

INDESIA



Informe Ética & IA Industrial

El uso responsable de la inteligencia artificial en el sector industrial



01

PRÓLOGO

Objetivo de este documento
y a quién está dirigido

Agradecimientos

Autores

02

INTRODUCCIÓN

Cómo definir la inteligencia artificial

Qué significa una IA ética

Diferencias y relación entre ética y normativa

Justificación de la necesidad
de un marco ético para la IA

Aspecto ético en el ámbito industrial

03

ESTADO DE LA CUESTIÓN DE LA ÉTICA EN IA

De la teoría a la realidad aplicada

Ejemplos de iniciativas globales

Referentes internacionales en ética de la IA

Marco de desarrollo GuIA para garantizar
una IA ética en la industria

Regulación, auditoría y certificación

Modelo de madurez de implementación
ética de IA

04

ANÁLISIS DE IMPACTO ÉTICO Y SOCIAL EN LA INDUSTRIA ESPAÑOLA

Propuesta de principios de IA
adoptados por IndesIA

Análisis de los casos de uso
de miembros IndesIA

Impacto de la regulación "AI Act"
en el sector industrial

Conclusiones del análisis

05

TOOLKIT DE IMPLEMENTACIÓN DE UNA IA ÉTICA EN EL SECTOR INDUSTRIAL

Toolkit de implementación e instrucciones
para su utilización

Herramienta cuestionario interactivo
y medida de impacto

Sistema de gobernanza de la IA
en empresa

Recursos de formación

Herramientas técnicas disponibles

06

CASO DE ÉXITO DE GOBERNANZA DE IA EN EMPRESA

Telefónica: Orígenes e IA responsable

Visión y principios de la IA

Factor diferenciador:
Gobernanza IA interna

Implementación:
De los principios éticos a la acción

Aprendizajes y recomendaciones
de mejores prácticas

07

CASO DE USO APLICADO PARA LA IMPLEMENTACIÓN DE IA ÉTICA EN LA INDUSTRIA

Contexto del caso
y consideraciones iniciales

Elementos de análisis

Relación con la gobernanza IA

08

CONCLUSIONES Y CONSEJOS PRÁCTICOS

01

PRÓLOGO

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

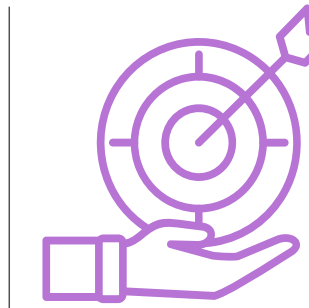
Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



1.1. Objetivo de este documento y a quién está dirigido

Este documento es un esfuerzo compartido de los miembros de **IndesIA**¹, la Asociación Industrial para el Impulso de la Economía del Dato y de la Inteligencia Artificial y **OdiselA**² (entidad colaboradora de IndesIA), el Observatorio del impacto social y ético de la inteligencia artificial. El documento tiene como objetivo presentar el estado del arte de la ética de inteligencia artificial aplicada al sector industrial, con un énfasis práctico y una serie de herramientas identificadas para una implantación responsable de la IA, en base a principios éticos especialmente adaptados al contexto industrial y los diferentes subsectores: energía, naval, telecomunicaciones, automoción, aeronáutica, construcción, e ingeniería.

Por lo tanto, este documento es una guía actualizada para cualquier organización adoptadora de inteligencia artificial que quiera aplicar la ética y la responsabilidad desde el diseño para sus implementaciones IA, en un contexto particularmente importante para la industria española y en un periodo decisivo para la regulación IA en Europa y la supervisión a nivel nacional de los algoritmos.

Notas

1. Indesia - www.indesia.org
2. OdiselA - www.odiseia.org

Prólogo

Introducción

Estado de la cuestión
de la ética en IA


Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



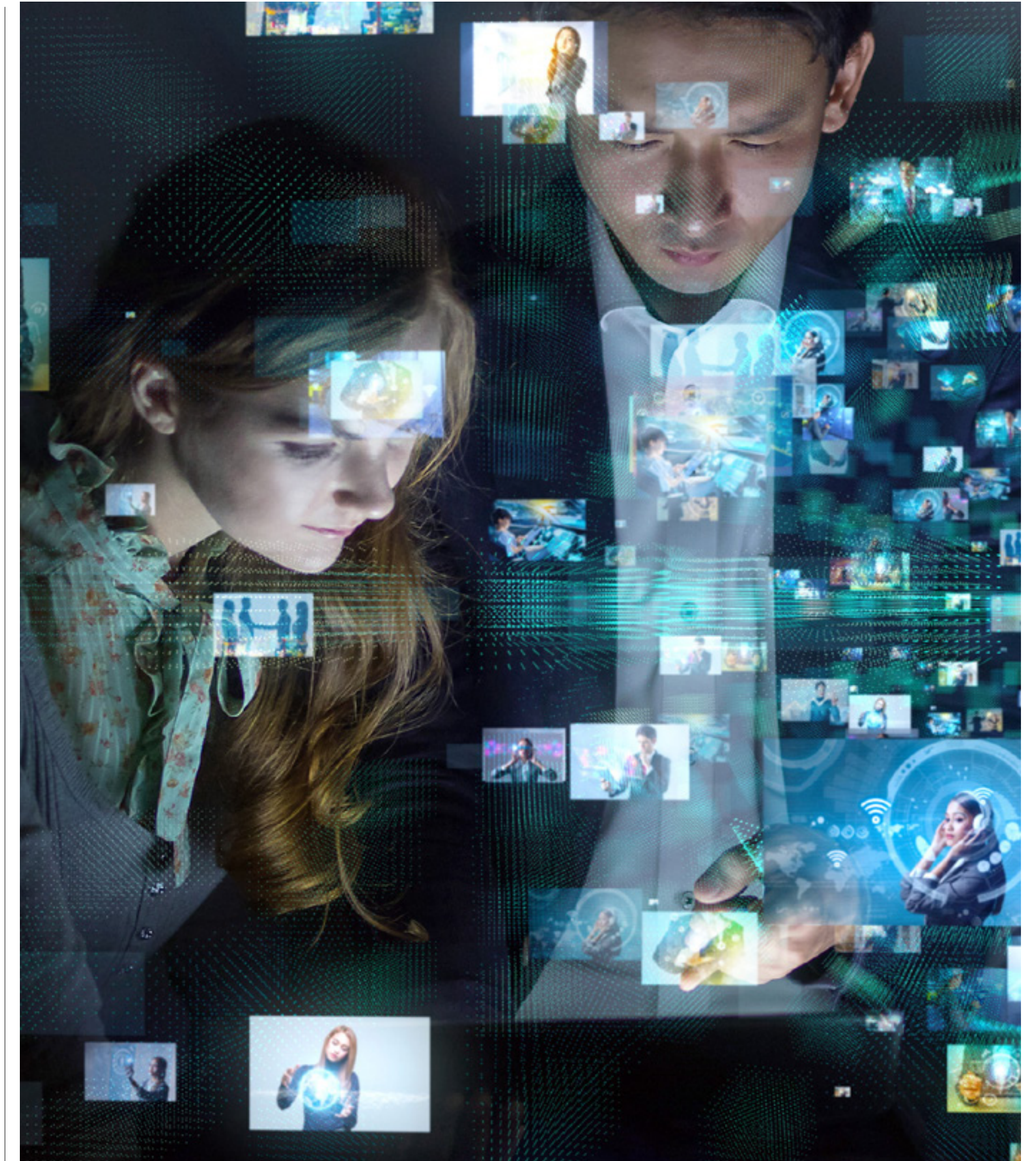
La IA revolucionará realmente nuestra forma de vivir y nuestra economía



1.2. Agradecimientos

El presente documento no sería posible sin el esfuerzo de diferentes actores, los cuales merecen un reconocimiento por las siguientes aportaciones:

- Los autores del informe, con una labor intensa de análisis, investigación, y divulgación llevada a cabo entre miembros de Microsoft, OdiselA, Repsol y Telefónica.
- Empresas promotoras de IndesIA (Airbus, Ferrovial, Gestamp, Navantia, Repsol, Técnicas Reunidas, Telefónica), las cuales han aportado sus casos de uso IA, así como la revisión del documento. Dichas aportaciones han permitido aterrizar los análisis con ejemplos reales y actuales de proyectos IA.



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

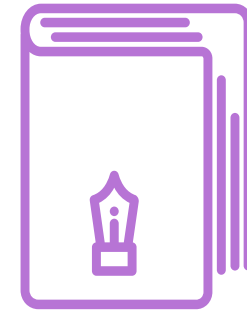
Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



1.3 Autores

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



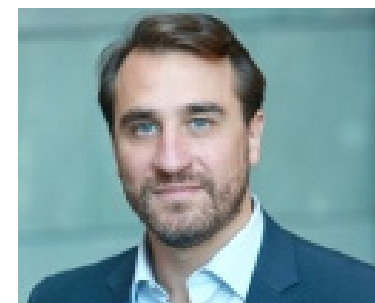
Juan Manuel Belloto Gómez

Asesor independiente
Director y coordinador general @OdiselA



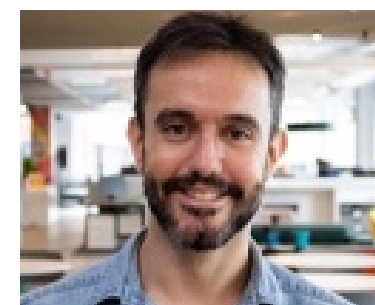
Richard (Dr.) Benjamins

Chief Data & AI Strategist @ Telefónica
Cofundador y Vicepresidente @ OdiselA



Juan José Casado Quintero

Chief Digital Officer @Repsol



Adrián González Sánchez

Miembro del Observatorio y Líder IA Industrial
@OdiselA Especialista de Datos & AI @ Microsoft



Luis Lamiable Carbonell

Responsable AI @ Telefónica



Alberto Pinedo Lapeña

National Technology Officer @ Microsoft Cloud Commission
VP + IA & Big Data Commission Member @ AMETIC



Rosa Romero Soler

Miembro del Observatorio y Codirectora Educación
@OdiselA Head of AI @ Cenit



Joaquina Salado Moraleda

Senior Manager Chief Data Office
Responsable AI@Telefónica



02

INTRODUCCIÓN

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos





Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

La inteligencia artificial (IA) está transformando la forma en que vivimos, producimos y trabajamos. Pero si bien la IA tiene el potencial de generar eficiencias e impulsar la producción económica, también plantea desafíos éticos y legales. Si se examinan y abordan los problemas de sostenibilidad que rodean a la IA, creemos que la tecnología tiene el potencial de servir tanto a las empresas como a la humanidad y crear oportunidades de inversión en el camino.

La IA revolucionará realmente nuestra forma de vivir y nuestra economía siempre y cuando la humanidad le otorgue su “licencia social” y para ello debe garantizarse un marco ético común de aplicación en estas nuevas tecnologías.

La disrupción de la inteligencia artificial en todos los ámbitos está presentando una serie de disfunciones debido a la velocidad de implantación y la falsa creencia del “cuanto antes mejor”, que no está otorgando espacio y tiempo para el debate ético, tan necesario como las evidencias muestran en este sentido.



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



2.1. Cómo definir la inteligencia artificial

Hoy en día, hay tantas definiciones diferentes de lo que la inteligencia artificial es o cómo debe ser definida, como alcances y capacidades se le presuponen a esta familia de tecnologías. Sin embargo, hay un cierto consenso con relación al hecho de que la inteligencia artificial es capaz de aumentar las capacidades humanas, incluidas aquellas cognitivas (tales como reconocimiento de imágenes, o transcripciones de audio) y las de automatización de la toma de decisiones.

Un sistema de IA es un sistema basado en máquinas capaz de influir sobre el entorno generando un resultado (predicciones, recomendaciones o decisiones) respecto a una serie de objetivos. Usa datos y aportaciones de máquinas y/o humanos para (i) percibir entornos virtuales y/o reales; (ii) sintetizar estas percepciones en modelos mediante el análisis de manera automática (e.g., con aprendizaje automático), o manualmente; y (iii) usar el modelo para formular opciones de resultados por inferencia. Los sistemas de IA están diseñados para operar con diversos niveles de autonomía. (OCDE 2019).

A lo largo de los años, ha habido discusiones en torno a lo que estas capacidades tecnológicas suponen para el futuro del trabajo, la privacidad individual (por el uso indebido de datos privados), y los riesgos potenciales en cuanto a sesgos o modelos no explicables. Todas estas reflexiones se engloban dentro de las consideraciones éticas de la inteligencia artificial.

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

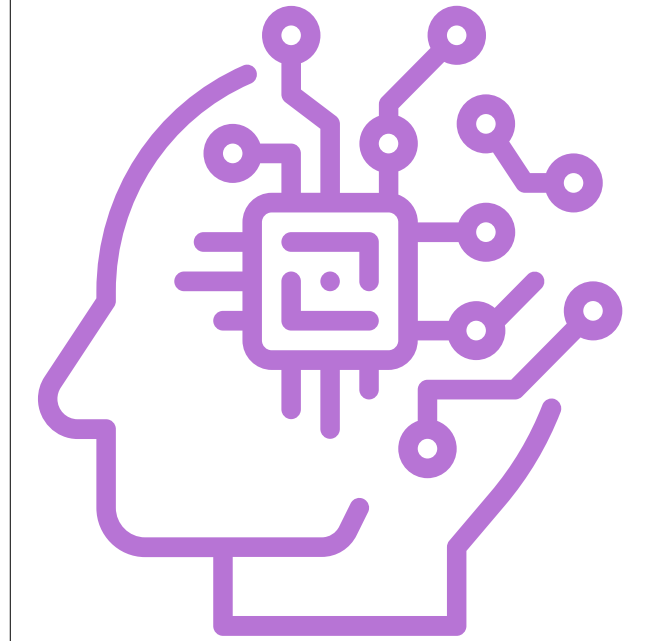
2.2. Qué significa una IA ética

La IA ética es parte de la familia de éticas aplicadas y su área de estudio se enfoca en el análisis del impacto positivo (o al menos no negativo) de las aplicaciones de inteligencia artificial. Es un área de estudio que ha cobrado relevancia a lo largo de los últimos años, motivada por casos específicos con consecuencias negativas o dañinas de los datos y la inteligencia artificial en el mundo, aunque la intención fuera buena o legítima.

Hablamos de una inteligencia artificial ética para hacer referencia a aquella IA que pretende evitar impactos o externalidades negativas derivadas de su uso. La tecnología es éticamente neutra, pero no así el uso que de ella se haga. La legislación vigente cada vez regula más obligaciones para desarrolladores y usuarios de sistemas de IA, así como derechos para posibles afectados. En tanto en cuanto exista dicha regulación ya no cabe hablar propiamente de ética de la IA para el conjunto de usos afectado por la regulación. El cumplimiento de la ética es un proceso voluntario, es decir nos referimos a obligaciones cuyo incumplimiento no implica una infracción sancionable por el Derecho. Dentro

de los márgenes de actuación que permite el Derecho, en ocasiones hay muchas opciones para el desarrollador o usuario del sistema de IA que pueden tener muy diversas consecuencias. Estas consecuencias, siendo la intención buena, pueden ser positivas o dañinas. La ética aplica cuando las actuaciones cumplen con el derecho, pero las consecuencias pueden ser negativas.

Como ya hemos visto, la inteligencia artificial es una tecnología transformadora con muchas oportunidades para mejorar nuestra vida, nuestra economía y nuestra sociedad. Pero también sabemos que su uso puede tener consecuencias negativas, aunque muchas veces no intencionadas. ¿Quién no conoce el caso del sistema de IA que discrimina a las personas de color en el ámbito jurídico en Estados Unidos³? ¿O el sistema inteligente que ayudaba a la contratación de personal y que discriminaba por género⁴? ¿O la influencia en procesos democráticos por algoritmos de recomendación que fomentan la polarización de la sociedad?



