistema de IA debería informar a los usuarios de que una decisión, contenido, recomendación o resultado es fruto de una decisión algorítmica?
 - En el caso de que el sistema de IA cuente con un bot de charla o un sistema conversacional, ¿son los usuarios finales humanos conocedores de que están interactuando con un agente no humano?

Acción humana:

- En el caso de que el sistema de IA se implante en el proceso de trabajo, ¿ha tenido usted en cuenta la asignación de tareas entre el sistema de IA y los trabajadores humanos para garantizar interacciones adecuadas y una supervisión y control humanas apropiadas?
 - ¿El sistema de IA mejora o aumenta las capacidades humanas?
 - ¿Se han adoptado medidas para evitar que los procesos de trabajo confíen o dependan en exceso del sistema de IA?

Supervisión humana:

- ¿Ha analizado cuál sería el nivel adecuado de control humano sobre el sistema de

IA específico y para el caso de uso concreto de que se trate?

- ¿Puede describir el nivel de control o implicación humana, si procede? ¿Quién es la persona que ostenta el control del sistema y cuáles son los momentos o herramientas para la intervención humana?
- ¿Ha establecido mecanismos y adoptado medidas para garantizar la posibilidad de dicho control supervisión humanos o para asegurar que las decisiones se tomen bajo la responsabilidad exclusiva de seres humanos?
- ¿Ha adoptado alguna medida para posibilitar la realización de auditorías y para solucionar cualquier problema relacionado con la gestión de la autonomía de la IA?
- En el caso de que exista un sistema de IA (o un caso de uso) autónomo o con capacidad de autoaprendizaje, ¿ha establecido mecanismos de control y supervisión más concretos?
 - ¿Qué tipo de mecanismos de detección y respuesta ha establecido para evaluar si algo puede salir mal?
 - ¿Se ha asegurado de disponer de un botón de desconexión o un procedimiento que permita abortar una operación en condiciones de seguridad en caso necesario? ¿Implica ese procedimiento que se aborta el proceso en su totalidad, en parte o la delegación del control a un ser humano?

2. Solidez técnica y seguridad

Resistencia a los ataques y seguridad:

- ¿Ha evaluado las posibles formas de ataque a las que puede ser vulnerable el sistema de IA?
 - En particular, ¿ha analizado los diferentes tipos y naturalezas de las vulnerabilidades, como la contaminación de los datos, la infraestructura física o los ciberataques?
- ¿Ha adoptado medidas o sistemas para garantizar la integridad del sistema de IA y su capacidad para resistir posibles ataques?

- ¿Ha evaluado el comportamiento de su sistema en situaciones o entornos imprevistos?
- ¿Ha analizado si su sistema se puede utilizar (y, en caso afirmativo, en qué medida) para diferentes fines? Si es así, ¿ha adoptado medidas adecuadas para prevenir su uso con fines no deseados (como, por ejemplo, la no divulgación de la investigación o despliegue del sistema)?

Plan de repliegue y seguridad general:

- ¿Se ha asegurado de que su sistema cuente con un plan de repliegue suficiente en el caso de que se enfrente a algún ataque malintencionado o a otro tipo de situación inesperada (por ejemplo, procedimientos técnicos de conmutación o formulación de preguntas a un ser humano antes de continuar)?
- ¿Ha analizado el nivel de riesgo que plantea el sistema de IA en el caso de uso concreto previsto?
 - ¿Ha introducido algún proceso para medir y evaluar los riesgos y la seguridad?
 - ¿Ha proporcionado la información necesaria en caso de que exista algún riesgo para la integridad física de las personas?
 - ¿Ha estudiado la posibilidad de contratar una póliza de seguro para hacer frente a los posibles daños que provoque el sistema de IA?
 - ¿Ha identificado los riesgos potenciales para la seguridad asociados a (otros) usos previsibles de la tecnología, incluidos los usos accidentales o malintencionados? ¿Existe algún plan para mitigar o gestionar esos riesgos?
- ¿Ha evaluado si es probable que el sistema de IA cause daños a los usuarios o a terceros? En caso afirmativo, ¿ha evaluado la probabilidad, el daño potencial, el público afectado y la gravedad de tales daños?
 - Si existe el riesgo de que el sistema de IA ocasione daños, ¿ha tenido en cuenta las leyes de responsabilidad civil y de protección de los consumidores? ¿Cómo?
 - ¿Ha analizado los efectos potenciales o el riesgo para la seguridad del medio ambiente o de la fauna?
 - ¿Ha tenido en cuenta en su análisis de riesgos si los problemas de seguridad o de la red (por ejemplo, los peligros para la ciberseguridad) plantean riesgos

para la seguridad o pueden causar daños debido a un comportamiento imprevisto del sistema de IA?

- ¿Ha estimado el efecto probable de un fallo de su sistema de IA que provoque que el sistema ofrezca resultados erróneos, quede fuera de servicio o proporcione resultados socialmente inaceptables (como, por ejemplo, prácticas discriminatorias)?
 - ¿Ha definido umbrales y mecanismos de gestión para los escenarios anteriores a fin de activar planes alternativos o de repliegue?
 - ¿Ha definido y ensayado planes de repliegue?

Precisión

- ¿Ha evaluado qué nivel y definición de precisión se requerirá en el contexto del sistema de IA y para el caso de uso previsto?
 - ¿Ha evaluado cómo se mide y garantiza la precisión?
 - ¿Ha adoptado medidas para garantizar que los datos utilizados sean exhaustivos y estén actualizados?
 - ¿Ha adoptado medidas para evaluar si es necesario disponer de datos adicionales, por ejemplo, para mejorar la precisión o eliminar sesgos?
- ¿Ha evaluado los daños que se ocasionarían si el sistema de IA realizara predicciones incorrectas?
- ¿Ha establecido algún mecanismo para medir si el sistema está realizando una cantidad inaceptable de predicciones erróneas?
- Si el sistema está realizando predicciones erróneas, ¿ha establecido una serie de pasos que permitan subsanar el problema?

Fiabilidad y reproducibilidad:

- ¿Ha diseñado una estrategia para supervisar y verificar que el sistema cumple los objetivos, el propósito y las aplicaciones previstas?
 - ¿Ha comprobado si es necesario tener en cuenta algún contexto o condición

particular para garantizar la reproducibilidad?

- ¿Ha introducido procesos o métodos de verificación para medir y garantizar los diferentes aspectos de la fiabilidad y la reproducibilidad?
- ¿Ha establecido algún proceso para describir las situaciones en las que un sistema de IA falla en determinados tipos de entornos?
- ¿Ha documentado y detallado claramente esos procesos para la verificación de la fiabilidad de los sistemas de IA?

¿Ha establecido algún mecanismo o comunicación para garantizar a los usuarios (finales) que el sistema de IA es fiable?

3. Gestión de la privacidad y de los datos

Respeto de la privacidad y de la protección de datos:

- Dependiendo del caso de uso, ¿ha establecido un mecanismo que permita notificar los problemas relacionados con la privacidad o la protección de datos en los procesos de recopilación de datos de los sistemas de IA (tanto con fines de formación como de funcionamiento) y su tratamiento?
- ¿Ha evaluado el tipo y alcance de los datos incluidos en sus bases de datos (por ejemplo, si estas contienen datos de carácter personal)?
- ¿Ha analizado formas de desarrollar el sistema de IA o de formar el modelo en las que no sea necesario utilizar datos personales o potencialmente sensibles (o que utilicen la mínima cantidad posible de este tipo de datos)?
- ¿Ha introducido mecanismos de aviso y control sobre los datos personales en función del caso de uso (como, por ejemplo, el consentimiento válido y la posibilidad de revocar el uso de dichos datos, cuando proceda)?
- ¿Ha tomado medidas para mejorar la privacidad, por ejemplo, a través de procesos como el encriptado, la anonimización y la agregación?
- En los casos en que exista una persona responsable de la privacidad de los datos, ¿la ha implicado desde una fase inicial del proceso?

Calidad e integridad de los datos:

- ¿Ha alineado su sistema con las normas potencialmente pertinentes (por ejemplo, ISO, IEEE) o ha adoptado protocolos generales para la gestión y gobernanza cotidianas de sus datos?
- ¿Ha establecido mecanismos de supervisión para la recopilación, almacenamiento, tratamiento y utilización de los datos?
- ¿Ha evaluado su grado de control sobre la calidad de las fuentes de datos externas utilizadas?
- ¿Ha instaurado procesos para garantizar la calidad y la integridad de sus datos? ¿Ha estudiado la posibilidad de introducir otros procesos? ¿Cómo está verificando que sus conjuntos de datos no son vulnerados ni objeto de ataques?

Acceso a los datos:

- ¿Qué protocolos, procesos y procedimientos se han seguido para gestionar y garantizar una adecuada gobernanza de los datos?
 - ¿Ha evaluado quién puede acceder a los datos de los usuarios y en qué circunstancias?
 - ¿Se ha asegurado de que esas personas poseen la cualificación para acceder a los datos, que se les exige acceder a ellos y que cuentan con las competencias necesarias para comprender los detalles de la política de protección de datos?
 - ¿Ha asegurado la existencia de un mecanismo de supervisión que permita registrar cuándo, dónde, cómo y quién accede a los datos, y con qué propósito?

4. Transparencia

Trazabilidad:

- ¿Ha adoptado medidas que puedan garantizar la trazabilidad? Esto puede conllevar la documentación de:
 - los métodos utilizados para diseñar y desarrollar el sistema algorítmico:

- o en el caso de un sistema de IA basado en reglas, se debería documentar el método de programación o la forma en que se creó el modelo;
- o en el caso de un sistema de IA basado en el aprendizaje, se debería documentar el método de formación del algoritmo, incluidos los datos de entrada que se recopilaron y seleccionaron y la forma en que se hizo;
- los métodos empleados para ensayar y validar el sistema algorítmico:
 - o en el caso de un sistema de IA basado en reglas, se deberían documentar los escenarios o casos de uso utilizados para los ensayos y la validación;
 - o en el caso de un modelo basado en el aprendizaje, se debería documentar la información sobre los datos utilizados para los ensayos y la validación;
- los resultados del sistema algorítmico:
 - o se deberían documentar los resultados del algoritmo o las decisiones adoptadas por este, así como otras posibles decisiones que se producirían en casos diferentes (por ejemplo, para otros subgrupos de usuarios).

Explicabilidad:

- ¿Ha evaluado en qué medida son comprensibles las decisiones y, por tanto, el resultado producido por el sistema de IA?
- ¿Se ha asegurado de que se pueda elaborar una explicación comprensible para todos los usuarios que puedan desearla sobre las razones por las que un sistema adoptó una decisión determinada que dio lugar a un resultado específico?
- ¿Ha evaluado en qué medida la decisión del sistema influye en los procesos de adopción de decisiones de la organización?
- ¿Ha evaluado por qué se desplegó ese sistema en particular en esa área concreta?
- ¿Ha evaluado el modelo de negocio del sistema (por ejemplo, de qué modo crea valor para la organización)?
- ¿Ha diseñado el sistema de IA teniendo en mente desde el principio la interpretabilidad?
 - ¿Ha investigado y tratado de utilizar el modelo más sencillo e interpretable posible para la aplicación en cuestión?

- ¿Ha evaluado si puede analizar sus datos relativos a la formación y los ensayos realizados?
¿Puede modificar y actualizar estos datos a lo largo del tiempo?
- ¿Ha evaluado si, tras la formación y el desarrollo del modelo, tiene alguna posibilidad de examinar su interpretabilidad o si dispone de acceso al flujo de trabajo interno del modelo?

Comunicación:

- ¿Ha informado a los usuarios (finales) —mediante cláusulas de exención de responsabilidad u otros medios— de que están interactuando con un sistema de IA y no con otro ser humano? ¿Ha etiquetado su sistema de IA como tal?
- ¿Ha establecido mecanismos para informar a los usuarios de las razones y criterios subyacentes a los resultados del sistema de IA?
 - ¿Se han comunicado claramente estos a los usuarios previstos?
 - ¿Ha establecido procesos que tengan en cuenta las opiniones de los usuarios y que utilicen dichas opiniones para adaptar el sistema?
 - ¿Ha informado sobre los riesgos potenciales o percibidos, como la posible existencia de sesgos?
 - ¿Ha tenido también en cuenta, según el caso de uso, la comunicación y la transparencia hacia otras audiencias, hacia terceros o hacia el público en general?
- ¿Ha dejado claro el propósito del sistema de IA y quién o qué podrá beneficiarse del producto o servicio que ofrezca este?
 - ¿Se han especificado y se ha informado claramente sobre los escenarios de utilización del producto, estudiando posibles métodos de comunicación alternativos para garantizar que dicha información sea comprensible y adecuada para los usuarios a los que se dirige?
 - Según el caso de uso, ¿ha tenido en cuenta la psicología humana y sus posibles limitaciones, como el riesgo de confusión, el sesgo de confirmación o la fatiga cognitiva?

- ¿Ha comunicado con claridad las características, limitaciones y posibles carencias del sistema de IA:
 - en caso de desarrollo: ¿a las personas encargadas de su despliegue en un producto o servicio?
 - en caso de despliegue: ¿a los usuarios finales o consumidores?

5. Diversidad, no discriminación y equidad

Necesidad de evitar sesgos injustos:

- ¿Se ha asegurado de que exista una estrategia o un conjunto de procedimientos para evitar crear o reforzar un sesgo injusto en el sistema de IA, tanto en relación con el uso de los datos de entrada como en lo referente al diseño del algoritmo?
 - ¿Ha evaluado y reconocido las posibles limitaciones que emanan de la composición de los conjuntos de datos utilizados?
 - ¿Ha tenido en cuenta la diversidad y representatividad de los usuarios en los datos? ¿Ha realizado ensayos para poblaciones específicas o casos de uso problemáticos?
 - ¿Ha investigado y utilizado las herramientas técnicas disponibles para mejorar su comprensión de los datos, el modelo y su rendimiento?
 - ¿Ha establecido procesos para verificar la existencia de posibles sesgos y llevar a cabo un seguimiento de estos durante las fases de desarrollo, despliegue y utilización del sistema?
- Dependiendo del caso de uso, ¿se ha asegurado de introducir un mecanismo que permita a otras personas informar sobre posibles problemas relacionados con la existencia de sesgos, discriminación o un rendimiento deficiente del sistema de IA?
 - ¿Ha estudiado vías y métodos de comunicación claros sobre cómo y a quién informar sobre este tipo de problemas?
 - ¿Ha tenido en cuenta no solo a los usuarios (finales) sino también a otras personas que puedan verse indirectamente afectadas por el sistema de IA?

- ¿Ha evaluado si existe la posibilidad de que las decisiones varíen, aunque las condiciones no cambien?
 - Si es así, ¿ha estudiado cuáles podrían ser las causas de ello?
 - En caso de variabilidad, ¿ha establecido algún mecanismo de medición o evaluación del impacto potencial de dicha variabilidad sobre los derechos fundamentales?
- ¿Se ha asegurado de utilizar una definición operativa adecuada de «equidad» para aplicarla en el diseño de sistemas de IA?
 - ¿Se trata de una definición de uso común? ¿Estudió otras definiciones antes de optar por la seleccionada?
 - ¿Ha instaurado análisis o parámetros cuantitativos para medir y poner a prueba la definición de equidad aplicada?
 - ¿Ha establecido mecanismos para garantizar la equidad en sus sistemas de IA? ¿Ha considerado otros posibles mecanismos?

Accesibilidad y diseño universal:

- ¿Se ha asegurado de que el sistema de IA se adapte a una amplia variedad de preferencias y capacidades individuales?
 - ¿Ha evaluado si las personas con discapacidad, con necesidades especiales o en riesgo de exclusión pueden utilizar el sistema de IA? ¿Cómo se integró este aspecto en el sistema y cómo se verifica su funcionamiento?
 - ¿Se ha asegurado de que la información sobre el sistema de IA también sea accesible para los usuarios de tecnologías asistenciales?
 - ¿Implicó o consultó a esta comunidad durante la fase de desarrollo del sistema de IA?
- ¿Ha tenido en cuenta el impacto de su sistema de IA en sus usuarios potenciales?
 - ¿Es el equipo involucrado en el desarrollo del sistema de IA representativo de la audiencia a la que va dirigido? ¿Es representativo de la población en general y tiene también en cuenta a otros grupos que pudieran verse afectados de manera tangencial por el sistema?

- ¿Ha evaluado la posibilidad de que haya personas o grupos que puedan verse afectados de forma desproporcionada por las implicaciones negativas del sistema?
- ¿Ha recabado la opinión de otros equipos o grupos representativos de diferentes contextos y experiencias?

Participación de las partes interesadas:

- ¿Ha estudiado la posibilidad de introducir algún mecanismo para incorporar la participación de diferentes partes interesadas en el desarrollo y la utilización del sistema de IA?
- ¿Ha allanado el camino para la introducción del sistema de IA en su organización, informando e implicando previamente a los trabajadores afectados y sus representantes?

6. Bienestar social y ambiental

Una IA sostenible y respetuosa con el medio ambiente:

- ¿Ha establecido mecanismos para medir el impacto ambiental del desarrollo, despliegue y utilización del sistema de IA (por ejemplo, energía consumida por cada centro de datos, tipo de energía utilizada por los centros de datos, etc.)?
- ¿Se ha asegurado de introducir medidas para reducir el impacto ambiental de su sistema de IA a lo largo de todo su ciclo de vida?

Impacto social:

- En el caso de que el sistema de IA interactúe directamente con seres humanos:
 - ¿Ha evaluado si el sistema de IA alienta a los humanos a establecer un vínculo y desarrollar la empatía con el sistema?
 - ¿Se ha asegurado de que el sistema indique claramente que su interacción social es simulada y que no tiene capacidad para «entender» ni «sentir»?
- ¿Se ha asegurado de que se entiendan correctamente los efectos sociales del

sistema de IA? Por ejemplo, ¿ha evaluado si existe un riesgo de pérdida de puestos de trabajo o de descualificación de la mano de obra? ¿Qué pasos se han dado para contrarrestar esos riesgos?

Sociedad y democracia:

- ¿Ha evaluado el impacto social global asociado al uso del sistema de IA más allá del que tenga sobre el usuario (final), como, por ejemplo, las partes interesadas que pueden verse indirectamente afectadas por dicho sistema?

7. Rendición de cuentas

Auditabilidad:

- ¿Ha establecido mecanismos para facilitar la auditabilidad del sistema por parte de agentes internos o independientes (garantizando, por ejemplo, la trazabilidad y registro de los procesos y resultados del sistema de IA)?

Minimización de efectos negativos y notificación de estos:

- ¿Ha llevado a cabo una evaluación de riesgos o de impacto del sistema de IA que tenga en cuenta a las diferentes partes interesadas que se vean afectadas por este de forma directa o indirecta?
- ¿Ha establecido marcos de formación y educación para el desarrollo de prácticas de rendición de cuentas?
 - ¿Qué trabajadores o partes del equipo están implicados en ello? ¿Trasciende la fase de desarrollo?
 - ¿Se explica también en esa formación el posible marco jurídico aplicable al sistema de IA?
 - ¿Ha considerado la posibilidad de crear una «junta de revisión ética de la IA» u otro mecanismo similar para debatir sobre las prácticas éticas y de rendición de cuentas en general, incluidas las posibles «zonas grises»?
- Además de las iniciativas o marcos internos para supervisar la ética y la rendición de cuentas, ¿se cuenta con algún tipo de orientación externa o se han establecido

también procesos de auditoría?

- ¿Existe algún proceso para que los trabajadores o agentes externos (por ejemplo, proveedores, consumidores, distribuidores/vendedores) informen sobre posibles vulnerabilidades, riesgos o sesgos en el sistema de IA o su aplicación?

Documentación de los equilibrios alcanzados:

- ¿Se ha establecido algún mecanismo para identificar los intereses y valores que implica el sistema de IA y los posibles equilibrios entre ellos?
- ¿Qué procesos ha seguido para decidir sobre los equilibrios necesarios? ¿Se ha asegurado de documentar la decisión sobre la búsqueda de dichos equilibrios?

Capacidad de obtener compensación:

- ¿Ha establecido un conjunto de mecanismos adecuado que permita obtener compensación en el caso de que se produzca cualquier daño o efecto adverso?
- ¿Se han instaurado mecanismos para proporcionar información a usuarios (finales) y a terceros sobre las oportunidades de obtener compensación?

Anexo B. Evaluación de los 100 indicadores según el modelo fundacional (FMTI)

Antes del modelo (upstream)	
Tamaño de los datos:	Para los datos utilizados en la construcción del modelo, ¿se revela el tamaño de los datos?
Fuentes de datos:	Para todos los datos utilizados en la construcción del modelo, ¿se revelan las fuentes de datos?
Creadores de datos:	Para todos los datos utilizados en la construcción del modelo, ¿hay alguna caracterización de las personas que crearon los datos?
Selección de fuentes de datos:	¿Se dan a conocer los protocolos de selección para incluir y excluir fuentes de datos?
Conservación de datos:	Para todas las fuentes de datos, ¿se divulgan los protocolos de conservación de dichas fuentes de datos?
Aumento de datos:	¿Se dan a conocer las medidas que toma el promotor para aumentar sus fuentes de datos?
Filtrado de datos nocivos:	Si los datos se filtran para eliminar contenidos nocivos, ¿hay una descripción del filtro asociado?
Datos protegidos por derechos de autor:	Para todos los datos utilizados en la construcción del modelo, ¿se revela el estado de los derechos de autor asociados?
Licencia de los datos:	Para todos los datos utilizados en la construcción del modelo, ¿se revela el estado de la licencia asociada?
Información personal en los datos:	Para todos los datos utilizados en la construcción del modelo, ¿se revela la inclusión o exclusión de información personal en dichos datos?
Uso de mano de obra humana:	¿Se dan a conocer las fases de la canalización de datos en las que interviene mano de obra humana?
Empleo de trabajadores de datos:	¿Se da a conocer la organización que emplea directamente a

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Antes del modelo (upstream)
las personas involucradas en la mano de obra de datos para cada fase de la canalización de datos?
Distribución geográfica de los trabajadores de datos: ¿Se divulga información geográfica sobre las personas implicadas en el trabajo de datos para cada fase de la cadena de datos?
Salarios: ¿Se dan a conocer los salarios de las personas que realizan trabajos con datos?
Instrucciones para la creación de datos: ¿Se divulgan las instrucciones dadas a las personas que realizan el trabajo de creación de datos?
Protecciones laborales: ¿Se dan a conocer las protecciones laborales de las personas que realizan trabajos con datos?
Terceros colaboradores: ¿Se dan a conocer los terceros que han participado o participan en el desarrollo del modelo?
Acceso a datos externos consultables: ¿Se proporciona a las entidades externas un acceso consultable a los datos utilizados para construir el modelo?
Acceso directo a datos externos: ¿Se proporciona a las entidades externas acceso directo a los datos utilizados para construir el modelo?
Uso del cálculo: ¿Se divulga el cómputo necesario para construir el modelo?
Duración del desarrollo: ¿Se da a conocer el tiempo necesario para construir el modelo?
Hardware informático: Para el hardware primario utilizado para construir el modelo, ¿se revelan la cantidad y el tipo de hardware?
Propietario del hardware: Para el hardware primario utilizado en la construcción del modelo, ¿se revela el propietario del hardware?
Consumo de energía: ¿Se da a conocer la cantidad de energía gastada en la construcción del modelo?
Emisiones de carbono: ¿Se indica la cantidad de carbono emitida (asociada a la energía utilizada) en la construcción del modelo?
Impacto medioambiental más amplio: ¿Además de las emisiones de carbono, se han revelado otros impactos medioambientales derivados de la construcción del modelo?
Etapas del modelo: ¿Se divulgan todas las etapas del proceso de desarrollo del modelo?
Objetivos del modelo: Para todas las etapas que se describen, ¿hay una descripción clara de los objetivos de aprendizaje asociados o una caracterización clara de la naturaleza de esta actualización del modelo?
Marcos básicos: ¿Se dan a conocer los marcos básicos utilizados para el desarrollo de modelos?
Dependencias adicionales: ¿Existen otras dependencias necesarias para construir el modelo además de los datos, el cálculo y el código?
Mitigación de la privacidad: ¿Se divulgan las medidas que toma el desarrollador para mitigar la presencia de IPI en los datos?
Mitigación de los derechos de autor: ¿Se divulgan las medidas que toma el desarrollador para mitigar la presencia de información protegida por derechos de autor en los datos?

Modelo
Modalidad de entrada: ¿Se dan a conocer las modalidades de entrada del modelo?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Modelo
Modalidad de salida: ¿Se dan a conocer las modalidades de salida del modelo?
Componentes del modelo: ¿Se divulgan todos los componentes del modelo?
Tamaño del modelo: Para todos los componentes del modelo, ¿se revela el tamaño del modelo asociado?
Arquitectura del modelo: ¿Se divulga la arquitectura del modelo?
Documentación centralizada del modelo: ¿Se incluye información clave sobre el modelo en un artefacto centralizado, como una ficha de modelo?
Protocolo de acceso externo al modelo: ¿Se divulga un protocolo para conceder a entidades externas acceso al modelo?
Acceso externo al modelo de caja negra: ¿Se proporciona acceso al modelo de caja negra a entidades externas?
Pleno acceso externo al modelo: ¿Se proporciona acceso completo al modelo a entidades externas?
Descripción de las capacidades: ¿Se describen las capacidades del modelo?
Demostración de capacidades: ¿Se demuestran las capacidades del modelo?
Evaluación de las capacidades: ¿Se evalúan rigurosamente las capacidades del modelo y se comunican los resultados de estas evaluaciones antes o al mismo tiempo que la publicación inicial del modelo?
Reproducibilidad externa de la evaluación de las capacidades: ¿Las evaluaciones de las capacidades del modelo son reproducibles por entidades externas?
Evaluación de las capacidades por terceros: ¿Las capacidades del modelo son evaluadas por terceros?
Descripción de las limitaciones: ¿Se describen las limitaciones del modelo?
Demostración de las limitaciones: ¿Se demuestran las limitaciones del modelo?
Evaluación de las limitaciones por terceros: ¿Pueden las limitaciones del modelo ser evaluadas por terceros?
Descripción de riesgos: ¿Se describen los riesgos del modelo?
Demostración de riesgos: ¿Están demostrados los riesgos del modelo?
Evaluación de daños no intencionados: ¿Se evalúan rigurosamente los riesgos del modelo en relación con los daños no intencionados, y se comunican los resultados de estas evaluaciones antes o al mismo tiempo que la publicación inicial del modelo?
Reproducibilidad externa de la evaluación del daño involuntario: ¿Son reproducibles por entidades externas las evaluaciones de los riesgos del modelo relacionados con los daños no intencionados?
Evaluación del daño intencionado: ¿Se han evaluado rigurosamente los riesgos del modelo en relación con los daños intencionados y se han comunicado los resultados de estas evaluaciones antes o al mismo tiempo que la publicación inicial del modelo?
Reproducibilidad externa de la evaluación del daño intencionado: ¿Las evaluaciones de los riesgos del modelo relacionados con el daño intencionado son reproducibles por entidades externas?
Evaluación de riesgos por terceros: ¿Los riesgos del modelo son evaluados por terceros?
Descripción de las mitigaciones: ¿Se divulgan las mitigaciones del modelo?
Demostración de mitigaciones: ¿Están demostradas las mitigaciones del modelo?
Evaluación de las mitigaciones: ¿Se evalúan rigurosamente las mitigaciones del modelo y se informa de los resultados de estas evaluaciones?
Reproducibilidad externa de la evaluación de las mitigaciones: ¿Las evaluaciones de mitigación del modelo son reproducibles por entidades externas?
Evaluación de las mitigaciones por terceros: ¿Pueden ser evaluadas las mitigaciones del modelo por terceros?