Notas

3. AI is sending people to jail and getting it wrong, MIT Technology Review.

4. Amazon scraps secret AI recruiting tool that showed bias against women, Reuters.

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

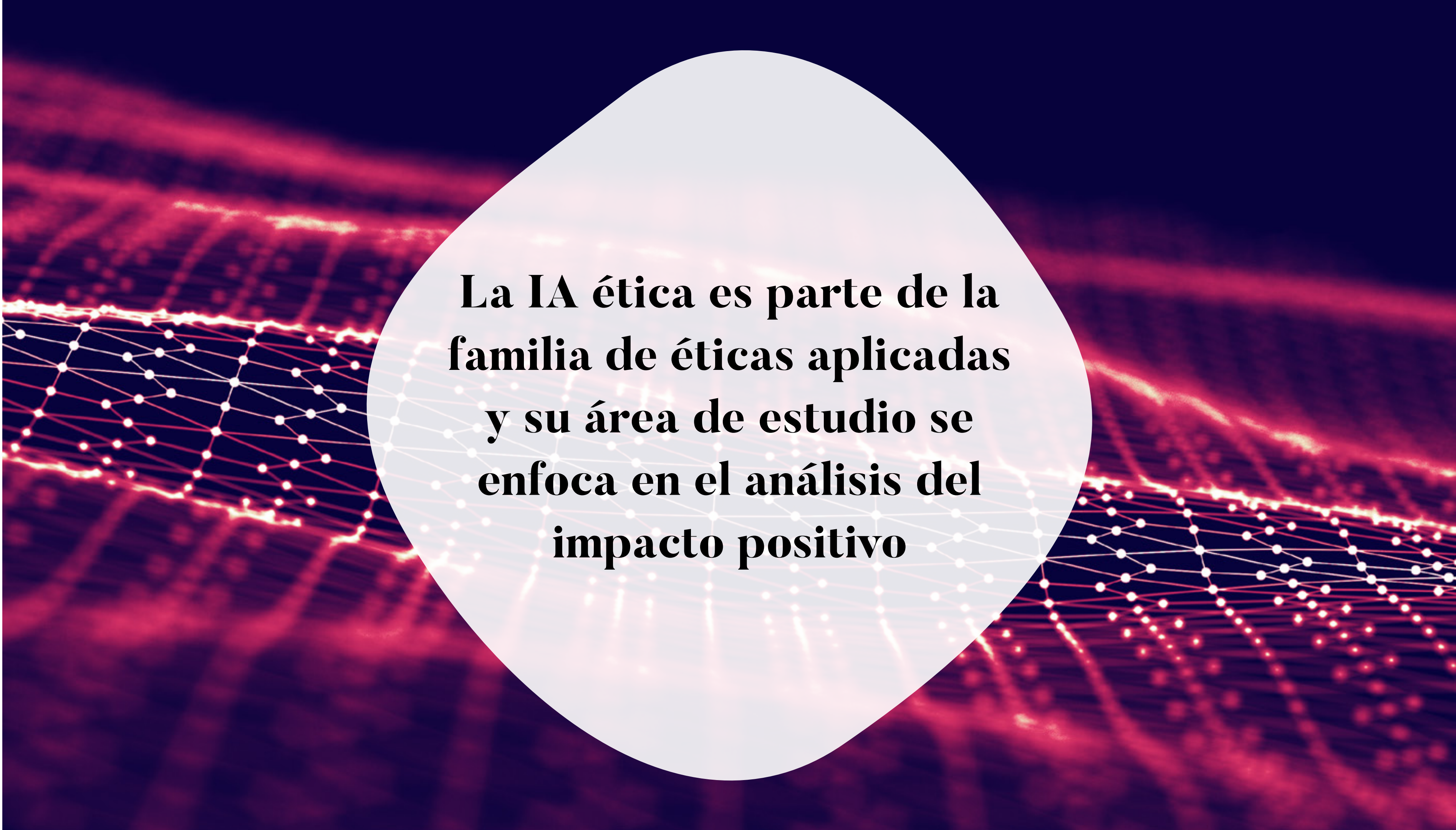
Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



La IA ética es parte de la familia de éticas aplicadas y su área de estudio se enfoca en el análisis del impacto positivo

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

¿Escándalos de privacidad⁵ o falta de transparencia? ¿Sesgos de género en la traducción automática? Y existe un largo etc.

Estas consecuencias negativas ocurren porque en un sistema de inteligencia artificial que aprende de los denominados datos masivos (big data), dichos datos pueden contener sesgos y una exactitud que no es 100 % perfecta. Es decir, siempre hay lugar para errores e imprecisiones, a lo que se suma que los algoritmos modernos son estructuras muy complejas difícilmente entendibles por las personas.

El objetivo de este documento es explicar de una manera práctica, a las empresas y organizaciones que quieran aplicar la inteligencia artificial para mejorar su actividad, cómo hacerlo de tal forma que evite las externalidades o consecuencias negativas derivadas de su uso.



Notas

5. Cambridge Analytica and Facebook: The Scandal and the Fallout So Far, NY Times.

Existen metodologías y procesos que pretenden minimizar las externalidades negativas derivadas del uso de IA y que pueden implementarse a través de normas éticas. Estas metodologías tratan de concretizar los pasos importantes cuando uno quiere aplicar la inteligencia artificial, incluyendo:

- **Seleccionar los principios de inteligencia artificial más adecuado para el sector en cuestión. Los riesgos no son iguales en el sector industrial (donde la robustez y seguridad es muy importante) y el sector sanitario (donde la explicabilidad y la privacidad es importante).**
- **Estimar el daño en caso de que algo vaya mal con la aplicación del sistema. No es lo mismo equivocarse en la recomendación de una película que en el diagnóstico de una enfermedad grave como el cáncer. El daño de un error de la inteligencia artificial se puede estimar teniendo en cuenta su gravedad, su escala y la probabilidad de que ocurra.**
- **Aplicar la metodología "inteligencia artificial ética desde el diseño". Esta metodología considera que es necesario unos principios de IA, una formación y concienciación a los empleados, un cuestionario con las preguntas adecuadas además de unas recomendaciones, herramientas técnicas, y un modelo de gobernanza. Si una organización sigue esta metodología, es probable que detecte y pueda mitigar los posibles riesgos.**

Una vez conocidos los riesgos y las posibles medidas correctivas, la organización tiene que decidir cómo actuar. Lo recomendable desde

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

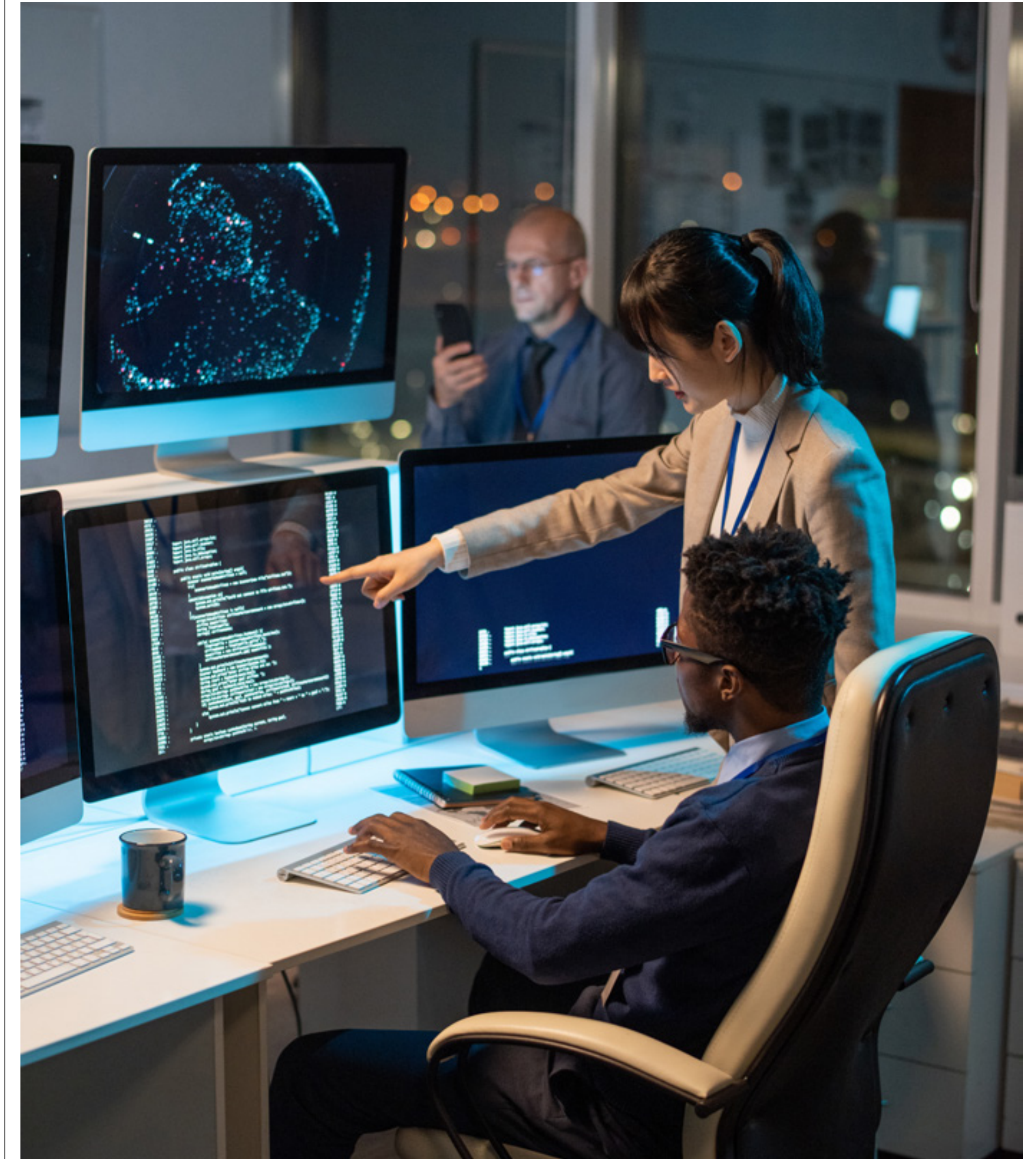
Conclusiones
y consejos prácticos

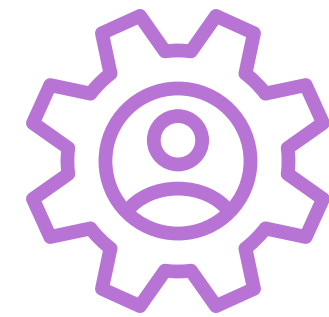
el punto de la ética aplicada a la IA sería que en caso de que no se pueda corregir y evitar los riesgos considerados inaceptables, no usar la inteligencia artificial para este caso de uso, independientemente de las normas “duras” que deba respetar.

En este momento, “termina” la parte ética y podría entrar la parte normativa (hard law). Una normativa podría exigir un proceso riguroso para aplicaciones de alto riesgo antes de que se pueda poner en producción. Y si no se cumple este proceso, imponer sanciones. La propuesta de la nueva regulación europea para la inteligencia artificial en 2021 (AI Act⁶) es la primera regulación horizontal que regulará los niveles de riesgo de los casos de uso IA, para las personas y la sociedad. Allí también entra una posible certificación que justificaría que, en casos de alto riesgo, se haya hecho todo lo posible para evitar riesgos inaceptables del uso de la inteligencia artificial.

Notas

6. Regulation Of The European Parliament And Of The Council Laying Down Harmonised Rules On Artificial Intelligence, European Commission.





2.3. Diferencias y relación entre ética

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

En el marco de la inteligencia artificial, la normativa y la ética se complementan. ¿Puede existir una sin la otra? ¿Se retroalimentan o son independientes? La ética puede considerarse que inspira a la creación de normas jurídicas y además puede servir para concretar el significado y contenido de dichas normas y las obligaciones que suponen, asimismo la ética marca las pautas esenciales a seguir en cada una de las fases del ciclo de vida de la inteligencia artificial.

La ética suele estar motivada por una convicción moral o cultural con respecto a lo que es correcto, como es el caso de la inteligencia artificial, donde los observatorios, empresas, y think tanks definen su concepto de una IA ética y positiva.

Las normas jurídicas que conforman el ordenamiento deben tenerse en cuenta a la hora de desarrollar protocolos o reglas éticas.

Un claro ejemplo de ello lo encontramos el Considerando 58 del Reglamento General de Protección de Datos de la Unión Europea⁷, a través del cual se establece que los sujetos obligados tienen que cumplir con el principio de transparencia, que exige que “toda información dirigida al público o al interesado sea concisa, fácilmente accesible y fácil de entender, y que se utilice un lenguaje claro y sencillo, y, además, en su caso, se visualice”, lo que conlleva que el principio ético de transparencia en el tratamiento de datos se vea reforzado con una norma imperativa fiscalizable por los poderes públicos y que, de incumplirse, se impongan las correspondientes sanciones.

El ser humano puede programar una solución de inteligencia artificial para la toma de decisiones con impacto sobre los negocios, la naturaleza y la propia vida humana. La inteligencia artificial no es y no puede ser

ajena a los principios y normas en evolución que rigen el comportamiento humano y mantienen o mejoran la estabilidad social.

El desarrollo de normas éticas supone un gran reto en la labor legislativa moderna en la medida en que muchas de estas normas éticas acaban positivizándose, convirtiéndose en normas duras que forman parte del ordenamiento jurídico y cuyo incumplimiento puede llevar aparejadas sanciones. De igual modo, la ética de la IA se acaba sustanciando en normativa que por lo general no es exigible jurídicamente como son los códigos de conducta, códigos sectoriales, normas profesionales, normalización técnica de la IA, etc. De hecho, normas jurídicas “duras” establecen marcos y promueven la adopción de códigos o estándares éticos así como sistemas de autorregulación.

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

2.4. Justificación de la necesidad de un marco ético para la IA

Una de las reflexiones por parte de equipos de inteligencia artificial es la conveniencia de concretar los conceptos éticos mediante recomendaciones, guidelines y modelos de gobierno que ayuden a las empresas a saber cómo hacerlo. Incluso implementar el cumplimiento de conceptos éticos mediante herramientas que permiten automatizar su gestión.

En ese sentido, 2022 nos ofrece una oportunidad increíble para poder adoptar enfoques responsables y éticos de manera proactiva, teniendo en cuenta que la regulación AI Act está en vías de ser aprobada e implementada, y que España (a través de la SEDIA) se ha ofrecido para lanzar un programa piloto para dicha regulación⁸.

Notas

8. Ministerio de Asuntos Económicos y Transformación Digital.



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



2.5. Aspecto ético en el ámbito industrial

El estado de madurez de inteligencia artificial de las empresas españolas y el potencial de esta en el sector industrial invita a reflexionar acerca de la importancia de una IA ética pero también de su adaptación al ámbito industrial.

No existe mucha literatura dedicada a este sector, ya sea a nivel nacional o internacional, pero el presente documento pretende coleccionar iniciativas, herramientas, y buenas prácticas de empresas del sector como un modelo de referencia para las organizaciones nacionales. Así mismo, el marco ético propuesto será orientado a la realidad de las empresas del sector industrial.

A lo largo del presente documento, se van a evaluar los principios éticos aplicables a los sectores de energía, ingeniería, telecomunicaciones, naval, automoción, aeronáutica y construcción. Estos principios, similares en forma a los incluidos en otros análisis éticos, van a ser medidos en base a su importancia relativa en el sector industrial. A modo de resumen anticipado, los aspectos de seguridad física o safety tienen una importancia vital, sobre todo en aplicaciones que afectan a máquinas o procesos industriales.

Prólogo

Introducción

**Estado de la cuestión
de la ética en IA**

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

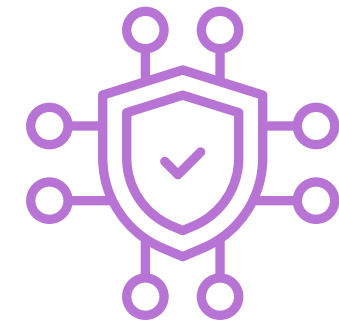
Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



03

ESTADO DE LA CUESTIÓN
DE LA ÉTICA EN IA



Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

El presente apartado tiene como objetivo realizar un recorrido sobre la situación actual de la IA ética aplicada a nivel mundial. Para ello, se realiza un breve recorrido por las iniciativas globales que hacen referencia a la IA ética, así como algunos marcos de trabajo y regulaciones aplicables.

3.1. De la teoría a la realidad aplicada

Mientras que la discusión sobre la ética en IA en el mundo académico ha estado activa en los últimos años, varias voces han expresado su preocupación por el impacto adverso de la IA tras recientes incidentes en el mundo real. El ritmo vertiginoso en el despliegue de los sistemas inteligentes en todos los ámbitos: cotidianos, culturales, industriales y educativos, sin regulación ni principios o guías que los rijan, ha provocado paradas abruptas que no están resultando de fácil reinicio.

Se observa a continuación que los primeros estudios y artículos científicos al respecto comenzaron, según Scopus⁹, en el año 2013 donde aparecen sólo 27 artículos que hacen referencia a la Ética de la Inteligencia Artificial.

Solo tres años más tarde, cuando aún no había producción científica significativa, comienzan a aparecer noticias impactantes sobre casos de sistemas de IA que plantean problemas éticos graves. Desde el caso del sistema de reclutamiento que discrimina por género en

Scopus, artículos relacionados
Búsqueda: Ethics & AI

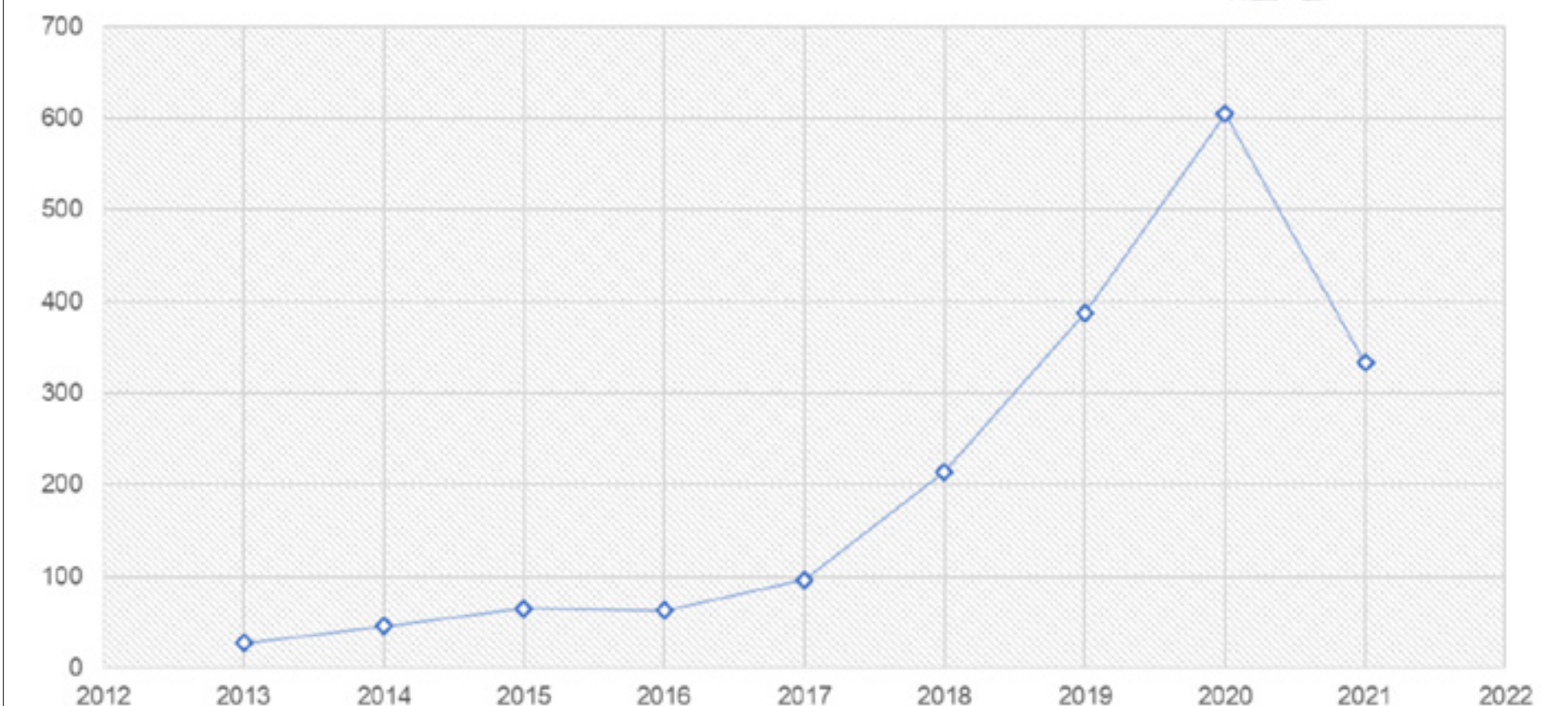


Figura 2 - Artículos relacionados a la ética y la IA, Scopus

Notas

9. Términos clave: #Ethics #Artificial Intelligence. Búsqueda realizada a través de la Biblioteca de la Universidad de Sevilla.