Modelo
Evaluación de la fiabilidad: ¿Se evalúa rigurosamente la fiabilidad del modelo y se publican los resultados de estas evaluaciones?
Reproducibilidad externa de la evaluación de la fiabilidad: ¿Las evaluaciones de fiabilidad son reproducibles por entidades externas?
Evaluación de la duración de la inferencia: ¿Se revela el tiempo necesario para la inferencia del modelo para una tarea claramente especificada en un conjunto de hardware claramente especificado?
Evaluación informática de la inferencia: ¿Se da a conocer el uso del cálculo para la inferencia del modelo en una tarea claramente especificada en un conjunto de hardware claramente especificado?

Después del modelo (downstream)
Toma de decisiones: ¿Se da a conocer el protocolo que sigue el desarrollador para decidir si publica o no un modelo?
Proceso de publicación: ¿Se ha divulgado una descripción del proceso de publicación del modelo?
Canales de distribución: ¿Se dan a conocer todos los canales de distribución?
Productos y servicios: ¿Divulga el promotor si alguno de los productos y servicios que ofrece dependen del modelo?
Detección de contenidos generados por máquinas: ¿Se dan a conocer mecanismos para detectar contenidos generados por este modelo?
Licencia del modelo: ¿Se divulga la licencia del modelo?
Condiciones del servicio: ¿Se dan a conocer las condiciones de servicio de cada canal de distribución?
Usuarios permitidos y prohibidos: ¿Se describe quién puede y quién no puede utilizar el modelo?
Usos permitidos, restringidos y prohibidos: ¿Se dan a conocer los usos permitidos, restringidos y prohibidos del modelo?
Aplicación de la política de uso: ¿Se da a conocer el protocolo de aplicación de la política de uso?
Justificación de las medidas coercitivas: ¿Reciben los usuarios una justificación cuando son objeto de una medida coercitiva por infringir la política de uso?
Mecanismo de apelación de infracciones de la política de uso: ¿Se ha divulgado un mecanismo para apelar posibles infracciones de la política de uso?
Comportamientos permitidos, restringidos y prohibidos de los modelos: ¿Se dan a conocer los comportamientos modelo permitidos, restringidos y prohibidos?
Aplicación de la política de comportamiento modelo: ¿Se da a conocer el protocolo de aplicación de la política de comportamiento modelo?
Interoperabilidad de las políticas de uso y de comportamiento modelo: ¿Se da a conocer la forma en que interoperan la política de uso y la política de comportamiento modelo?
Interacción del usuario con el sistema de IA: En el caso de los canales de distribución con interfaces orientadas al usuario, ¿se notifica a los usuarios (i) que están interactuando con un sistema de IA, (ii) el modelo de base específico con el que están interactuando y (iii) que los resultados son generados por máquinas?
Exención de responsabilidad sobre el uso: En el caso de los canales de distribución con interfaces orientadas al usuario, ¿se facilitan a los usuarios descargos de responsabilidad sobre el uso del

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Después del modelo (downstream)
modelo?
Política de protección de datos de los usuarios: ¿Se dan a conocer los protocolos sobre cómo el desarrollador almacena, accede y comparte los datos de los usuarios?
Usos permitidos y prohibidos de los datos de los usuarios ¿Se dan a conocer los usos permitidos y prohibidos de los datos de los usuarios?
Protocolo de acceso a los datos de uso: ¿Se divulga un protocolo para conceder a entidades externas acceso a los datos de uso?
Protocolo de versiones: ¿Existe una versión divulgada y un protocolo de versiones para el modelo?
Registro de cambios: ¿Existe un registro de cambios divulgado para el modelo?
Política de amortización: ¿Existe una política de depreciación divulgada para el desarrollador?
Mecanismo de retroalimentación: ¿Se divulga un mecanismo de retroalimentación?
Resumen de las reacciones: ¿se publica un informe o resumen sobre las reacciones que ha recibido el promotor o sobre la forma en que ha respondido a ellas?
Consultas gubernamentales: ¿Se divulga un resumen de las consultas gubernamentales relacionadas con el modelo recibidas por el promotor?
Mecanismo de seguimiento: Para cada canal de distribución, ¿se divulga un mecanismo de supervisión para el seguimiento del uso del modelo?
Aplicaciones posteriores: En todas las formas de uso posterior, ¿depende el número de aplicaciones del modelo de cimentación divulgado?
Sectores del mercado afectados: En todas las aplicaciones posteriores, ¿se divulga la fracción de aplicaciones correspondiente a cada sector del mercado?
Personas afectadas: En todas las formas de uso posterior, ¿se divulga el número de individuos afectados por el modelo de fundación?
Informes de uso: ¿Se divulga un informe de uso que ofrezca estadísticas de uso que describan el impacto del modelo en los usuarios?
Estadísticas geográficas: En todas las formas de uso descendente, ¿se divulgan las estadísticas de uso de los modelos por zonas geográficas?
Mecanismo de reparación: ¿Se ha revelado algún mecanismo para ofrecer reparación a los usuarios por los daños sufridos?
Documentación centralizada para uso posterior: ¿La documentación para uso posterior está centralizada en un artefacto centralizado?
Documentación sobre el uso intermedio responsable: ¿Se divulga la documentación sobre el uso posterior responsable?