	<p>Amazon (2018), el caso de Facebook y Cambridge Analítica (2018), y más recientemente los casos de moratoria en la comercialización de los sistemas de reconocimiento facial de Amazon, Microsoft o IBM (2020), así como la polémica de la investigadora de Timnit Gebru en Google (2020).</p>	<p>sobre Ética de Sistemas Autónomos e Inteligentes junto con sus Grupos de Trabajo de Estándares IEEE P7000™, que fueron marcados bajo el concepto de EAD (La Iniciativa Global IEEE sobre Ética de Sistemas Autónomos e Inteligentes, 2019).</p>
Prólogo	<p>Tradicionalmente, el “gap” entre aquellas cuestiones o sistemas de carácter revolucionario objeto de estudio por el sector académico e investigador y su puesta en marcha en el mundo real ha permitido controlar el proceso e ir adecuando regulación, normativa y metodología para minimizar su potencial impacto social negativo.</p>	<p>Como expuso Virginia Dignum en 2017 durante Conferencia conjunta internacional IJCAI sobre IA¹⁰, Responsible autonomy (2017): “se necesitan medios para integrar los valores morales, sociales y legales con los desarrollos tecnológicos en IA, tanto durante el proceso de diseño como formando parte de los algoritmos de deliberación empleados por estos sistemas”.</p>
Introducción		
Estado de la cuestión de la ética en IA	<p>No ha sido así con los sistemas inteligentes que tratados como cualquier otro sistema tecnológico no contemplaron desde el inicio su impacto social y el posible efecto negativo de la retroalimentación del impacto generado. La IA difiere de tradicionales sistemas tecnológicos al ampliar tanto su alcance como la escala de su impacto y la repercusión del uso que se hace de ella en su propio diseño como sistema inteligente. En concreto, la IA que se basa en el aprendizaje automático se diferencia por ser indeterminista, por cometer errores, por poder ampliar los sesgos existentes en los datos, y por tener capacidad de autonomía.</p>	<p>El siguiente apartado un resumen de la situación actual de los estudios y avances en materia de IA ética y responsable.</p>
Análisis de impacto ético y social en la industria española		
Toolkit de implementación de una IA ética en el sector industrial		
Caso de éxito de gobernanza de IA en empresa		
Caso de uso aplicado para la implementación de IA ética en la industria	<p>Los intentos de llevar la discusión académica en curso sobre el terreno se han plasmado en forma de directrices y principios que carecen de prácticas para implementarlos (Morley et al. 2019). De estas pautas, quizás las más destacadas han sido las pautas de IEEE para el diseño éticamente alineado (el IEEE Iniciativa Global de Ética de Sistemas Autónomos e Inteligentes 2019), nacida de la Iniciativa Global IEEE</p>	
Conclusiones y consejos prácticos		

Notas

10. IJCAI - International Joint Conferences on Artificial Intelligence Organization.

INICIATIVAS GLOBALES

3.2. Ejemplos de iniciativas globales

Han sido muchas las organizaciones que han elaborado declaraciones de los valores o principios que deberían guiar el desarrollo y despliegue de la IA en la sociedad. A continuación, se presenta un extracto de algunas de las más relevantes, con mención especial a aquellas más avanzadas en el ámbito tecnológico-industrial:



El 24 de noviembre de 2021, la Conferencia General de la UNESCO adoptó la Recomendación sobre la ética de la inteligencia artificial en su 41ª reunión. La UNESCO se embarcó en un proceso de dos años para elaborar este primer instrumento normativo mundial sobre la ética de la inteligencia artificial en forma de recomendación, tras la decisión de su Conferencia General en su 40ª reunión en noviembre de 2019.

- **En 2020, la atención se centró en la preparación del proyecto de texto de la Recomendación con la asistencia de un Grupo de**

expertos ad hoc (AHEG). Esta fase incluyó consultas inclusivas y multidisciplinares con una amplia gama de partes interesadas, las cuales fueron extremadamente importantes para garantizar que el texto fuera lo más inclusivo posible.

- **Hacia fines de 2020 y en 2021, el foco estuvo en un proceso intergubernamental y en las negociaciones sobre el proyecto de texto para producir una versión final de la Recomendación para su posible adopción por parte de la Conferencia General de la UNESCO en su 41ª reunión a fines de 2021 en París.**

El trabajo de la UNESCO sobre la Recomendación se basa en el estudio preliminar sobre la ética de la inteligencia artificial de la Comisión Mundial de Ética del Conocimiento Científico y la Tecnología de la UNESCO (COMEST). Este estudio enfatiza que actualmente ningún instrumento global cubre todos los campos que guían el desarrollo y la aplicación de la IA con un enfoque centrado en el ser humano. La innovación que aporta el trabajo de UNESCO es que no sólo define los principios éticos para el uso de la IA, sino también sugiere recomendaciones políticas para respetar estos principios. En este sentido, va un paso más que la mayoría de las iniciativas similares.

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

INICIATIVAS GLOBALES



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

Se trata de una organización benéfica y de divulgación que trabaja para garantizar que las tecnologías más poderosas del mañana sean beneficiosas para la humanidad. Considera que con tecnologías tan poderosas como la futura inteligencia artificial, planificar con anticipación es una mejor estrategia que aprender de los errores, por lo que apoya la investigación y otros esfuerzos destinados a evitar problemas. En enero de 2017, durante la segunda conferencia del FLI se establecieron los Principios de Asilomar¹¹. Estos principios han sido refrendados por las personalidades más ilustres en el campo de la investigación sobre IA, y se establecieron para guiar el desarrollo continuo y ofrecer oportunidades que ayuden y empoderen a las personas en las décadas y siglos venideros.

- Personas de referencia en el ámbito de IA como son Elon Musk (Tesla), Larry Page (Google), Sam Altman (OpenAI), Sam Teller (SpaceX), Yoshua Bengio, Yann LeCun, Stuart Russell (Berkeley), Sam Harris, Demis Hassabis (DeepMind), Kate Crawford (Microsoft / MIT) y Francesca Rossi (IBM /Padova) participaron en las conferencias y en los grupos de trabajos que establecieron y firmaron los principios de Asilomar.

Notas

¹¹. Principios Asilomar de la IA.



3.3. Referentes internacionales en ética de la IA

Con el objetivo de hacer frente a los desafíos políticos y sociales ligados a la inteligencia artificial, gobiernos de todo el mundo han elaborado estrategias políticas de IA en las que establecen sus objetivos generales y áreas clave de actividad. Entre ellas, destacan iniciativas pioneras como las de Canadá y Singapur en 2017, y las iniciativas españolas entre 2019 (Estrategia de I+D en Inteligencia Artificial¹²) y 2020 (Estrategia Nacional de Inteligencia Artificial de la SEDIA¹³):

De cada uno de estos países se ha obtenido, por una parte, sus objetivos políticos y por otra las áreas en las que están centrando principalmente sus esfuerzos de IA. Este mapeo tiene una doble intención; la de localizar aquellos objetivos y áreas convergentes por la mayoría de los países y la de identificar qué países tienen una mayor similitud con España, al menos en sus inicios, a la hora de elaborar sus estrategias nacionales de IA. Además, ayudará a identificar áreas específicas de interés para este estudio, como pueda ser el caso de Alemania, Reino Unido o Bélgica en lo relativo a las acciones dirigidas al establecimiento de alianzas entre el Estado y la Industria.



Figura 2 - Estrategias nacionales de IA, Holoniq¹⁴

Notas

- 12. Spanish RDI Strategy in AI, Ministerio de Ciencia, Innovación y Universidades (2019).
- 13. ENIA, Secretaría de Estado de Digitalización e Inteligencia Artificial (2020).
- 14. The 2020 AI Strategy Landscape.



Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos



En lo relativo a las estrategias políticas, obtenemos el siguiente mapeo de objetivos:

Objetivos políticos de las estrategias de IA	Alemania	Canadá	Estados Unidos	Francia	Japón	Países Bajos	Reino Unido	Suecia	Suiza	Corea del Sur	China	Singapur	EE.UU.
Liderazgo internacional y Cooperación	✓			✓			✓				✓		✓
Impacto social y ético	✓	✓	✓		✓		✓	✓	✓	✓	✓		✓
Sostenibilidad	✓		✓		✓	✓						✓	
Democratización y diálogo social	✓		✓		✓							✓	✓
Competitividad, innovación e investigación	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓
Sector Público y Administración. Servicios y eficiencia		✓	✓	✓									✓
Potenciación y atracción del talento			✓		✓					✓	✓		
Productividad y crecimiento económico e industrial	✓		✓		✓		✓	✓	✓	✓	✓		✓
Función de control por parte del estado						✓							
Similitud con España	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Figura 3 - Objetivos políticos de las estrategias de IA¹⁵

En él podemos explorar las similitudes en cuanto a objetivos políticos por parte de los diferentes países seleccionados. Por ejemplo, se observan ejes estratégicos comunes como la competitividad, la innovación y la investigación, el impacto social y ético de la IA, y la necesidad de mejora de la productividad y crecimiento económico e industrial.

Respecto a España, podemos concluir que naciones como son Alemania, Canadá, Holanda, Australia, China o Japón manifiestan una alta similitud en cuanto a objetivos estratégicos con nuestra nación. Más adelante se ampliará esta similitud en lo relativo a las áreas de estudio sobre IA y en particular sobre ética de la IA aplicada a la industria por parte de estos países para poder establecer los modelos o referentes para este estudio en particular.



Notas

15. Cornerstones of Political AI Strategies.

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

A continuación, se presenta una visión similar centrada en las áreas de actividad de cada país. Este nuevo mapa permite distinguir:

- Las áreas de mayor intensidad y que por tanto están siendo abordadas mundialmente en las estrategias de IA a nivel nacional. Concretamente, las de regulación, seguridad y legislación, y transferencia de conocimiento, aplicación y emprendimiento.
- Áreas de desarrollo lingüístico para el uso de metamodelos del lenguaje, que son especialmente destacables en España, China y Noruega.
- Área de "Alianzas estado e industria" que podría ser interesante para llevar a cabo un análisis detallado de cada una de estas iniciativas por cuanto puedan aplicar a España y a su industria.



Notas

16. Cornerstones of Political AI Strategies.



Áreas de actividad

	Francia	Reino Unido	Alemania	Italia	Países Bajos	China	India	Corea del Sur	EEUU
Expansión de la investigación	✓		✓	✓	✓			✓	✓
Transformación del tejido productivo y empleo	✓		✓	✓		✓		✓	✓
Regulación, seguridad y legislación	✓	✓	✓	✓	✓	✓	✓	✓	✓
Transferencia de conocimiento, aplicación y emprendimiento	✓	✓	✓	✓	✓	✓	✓	✓	✓
Diálogo internacional y público	✓		✓			✓	✓	✓	✓
Capital humano: Educación y aprendizaje a lo largo de la vida		✓	✓	✓	✓		✓	✓	✓
Desafíos sociales		✓	✓	✓	✓		✓	✓	✓
Corpus lingüísticos ampliados para modelos lingüísticos			✓					✓	
Infraestructura tecnológica y Smart city		✓	✓	✓	✓	✓	✓	✓	✓
Recursos naturales y medioambiente			✓		✓			✓	
Salud, envejecimiento y discapacidad	✓				✓	✓		✓	✓
Defensa					✓			✓	✓
Alianzas estado e industria	✓		✓			✓	✓	✓	✓

Figura 4 - Áreas de actividad por país¹⁶

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

De este análisis sobre objetivos estratégicos y áreas de actividad puede concluirse:

- El impacto social y ético de la IA es una preocupación global y que para alcanzar la mejora en la productividad y el crecimiento económico e industrial deberá ser vital la integración de ambos objetivos en la Ética de la IA para la industria.
- Destacan países con alta similitud desde el punto de vista estratégico con España que podrán servir al país como referentes destacando entre ellos: Alemania, Canadá o Australia.
- El área de actividad sobre alianzas entre el estado y la industria también presenta excelentes candidatos con los que España podrá compartir iniciativas. De nuevo serán candidatos: Alemania, Reino Unido o Japón.

En el siguiente gráfico se refleja la relación positiva (color azul) o negativa (color naranja) entre los artículos y citas en revistas de investigación científica acerca de la ética de la IA y su foco en la industria o campo de la ingeniería. Como se puede observar, mientras que Estados Unidos, Francia y Reino Unido tienen una producción sobre ética de la IA importante, su relación sobre estos estudios en el ámbito de la industria está muy por debajo de países como Alemania, Canadá o Australia.

Notas

17. Términos clave: #Ethics #Artificial Intelligence #Industry. Búsqueda realizada a través de la Biblioteca de la Universidad de Sevilla.

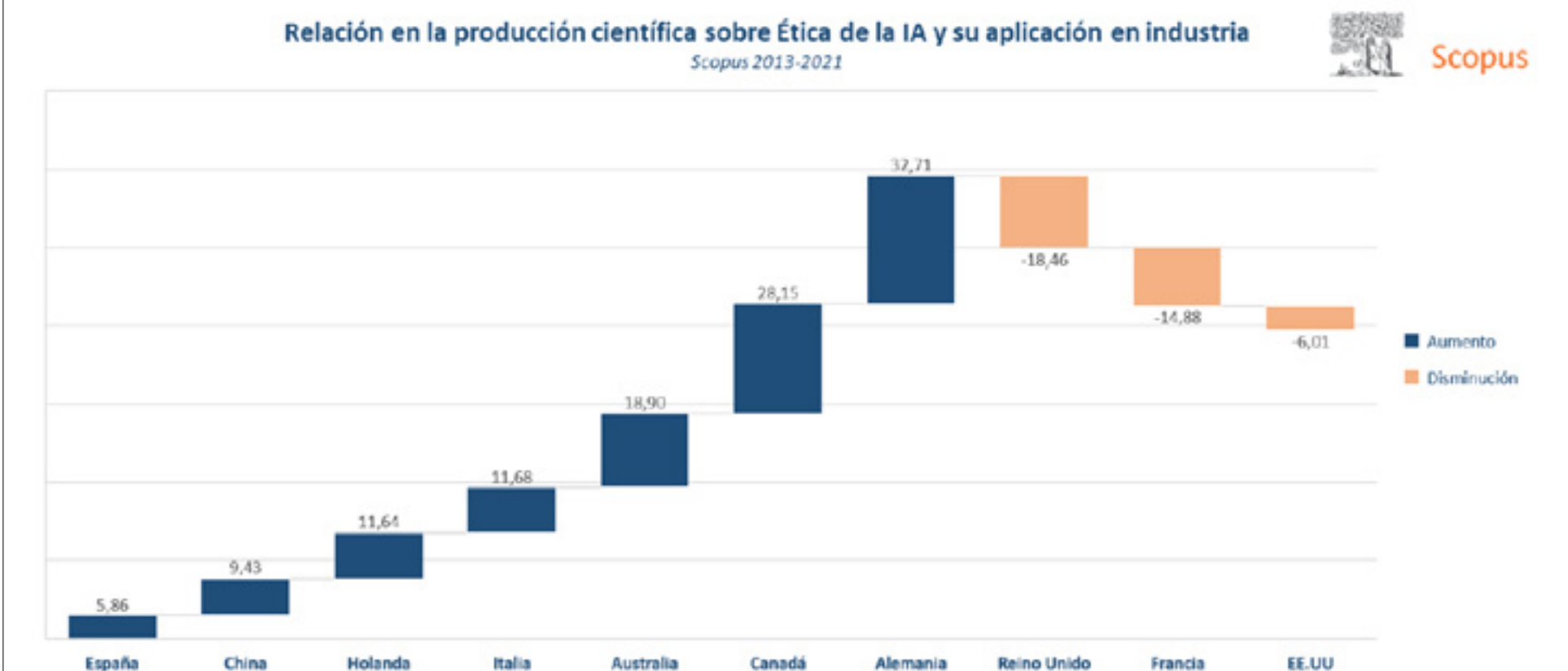


Figura 5 - Investigación en ética en IA y aplicación industrial (2013-21), Scopus¹⁷

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

3.4. Marco de desarrollo GuIA para garantizar una IA ética en la industria

La IA centrada en el ser humano, debe guiar cualquier marco de desarrollo que se construya y así exigir que los sistemas inteligentes no solo sean explicativos, seguros y robustos, sino que además cuenten con los principios europeos como la ausencia de discriminación y la protección de la privacidad. Y esta virtud humanista debe integrar al sector industrial amparando sus acciones y operaciones y apoyando sus decisiones por el bien propio y de la sociedad a la que presta sus servicios y a la que proporciona sus productos.