Anexo C. – Formulario del marco de trabajo

Requisito	Subnivel	Código de pregunta	Pregunta
1. Agentes humanos y supervisión	1.1 Derechos fundamentales	1.1.1	¿Ha llevado a cabo una evaluación de impacto sobre los derechos humanos fundamentales para identificar y documentar las posibles compensaciones antes del desarrollo del sistema y ha analizado las posibles contramedidas para reducir o justificar los riesgos identificados?
		1.1.2	¿Ha establecido mecanismos para recibir información sobre los sistemas de IA que pueden vulnerar los derechos humanos fundamentales?
		1.1.3	¿Existen indicadores específicos para verificar el cumplimiento de los derechos fundamentales de la UE en el modelo, servicio o producto?
	1.2 Agentes humanos	1.2.1	¿Ha evaluado cuál es el valor añadido específico del sistema de AI para ayudar al usuario final humano en sus tareas y objetivos?
		1.2.2	¿Ha evaluado la asignación de tareas entre el sistema de AI y las personas para lograr una interacción significativa?
		1.2.3	¿Comunica el sistema de IA a los usuarios (finales) que está interactuando con un agente no humano y que una decisión, contenido, consejo o resultado es fruto de una decisión algorítmica?
		1.2.4	¿Se han proporcionado a los usuarios los conocimientos y herramientas necesarios para comprender e interactuar con los sistemas de IA de forma satisfactoria?
		1.2.5	¿Se ha permitido a los usuarios cuestionar razonablemente el sistema y evitar una dependencia excesiva del mismo?
		1.2.6	¿Se permite a los usuarios modificar una decisión basada únicamente en un tratamiento automatizado cuando ésta produzca efectos jurídicos en los usuarios o les afecte de forma significativa?
		1.2.7	¿Se han designado claramente los responsables y encargados del tratamiento de datos, y conocen sus obligaciones relacionadas con el control y la supervisión humana (RGPD-Art. 24, 28)?
		1.2.8	¿Existe un Delegado de Protección de Datos (DPD) designado (RGPD-Art. 37, 38)?
		1.2.9	¿Se realizan evaluaciones de impacto en la protección de datos para identificar riesgos asociados con procesos automatizados que afectan a los usuarios (RGPD-Art.35)?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta
		1.2.10	¿Se verifica la eficacia de los agentes humanos en la supervisión del modelo, producto o servicio?
		1.2.11	¿Los agentes humanos tienen el control sobre el sistema producto o servicio de IA?
		1.2.12	¿Los agentes humanos pueden intervenir en caso de fallo del sistema de IA?
	1.3 Supervisión humana	1.3.1	¿Existe un nivel adecuado de control humano como HITL (Human-in-the-Loop), HOTL (Human-on-the-Loop), y HIC (Human-in-Control) con mecanismos y medidas específicos para este sistema, producto o servicio de IA y su caso de uso ?
		1.3.2	¿Ha previsto un botón de parada o un procedimiento para interrumpir una operación de la IA en caso necesario?
		1.3.3	¿Se aplican medidas técnicas y organizativas para garantizar que los datos personales se procesen de manera lícita (conforme a la ley) y transparente respetando los principios de protección de datos (RGPD-Art.25)?
		1.3.4	¿Se obtiene el consentimiento explícito del interesado para decisiones automatizadas (RGPD-Art-6, 9)?
		1.3.5	¿Existen protocolos claros para que humanos supervisen el modelo, producto o servicio de IA durante su operación?
		1.3.6	¿La supervisión humana permite detener el sistema de IA en cualquier momento de operación?
2. Solidez técnica y seguridad	2.1 Resistencia a los ataques y seguridad	2.1.1	¿Ha tenido en cuenta distintos tipos y naturalezas de vulnerabilidades, como la contaminación de datos, la fuga de modelos, la infraestructura física o los ciberataques, y ha adoptado las contramedidas respectivas?
		2.1.2	¿Ha comprobado cómo se comporta su sistema, producto o servicio de IA en situaciones y entornos inesperados?
		2.1.3	¿Ha tomado medidas preventivas adecuadas contra el posible uso indebido del modelo, producto o servicio de IA (como bots maliciosos, desinformación, creación de virus, etc.)?
		2.1.4	¿El modelo, producto o servicio de IA cuenta con estrategias para resistir ataques externos?
		2.1.5	¿Se realizan auditorías periódicas para garantizar la seguridad del sistema, producto o servicio de IA?
		2.1.6	¿Las auditorías programadas garantizan la resistencia a ataques en seguridad?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta
	2.2 Plan de emergencia y seguridad general	2.1.7	¿La documentación de resistencia a ataques respalda su efectividad en el marco de seguridad?
		2.2.1	¿Ha calculado el impacto probable de un fallo de su sistema de IA cuando proporciona resultados erróneos, deja de estar disponible o proporciona resultados socialmente inaceptables (por ejemplo, discriminación)?
		2.2.2	¿Existe un plan para mitigar o gestionar los riesgos de resultados erróneos o socialmente inaceptables?
		2.2.3	¿Ha definido umbrales y establecido procedimientos de gobernanza para activar planes alternativos/de emergencia?
		2.2.4	¿Ha comprobado y considerado el nivel de riesgo que tendría el sistema de IA funcionando con el plan alternativo?
		2.2.5	¿Dispone de una póliza de seguros para hacer frente a posibles daños (a los usuarios, al medioambiente, etc) causados por el sistema, producto o servicio de IA?
		2.2.6	¿Ha considerado y tenido en cuenta las normas de responsabilidad y protección de los consumidores en el desarrollo de producto, o servicio de IA?
		2.2.7	¿Tiene implementados planes de contingencia para proteger la infraestructura de hardware, software y datos de su sistema de IA (RGPD-Art.32)?
		2.2.8	¿Tiene implementados mecanismos de reversión a versiones previas o plan de fallback para su sistema de IA(RGPD-Art.32)?
		2.2.9	¿El sistema de IA puede recuperar su funcionalidad tras un fallo crítico?
	2.3 Precision	2.3.1	¿Ha evaluado qué nivel y definición de precisión se requerirían en el contexto del sistema de IA y del caso de uso?
		2.3.2	¿Ha comprobado qué perjuicios se causarían si el sistema de IA realiza predicciones inexactas?
		2.3.3	¿Ha establecido métodos para medir si su sistema está realizando una cantidad inaceptable de predicciones inexactas?
		2.3.4	¿Tiene implementados mecanismos para garantizar la exactitud de los datos procesados (RGPD-Art.5)?
		2.3.5	¿El modelo garantiza niveles consistentes de precisión durante su operación?
		2.3.6	¿El sistema de IA ha sido probado en múltiples escenarios para garantizar su precisión?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta
	2.4 Fiabilidad y reproducibilidad	2.3.7	¿Los resultados del sistema han sido validados con datos reales?
		2.4.1	¿Ha establecido y documentado una estrategia para supervisar y comprobar si el sistema de IA cumple los objetivos, propósitos y aplicaciones previstos en el diseño e implementación?
		2.4.2	¿Ha documentado contextos específicos o condiciones particulares necesarias para garantizar la reproductibilidad del sistema de IA?
		2.4.3	¿Tiene mecanismos y procesos para garantizar la fiabilidad del sistema de IA (RGPD-Art.32)?
		2.4.4	¿Los resultados del modelo, producto o servicio de IA, son reproducibles en diferentes entornos?
3. Privacidad y gobernanza de datos	3.1 Respeto de la intimidad y protección de datos	3.1.1	En función del caso de uso, ¿ha establecido un mecanismo que permita a otras personas señalar problemas relacionados con la privacidad o la protección de datos en la recopilación y procesamiento de datos del sistema, producto o servicio de IA?
		3.1.2	¿Consideró formas de desarrollar el sistema de IA (o de entrenar el modelo, producto o servicio), con y sin un uso mínimo de datos potencialmente sensibles personales?
		3.1.3	¿Incorporó en su sistema, producto o servicio de IA, mecanismos de notificación y control de los datos personales en función del caso de uso (como el consentimiento válido y la posibilidad de revocarlo, cuando proceda)?
		3.1.4	¿Ha tomado medidas para mejorar la privacidad, como el cifrado, la anonimización y la agregación?
		3.1.5	En caso de que exista un responsable de la protección de datos (RGPD), ¿se ha implicado a esta persona en una fase temprana del proceso de diseño e implantación?
		3.1.6	¿El personal de su empresa está capacitado para garantizar la transparencia en el tratamiento de datos personales? (RGPD-Art.12)
		3.1.7	¿Se cumplen normativas de privacidad (RGPD), en los datos utilizados por el modelo, producto o servicio de IA?
		3.1.8	¿Se protege la confidencialidad de los datos personales utilizados por el sistema?
		3.1.9	¿Las auditorías programadas garantizan la práctica adecuada de procesamiento de datos en lo relacionado a privacidad?
		3.1.10	¿La documentación de procesamiento de datos respalda su efectividad en lo relacionado a privacidad?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta
	3.2 Calidad e integridad de los datos	3.2.1	¿Ha establecido mecanismos de supervisión para la recogida, el almacenamiento, el tratamiento y el uso de los datos?
		3.2.2	¿Ha evaluado en qué medida controla la calidad de las fuentes de datos externas utilizadas para su modelo, producto o servicio de IA?
		3.2.3	¿Existen en su organización procesos para rectificar datos personales inexactos? (RGPD-Art.16)
		3.2.4	¿Existen procesos para verificar la integridad de los datos utilizados por el sistema?
	3.3 Acceso a los datos	3.3.1	¿Utiliza protocolos, procesos y procedimientos para gestionar y garantizar una gobernanza adecuada de los datos?
		3.3.2	¿Ha evaluado quién puede acceder a los datos de los usuarios y en qué circunstancias?
		3.3.3	¿Se ha asegurado de que estas personas están cualificadas con las competencias necesarias para comprender los detalles de la política de protección de datos?
		3.3.4	¿Ha garantizado un mecanismo de supervisión para registrar cuándo, dónde, cómo, por quién y con qué fin específico se ha accedido a los datos?
		3.3.5	¿Los interesados tienen garantías de que se ejecutará su derecho de acceso de sus datos personales y como son tratados? (RGPD-Art.15)
		3.3.6	¿Existen restricciones claras para limitar el acceso a los datos sensibles al modelo, producto o servicio de IA?
		3.3.7	¿Existen auditorías para controlar el acceso a los datos del sistema?
4. Transparencia	4.1 Trazabilidad	4.1.1	¿Ha establecido medidas que garanticen la trazabilidad de los métodos de diseño, desarrollo y validación del sistema algorítmico?
		4.1.2	¿Ha establecido medidas que garanticen la trazabilidad de los resultados del sistema algorítmico?
		4.1.3	¿Su organización verifica que los procesos de trazabilidad se cumplen? (RGPD-Art.30)
		4.1.4	¿Es posible rastrear el origen de los datos utilizados en el modelo?
		4.1.5	¿Es posible rastrear las decisiones tomadas por el sistema de IA?
	4.2 Explicabilidad	4.2.1	¿Ha evaluado hasta qué punto las decisiones y, por tanto, los resultados tomados por el sistema de IA pueden ser comprendidos por sus propios constructores?
		4.2.2	¿Ha garantizado una explicación de por qué el sistema tomó una determinada decisión que dio lugar a un determinado resultado que todos los usuarios puedan entender?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta
		4.2.3	¿Puede cambiar y actualizar la forma de entrenar y probar los datos a lo largo del tiempo?
		4.2.4	¿Su empresa tiene políticas para garantizar la explicabilidad de las decisiones automatizadas? (RGPD-Art.22)
		4.2.5	¿Existen procedimientos para proporcionar explicaciones claras sobre el tratamiento automatizado de datos? (RGPD-Art.22)
		4.2.6	¿Se proporciona una explicación clara sobre cómo el modelo toma decisiones?
		4.2.7	¿Las decisiones del sistema de IA son explicables a los usuarios?
		4.2.8	¿Las auditorías programadas garantizan la práctica adecuada de explicabilidad con relación a la transparencia?
		4.2.9	¿La documentación de explicabilidad respalda su efectividad en el marco de transparencia?
	4.3 Comunicación	4.3.1	¿Ha establecido mecanismos para informar de forma clara a los destinatarios o usuarios (finales) sobre las razones y los criterios en los que se basan los resultados del sistema de IA?
		4.3.2	¿Ha establecido procesos que tengan en cuenta los comentarios de los usuarios y los utilicen para adaptar /mejorar el sistema?
		4.3.3	¿Se comunicaron a los posibles usuarios, los riesgos potenciales o percibidos, como la parcialidad, la fiabilidad o las posibles deficiencias?
		4.3.4	Según el caso de uso, ¿ha considerado la comunicación de los riesgos potenciales o percibidos, hacia otras audiencias, terceros o el público en general?
		4.3.5	¿Existen métodos para comunicar a los interesados sobre el uso de sus datos personales? (RGPD-Art.12)
		4.3.6	¿Existen procesos para verificar la claridad y comprensión de las notificaciones enviadas a los interesados, en relación con el uso de sus datos (RGPD-Art.12)?
		4.3.7	¿Existen estrategias para mejorar la comunicación al usuario final, sobre las políticas de tratamiento de datos? (RGPD-Art.14)
		4.3.8	¿Existen procedimientos para gestionar las solicitudes de información realizadas por los interesados? (RGPD-Art.15)
		4.3.9	¿El modelo comunica riesgos y beneficios de manera transparente al público?
		4.3.10	¿El sistema informa claramente sobre sus limitaciones y capacidades?
		4.3.11	¿Los usuarios son informados si están interactuando con un sistema de IA?
		5.1.1	¿Se verifica la equidad del modelo para beneficiar a todos los grupos de usuarios?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta
5. Diversidad, no discriminación y equidad	5.1 Evitar el sesgo injusto	5.1.2	¿Las auditorías programadas garantizan la práctica adecuada de mitigación de sesgos?
		5.1.3	¿La documentación de mitigación de sesgos respalda su efectividad en lo relacionado a evitar sesgos?
		5.1.4	¿Existen medidas para identificar y eliminar posibles sesgos en los procesos automatizados? (RGPD-Art.22)
		5.1.5	¿Existen auditorías para garantizar la equidad en las decisiones automatizadas? (RGPD-Art.35)
		5.1.6	¿Existen procesos internos para mitigar los riesgos de sesgos injustos en los sistemas de IA? (RGPD-Art.25)
		5.1.7	¿Se realizan capacitaciones para los equipos responsables (personal a cargo) para evitar sesgos en el diseño de sistemas? (RGPD-Art.24)
		5.1.8	¿Ha evaluado las posibles limitaciones de sesgo derivadas de la composición de los conjuntos de datos utilizados (por ejemplo, diversidad, representatividad)?
		5.1.9	¿Ha realizado pruebas para poblaciones específicas o casos de uso problemáticos?
		5.1.10	¿Cómo se comprueba y controla el posible sesgo durante la fase de desarrollo, despliegue y uso del sistema?
		5.1.11	¿Ha tenido en cuenta, además de a los usuarios (finales), a otras personas potencialmente afectadas de forma indirecta por el sistema, producto o servicio de IA?
		5.1.12	¿Se ha asegurado de contar con una definición operativa adecuada de "sesgo" que aplique en el diseño de sistemas de IA?
	5.2 Accesibilidad y diseño universal	5.2.1	¿El modelo cumple con estándares internacionales de accesibilidad (cumplimiento técnico)?
		5.2.2	¿El sistema está diseñado para ser accesible a personas con discapacidades (usabilidad real)?
		5.2.3	¿Las auditorías programadas garantizan la práctica adecuada de diversidad cultural en inclusión?
		5.2.4	¿La documentación de diversidad cultural respalda su efectividad en el marco de inclusión?
		5.2.5	¿Las auditorías programadas garantizan la práctica adecuada de tratamiento igualitario en equidad?
		5.2.6	¿La documentación de tratamiento igualitario respalda su efectividad en el marco de equidad?
		5.2.7	¿Ha evaluado si el sistema de IA es utilizable por personas con necesidades especiales o con discapacidad o en riesgo de exclusión?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta
		5.2.8	¿Ha considerado el reglamento general de protección de datos y el reglamento de IA, durante la fase de desarrollo del sistema de IA?
		5.2.9	¿Ha tenido en cuenta el impacto de su sistema de IA en el público potencial de usuarios específicos?
		5.2.10	¿Ha evaluado si el equipo que participa en la creación del sistema de IA es representativo de la población destinataria en general, teniendo en cuenta también otros grupos que podrían verse afectados tangencialmente?
		5.2.11	¿Ha evaluado si podría haber personas o grupos que se vieran desproporcionadamente afectados por las consecuencias negativas por la implementación y uso del sistema IA?
		5.2.12	¿La empresa incorporó los principios de diseño universal en los procesos tecnológicos relacionados al sistema, producto o servicio de IA? (RGPD-Art.25)
		5.2.13	¿Existen herramientas implementadas en el modelo, producto o servicio de IA, para asegurar la accesibilidad de los datos personales a los interesados? (RGPD-Art.15)
	5.3 Participación de las partes interesadas	5.3.1	¿Ha previsto algún mecanismo para incluir la participación de partes interesadas con distintas necesidades en el desarrollo y el uso del sistema de IA?
		5.3.2	¿Recibió información de diferentes grupos multi-disciplinarios, que representaran diferentes contextos y experiencias para desarrollar el sistema, producto o servicio de IA?
		5.3.3	¿Recibió comentarios sobre el uso del sistema por parte de otros equipos o grupos que representaron diferentes contextos y experiencias?
		5.3.4	¿Se consultó a las partes interesadas durante el diseño del sistema?
		5.3.5	¿Existen mecanismos para involucrar a los interesados en las decisiones relacionadas con el tratamiento de datos? (RGPD-Art.12)
		5.3.6	¿Toma en cuenta las opiniones de los interesados en los procesos de toma de decisiones? (RGPD-Art.12)
		5.3.7	¿Dispone de métodos de retroalimentación para los interesados en relación a las políticas de tratamiento de datos? (RGPD-Art.13)
6. Bienestar social y medioambiental	6.1 IA sostenible y respetuosa con el medio ambiente	6.1.1	¿Ha establecido mecanismos para medir el impacto medioambiental del desarrollo, despliegue y uso del sistema de IA (por ejemplo, el tipo de energía, cantidad, climatización utilizadas por los centros de datos)?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta
		6.1.2	¿Ha tomado medidas para reducir el impacto ambiental del ciclo de vida de su sistema de IA?
		6.1.3	¿Se han implementado políticas para garantizar que su modelo, producto o servicio de IA sea sostenible y respetuoso con el medio ambiente? (RGPD-Art.5)
		6.1.4	¿Dispone de estrategias para minimizar el consumo energético de su modelo, producto o servicio de IA? (RGPD-Art.25)
		6.1.5	¿Se evalúan los impactos medioambientales asociados con el uso del modelo, producto o servicio de IA? (RGPD-Art.35)
		6.1.6	¿Existen sistemas para el monitoreo de la sostenibilidad de las operaciones del modelo, producto o servicio IA? (RGPD-Art.32)
		6.1.7	¿El modelo incorpora estrategias para minimizar el impacto ambiental del modelo, producto o servicio de IA?
	6.2 Impacto social	6.2.1	¿Ha evaluado si su sistema de IA anima a los humanos a desarrollar apego y empatía hacia el sistema?
		6.2.2	¿Se ha asegurado de que el sistema de IA señala claramente que su interacción social es simulada y que no tiene capacidades de "entender" y "sentir"?
		6.2.3	¿Se ha asegurado de que se comprenden bien por parte de los usuarios, las repercusiones sociales positivas y negativas del sistema de IA?
		6.2.4	¿Adopta medidas para garantizar un impacto social positivo del tratamiento de datos personales en su sistema de IA? (RGPD-Art.35)
		6.2.5	¿Su organización ha implementado medidas proactivas, para evaluar y mejorar los efectos sociales de usar su modelo, producto o sistema de IA? (RGPD-Art.24)
		6.2.6	¿Su organización participa activamente, en la evaluación de impacto de sus actividades con el uso de datos personales? (RGPD-Art.13)
		6.2.7	¿Se utilizan indicadores sociales para medir el impacto del modelo, producto o servicio?
		6.2.8	¿El sistema de IA contribuye positivamente al bienestar social en general?
		6.2.9	¿Las auditorías programadas garantizan la práctica adecuada de respeto a derechos humanos que permitan cumplir la normativa legal?
		6.2.10	¿La documentación de respeto a derechos humanos respalda su efectividad en lo relacionado a cumplir con leyes?

Requisito	Subnivel	Código de pregunta	Pregunta
7. Rendición de cuentas	7.1 Auditoría	7.1.1	¿Se ha asegurado de que el sistema de IA pueda ser auditado por un organismo interno autónomo?
		7.1.2	¿Mantiene registros de auditoría para evaluar el cumplimiento normativo? (RGPD-Art.30)
		7.1.3	¿El modelo es auditado de manera regular para garantizar la transparencia?
		7.1.4	¿Existen registros claros para facilitar las auditorías del sistema?
		7.1.5	¿Las auditorías programadas aseguran la correcta implementación de prácticas relacionadas con la rendición de cuentas?
		7.1.6	¿La documentación generada en las auditorías demuestra su eficacia en el contexto de la rendición de cuentas?
	7.2 Minimizar y notificar el impacto negativo	7.2.1	¿Ha evaluado qué trabajadores o ramas del equipo participan en todas las fases del sistema de IA (desde el diseño hasta el uso)?
		7.2.2	¿Se brindó capacitación y educación a los diferentes grupos interesados para fomentar el desarrollo de prácticas relacionadas con la rendición de cuentas?
		7.2.3	¿Se capacita también en el posible marco jurídico aplicable al sistema de IA?
		7.2.4	¿Ha previsto algún tipo de orientación externa para supervisar la ética y la responsabilidad del sistema de IA, además de las iniciativas internas?
		7.2.5	¿Dispone de procedimientos para detectar y notificar impactos negativos relacionados con el tratamiento de datos? (RGPD-Art.33)
		7.2.6	¿Posee mecanismos para mitigar riesgos derivados del tratamiento de datos? (RGPD-Art.32)
		7.2.7	¿Cuenta con procedimientos de comunicación establecidos para notificar sobre incidentes de seguridad? (RGPD- Art.34)
		7.2.8	¿Los usuarios son notificados sobre posibles impactos negativos identificados en el modelo, producto o servicio de IA?
		7.2.9	¿Las auditorías programadas garantizan la práctica adecuada de filtrado de temas sensibles en evitar contenido inapropiado?
		7.2.10	¿La documentación sobre el filtrado de temas sensibles garantiza su eficacia en la prevención de contenido inapropiado?
	7.3 Documentación de los compromisos	7.3.1	¿Se ha implementado un mecanismo para identificar los intereses y valores relevantes asociados al sistema de IA, así como las posibles compensaciones entre ellos?

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta
		7.3.2	¿Mantiene registros para documentar compromisos relacionados con la protección de datos? (RGPD-Art.30)
		7.3.3	¿Dispone y almacena documentación que acredita su cumplimiento con las obligaciones del RGPD? (RGPD-Art.24)
		7.3.4	¿Se documentan los compromisos éticos relacionados con el uso del modelo?
		7.3.5	¿El sistema incluye documentación sobre compromisos éticos?
		7.3.6	¿Las auditorías programadas garantizan la práctica adecuada de resultados fiables en veracidad?
		7.3.7	¿La documentación sobre resultados fiables garantiza su efectividad en términos de veracidad?
	7.4 Capacidad de reparación	7.4.1	¿Ha establecido un conjunto adecuado de mecanismos que permitan la reparación en caso de que se produzca algún daño o impacto adverso?
		7.4.2	¿Cuenta con mecanismos para comunicar a los usuarios finales y a terceros las opciones disponibles para ejercer sus derechos o presentar sus quejas?
		7.4.3	¿Dispone de mecanismos para garantizar la capacidad de reparación en caso de incumplimientos? (RGPD-Art.32)
		7.4.4	¿Cuenta con estrategias para indemnizar o reparar a los afectados en caso de incidentes de brechas de seguridad que impliquen riesgos significativos para sus derechos y libertades ? (RGPD-Art.34)
		7.4.5	¿Existen procedimientos para gestionar las reclamaciones de los interesados en caso de incidentes de seguridad? (RGPD-Art.24)