Tras un estudio exhaustivo de las iniciativas citadas en el apartado 2.2 tanto globales, internacionales como europeas, y en colaboración con PwC España, OdiselA ha desarrollado un esquema conceptual, una estructura previa sobre ética de la IA denominado GuIA¹⁸ que permitirá a cualquier empresa, organización u organismo público

acometer con seguridad sus sistemas inteligentes, minimizando los riesgos en sus servicios y/o productos derivados ante posibles efectos adversos del uso de esta poderosa tecnología. GuIA aspira a ser el framework español de ética de la IA al cual acudir para concretar las perspectivas técnicas, normativas, organizativas y de gobierno de las buenas prácticas en ética de la IA.



Figura 6 - Framework GuIA, OdiselA

Notas

18. Iniciativa GuIA.

Prólogo

Introducción

**Estado de la cuestión
de la ética en IA**

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

En este sentido, las empresas de IndesIA son el entorno adecuado para comenzar a aplicar dicho framework, donde gracias a los antecedentes, teorías consensuadas, regulaciones existentes y alineamiento de principios fundamentales sobre los que se asienta GuIA, se puedan llegar a establecer estándares, reglas, protocolos de actuación y características organizativas comunes que puedan ponerse en práctica y permitan ir definiendo especificaciones propias y verticalizadas a ámbitos industriales tales como la producción, comercialización y desarrollo de infraestructuras energéticas, las telecomunicaciones, la aeronáutica, la automoción, la construcción o la ingeniería naval.

Las compañías que forman IndesIA, presentes y futuras, no partirán de cero, iniciarán este camino reutilizando el conocimiento de OdiselA, organización experta y reconocida, y aportarán al proyecto de la ética

de la IA en la industria en España toda su experiencia para erigirse como referentes en la industria. Reutilizar, minimizar el time to market y ahorrar costes como visión empresarial de la ética de la IA debe ser misión de IndesIA así como gestionar las expectativas del resto de compañías en su acercamiento al mundo de la inteligencia artificial aplicada a la industria.

Como se ha señalado previamente, no se puede perder de vista que fallos no previstos vinculados a la IA tienen un impacto directo, veloz y negativo sobre la reputación empresarial. IndesIA con la colaboración de OdiselA trabaja en las iniciativas para establecer el marco general de la ética de la IA en la industria y asesorarán en la contextualización de este marco para cada línea de actividad, sector industrial o incluso compañía que así lo requiera en búsqueda de la normalización y estandarización de los sistemas inteligentes industriales.



Prólogo

Introducción

**Estado de la cuestión
de la ética en IA**

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

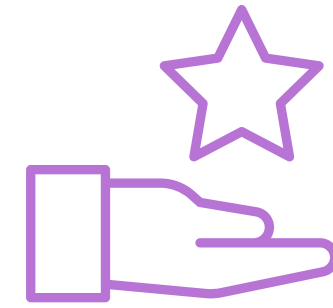
Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



**El impacto social y
ético de la IA es una
preocupación global y que
para alcanzar la mejora
en la productividad y el
crecimiento económico e
industrial deberá ser vital
la integración de ambos.**



Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

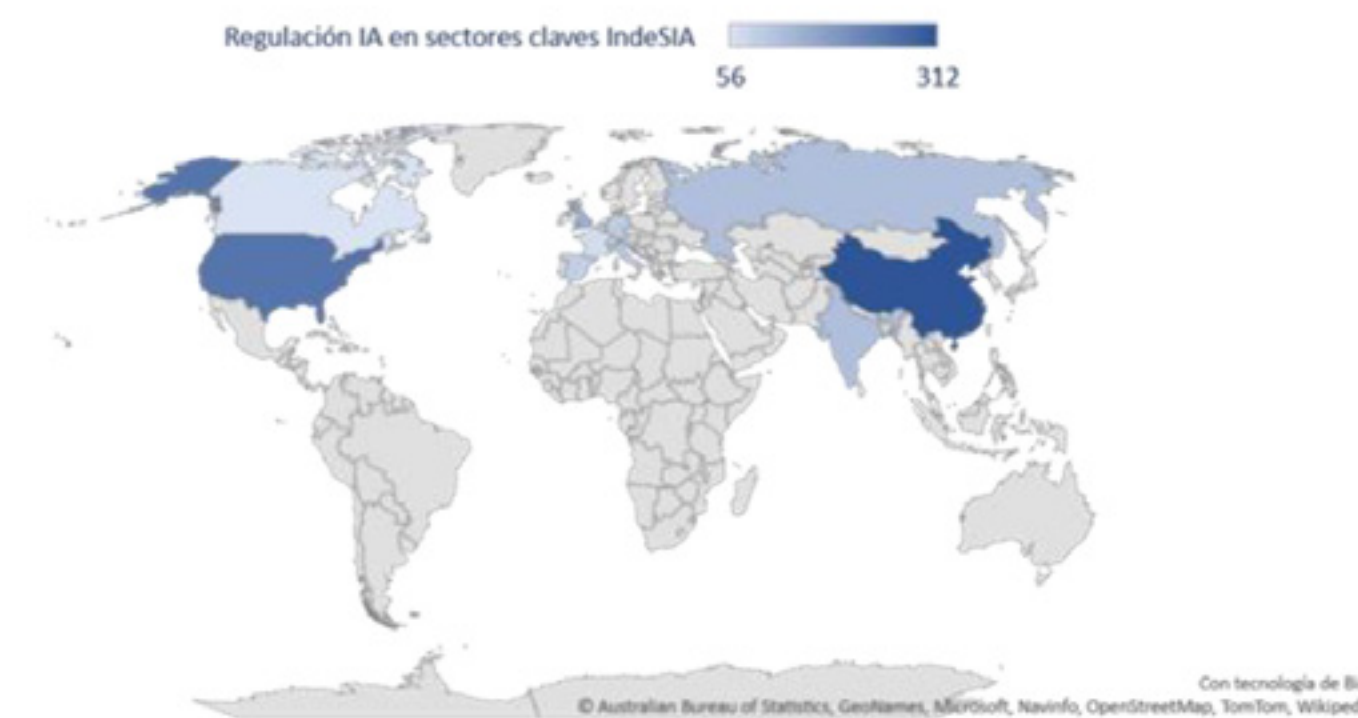
Conclusiones y consejos prácticos

3.5. Regulación, auditoría y certificación

La creciente preocupación por el daño potencial de los sistemas inteligentes y la evidente necesidad de ir transitando de las recomendaciones a la regulación se refleja en la producción de artículos científicos que con mayor intensidad se han producido desde el año 2019 en adelante a escala mundial.

Como puede observarse en la siguiente figura, destaca la producción sobre temas regulatorios de la Unión Europea y en particular, la relativa a aquellos sectores clave de interés para IndesIA.

Producción científica por país sobre regulación IA. 2013-2022



Regulación IA en sectores claves Indesia 2013-2022

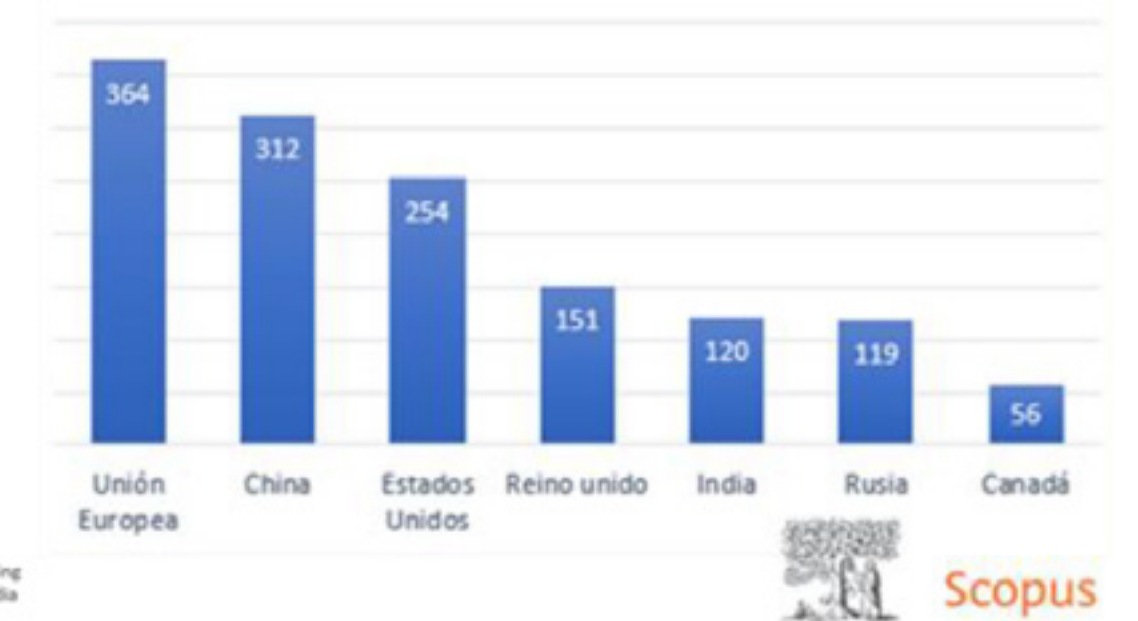


Figura 7 - Producción científica sobre regulación en IA en sectores Indesia, Scopus¹⁹

Los marcos de referencia para la evaluación ética de la IA son un listón más alto si cabe que el de la ley, si bien cabe enfatizar que el marco legal es a su vez móvil y evolutivo, y que depende ante todo de que la percepción ciudadana sobre los riesgos asociados a una innovación tecnológica trascienda del debate público y mediático al político, activándose con ello la acción legislatora.

Notas

19. Términos clave: #Ethics #Artificial Intelligence #Industry #regulation #audit. Filtro por sector industrial: Aeronáutico, Automoción, Construcción, Energético, Ingeniería, Naval y Telecomunicaciones. Búsqueda realizada a través de la Biblioteca de la Univ. de Sevilla.

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

En febrero de 2020 la Comisión Europea publicó el White Paper on AI²⁰ y llevó a cabo una consulta pública al respecto de esta tecnología, cuyas respuestas -más de 1.200 por parte tanto de ciudadanía como de entidades- arrojaron, entre otros los siguientes resultados:

- El 90% de las personas que respondieron creían que la IA puede atender contra los derechos fundamentales.
- El 87% opinaba que la IA puede llevar a resultados discriminatorios.
- El 78% pensaba que la IA puede desencadenar decisiones y acciones cuyos motivos no pueden ser explicados.
- El 70% afirmaba que sus resultados no son siempre acertados.



Notas

20. On AI - A European approach to excellence and trust.

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

3.5. 1 La futura AI Act de la Unión Europea

El 21 de abril de 2021, la Comisión Europea publicó su propuesta de nueva Ley de Inteligencia Artificial. La AI Act se basa en varias iniciativas y publicaciones recientes que colectivamente han presagiado la legislación de la UE sobre IA. Por ejemplo, en las Directrices éticas para una IA confiable de 2019, el Grupo de Expertos de Alto Nivel sobre IA, AI HLEG, estipula que los sistemas de IA deben ser éticos, legales y técnicamente robustos. La Resolución del Parlamento Europeo, de 20 de octubre de 2020 además elabora sobre Régimen de responsabilidad civil en materia de inteligencia artificial²¹.

La AI Act supone el primer intento de establecer un marco jurídico general para el desarrollo y uso de sistemas de IA en la UE, sin perjuicio de la legislación ya existente que ya afecta a muchos sistemas de IA, como especialmente el RGPD. Al igual que el RGPD cuyos efectos se expanden mundialmente, el alcance del AIA no solo afectará a los

sistemas de IA producidos en la UE, sino que se aplicará asimismo en los sistemas usados en la UE de forma independiente a sus proveedores u origen. Este hecho es lo que se ha denominado el “efecto Bruselas” por el cual las multinacionales eligen converger sus prácticas internacionales con las leyes de UE simplemente porque es práctico hacerlo.



Notas

²¹. Resolución del Parlamento Europeo.

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

La característica más distintiva de la legislación de la UE propuesta es su enfoque proporcionado y basado en el riesgo. Proporcionado en el sentido en el que promueve la adopción de los sistemas de IA y aborda, o al menos gestiona, los desafíos asociados con tales tecnologías. Siendo por tanto un buen punto de partida para convertir consideraciones éticas sobre aplicaciones de alto riesgo en normativa. Dicho enfoque basado en riesgos clasifica los sistemas de IA en cuatro tipologías:

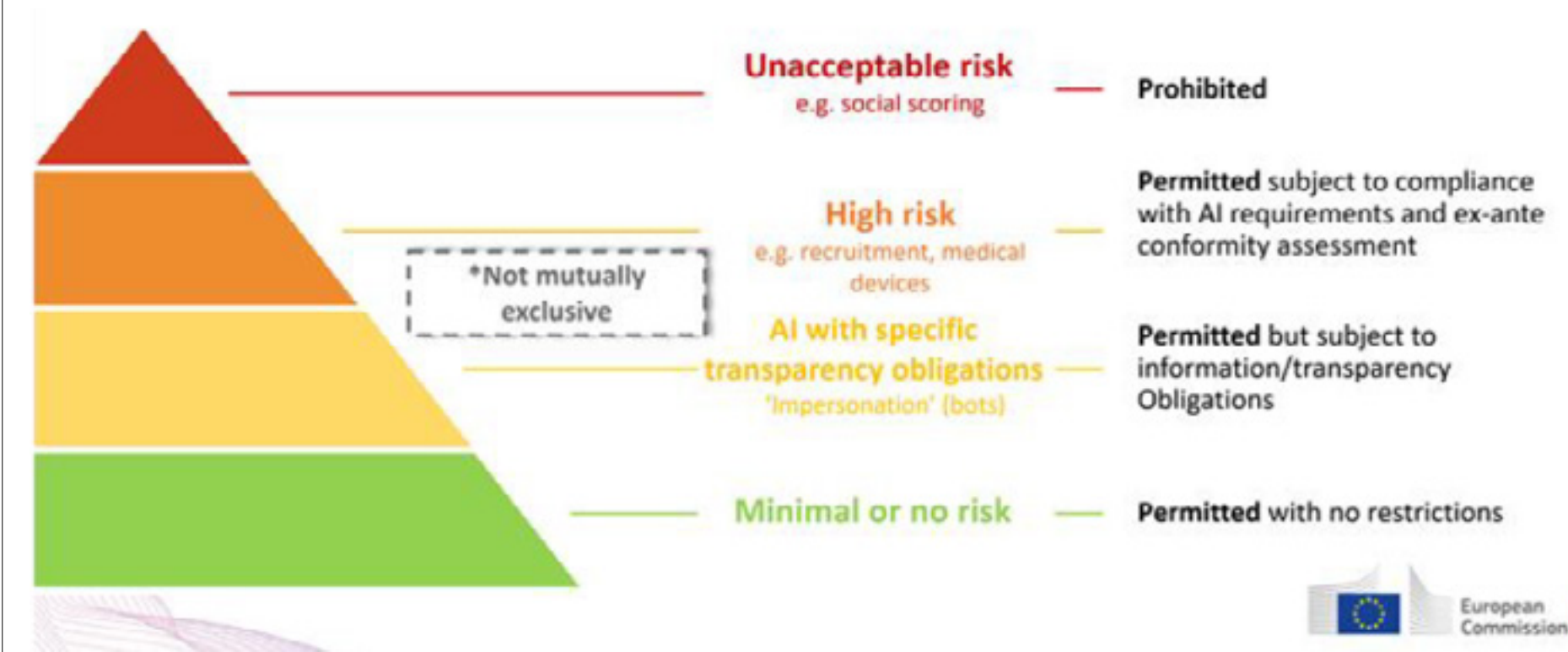


Figura 8 - The risk-based approach to AI governance, European Commission²²

Los sistemas de algo riesgo incluyen sistemas de IA que se aplican en las siguientes áreas:

1. Identificación biométrica y categorización de personas físicas.
2. Gestión y explotación de infraestructuras críticas.
3. Educación y formación profesional.
4. Empleo, gestión de trabajadores y acceso al autoempleo.
5. Acceso y disfrute de los servicios privados esenciales y de los servicios y prestaciones públicas.
6. Aplicación de la ley.
7. Gestión de la migración, el asilo y el control de fronteras.
8. Administración de justicia y procesos democráticos.

Notas

22. Spiceworks.

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

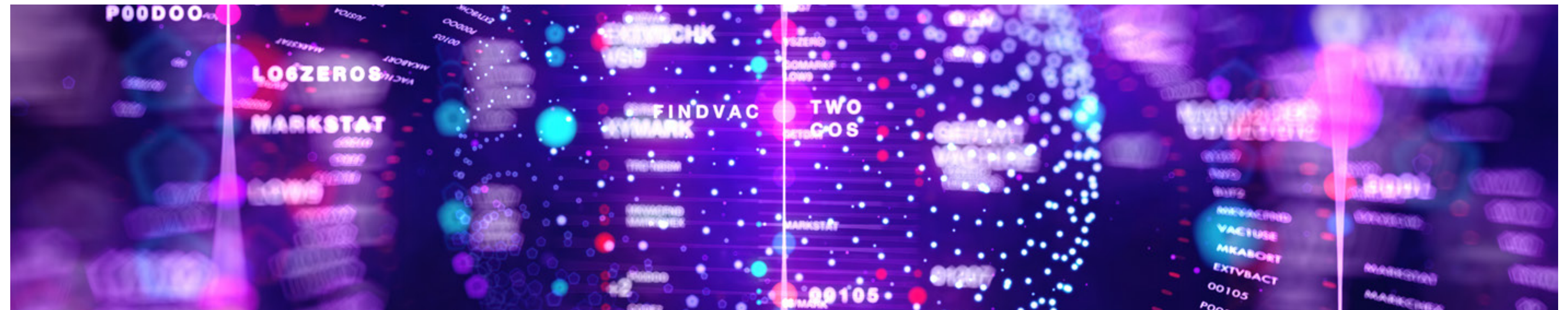
Conclusiones y consejos prácticos

Para los sistemas de IA de alto riesgo, la AI Act introduce varios requisitos de obligatorio cumplimiento, incluyendo:

- Sistema de gestión de riesgos.
- Datos y gobernanza de datos
- Documentación técnica
- Registros
- Transparencia y comunicación de información a los usuarios
- Vigilancia humana
- Precisión, solidez y ciberseguridad

La evaluación de la conformidad que los proveedores deben realizar antes de comercializar estos sistemas (AI Act, artículo 43). Una vez que el citado sistema ha demostrado la conformidad con la AI Act, y ha recibido el llamado mercado CE, puede desplegarse en el mercado interior de la UE y moverse libremente dentro de él (AI Act, artículo 44). Hay tres formas diferentes de llevar a cabo estas evaluaciones y vienen determinadas por la naturaleza del sistema:

- Sistemas incorporados a su vez como componentes en otros productos que están regulados. Estos están sujetos a evaluaciones de conformidad "ex ante" de terceros en virtud de la legislación vigente en materia de seguridad del sector al que corresponda el mencionado producto. La razón de ello es evitar la duplicación de cargas administrativas y mantener funciones y responsabilidades claras, garantizando al mismo tiempo una fuerte coherencia entre los diferentes capítulos de la legislación de la UE. Sin embargo, también implica que no se llevarán a cabo evaluaciones de conformidad específicas de IA.



Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

• Sistemas independientes. Estos incluyen los sistemas de IA utilizados en el reclutamiento, la determinación del acceso a las instituciones educativas o el “profiling” de personas para la aplicación de la ley, entre otros. En este caso, dos opciones para realizar evaluaciones de conformidad ex ante:

1. Realizar evaluaciones de la conformidad ex ante basadas en el control interno.
2. Implicar a una auditoría externa, es decir, un “organismo notificado”, para evaluar su sistema de gestión de la calidad y la documentación técnica (AI Act, artículo 43).



En cualquiera de los casos, estas auditorías sean internas o externas deben cumplir las normas y particularmente cumplir con las siguientes obligaciones:

- Documentar el propósito previsto del sistema de IA en cuestión.
- Proporcionar instrucciones detalladas para el usuario.
- Divulgar los métodos utilizados para desarrollar el sistema.
- Justificar las elecciones de diseño críticas realizadas por el proveedor.

La Comisión Europea a través de la AI Act propone por una parte lo que se denomina una “gobernanza dura” de la IA por cuanto se refiere a una gobernanza sometida a obligaciones legalmente vinculantes. A este tipo de gobernanza aplican los mecanismos mencionados anteriormente sobre los sistemas de alto riesgo: las evaluaciones ex ante y el seguimiento expost. Mientras que, de forma complementaria, propone una “gobernanza blanda” para el resto de los sistemas de IA. En este sentido la AI Act fomenta la adopción voluntaria de códigos de conducta y apoya la aparición de mecanismos complementarios de gobernanza blanda. Se trata de una fórmula inteligente que permite establecer un modelo de gobierno de la IA mientras se desarrolla la legislación en sectores que experimentan, como es el caso de la IA, una innovación tecnológica de ritmo rápido.

En sí misma, la legislación propuesta de la UE constituye una gobernanza estricta. Sin embargo, la AI Act también deja espacio para la gobernanza

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

blanda en general y la auditoría posterior al cumplimiento, basada en la ética, en particular. Esto deja a las organizaciones individuales la libertad de elaborar sus propios principios rectores, adoptar directrices recomendadas por la Junta Europea de Inteligencia Artificial o declarar la adhesión a cualquier otro conjunto de normas relevantes para su industria específica o caso de uso.

3.5. 2 Caso IEEE CertifAIEd

IEEE SA, Institute of Electrical and Electronics Engineers, asociación mundial de ingenieros dedicada a la normalización y el desarrollo en áreas técnicas, viene desarrollando un proceso de certificación de sistemas de IA denominado IEEE CertifAIEd²³. Dicho proceso se basa en el trabajo del Programa de Certificación Ética para Sistemas Autónomos e Inteligentes (ECPAIS) del mismo IEEE.

IEEE CertifAIEd define un procedimiento de evaluación de riesgo y conformidad a seguir para un sistema o servicio de IA específico. La institución que busca la certificación para su producto, servicio o sistema proporciona información detallada a una autoridad de confianza de acuerdo con las pautas de IEEE CertifAIEd, es decir participa de un proceso de auditoría externa. Después de completar con éxito la auditoría, la institución recibe un certificado que indica que el producto cumple con los criterios éticos establecidos por IEEE CertifAIEd.

Notas

23. IEEE The Mark of AI Ethics.

Los criterios IEEE CertifAIEd consisten en un conjunto de especificaciones detalladas para la evaluación, valoración y certificación de las propiedades éticas de los productos, servicios y sistemas de IA. IEEE CertifAIEd define cuatro conjuntos de criterios para la evaluación de las propiedades éticas de los sistemas autónomos e inteligentes: Transparencia, responsabilidad, sesgo algorítmico e intimidad.



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

**La Comisión
Europea a través
de la AI Act
propone por una
parte lo que se
denomina una
“gobernanza
dura” de la IA**

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

Cada uno de estos es de mayor o menor relevancia para un sistema o servicio específico y su dominio de aplicación. En consecuencia, se puede seleccionar cualquier subconjunto o todos ellos para realizar la evaluación de un determinado producto.

A finales de 2021, en la ciudad de Viena se llevó a cabo una experiencia piloto para la certificación de un sistema de IA del grupo empresarial Wiener Stadtwerke. Este grupo público empresarial brinda servicios en las áreas de transporte público, electricidad, gas natural, calefacción, telecomunicaciones, estacionamiento, servicios funerarios y cementerios a más de dos millones de clientes en la región metropolitana de Viena. Con más de 15.000 empleados y una facturación anual de más de tres mil millones de euros, el grupo Wiener Stadtwerke pertenece a las 25 empresas más grandes de Austria.

En los últimos años, el grupo Wiener Stadtwerke ha explorado varias ideas para usar la tecnología de IA en proyectos piloto, siempre adhiriéndose al objetivo general de brindar servicios de alta calidad de manera eficiente a los ciudadanos de Viena. Uno de estos fue seleccionado para una evaluación ética exhaustiva en el piloto IEEE CertifAIED con IEEE. Este es un sistema de clasificación de correo electrónico (ECS), que se utiliza para asignar automáticamente categorías a las solicitudes entrantes del servicio de atención al cliente.

El primer paso en el proceso de evaluación fue explicar detalladamente el sistema y su contexto a un panel de cinco expertos de IEEE, incluidos los antecedentes y los objetivos del proyecto, la arquitectura y las interfaces del sistema, el componente de aprendizaje automático y

los datos utilizados para el modelo. formación, así como los efectos del nuevo sistema en las personas y procesos de la organización.

En base a esta información, se realizó una evaluación de riesgos según el IEEE CertifAIED. Para cada uno de los 26 valores éticos, el panel de expertos calificó la probabilidad de que el ECS socave ese valor ético, considerando escenarios potenciales concretos en el despliegue del sistema en el contexto de Wiener Stadtwerke. Los resultados de este análisis de riesgo se utilizaron para determinar el más relevante de los cuatro conjuntos de criterios IEEE CertifAIED para la aplicación: responsabilidad en el caso de ECS.



Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

A continuación, IEEE proporcionó una lista de 43 criterios éticos con breves definiciones a Wiener Stadtwerke, quien luego debía proporcionar evidencia para cada criterio, mostrando que la cuestión o cuestión ética respectiva se aborda adecuadamente en el sistema y su contexto. Estos criterios van desde aspectos bastante técnicos, como el análisis de errores, el ajuste de hiper parámetros y la mitigación de falsos positivos, hasta aspectos más relacionados con la gobernanza relacionados con la organización, como la adopción de un enfoque en capas, evitar la inacción, la demora y la indiferencia, y la autoridad humana y autonomía.

Para cada uno de los 43 criterios, Wiener Stadtwerke proporcionó evidencia en forma de documentación técnica, arquitectura del sistema y detalles de implementación de software, capturas de pantalla, diapositivas y actas de reuniones, informes internos y públicos, documentos de estrategia, definiciones de procesos y funciones, organigramas, etc., dando detalles completos para el criterio respectivo. Se compiló un documento llamado Caso de ética, donde Wiener Stadtwerke afirma que su sistema y su uso no presentan problemas éticos, brindando información general sobre el sistema, sus antecedentes, alcance, etc., así como todas las evidencias para los 43 criterios de rendición de cuentas. Este caso de ética se envió luego a IEEE para su evaluación.

Finalmente, IEEE entregó un informe de evaluación a Wiener Stadtwerke. Esto incluyó comentarios específicos para cada uno de los 43 criterios de los miembros del panel de expertos, indicando en qué medida se consideró cumplido el criterio respectivo y qué se podría hacer para mejorar aún más en el área respectiva. También

incluyó una confirmación general de que el caso de ética presentado justifica el reconocimiento y la certificación a través del programa IEEE CertifAIEd para el sistema de clasificación de correo electrónico de Wiener Stadtwerke.



Prólogo

Introducción

**Estado de la cuestión
de la ética en IA**

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

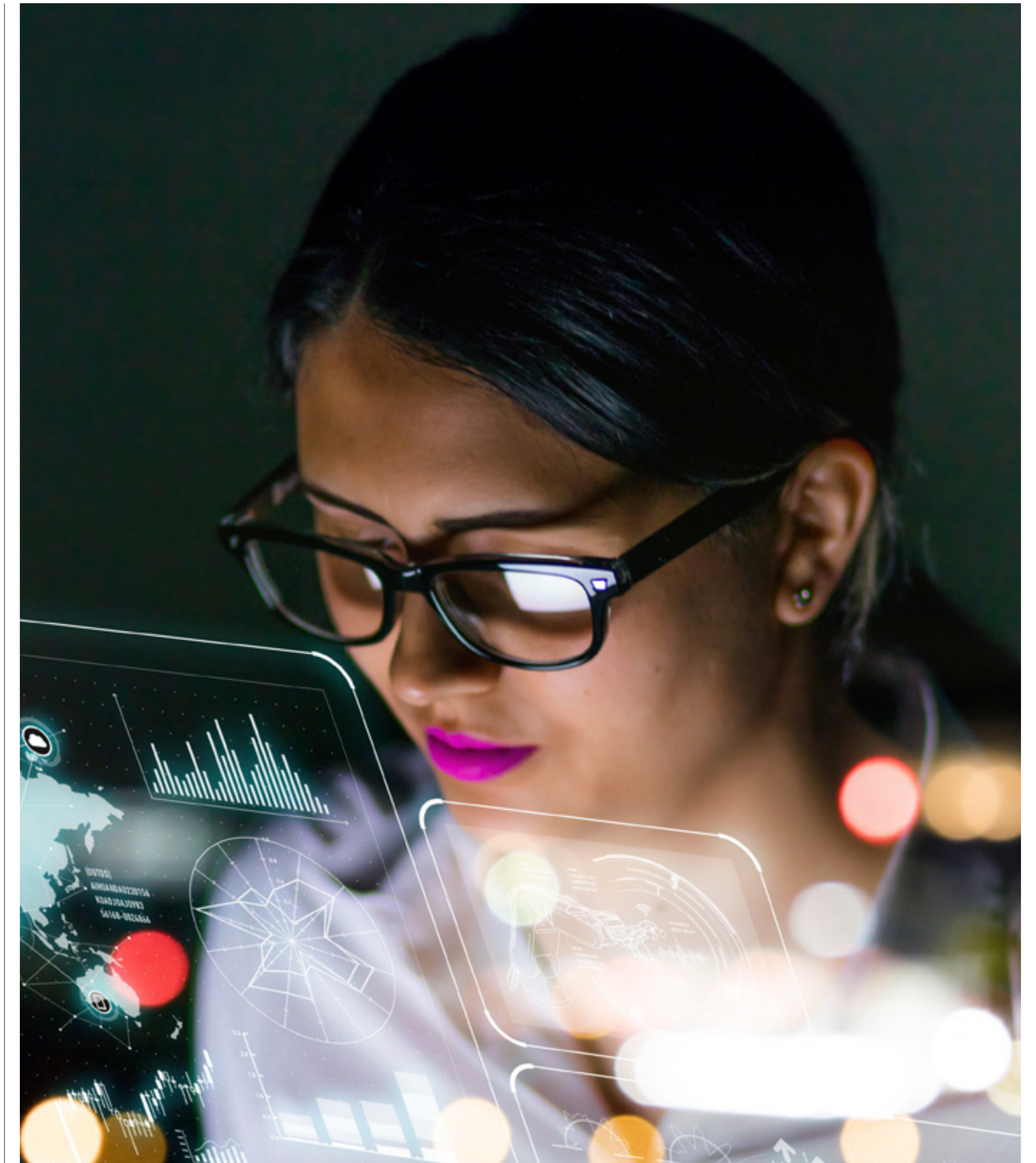
Conclusiones
y consejos prácticos

3.5.3 Norma ISO/IEC 42001

La ISO/IEC 42001, *Information Technology — AI — Management system*²⁴ especifica los requisitos y proporciona orientación para establecer, implementar, mantener y mejorar continuamente un sistema de gestión de inteligencia artificial en el contexto de una organización. Está destinada a ser utilizada por una organización que proporcione o utilice productos o servicios que empleen inteligencia artificial y ayuda a la organización a desarrollar o utilizar la inteligencia artificial de forma responsable en la consecución de sus objetivos, así como cumplir los requisitos reglamentarios aplicables, las obligaciones relacionadas con las partes interesadas y las expectativas de éstas.

Notas

24. ISO/IEC 42001.



Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

3.6. Modelo de madurez de implementación ética de IA

Hay distintos niveles de madurez en las organizaciones cuando hablamos del uso ético de la IA, en función del grado de implementación en la empresa: aquellos sectores clave de interés para IndesIA.



No ética

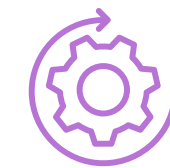
En el nivel más bajo, la organización no considera aspectos éticos de los sistemas IA en las decisiones relevantes. Solo se consideran aspectos de negocio. La mayoría de las empresas en España se encuentran en este nivel.



Principios

En este segundo nivel, la organización ha definido y publicado los principios éticos que quiere respetar en las decisiones sobre el uso de la IA para su negocio. Pero, no ha definido cómo implementarlo en su organización, es decir, lo deja al criterio de cada uno para respetar estos principios y no tiene

manera de verificar si lo está haciendo bien o no. Muchas de las empresas grandes (IBEX) y las internacionales se encuentran en este nivel.



Herramientas

En este nivel, aparte de los principios éticos, existen herramientas que ayudan a los empleados a tratar los aspectos éticos del uso de la IA. Puede existir, por ejemplo, un cuestionario con preguntas éticas que se tiene que completar para cada producto o proyecto que usa IA. O herramientas para detectar y/o mitigar sesgos en los datos o los algoritmos. Pero aún falta una coordinación completa en la organización que asegure el cumplimiento ético.



Gobierno

El nivel más maduro implica que la ética de la IA está integrada en toda la organización con claras responsabilidades, roles y procesos. Requiere una conversación transversal de la organización

para acordar cómo se relacionan los distintos departamentos implicados. Existe un modelo de gobernanza con un modelo operativo que deja claro en cada momento qué hay que hacer, quién es el responsable, y cómo actuar en caso de problemas (es decir, cómo y cuándo escalar). En este nivel, las organizaciones tienen roles específicos y comités éticos.

El capítulo 4 del documento se va a enfocar en una propuesta de marco ético para el sector industrial español, que tiene en cuenta estos niveles.

En este sentido, IndesIA ha preparado otro informe orientado a la adopción de IA en empresa en el que se pueden explorar los niveles de madurez. Un resumen de dicho informe está disponible en <https://www.indesia.org/framework-indesia>.

04

ANÁLISIS DE IMPACTO ÉTICO Y SOCIAL EN LA INDUSTRIA ESPAÑOLA

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

**Análisis de impacto
ético y social en la
industria española**

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

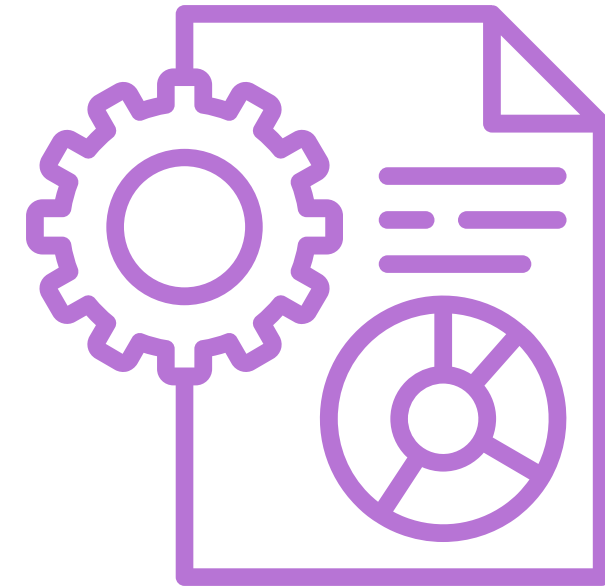
**Análisis de impacto
ético y social en la
industria española**

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



Como parte de la evaluación de impacto ético y social de IndesIA, se han analizado diferentes iniciativas y casos de uso de IA, con el objetivo de crear una versión consolidada de los principios éticos aplicables a la IA en el sector industrial.

La propuesta de dichos principios tiene en cuenta las cuestiones presentadas en el capítulo 2, su adaptación en función del análisis de los casos de las empresas de IndesIA (3.2), así como la valoración de la importancia de cada principio en el sector industrial.