Anexo D. – Formulario llenado con la empresa

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
1. Agentes humanos y supervisión	1.1 Derechos fundamentales	1.1.1	¿Ha llevado a cabo una evaluación de impacto sobre los derechos humanos fundamentales para identificar y documentar las posibles compensaciones antes del desarrollo del sistema y ha analizado las posibles contramedidas para reducir o justificar los riesgos identificados?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		1.1.2	¿Ha establecido mecanismos para recibir información sobre los sistemas de IA que pueden vulnerar los derechos humanos fundamentales?	1. La acción nunca se ha realizado	1	
		1.1.3	¿Existen indicadores específicos para verificar el cumplimiento de los derechos fundamentales de la UE en el modelo, servicio o producto?	1. La acción nunca se ha realizado	1	
	1.2 Agentes humanos	1.2.1	¿Ha evaluado cuál es el valor añadido específico del sistema de AI para ayudar al usuario final humano en sus tareas y objetivos?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		1.2.2	¿Ha evaluado la asignación de tareas entre el sistema de AI y las personas para lograr una interacción significativa?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en	3	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
				situaciones particulares no se realizan.		
		1.2.3	¿Comunica el sistema de IA a los usuarios (finales) que está interactuando con un agente no humano y que una decisión, contenido, consejo o resultado es fruto de una decisión algorítmica?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		1.2.4	¿Se han proporcionado a los usuarios los conocimientos y herramientas necesarios para comprender e interactuar con los sistemas de IA de forma satisfactoria?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		1.2.5	¿Se ha permitido a los usuarios cuestionar razonablemente el sistema y evitar una dependencia excesiva del mismo?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		1.2.6	¿Se permite a los usuarios modificar una decisión basada únicamente en un tratamiento automatizado cuando ésta produzca efectos jurídicos en los usuarios o les afecte de forma significativa?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		1.2.7	¿Se han designado claramente los responsables y encargados del tratamiento de datos, y conocen sus obligaciones relacionadas con	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma	5	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
			el control y la supervisión humana (RGPD-Art. 24, 28)?	coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.		
		1.2.8	¿Existe un Delegado de Protección de Datos (DPD) designado (RGPD-Art. 37, 38)?	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.	5	
		1.2.9	¿Se realizan evaluaciones de impacto en la protección de datos para identificar riesgos asociados con procesos automatizados que afectan a los usuarios (RGPD-Art.35)?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		1.2.10	¿Se verifica la eficacia de los agentes humanos en la supervisión del modelo, producto o servicio?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		1.2.11	¿Los agentes humanos tienen el control sobre el sistema producto o servicio de IA?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		1.2.12	¿Los agentes humanos pueden intervenir en caso de fallo del sistema de IA?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
	1.3 Supervisión humana	1.3.1	<p>¿Existe un nivel adecuado de control humano como HITL (Human-in-the-Loop), HOTL (Human-on-the-Loop), y HIC (Human-in-Control) con mecanismos y medidas específicos para este sistema, producto o servicio de IA y su caso de uso ?</p> <p>HITL (human-in-the-loop): capacidad de intervención humana en cada ciclo de decisión del sistema, lo que en muchos casos no es posible ni deseable. HOTL (human-on-the-loop): capacidad de intervención humana durante el ciclo de diseño del sistema y supervisión de su funcionamiento. HIC (human-in-command), capacidad para supervisar la actividad general del sistema de IA y capacidad para decidir cuándo y cómo utilizar el sistema en una situación concreta (es decir, no utilizarlo)</p>	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		1.3.2	¿Ha previsto un botón de parada o un procedimiento para interrumpir una operación de la IA en caso necesario?	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
		1.3.3	¿Se aplican medidas técnicas y organizativas para garantizar que los datos personales se procesen de manera lícita (conforme a la ley) y transparente respetando los principios de protección de datos (RGPD-Art.25)?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		1.3.4	¿Se obtiene el consentimiento explícito del interesado para decisiones automatizadas (RGPD-Art-6, 9)?	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.	5	
		1.3.5	¿Existen protocolos claros para que humanos supervisen el modelo, producto o servicio de IA durante su operación?	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.	5	
		1.3.6	¿La supervisión humana permite detener el sistema de IA en cualquier momento de operación?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
2. Solidez técnica y seguridad	2.1 Resistencia a los ataques y seguridad	2.1.1	¿Ha tenido en cuenta distintos tipos y naturalezas de vulnerabilidades, como la contaminación de datos, la fuga de	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
			modelos, la infraestructura física o los ciberataques, y ha adoptado las contramedidas respectivas?			
		2.1.2	¿Ha comprobado cómo se comporta su sistema, producto o servicio de IA en situaciones y entornos inesperados?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		2.1.3	¿Ha tomado medidas preventivas adecuadas contra el posible uso indebido del modelo, producto o servicio de IA (como bots maliciosos, desinformación, creación de virus, etc)?	1. La acción nunca se ha realizado	1	
		2.1.4	¿El modelo, producto o servicio de IA cuenta con estrategias para resistir ataques externos?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		2.1.5	¿Se realizan auditorías periódicas para garantizar la seguridad del sistema, producto o servicio de IA?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		2.1.6	¿Las auditorías programadas garantizan la resistencia a ataques en seguridad?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		2.1.7	¿La documentación de resistencia a ataques respalda su efectividad en el marco de seguridad?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
	2.2 Plan de emergencia y seguridad general	2.2.1	¿Ha calculado el impacto probable de un fallo de su sistema de IA cuando proporciona resultados erróneos, deja de estar disponible o proporciona resultados socialmente inaceptables (por ejemplo, discriminación)?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		2.2.2	¿Existe un plan para mitigar o gestionar los riesgos de resultados erróneos o socialmente inaceptables?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		2.2.3	¿Ha definido umbrales y establecidos procedimientos de gobernanza para activar planes alternativos/de emergencia?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		2.2.4	¿Ha comprobado y considerado el nivel de riesgo que tendría el sistema de IA funcionando con el plan alternativo?	1. La acción nunca se ha realizado	1	
		2.2.5	¿Dispone de una póliza de seguros para hacer frente a posibles daños (a los usuarios, al medioambiente, etc) causados por el sistema, producto o servicio de IA?	1. La acción nunca se ha realizado	1	
		2.2.6	¿Ha considerado y tenido en cuenta las normas de responsabilidad y protección de los consumidores en el desarrollo de producto, o servicio de IA?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
		2.2.7	¿Tiene implementados planes de contingencia para proteger la infraestructura de hardware, software y datos de su sistema de IA (RGPD-Art.32)?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		2.2.8	¿Tiene implementados mecanismos de reversión a versiones previas o plan de fallback para su sistema de IA(RGPD-Art.32)?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		2.2.9	¿El sistema de IA puede recuperar su funcionalidad tras un fallo crítico?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
	2.3 Precisión	2.3.1	¿Ha evaluado qué nivel y definición de precisión se requerirían en el contexto del sistema de IA y del caso de uso?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		2.3.2	¿Ha comprobado qué perjuicios se causarían si el sistema de IA realiza predicciones inexactas?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		2.3.3	¿Ha establecido métodos para medir si su sistema está realizando una cantidad inaceptable de predicciones inexactas?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		2.3.4	¿Tiene implementados mecanismos para garantizar la exactitud de los datos procesados (RGPD-Art.5)?	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma	5	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
				coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.		
		2.3.5	¿El modelo garantiza niveles consistentes de precisión durante su operación?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		2.3.6	¿El sistema de IA ha sido probado en múltiples escenarios para garantizar su precisión?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		2.3.7	¿Los resultados del sistema han sido validados con datos reales?	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.	5	
	2.4 Fiabilidad y reproducibilidad	2.4.1	¿Ha establecido y documentado una estrategia para supervisar y comprobar si el sistema de IA cumple los objetivos, propósitos y aplicaciones previstos en el diseño e implementación?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		2.4.2	¿Ha documentado contextos específicos o condiciones particulares necesarias para garantizar la reproductibilidad del sistema de IA?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
3. Privacidad y gobernanza de datos		2.4.3	¿Tiene mecanismos y procesos para garantizar la fiabilidad del sistema de IA (RGPD-Art.32)?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		2.4.4	¿Los resultados del modelo, producto o servicio de IA, son reproducibles en diferentes entornos?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
	3.1 Respeto de la intimidad y protección de datos	3.1.1	En función del caso de uso, ¿ha establecido un mecanismo que permita a otras personas señalar problemas relacionados con la privacidad o la protección de datos en la recopilación y procesamiento de datos del sistema, producto o servicio de IA?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		3.1.2	¿Consideró formas de desarrollar el sistema de IA (o de entrenar el modelo, producto o servicio), con y sin un uso mínimo de datos potencialmente sensibles personales?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		3.1.3	¿Incorporó en su sistema, producto o servicio de IA, mecanismos de notificación y control de los datos personales en función del caso de uso (como el consentimiento válido y la	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
			posibilidad de revocarlo, cuando proceda)?			
		3.1.4	¿Ha tomado medidas para mejorar la privacidad, como el cifrado, la anonimización y la agregación?	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.	5	
		3.1.5	En caso de que exista un responsable de la protección de datos (RPD), ¿se ha implicado a esta persona en una fase temprana del proceso de diseño e implantación?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		3.1.6	¿El personal de su empresa está capacitado para garantizar la transparencia en el tratamiento de datos personales? (RGPD-Art.12)	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		3.1.7	¿Se cumplen normativas de privacidad (RGPD), en los datos utilizados por el modelo, producto o servicio de IA?	1. La acción nunca se ha realizado	1	
		3.1.8	¿Se protege la confidencialidad de los datos personales utilizados por el sistema?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
		3.1.9	¿Las auditorías programadas garantizan la práctica adecuada de procesamiento de datos en lo relacionado a privacidad?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		3.1.10	¿La documentación de procesamiento de datos respalda su efectividad en lo relacionado a privacidad?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
	3.2 Calidad e integridad de los datos	3.2.1	¿Ha establecido mecanismos de supervisión para la recogida, el almacenamiento, el tratamiento y el uso de los datos?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		3.2.2	¿Ha evaluado en qué medida controla la calidad de las fuentes de datos externas utilizadas para su modelo, producto o servicio de IA?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		3.2.3	¿Existen en su organización procesos para rectificar datos personales inexactos? (RGPD-Art.16)	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.	5	
		3.2.4	¿Existen procesos para verificar la integridad de los datos utilizados por el sistema?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
	3.3 Acceso a los datos	3.3.1	¿ Utiliza protocolos, procesos y procedimientos para gestionar y garantizar una gobernanza adecuada de los datos?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		3.3.2	¿Ha evaluado quién puede acceder a los datos de los usuarios y en qué circunstancias?	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.	5	
		3.3.3	¿Se ha asegurado de que estas personas están cualificadas con las competencias necesarias para comprender los detalles de la política de protección de datos?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		3.3.4	¿Ha garantizado un mecanismo de supervisión para registrar cuándo, dónde, cómo, por quién y con qué fin específico se ha accedido a los datos?	1. La acción nunca se ha realizado	1	
		3.3.5	¿Los interesados tienen garantías de que se ejecutará su derecho de acceso de sus datos personales y como son tratados? (RGPD-Art.15)	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		3.3.6	¿Existen restricciones claras para limitar el acceso a los datos sensibles al modelo, producto o servicio de IA?	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma	5	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
4. Transparencia	4.1 Trazabilidad			coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.		
		3.3.7	¿Existen auditorías para controlar el acceso a los datos del sistema?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		4.1.1	¿Ha establecido medidas que garanticen la trazabilidad de los métodos de diseño, desarrollo y validación del sistema algorítmico?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		4.1.2	¿Ha establecido medidas que garanticen la trazabilidad de los resultados del sistema algorítmico?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		4.1.3	¿Su organización verifica que los procesos de trazabilidad se cumplen? (RGPD-Art.30)	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		4.1.4	¿Es posible rastrear el origen de los datos utilizados en el modelo?	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.	5	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
	4.2 Explicabilidad	4.1.5	¿Es posible rastrear las decisiones tomadas por el sistema de IA?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		4.2.1	¿Ha evaluado hasta qué punto las decisiones y, por tanto, los resultados tomados por el sistema de IA pueden ser comprendidos por sus propios constructores?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		4.2.2	¿Ha garantizado una explicación de por qué el sistema tomó una determinada decisión que dio lugar a un determinado resultado que todos los usuarios puedan entender?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		4.2.3	¿Puede cambiar y actualizar la forma de entrenar y probar los datos a lo largo del tiempo?	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.	5	
		4.2.4	¿Su empresa tiene políticas para garantizar la explicabilidad de las decisiones automatizadas? (RGPD-Art.22)	1. La acción nunca se ha realizado	1	
		4.2.5	¿Existen procedimientos para proporcionar explicaciones claras sobre el tratamiento automatizado de datos? (RGPD-Art.22)	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
		4.2.6	¿Se proporciona una explicación clara sobre cómo el modelo toma decisiones?	1. La acción nunca se ha realizado	1	
		4.2.7	¿Las decisiones del sistema de IA son explicables a los usuarios?	1. La acción nunca se ha realizado	1	
		4.2.8	¿Las auditorías programadas garantizan la práctica adecuada de explicabilidad en relación a la transparencia?	1. La acción nunca se ha realizado	1	
		4.2.9	¿La documentación de explicabilidad respalda su efectividad en el marco de transparencia?	1. La acción nunca se ha realizado	1	
	4.3 Comunicación	4.3.1	¿Ha establecido mecanismos para informar de forma clara a los destinatarios o usuarios (finales) sobre las razones y los criterios en los que se basan los resultados del sistema de IA?	1. La acción nunca se ha realizado	1	
		4.3.2	¿Ha establecido procesos que tengan en cuenta los comentarios de los usuarios y los utilicen para adaptar /mejorar el sistema?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		4.3.3	¿Se comunicaron a los posibles usuarios, los riesgos potenciales o percibidos, como la parcialidad, la fiabilidad o las posibles deficiencias?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
		4.3.4	Según el caso de uso, ¿ha considerado la comunicación de los riesgos potenciales o percibidos, hacia otras audiencias, terceros o el público en general?	1. La acción nunca se ha realizado	1	
		4.3.5	¿Existen métodos para comunicar a los interesados sobre el uso de sus datos personales? (RGPD-Art.12)	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		4.3.6	¿Existen procesos para verificar la claridad y comprensión de las notificaciones enviadas a los interesados, en relación con el uso de sus datos (RGPD-Art.12)?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		4.3.7	¿Existen estrategias para mejorar la comunicación al usuario final, sobre las políticas de tratamiento de datos? (RGPD-Art.14)	1. La acción nunca se ha realizado	1	
		4.3.8	¿Existen procedimientos para gestionar las solicitudes de información realizadas por los interesados? (RGPD-Art.15)	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		4.3.9	¿El modelo comunica riesgos y beneficios de manera transparente al público?	1. La acción nunca se ha realizado	1	
		4.3.10	¿El sistema informa claramente sobre sus limitaciones y capacidades?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en	3	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
				situaciones particulares no se realizan.		
		4.3.11	¿Los usuarios son informados si están interactuando con un sistema de IA?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
5. Diversidad, no discriminación y equidad	5.1 Evitar el sesgo injusto	5.1.1	¿Se verifica la equidad del modelo para beneficiar a todos los grupos de usuarios?	1. La acción nunca se ha realizado	1	
		5.1.2	¿Las auditorías programadas garantizan la práctica adecuada de mitigación de sesgos?	1. La acción nunca se ha realizado	1	
		5.1.3	¿La documentación de mitigación de sesgos respalda su efectividad en lo relacionado a evitar sesgos?	1. La acción nunca se ha realizado	1	
		5.1.4	¿Existen medidas para identificar y eliminar posibles sesgos en los procesos automatizados? (RGPD-Art.22)	1. La acción nunca se ha realizado	1	
		5.1.5	¿Existen auditorías para garantizar la equidad en las decisiones automatizadas? (RGPD-Art.35)	1. La acción nunca se ha realizado	1	
		5.1.6	¿Existen procesos internos para mitigar los riesgos de sesgos injustos en los sistemas de IA? (RGPD-Art.25)	1. La acción nunca se ha realizado	1	
		5.1.7	¿Se realizan capacitaciones para los equipos responsables (personal a cargo) para evitar sesgos en el diseño de sistemas? (RGPD-Art.24)	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
		5.1.8	¿Ha evaluado las posibles limitaciones de sesgo derivadas de la composición de los conjuntos de datos utilizados (por ejemplo, diversidad, representatividad)?	1. La acción nunca se ha realizado	1	
		5.1.9	¿Ha realizado pruebas para poblaciones específicas o casos de uso problemáticos?	1. La acción nunca se ha realizado	1	
		5.1.10	¿Cómo se comprueba y controla el posible sesgo durante la fase de desarrollo, despliegue y uso del sistema?	1. La acción nunca se ha realizado	1	
		5.1.11	¿Ha tenido en cuenta, además de a los usuarios (finales), a otras personas potencialmente afectadas de forma indirecta por el sistema, producto o servicio de IA?	1. La acción nunca se ha realizado	1	
		5.1.12	¿Se ha asegurado de contar con una definición operativa adecuada de "sesgo" que aplique en el diseño de sistemas de IA?	1. La acción nunca se ha realizado	1	
	5.2 Accesibilidad y diseño universal	5.2.1	¿El modelo cumple con estándares internacionales de accesibilidad (cumplimiento técnico)?	1. La acción nunca se ha realizado	1	
		5.2.2	¿El sistema está diseñado para ser accesible a personas con discapacidades (usabilidad real)?	1. La acción nunca se ha realizado	1	
		5.2.3	¿Las auditorías programadas garantizan la práctica adecuada de diversidad cultural en inclusión?	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
		5.2.4	¿La documentación de diversidad cultural respalda su efectividad en el marco de inclusión?	1. La acción nunca se ha realizado	1	
		5.2.5	¿Las auditorías programadas garantizan la práctica adecuada de tratamiento igualitario en equidad?	1. La acción nunca se ha realizado	1	
		5.2.6	¿La documentación de tratamiento igualitario respalda su efectividad en el marco de equidad?	1. La acción nunca se ha realizado	1	
		5.2.7	¿Ha evaluado si el sistema de IA es utilizable por personas con necesidades especiales o con discapacidad o en riesgo de exclusión? NOTA: la exclusión puede provenir de la falta de acceso a la tecnología, la conexión o la falta de conocimientos. ¿Cómo se diseñó esto en el sistema y cómo se verifica?	1. La acción nunca se ha realizado	1	
		5.2.8	¿Ha considerado el reglamento general de protección de datos y el reglamento de IA, durante la fase de desarrollo del sistema de IA?	1. La acción nunca se ha realizado	1	
		5.2.9	¿Ha tenido en cuenta el impacto de su sistema de IA en el público potencial de usuarios específicos?	1. La acción nunca se ha realizado	1	
		5.2.10	¿Ha evaluado si el equipo que participa en la creación del sistema de IA es representativo de la	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
			población destinataria en general, teniendo en cuenta también otros grupos que podrían verse afectados tangencialmente?			
		5.2.11	¿Ha evaluado si pudiera haber personas o grupos que se vieran desproporcionadamente afectados por las consecuencias negativas por la implementación y uso del sistema IA?	1. La acción nunca se ha realizado	1	
		5.2.12	¿La empresa incorporó los principios de diseño universal en los procesos tecnológicos relacionados al sistema, producto o servicio de IA? (RGPD-Art.25)	5. Existe un sistema de medidas y procesos para llevar a cabo la acción de forma coherente. Se llevan a cabo acciones de revisión y mejora continua del sistema.	5	
		5.2.13	¿Existen herramientas implementadas en el modelo, producto o servicio de IA, para asegurar la accesibilidad de los datos personales a los interesados? (RGPD-Art.15)	1. La acción nunca se ha realizado	1	
	5.3 Participación de las partes interesadas	5.3.1	¿Ha previsto algún mecanismo para incluir la participación de partes interesadas con distintas necesidades en el desarrollo y el uso del sistema de IA?	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
		5.3.2	¿Recibió información de diferentes grupos multi-disciplinarios, que representaran diferentes contextos y experiencias para desarrollar el sistema, producto o servicio de IA?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		5.3.3	¿Recibió comentarios sobre el uso del sistema por parte de otros equipos o grupos que representaron diferentes contextos y experiencias?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		5.3.4	¿Se consultó a las partes interesadas durante el diseño del sistema?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		5.3.5	¿Existen mecanismos para involucrar a los interesados en las decisiones relacionadas con el tratamiento de datos? (RGPD-Art.12)	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		5.3.6	¿Toma en cuenta las opiniones de los interesados en los procesos de toma de decisiones? (RGPD-Art.12)	1. La acción nunca se ha realizado	1	
		5.3.7	¿ Dispone de métodos de retroalimentación para los interesados en relación a las políticas de tratamiento de datos? (RGPD-Art.13)	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
6. Bienestar social y medioambiental	6.1 IA sostenible y respetuosa con el medio ambiente	6.1.1	¿Ha establecido mecanismos para medir el impacto medioambiental del desarrollo, despliegue y uso del sistema de IA (por ejemplo, el tipo de energía, cantidad, climatizaciones utilizadas por los centros de datos)?	1. La acción nunca se ha realizado	1	
		6.1.2	¿Ha tomado medidas para reducir el impacto ambiental del ciclo de vida de su sistema de IA?	1. La acción nunca se ha realizado	1	
		6.1.3	¿Se han implementado políticas para garantizar que su modelo, producto o servicio de IA sea sostenible y respetuoso con el medio ambiente? (RGPD-Art.5)	1. La acción nunca se ha realizado	1	
		6.1.4	¿Dispone de estrategias para minimizar el consumo energético de su modelo, producto o servicio de IA? (RGPD-Art.25)	1. La acción nunca se ha realizado	1	
		6.1.5	¿Se evalúan los impactos medioambientales asociados con el uso del modelo, producto o servicio de IA? (RGPD-Art.35)	1. La acción nunca se ha realizado	1	
		6.1.6	¿Existen sistemas para el monitoreo de la sostenibilidad de las operaciones del modelo, producto o servicio IA? (RGPD-Art.32)	1. La acción nunca se ha realizado	1	
		6.1.7	¿El modelo incorpora estrategias para minimizar el impacto	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
	6.2 Impacto social		ambiental del modelo, producto o servicio de IA?			
		6.2.1	¿Ha evaluado si su sistema de IA anima a los humanos a desarrollar apego y empatía hacia el sistema?	1. La acción nunca se ha realizado	1	
		6.2.2	¿Se ha asegurado de que el sistema de IA señala claramente que su interacción social es simulada y que no tiene capacidades de "entender" y "sentir"?	1. La acción nunca se ha realizado	1	
		6.2.3	¿Se ha asegurado de que se comprenden bien por parte de los usuarios, las repercusiones sociales positivas y negativas del sistema de IA?	1. La acción nunca se ha realizado	1	
		6.2.4	¿Adopta medidas para garantizar un impacto social positivo del tratamiento de datos personales en su sistema de IA? (RGPD-Art.35)	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		6.2.5	¿Su organización ha implementado medidas proactivas, para evaluar y mejorar los efectos sociales de usar su modelo, producto o sistema de IA? (RGPD-Art.24)	1. La acción nunca se ha realizado	1	
		6.2.6	¿ Su organización participa activamente, en la evaluación de impacto de sus actividades con el	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
			uso de datos personales? (RGPD-Art.13)			
		6.2.7	¿Se utilizan indicadores sociales para medir el impacto del modelo, producto o servicio?	1. La acción nunca se ha realizado	1	
		6.2.8	¿El sistema de IA contribuye positivamente al bienestar social en general?	1. La acción nunca se ha realizado	1	
		6.2.9	¿Las auditorías programadas garantizan la práctica adecuada de respeto a derechos humanos que permitan cumplir la normativa legal?	1. La acción nunca se ha realizado	1	
		6.2.10	¿La documentación de respeto a derechos humanos respalda su efectividad en lo relacionado a cumplir con leyes?	1. La acción nunca se ha realizado	1	
7. Rendición de cuentas	7.1 Auditoría	7.1.1	¿Se ha asegurado de que el sistema de IA pueda ser auditado por un organismo interno autónomo?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		7.1.2	¿Mantiene registros de auditoría para evaluar el cumplimiento normativo? (RGPD-Art.30)	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		7.1.3	¿El modelo es auditado de manera regular para garantizar la transparencia?	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
		7.1.4	¿Existen registros claros para facilitar las auditorías del sistema?	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		7.1.5	¿Las auditorías programadas aseguran la correcta implementación de prácticas relacionadas con la rendición de cuentas?	1. La acción nunca se ha realizado	1	
		7.1.6	¿La documentación generada en las auditorías demuestra su eficacia en el contexto de la rendición de cuentas?	1. La acción nunca se ha realizado	1	
	7.2 Minimizar y notificar el impacto negativo	7.2.1	¿Ha evaluado qué trabajadores o ramas del equipo participan en todas las fases del sistema de IA (desde el diseño hasta el uso)?	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		7.2.2	¿Se brindó capacitación y educación a los diferentes grupos interesados para fomentar el desarrollo de prácticas relacionadas con la rendición de cuentas?	1. La acción nunca se ha realizado	1	
		7.2.3	¿Se capacita también en el posible marco jurídico aplicable al sistema de IA?	1. La acción nunca se ha realizado	1	
		7.2.4	¿Ha previsto algún tipo de orientación externa para supervisar la ética y la	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
			responsabilidad del sistema de IA, además de las iniciativas internas?			
		7.2.5	¿Dispone de procedimientos para detectar y notificar impactos negativos relacionados con el tratamiento de datos? (RGPD-Art.33)	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		7.2.6	¿Posee mecanismos para mitigar riesgos derivados del tratamiento de datos? (RGPD-Art.32)	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		7.2.7	¿Cuenta con procedimientos de comunicación establecidos para notificar sobre incidentes de seguridad? (RGPD- Art.34)	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		7.2.8	¿Los usuarios son notificados sobre posibles impactos negativos identificados en el modelo, producto o servicio de IA?	1. La acción nunca se ha realizado	1	
		7.2.9	¿Las auditorías programadas garantizan la práctica adecuada de filtrado de temas sensibles en evitar contenido inapropiado?	1. La acción nunca se ha realizado	1	
		7.2.10	¿La documentación sobre el filtrado de temas sensibles garantiza su eficacia en la prevención de contenido inapropiado?	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
	7.3 Documentación de los compromisos	7.3.1	¿Se ha implementado un mecanismo para identificar los intereses y valores relevantes asociados al sistema de IA, así como las posibles compensaciones entre ellos?	1. La acción nunca se ha realizado	1	
		7.3.2	¿Mantiene registros para documentar compromisos relacionados con la protección de datos? (RGPD-Art.30)	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		7.3.3	¿Dispone y almacena documentación que acredita su cumplimiento con las obligaciones del RGPD? (RGPD-Art.24)	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
		7.3.4	¿Se documentan los compromisos éticos relacionados con el uso del modelo?	1. La acción nunca se ha realizado	1	
		7.3.5	¿El sistema incluye documentación sobre compromisos éticos?	1. La acción nunca se ha realizado	1	
		7.3.6	¿Las auditorías programadas garantizan la práctica adecuada de resultados fiables en veracidad?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		7.3.7	¿La documentación sobre resultados fiables garantiza su efectividad en términos de veracidad?	1. La acción nunca se ha realizado	1	