4.1. Propuesta de principios de IA adoptados por IndesIA

Se ha agregado una lista de principios en base al análisis de 27 iniciativas en todo el mundo como parte del marco GuIA de OdiselA, con un análisis de la relevancia de cada principio para el sector industrial. El grupo de iniciativas analizadas incluye ejemplos de gobierno, academia, empresas, etc.

Nota: “Core Ética IA” significa que el principio en cuestión es importante para cualquier proyecto de IA, independientemente de la industria. El resto, tienen un foco importante para el sector industrial que nos ocupa y son los más aplicables a los casos de uso de IndesIA, analizados en el capítulo 3.2.



Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

Principio	Explicación	Foco industrial
Privacidad y Gobernanza	Salvaguarda de la privacidad del individuo . Incluye informar cómo se usarán los datos, para qué objetivos, recolección de consentimiento, derecho a la rectificación y eliminación. Se diferencia entre Privacidad y Privacidad by design.	Core Ética IA (relevante para cualquier sector)
Rendición de cuentas	Responsabilidad y rendición de cuentas se refieren a la asignación de responsabilidades y la posibilidad de auditarlas para los problemas que pueden surgir con el desarrollo y uso de la IA. Incluyen tanto responsabilidad retrospectiva como medidas prospectivas, y formas específicas de lidiar con errores técnicos y organizacionales.	Core Ética IA
Seguridad y Robustez	Prevención de accidentes y operaciones no esperadas . Incluye robustez en la exactitud y reproducibilidad de los resultados del sistema. Incluye evaluación detallada para conocer las principales amenazas y resiliencia a estas.	Foco en safety para aplicaciones industriales
Transparencia y Explicabilidad	Explicabilidad e interpretabilidad de los algoritmos , incluye el modelo y los datos usados.	Foco en aplicaciones industriales de cara a clientes finales

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

Principio	Explicación	Foco industrial
Justicia e Inclusión	Justicia en los aspectos de ecuanimidad algorítmica e inclusión. Incluye prevención y evaluación de sesgos. Se enfatiza el incluir en los procesos consideraciones éticas amplias. Por ejemplo: inclusividad, igualdad, diversidad.	Core Ética IA
Supervisión Humana	En la toma de decisiones autónomas con impacto en las personas , un humano siempre debe formar parte, de una u otra forma, de la decisión.	Foco en definir una supervisión adecuado según el riesgo e impacto del sistema
Derechos Humanos y Futuro del Trabajo	Incluye promoción de los derechos humanos , foco en el beneficio hacia la sociedad , valores humanos para la prosperidad humana , y acceso a la tecnología	Foco en el impacto de la automatización en el trabajo y la posible vigilancia de trabajadores
Sostenibilidad e Impacto Ambiental	Se refiere a la justicia intergeneracional y describe la obligación hacia futuras generaciones para asegurar y preservar sus condiciones de vida. Incluye la creación de infraestructuras que economicen recursos energéticos, la eliminación y reparación de hardware usado para la IA y los efectos positivos que los sistemas de IA tengan en el medioambiente.	Foco en impacto medioambiental de las aplicaciones industriales

PRINCIPIOS ÉTICOS

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

**Análisis de impacto
ético y social en la
industria española**

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

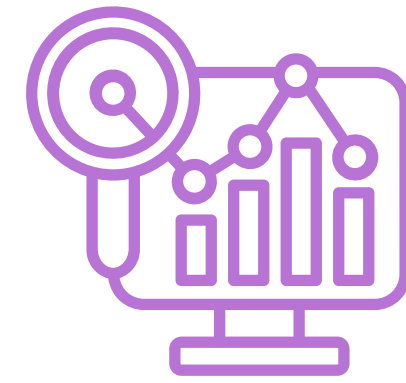
De manera ilustrativa, se añaden los principios éticos adoptados por algunas de las empresas fundadoras de IndesIA:



- Fairness
- Reliability & Safety
- Privacy & Security
- Inclusiveness
- Transparency
- Accountability



- Fair AI
- Transparent and Explainable AI
- Human-centric AI
- Privacy and Security by Design
- Working with partners and third parties



4.2. Análisis de los casos de uso de miembros IndesIA

Con el fin de identificar los principios de IA ética más relevantes en el sector industrial, se analizaron una serie de casos de uso atendiendo al posible impacto de estos sobre los principios mencionados anteriormente. Los casos de usos fueron clasificados en función de su impacto sobre máquinas y/o personas; y se identificó su nivel de riesgo en virtud de la propuesta de regulación (AI Act) de la Comisión Europea (CE). La diferencia entre afectar a “máquinas” o “personas” es relevante porque la ética es relevante para el impacto en personas, no en máquinas.



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

**Análisis de impacto
ético y social en la
industria española**

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

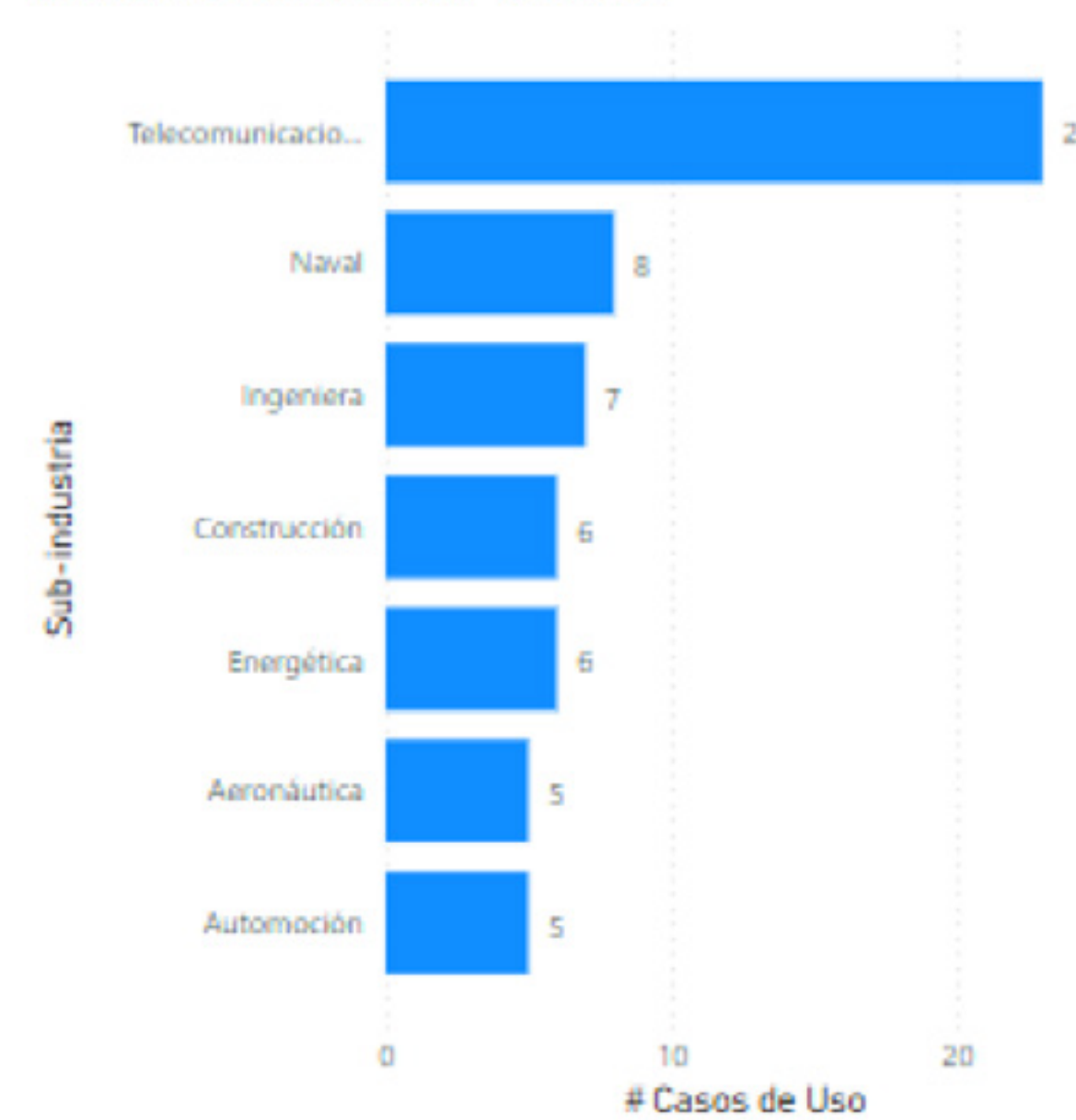
Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

Así, se analizó un total de 60 casos de uso, abarcando las sub-industria de telecomunicaciones, naval, aeronáutica, automoción, construcción, energética e ingeniería. Como se aprecia en la siguiente imagen, el mayor subgrupo de casos analizados corresponde con el sector de telecomunicaciones (esto es debido a que había más casos de usos disponibles para su análisis; no necesariamente por la existencia de más casos de uso en dicho sector).

Respecto a la orientación de los casos, el número de estos que se refieren a máquinas/sistemas es ligeramente inferior al de aquellos que tienen un impacto sobre las personas, aunque si quitamos los casos de uso de Telefónica del análisis, el número de casos de uso que afectan a máquinas es significativamente mayor. Además, existe un número reducido de casos que están orientados tanto a máquinas como a individuos.

Casos de Uso por Sub-industria



Tipos de Casos IA por Orientación

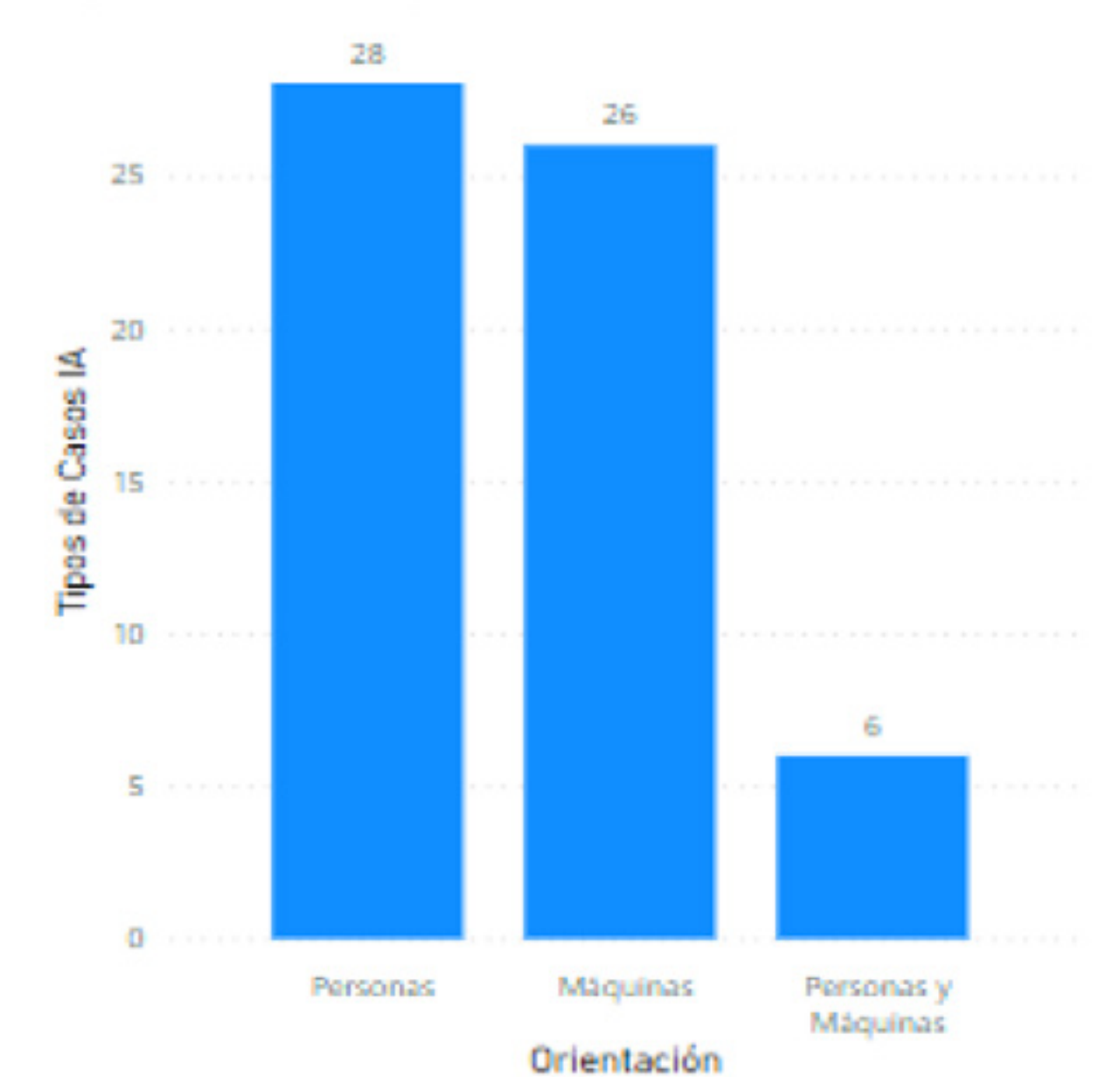


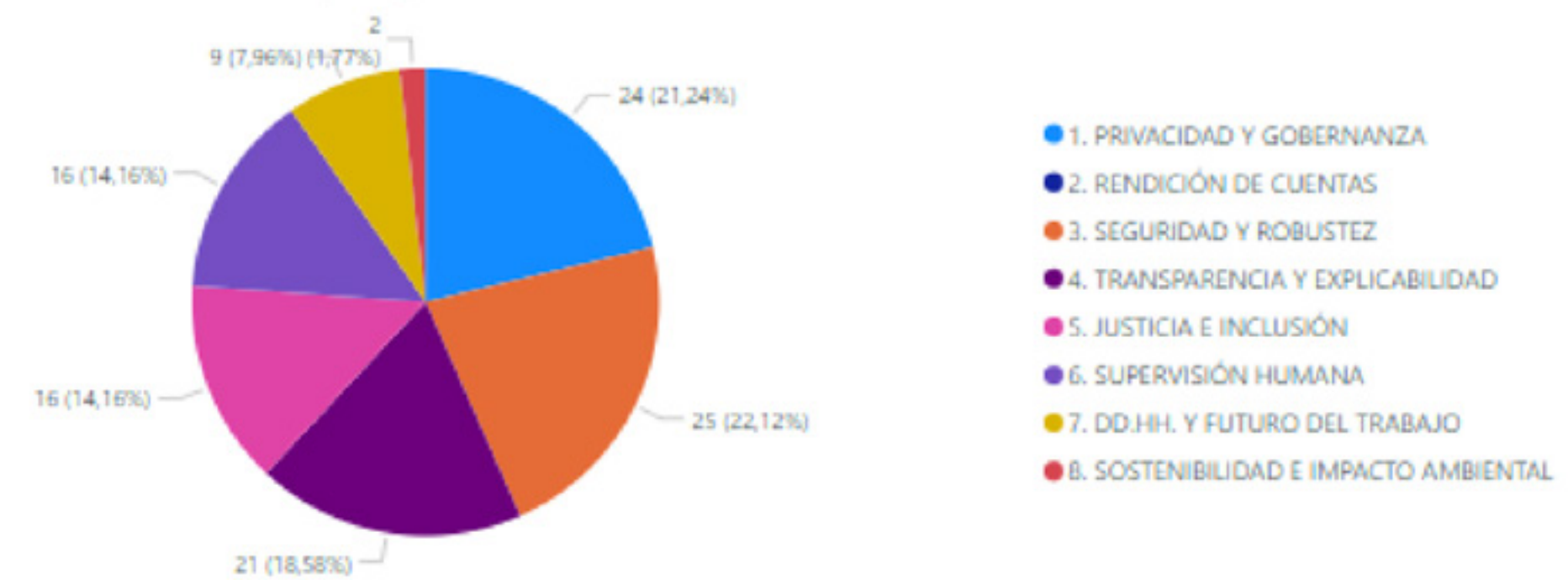
Figura 9 - Tipos de caso de uso IA por sub-industria y orientación



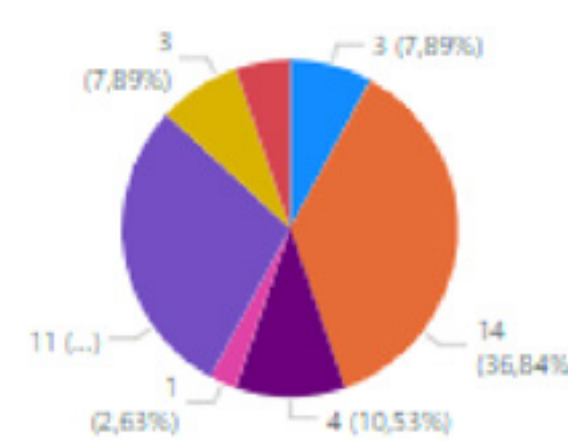
El enfoque máquina/persona también se tuvo en cuenta a la hora de analizar el impacto de los casos sobre los ocho principios de IA ética considerados. Como se aprecia en la imagen, atendiendo al total general de casos, destacan los principios de seguridad y robustez (25); privacidad y gobernanza (24); y transparencia y explicabilidad (21); seguidos de justicia e inclusión (16) y supervisión humana (16). También, llama la atención que ningún caso fue identificado como posible riesgo para el principio de rendición de cuentas. Esto se debe a que el principio de rendición de cuentas es algo que depende más del modelo de gobernanza de la organización que de un caso concreto y no se consideró este principio en el análisis.