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Requisito	Subnivel	Código de pregunta	Pregunta	Respuesta - Selección	Puntuación - Valor	Comentarios
	7.4 Capacidad de reparación	7.4.1	¿Ha establecido un conjunto adecuado de mecanismos que permitan la reparación en caso de que se produzca algún daño o impacto adverso?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		7.4.2	¿Cuenta con mecanismos para comunicar a los usuarios finales y a terceros las opciones disponibles para ejercer sus derechos o presentar sus quejas?	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		7.4.3	¿Dispone de mecanismos para garantizar la capacidad de reparación en caso de incumplimientos? (RGPD-Art.32)	3. Existen medidas o procesos para llevar a cabo la acción. Sólo en situaciones particulares no se realizan.	3	
		7.4.4	¿Cuenta con estrategias para indemnizar o reparar a los afectados en caso de incidentes de brechas de seguridad que impliquen riesgos significativos para sus derechos y libertades ? (RGPD-Art.34)	2. La acción se lleva a cabo ocasionalmente, dependiendo de la decisión de la propia persona.	2	
		7.4.5	¿Existen procedimientos para gestionar las reclamaciones de los interesados en caso de incidentes de seguridad? (RGPD-Art.24)	4. Existen medidas o procesos para llevar a cabo la acción de forma coherente.	4	
TOTAL					422	

Anexo E. Informe para la Empresa

RESUMEN EJECUTIVO

La responsabilidad en el uso y generación de IA es importante para su empresa, razón por la cual ha accedido a participar en este estudio.

Se realizó una encuesta que consistió en 181 preguntas (171 técnicas y 10 sobre la empresa) a la cual asistieron delegados de los tres departamentos involucrados con IA: innovación, data y analítica y desarrollo e innovación.

Se obtuvieron 3 índices de IA responsable:

- Global, considerando todas las áreas y subáreas evaluadas: 422 que corresponde a nivel 2 – básico.
- Ponderado, considerando el mismo aporte de todas las áreas: 2,42 que corresponde a nivel 2 – básico.
- Por áreas, considerando el cumplimiento de cada subárea: se identificaron subáreas desarrolladas y áreas en nivel 1, por lo que la empresa debe prestar atención a las recomendaciones para afianzar el nivel 2- básico en todas las áreas, sin embargo, la meta es llevar paso a paso a la empresa al nivel 4, que constituye el mínimo nivel para considerar a una IA como responsable.

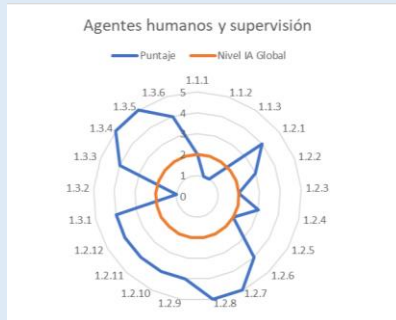
ANÁLISIS DE LAS ÁREAS EVALUADAS

Se evaluaron 7 áreas de responsabilidad en IA cuyos resultados fueron los siguientes:

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Resultado de la evaluación de cada área

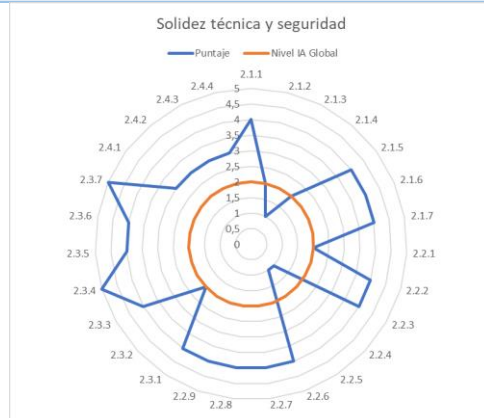
Acciones para afianzar la empresa en nivel 2 y posteriormente llegar al menos a nivel 4 de IA responsable



Derechos fundamentales:

- Establecer mecanismos para informar que el sistema de IA pudiera vulnerar los derechos fundamentales.
- Establecer indicadores específicos para verificar el cumplimiento de los derechos fundamentales en el sistema de IA.

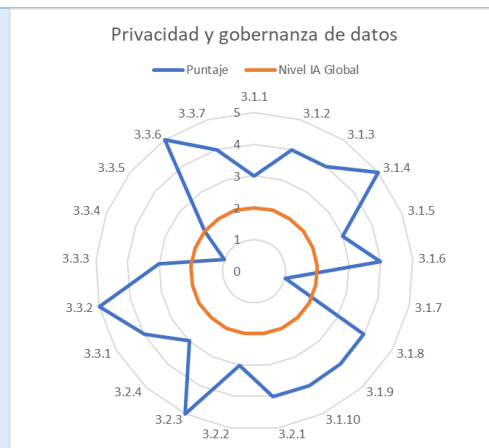
Supervisión humana: Establecer mecanismos o procedimientos para detener la operación del sistema de IA en caso necesario.



Resistencia a ataques y seguridad: Adoptar medidas preventivas para prevenir el uso indebido del sistema de IA.

Plan de emergencia y seguridad:

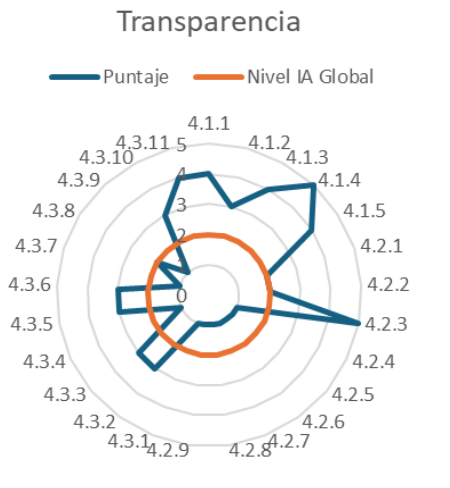
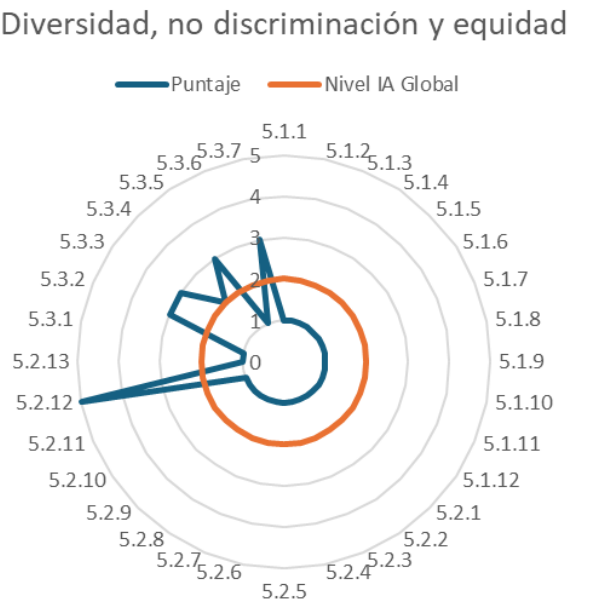
- Comprobar y considerar el nivel de riesgo del sistema de IA cuando opera con el plan alternativo.
- Disponer de póliza de seguros para enfrentar posibles daños a los usuarios o al medio ambiente causados por el servicio de IA.



Respeto a la intimidad: Cumplir con las normativas de privacidad (derechos de los interesados, tratamiento adecuado de datos, obtención del consentimiento del interesado) en los datos utilizados en el servicio de IA.

Acceso a los datos: Establecer un mecanismo de supervisión para registrar cómo, quién, dónde y finalidad se accedieron a los datos.

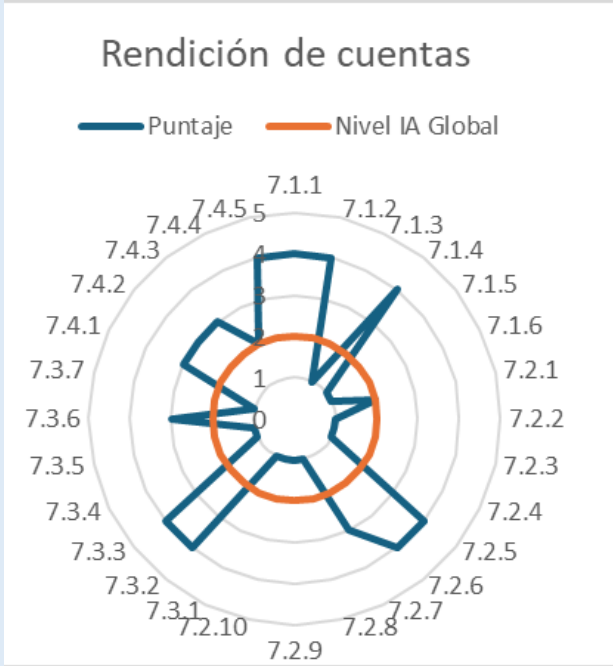
Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Resultado de la evaluación de cada área	Acciones para afianzar la empresa en nivel 2 y posteriormente llegar al menos a nivel 4 de IA responsable
<p>Transparencia</p>  <p>— Puntaje — Nivel IA Global</p>	<p>Explicabilidad:</p> <ul style="list-style-type: none"> • Establecer políticas para garantizar la explicabilidad de las decisiones automatizadas. • Establecer procedimientos para proporcionar explicaciones claras sobre el tratamiento automatizado de los datos. • Explicar claramente como toma decisiones el modelo de IA. • Hacer que las decisiones del sistema de IA sean explicables a los usuarios. • Realizar auditorías programadas para garantizar la explicabilidad en relación a la transparencia del modelo de IA. <p>Comunicación:</p> <ul style="list-style-type: none"> • Establecer mecanismos para informar a los usuarios finales los mecanismos que utilizan los sistemas de IA para entregar los resultados. • Considerar comunicar a otras audiencias los riesgos potenciales o percibidos del uso del sistema de IA. • Establecer estrategias para comunicar al usuario final sobre la política de tratamiento de datos. • Considerar comunicar riesgos y beneficios del uso del sistema de IA al público.
<p>Diversidad, no discriminación y equidad</p>  <p>— Puntaje — Nivel IA Global</p>	<p>Evitar el sesgo injusto:</p> <ul style="list-style-type: none"> • Verificar la equidad del sistema de IA • Establecer auditorías programadas que garanticen prácticas de mitigación de sesgo. • Disponer de documentación efectiva para evitar sesgos. • Adoptar medidas para identificar y eliminar sesgos en procedimientos automatizados. • Establecer auditorías que garanticen la equidad en las decisiones automatizadas. • Establecer procesos internos para mitigar los riesgos de sesgos. • Capacitar al personal para evitar sesgo en el diseño de los sistemas de IA • Evaluar las limitaciones de sesgo por la composición de los conjuntos de datos. • Hacer pruebas de sesgo en poblaciones específicas afectadas por sesgo, • Comprobar y controlar el sesgo durante las etapas de desarrollo, despliegue y uso del sistema de IA • Tomar en cuenta a otros grupos de interés (diferente de los usuarios) que se verán afectados indirectamente por el sistema de IA • Definición operativa de “sesgo” en el diseño de sistemas de IA. <p>Accesibilidad y diseño universal</p>

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Resultado de la evaluación de cada área	Acciones para afianzar la empresa en nivel 2 y posteriormente llegar al menos a nivel 4 de IA responsable
	<ul style="list-style-type: none"> • Cumplir técnicamente los estándares internacionales de accesibilidad. • Diseñar los sistemas con usabilidad real para personas discapacitadas. • Establecer auditorías que garanticen la práctica adecuada de diversidad cultural, inclusión, tratamiento igualitario en equidad. • Realizar documentación de diversidad cultural que respalde la inclusión, tratamiento igualitario, equidad. • Evaluar si el sistema de IA es utilizable por personas en riesgo de exclusión o con necesidades especiales • Considerar en la fase de desarrollo lo señalado en el reglamento general de protección de datos y reglamento de IA de la UE. • Considerar el impacto del sistema de IA en los usuarios potenciales • Verificar que el equipo que participa en la creación del sistema de IA es representativo de la población destinataria final. • Evaluar si hay personas o grupos que se verán desproporcionalmente afectados por las consecuencias negativas de la implementación y uso del sistema de IA. • Incluir herramientas en el sistema de IA que permitan a los interesados acceder a sus datos personales <p>Participación de las partes interesadas</p> <ul style="list-style-type: none"> • Incluir mecanismos para incluir partes interesadas con distintas necesidades durante el desarrollo y uso del sistema de IA.
<p>Bienestar social y medioambiental</p>	<p>IA sostenible y respetuosa con el medio ambiente</p> <ul style="list-style-type: none"> • Establecer mecanismos para medir el impacto en el medio ambiente del desarrollo, despliegue y uso del sistema de IA • Tomar medidas para reducir el impacto ambiental del sistema de IA durante su ciclo de vida. • Implementar políticas para que el sistema de IA sea sostenible y respetuoso con el medio ambiente. • Desarrollar estrategias para minimizar el consumo energético del sistema de IA. • Diseñar monitoreos de sostenibilidad de las operaciones del sistema de IA. • Incluir en el sistema de IA estrategias para reducir su impacto ambiental. <p>Impacto social</p> <ul style="list-style-type: none"> • Evaluar si el sistema de IA anima al hombre a desarrollar apego o empatía hacia el sistema • El sistema de IA debe señalar que la interacción social es simulada, no entiende o siente.

Definición de Controles, Procesos y Estructuras para el Diseño de un sistema de IA responsable

Resultado de la evaluación de cada área	Acciones para afianzar la empresa en nivel 2 y posteriormente llegar al menos a nivel 4 de IA responsable																																																										
 <p>Rendición de cuentas</p> <p>— Puntaje — Nivel IA Global</p> <p>The radar chart displays scores for 20 categories (7.1.1 to 7.4.5) on a scale of 0 to 5. The 'Nivel IA Global' is marked at 1. The 'Puntaje' (score) is shown as a blue line connecting the data points for each category. The scores are generally higher than the global level, with some categories reaching 4 or 5.</p> <table border="1"> <thead> <tr> <th>Categoría</th> <th>Puntaje</th> </tr> </thead> <tbody> <tr><td>7.1.1</td><td>4.5</td></tr> <tr><td>7.1.2</td><td>4.5</td></tr> <tr><td>7.1.3</td><td>4.5</td></tr> <tr><td>7.1.4</td><td>4.5</td></tr> <tr><td>7.1.5</td><td>4.5</td></tr> <tr><td>7.1.6</td><td>4.5</td></tr> <tr><td>7.2.1</td><td>4.5</td></tr> <tr><td>7.2.2</td><td>4.5</td></tr> <tr><td>7.2.3</td><td>4.5</td></tr> <tr><td>7.2.4</td><td>4.5</td></tr> <tr><td>7.2.5</td><td>4.5</td></tr> <tr><td>7.2.6</td><td>4.5</td></tr> <tr><td>7.2.7</td><td>4.5</td></tr> <tr><td>7.2.8</td><td>4.5</td></tr> <tr><td>7.2.9</td><td>4.5</td></tr> <tr><td>7.2.10</td><td>4.5</td></tr> <tr><td>7.3.1</td><td>4.5</td></tr> <tr><td>7.3.2</td><td>4.5</td></tr> <tr><td>7.3.3</td><td>4.5</td></tr> <tr><td>7.3.4</td><td>4.5</td></tr> <tr><td>7.3.5</td><td>4.5</td></tr> <tr><td>7.3.6</td><td>4.5</td></tr> <tr><td>7.3.7</td><td>4.5</td></tr> <tr><td>7.4.1</td><td>4.5</td></tr> <tr><td>7.4.2</td><td>4.5</td></tr> <tr><td>7.4.3</td><td>4.5</td></tr> <tr><td>7.4.4</td><td>4.5</td></tr> <tr><td>7.4.5</td><td>4.5</td></tr> </tbody> </table>	Categoría	Puntaje	7.1.1	4.5	7.1.2	4.5	7.1.3	4.5	7.1.4	4.5	7.1.5	4.5	7.1.6	4.5	7.2.1	4.5	7.2.2	4.5	7.2.3	4.5	7.2.4	4.5	7.2.5	4.5	7.2.6	4.5	7.2.7	4.5	7.2.8	4.5	7.2.9	4.5	7.2.10	4.5	7.3.1	4.5	7.3.2	4.5	7.3.3	4.5	7.3.4	4.5	7.3.5	4.5	7.3.6	4.5	7.3.7	4.5	7.4.1	4.5	7.4.2	4.5	7.4.3	4.5	7.4.4	4.5	7.4.5	4.5	<ul style="list-style-type: none"> Asegurarse que los usuarios finales entienden bien las repercusiones sociales positivas y negativas de su interacción con el sistema de IA. Implementar medidas proactivas para evaluar los efectos sociales del uso del sistema de IA. Evaluar el impacto del uso de datos personales por parte de la empresa. Utilizar indicadores sociales para medir el impacto de sistema de IA. Verificar si el sistema de IA contribuye al bienestar general del hombre. Establecer auditorías que confirmen el respeto a los derechos humanos por parte del sistema de IA. Realizar documentación que confirme que el sistema de IA respeta los derechos humanos. <p>Auditorías</p> <ul style="list-style-type: none"> Auditar regularmente el modelo o sistema de IA para garantizar transparencia. Auditar la implementación de prácticas de rendición de cuentas. Generar documentación que confirme que se rinden cuentas respecto al sistema de IA. <p>Minimizar y notificar el impacto negativo</p> <ul style="list-style-type: none"> Capacitar a los diferentes grupos de interesados en prácticas de rendición de cuentas. Capacitar a las partes interesadas en el marco jurídico aplicable al sistema de IA. Prever orientación externa para supervisar la ética y responsabilidad del sistema de IA. Notificar a los usuarios de los impactos negativos del sistema de IA. Auditar que se filtren temas sensibles y evite contenido inapropiado. Generar documentación que confirme que el sistema de IA es eficaz previniendo contenido inapropiado. <p>Documentar los compromisos y compensaciones</p> <ul style="list-style-type: none"> Identificar intereses y valores asociados al sistema de IA Documentar los compromisos éticos relacionados con el uso del sistema de IA Incluir documentación ética en el sistema de IA. Verificar la veracidad de la documentación sobre resultados fiables
Categoría	Puntaje																																																										
7.1.1	4.5																																																										
7.1.2	4.5																																																										
7.1.3	4.5																																																										
7.1.4	4.5																																																										
7.1.5	4.5																																																										
7.1.6	4.5																																																										
7.2.1	4.5																																																										
7.2.2	4.5																																																										
7.2.3	4.5																																																										
7.2.4	4.5																																																										
7.2.5	4.5																																																										
7.2.6	4.5																																																										
7.2.7	4.5																																																										
7.2.8	4.5																																																										
7.2.9	4.5																																																										
7.2.10	4.5																																																										
7.3.1	4.5																																																										
7.3.2	4.5																																																										
7.3.3	4.5																																																										
7.3.4	4.5																																																										
7.3.5	4.5																																																										
7.3.6	4.5																																																										
7.3.7	4.5																																																										
7.4.1	4.5																																																										
7.4.2	4.5																																																										
7.4.3	4.5																																																										
7.4.4	4.5																																																										
7.4.5	4.5																																																										

RECOMENDACIONES FINALES

- Educar a todo el personal involucrado en la concepción, diseño, desarrollo, implementación y uso de los sistemas de IA en todo lo relacionado a IA responsable, a fin de que cada persona identifique la parte del cumplimiento que se encuentra dentro de sus actividades y lo realice como parte de sus tareas diarias.
- Designar a un delegado de los departamentos relacionados con IA para que conformen el equipo de IA responsable y sean los encargados de velar por el tema dentro de su departamento.
- Atender primero las recomendaciones de las áreas: *7. Rendición de cuentas*, *6. Bienestar social y medio ambiental* y *5. Diversidad, no discriminación y equidad* ya que tienen el mayor número de subáreas con respuesta “La acción nunca se ha realizado”, para que todas las áreas y subáreas de IA responsable lleguen a nivel 2 y de ahí implementar las acciones necesarias para alcanzar el nivel 4 que es el nivel mínimo para señalar que se trata de una IA responsable.
- Realizar evaluaciones periódicas del estado de la IA responsable en las todas las áreas ya que existen subáreas que tienen nivel 1 y todas deben subir a nivel 2.
- Los resultados obtenidos muestran que la entidad financiera da prioridad al desarrollo de productos y servicios de IA, por sobre el cumplimiento de las recomendaciones existentes, tal vez por falta de conocimiento de las regulaciones existentes o porque no son obligatorios a la fecha del estudio.
- El nivel más alto alcanzado fue en el área de “Privacidad y Gobernanza del Dato” y el área con menor cumplimiento es “Bienestar social y medioambiental”.