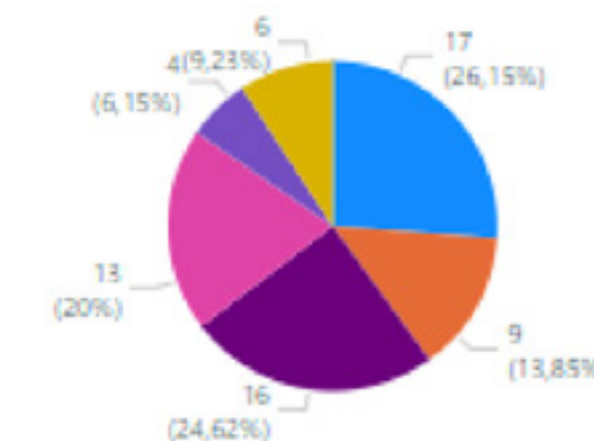
Impacto general de los casos sobre principios éticos



Foco máquinas/procesos



Foco personas



Foco personas y máquinas

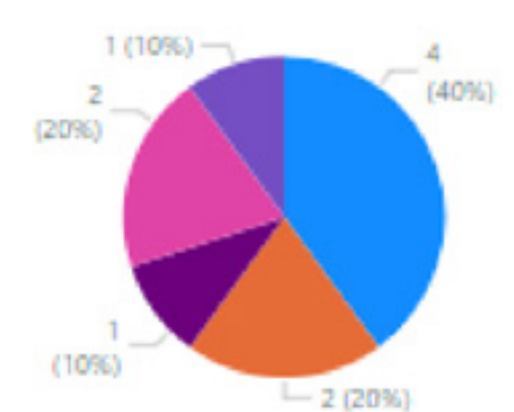


Figura 10 - Impacto general de los casos de uso en base a principios éticos

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

**Análisis de impacto
ético y social en la
industria española**

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



La preponderancia de los principios se entiende mejor si se observan los diagramas individuales (según orientación). Así, en el caso de las aplicaciones referidas a máquinas, destacan los principios de seguridad, robustez y de supervisión humana. La importancia del primero se entiende por las consecuencias de un fallo o accidente en el contexto industrial, que pueden suponer graves consecuencias; y respecto al segundo, su relevancia viene motivada por la posible autonomía de estos sistemas o, mejor dicho, por la potencial disminución de autonomía de la persona o su sobredependencia frente al sistema. En los casos de uso que afectan a componentes críticos (de alto riesgo), es importante mantener una cierta intervención humana, es decir, limitar la autonomía del sistema.

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

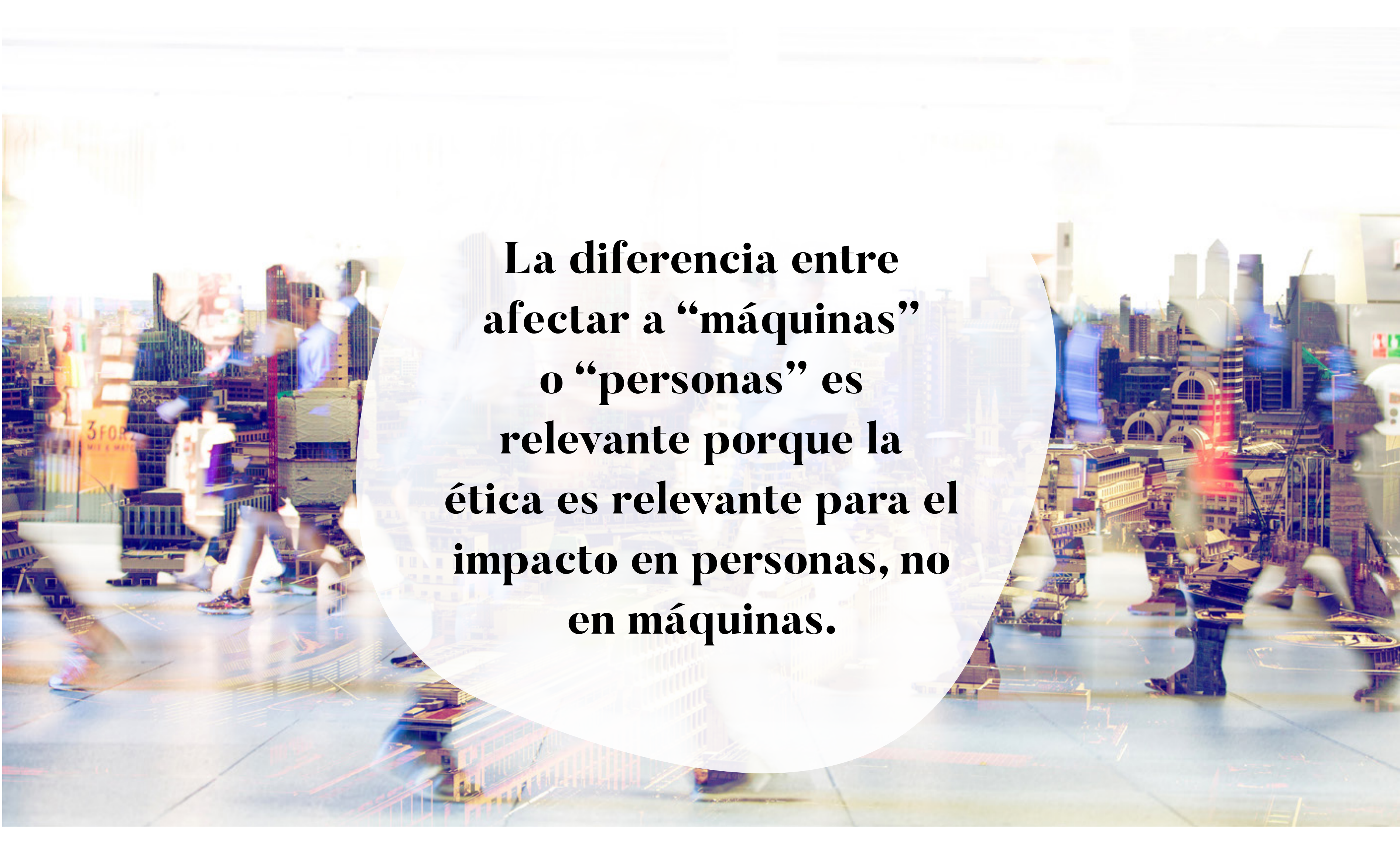
**Análisis de impacto
ético y social en la
industria española**

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



**La diferencia entre
afectar a “máquinas”
o “personas” es
relevante porque la
ética es relevante para el
impacto en personas, no
en máquinas.**

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

Por el contrario, si se consideran los casos orientados a personas, predominan los siguientes principios: privacidad y gobernanza, debido a la habitual cesión de datos de individuos para nutrir los sistemas; transparencia y explicabilidad, por la necesidad de ser transparentes frente a los usuarios sobre qué datos usa el sistema y para qué; y justicia e inclusión, ya que es probable que el sistema refleje los sesgos presentes en los datos (i.e., prejuicios, falta de representatividad, discriminación, etc.) si estos no han sido debidamente tratados. Respecto al grupo de casos que impactan sobre máquinas y personas, se aprecia una combinación de los dos subgrupos de principios anteriores, ya que destacan los de privacidad y gobernanza; seguridad y robustez; y justicia e inclusión.

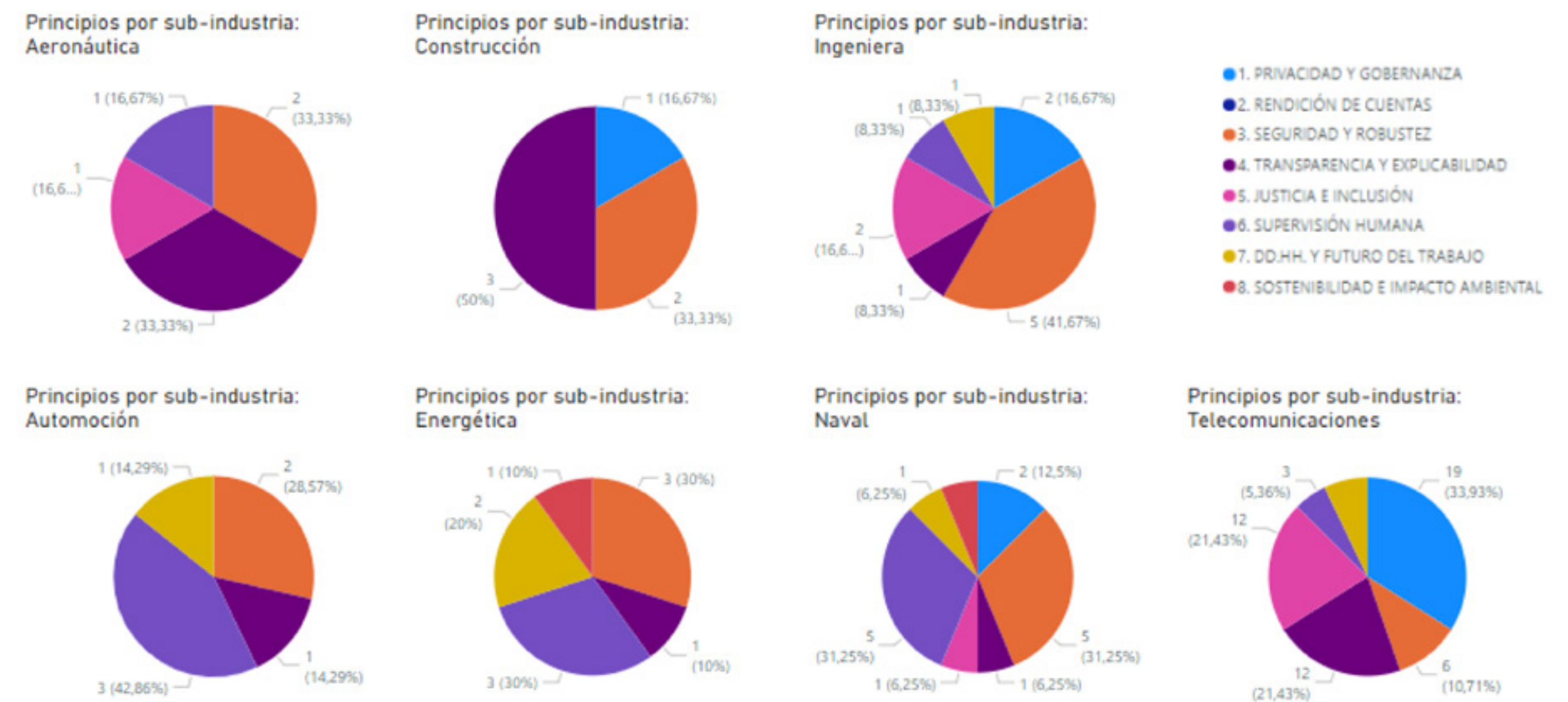


Figura 11 - Importancia de los principios éticos por sub-industria

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

**Análisis de impacto
ético y social en la
industria española**

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



Respecto al impacto de los casos por industria, destaca el principio de seguridad y robustez en la mayoría de los sectores, debido a que la mayoría de los casos de uso afectan a máquinas. No obstante, si se atiende al sector de las telecomunicaciones (donde hemos analizado más casos de uso), este principio queda relegado por los de privacidad y gobernanza; justicia e inclusión; y transparencia y explicabilidad por tener más casos de uso que afectan a las personas. Otro hecho llamativo es la similitud entre los principios impactados en el sector de la automoción y el energético, donde destacan el principio de supervisión humana y el de DD.HH. y futuro del trabajo (además de seguridad y robustez).

Además de esta evaluación sobre los principios más relevantes por sector industrial, se identificaron también los sectores que mayor impacto tienen sobre cada principio de IA ética. El impacto de los sectores sobre los principios se evaluó en base al recuento de casos de uso de cada sector que presentan un impacto sobre el principio en cuestión.

- Prólogo
- Introducción
- Estado de la cuestión de la ética en IA
- Análisis de impacto ético y social en la industria española**
- Toolkit de implementación de una IA ética en el sector industrial
- Caso de éxito de gobernanza de IA en empresa
- Caso de uso aplicado para la implementación de IA ética en la industria
- Conclusiones y consejos prácticos

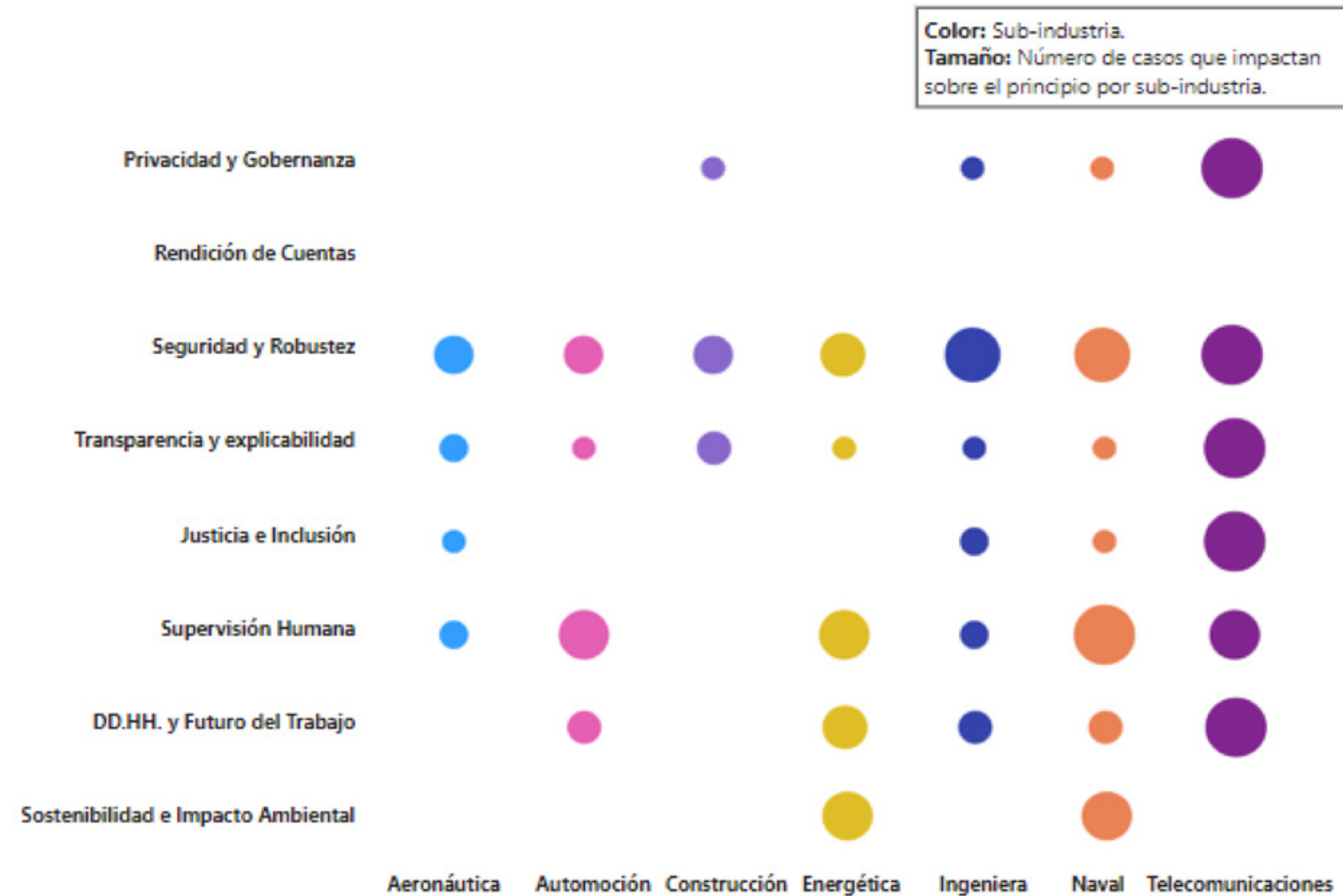


Figura 12 - Importancia de los principios éticos por sub-industria (visual)

Los resultados se pueden observar en la siguiente visualización. No resulta extraño que el sector de telecomunicaciones sea el más relevante en la mayoría de los principios, debido a que abarca un mayor número de casos. Sin embargo, esta sub-industria no presenta el mayor impacto sobre los principios de supervisión humana, que corresponde al sector naval. También destaca el sector energético, con un impacto intermedio sobre los principios de seguridad y robustez; supervisión humana; y DD.HH. y futuro del trabajo. El hecho que “rendición de cuentas” no aparece es, como comentamos antes, debido a que no disponemos de información al respecto en el caso de uso ya que aplica más a la organización y no a casos de uso individuales.



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

**Análisis de impacto
ético y social en la
industria española**

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



4.3. Impacto de la regulación “AI Act” en el sector industrial

Con vistas a la futura regulación de IA en Europa (AI Act), para tener una primera idea sobre qué casos de uso del sector industrial podrían ser clasificados como “alto riesgo”, adicionalmente, los casos de uso fueron clasificados atendiendo a las categorías de riesgo (inaceptable, alto, limitado, bajo) propuestas por la CE.

Como hemos vistos antes, los casos de uso que pertenecerán a la categoría de alto riesgo son los componentes de seguridad de productos ya regulados sectorialmente, y los sistemas de IA en los siguientes ámbitos: identificación biométrica y categorización de personas físicas, gestión y explotación de infraestructuras críticas, educación y formación profesional, empleo, gestión de trabajadores y acceso al autoempleo, acceso y disfrute de los servicios privados esenciales y de los servicios y prestaciones públicas, aplicación de la ley, gestión de la migración, el asilo y el control de fronteras, administración de justicia y procesos democráticos.



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

**Análisis de impacto
ético y social en la
industria española**

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



Como es de esperar, ningún caso pertenece al grupo de riesgo inaceptable. De hecho, la mayoría de los casos de uso corresponde al grupo de riesgo bajo, como se aprecia en la siguiente gráfica. Tan solo el número de casos de alto riesgo es superior al resto en el sector naval; y está empatado con el grupo de bajo riesgo en la sub-industria aeronáutica y en el grupo de casos orientados a máquinas. De hecho, en este último caso abarca el mayor subgrupo de casos de alto riesgo (13), lo que se debe a que los sistemas de IA en máquinas a menudo actúan como o influyen sobre componentes de seguridad (que según la CE, pertenecen a la categoría de alto riesgo). Respecto a los casos de riesgo limitados, el mayor número de estos (6) se encuentra en el grupo de casos orientados a personas; lo que se explica ya que los chatbots y asistentes virtuales, que son los clasificados como riesgo limitado principalmente, son sistemas orientados a personas.

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

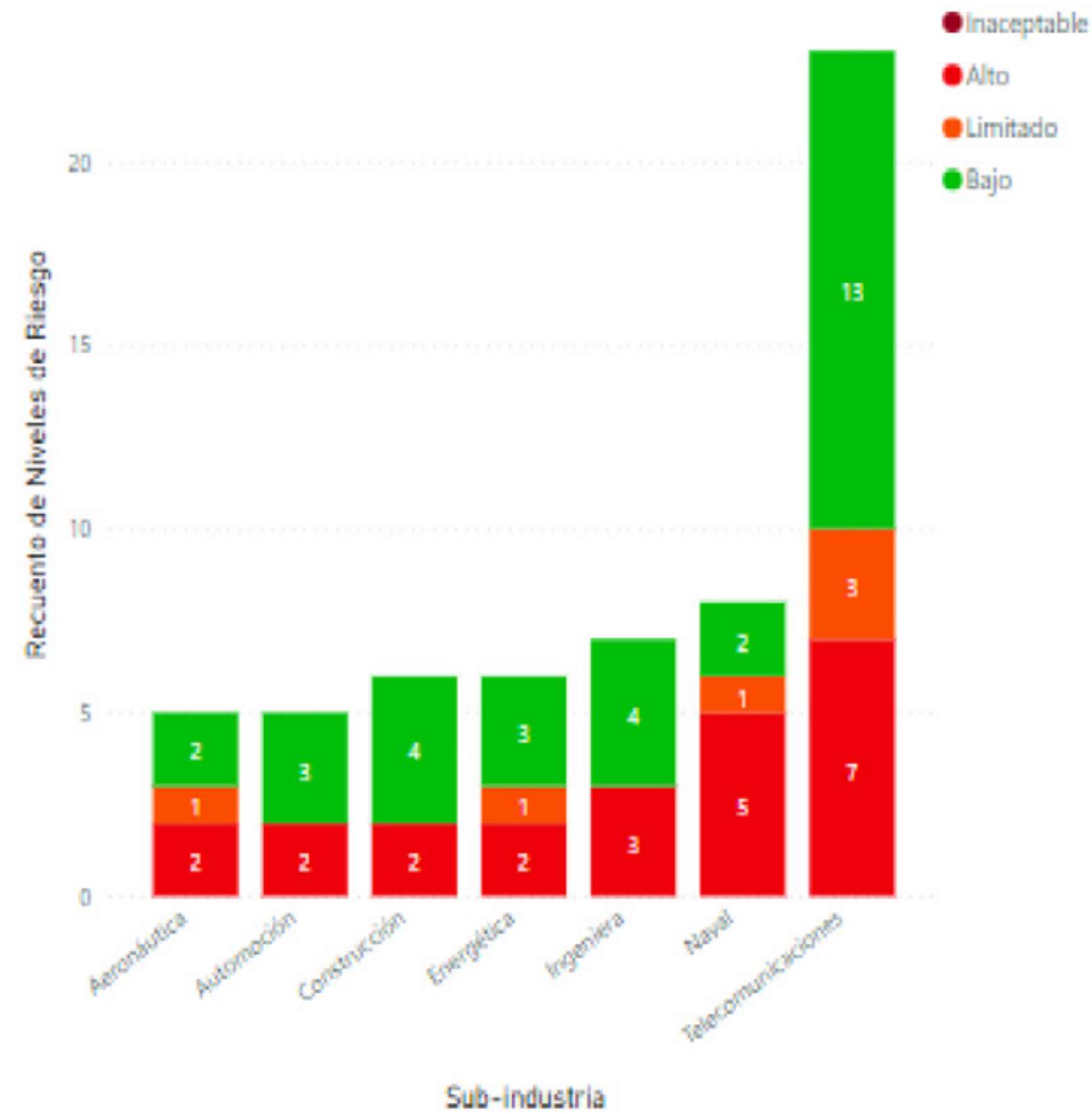
Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

Media Niveles de Riesgo IA por Sub-industria



Niveles de Riesgo según AI Act

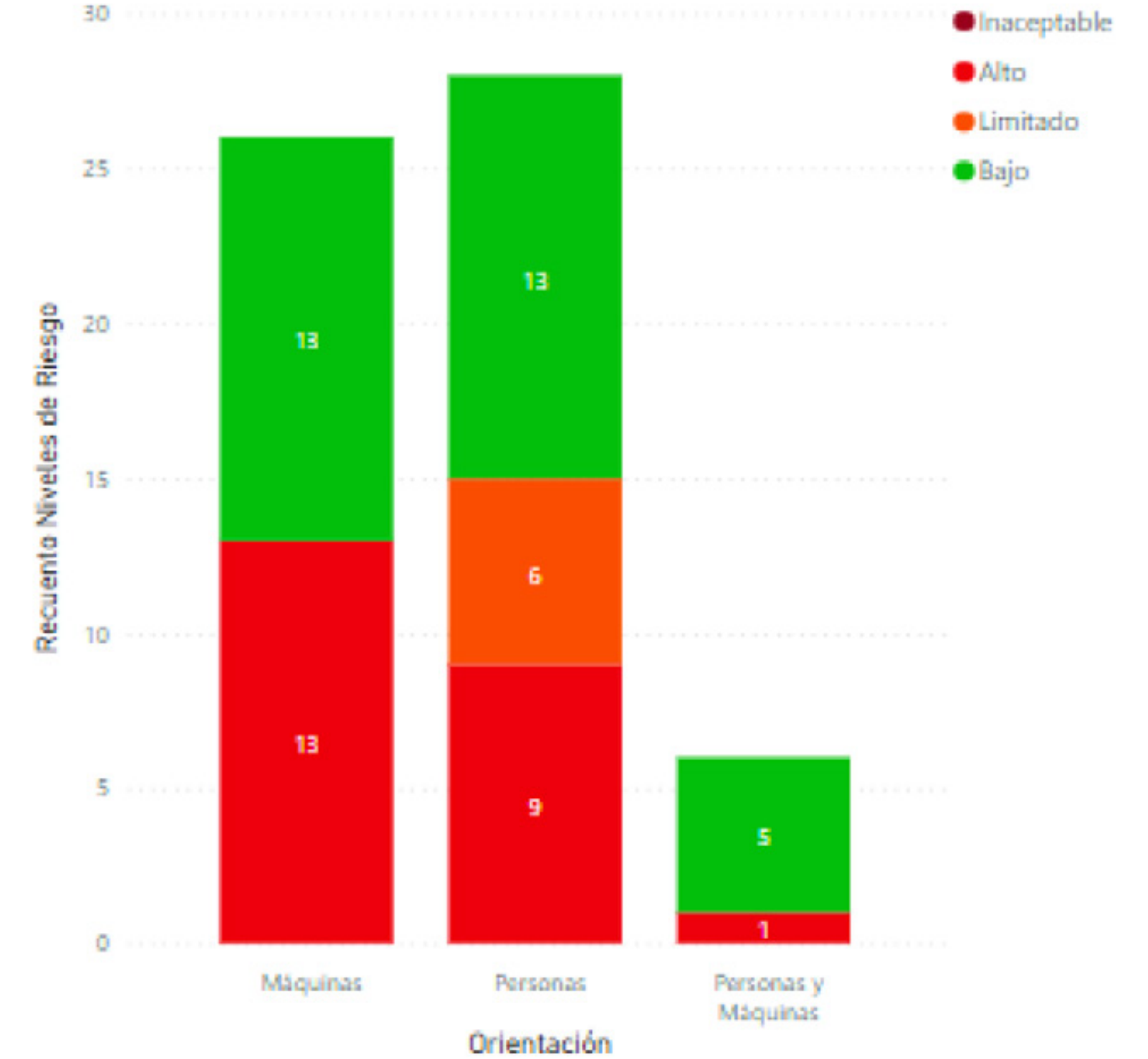


Figura 13 - Niveles de riesgo de IA Act por sub-industria

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

Considerando los casos a nivel individual, conviene resaltar una serie de observaciones. En primer lugar, sobre aquellos casos de uso en los que el sistema de IA estaba destinado a las labores de inspección y mantenimiento, estos fueron clasificados como de alto riesgo dado la función del sistema como componente de seguridad. Ello con independencia de que el resultado del sistema fuera un input más para el operario que lleva a cabo la inspección, es decir, independientemente de que el sistema tomase la última decisión al respecto o no. Esto puede parecer exagerado, pero existe el problema probado de la sobreconfianza de las personas en las recomendaciones de los sistemas de IA.

Otro aspecto a destacar tiene que ver con casos concretos como el del gemelo digital. Si un sistema de este tipo es usado para gestionar y mantener infraestructura crítica, entonces el caso fue considerado como de alto riesgo dada la función de seguridad que cumple este gemelo digital. Por el contrario, si el gemelo es tan solo empleado en fases de diseño y experimentación, entonces se puede decir que el riesgo disminuye significativamente (i.e., caso de riesgo bajo).

Por otro lado, en relación a los casos basados en la optimización logística, estos fueron identificados como de bajo riesgo. Sin embargo, se debe velar por que la optimización de las rutas no sea demasiado estricta (e.g., que el operario deba cumplir una entrega en muy poco tiempo), ya que en caso contrario podría producirse accidentes y, por tanto, el riesgo sería mayor, como podemos leer en algunas noticias²⁵.

Aunque la nueva regulación de IA no entrará en vigor en los próximos meses, y una vez en vigor habrá seguramente un periodo de adaptación de dos años, España va a hacer una prueba regulatoria de esta regulación en el cuál las empresas pueden participar de manera voluntaria.



Notas

25. Suceso de Amazon (2021).

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

**Análisis de impacto
ético y social en la
industria española**

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

4.4. Conclusiones del análisis

A modo de conclusión, el número de casos por orientación a priori evidencia un mayor peso del foco en personas. Sin embargo, si se excluye al sector de las telecomunicaciones del recuento, el número de casos orientados a máquinas es superior: 21 casos de máquinas frente a 15 de personas, algo que creemos es típico para el sector industrial. Respecto a los principios éticos más impactados, destacan el de seguridad y robustez y el de supervisión humana en el grupo orientado a máquinas; y los de privacidad y gobernanza, transparencia y explicabilidad, y justicia e inclusión en el caso de los sistemas de IA con foco en personas.

Por último, en relación con los niveles de riesgo definidos en el AI Act, un número importante de casos pertenece a la categoría de riesgo bajo; y la de riesgo alto, cobra importancia en los casos orientados a máquinas, sobre todo cuando tratan de componentes de seguridad.

Un objetivo del análisis ha sido también ver cómo la inteligencia artificial en el sector industrial afecta al futuro del trabajo dado que la automatización es una aplicación importante. Pero en los casos de uso analizados sólo lo hemos visto en los subsectores automoción y energía.



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

**Toolkit de
implementación de una
IA ética en el sector
industrial**

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



05

TOOLKIT DE
IMPLEMENTACIÓN DE
UNA IA ÉTICA EN EL
SECTOR INDUSTRIAL

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

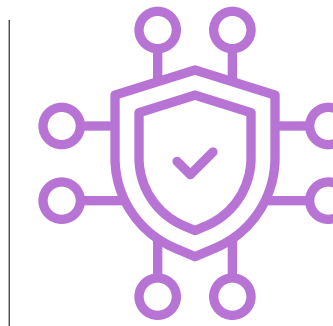
Análisis de impacto ético
y social en la industria
española

**Toolkit de
implementación de una
IA ética en el sector
industrial**

Caso de éxito de gobernanza
de IA en empresa

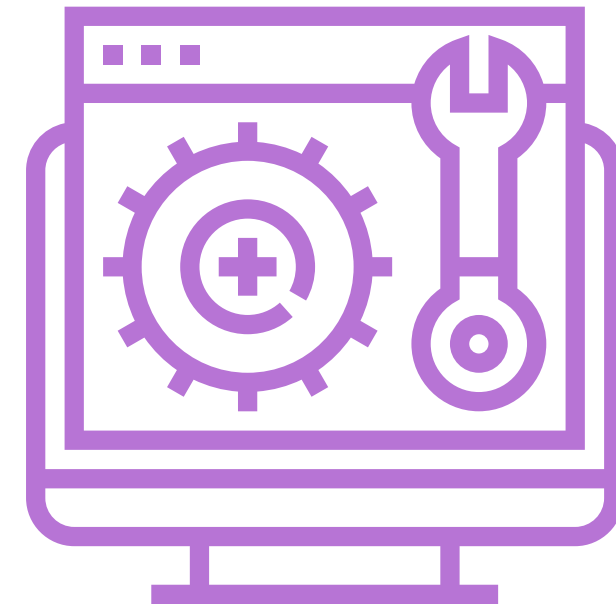
Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



El objetivo principal de este documento de IndesIA es crear un modelo de referencia que permita a las empresas del sector industrial español adoptar tecnologías de inteligencia artificial con un enfoque responsable.

Este capítulo consiste en una propuesta de toolkit de implementación de la IA ética en empresa, basada en el análisis general del estado de la cuestión (capítulo 2), y el análisis adaptado al contexto industrial español (capítulo 3) previamente presentados. Dicho toolkit incluye una serie de herramientas para facilitar el proceso de evaluación de los casos de uso de IA.

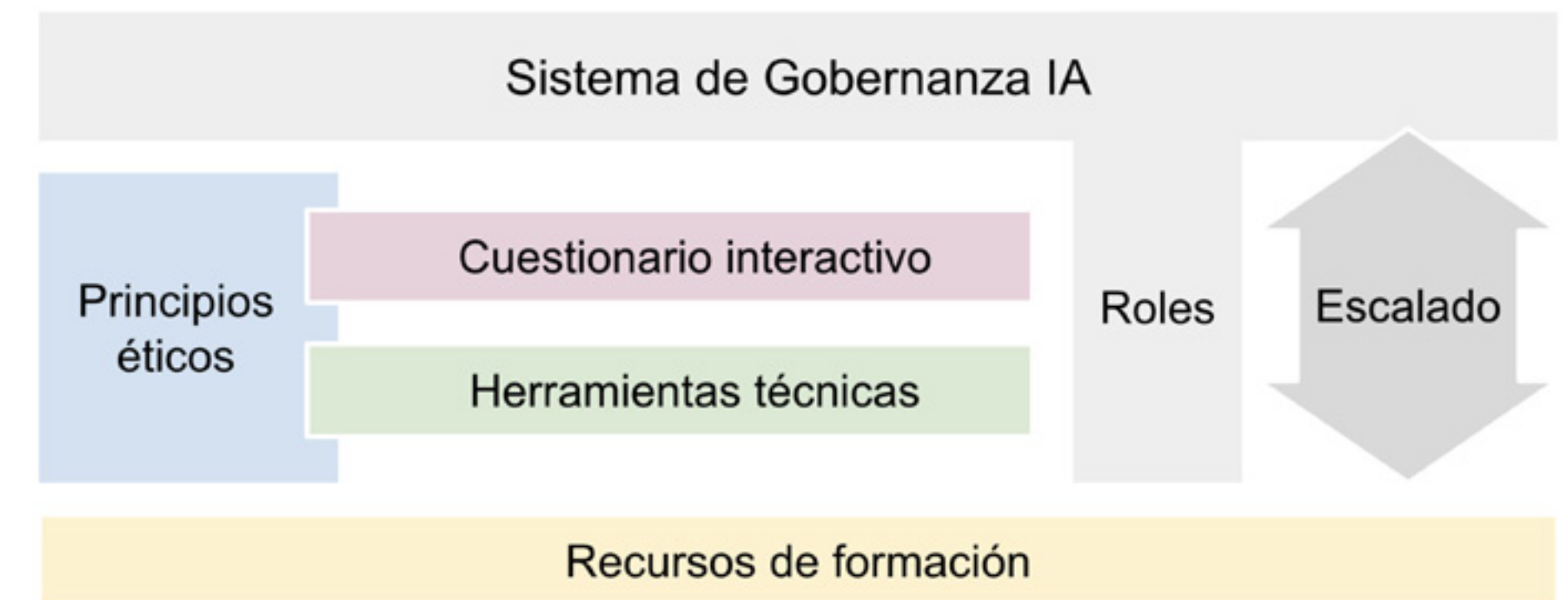


5.1. Toolkit de implementación e instrucciones para su utilización

El toolkit de implementación de IA ética de IndesIA y OdiselA incluye los siguientes elementos:

- Lista de principios éticos de IA adaptados al sector industrial, y presentados en la sección 3.1 de este documento, con especial énfasis en aquellos más aplicables al tipo de proyecto IA en el sector industrial.
- Herramientas técnicas (por principio ético) para un análisis aplicado a nivel de proyecto y datos disponibles, en la sección 4.5.
- Herramienta cuestionario interactivo (por producto o proyecto) de la sección 4.2, para el análisis de impacto ético por proyecto, en base a la lista de principios, además de recomendar algunas acciones para mitigar el riesgo ético. Asimismo, esta herramienta genera un scoring de la medida de impacto en función de los resultados del cuestionario.

- Sistema de gobernanza de IA (por empresa) de la sección 4.3, con una definición de roles y enfoques de escalado para gestión de casos críticos, y una selección de enfoques alternativos en función de la empresa (4.3.d).
- Recursos de formación de ética en IA, seleccionados en base al enfoque industrial y recursos disponibles de miembros IndesIA en la sección 4.4.



El uso responsable
de la IA en el sector
industrial

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

**Toolkit de
implementación de una
IA ética en el sector
industrial**

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



**El objetivo es
que las empresas
entiendan y
adopten los
diferentes
elementos del
toolkit**

El objetivo es que las empresas entiendan y adopten los diferentes elementos del toolkit, y que los tomen como referencia para adaptarlos al contexto particular de la empresa. Esto quiere decir, dependiendo del tipo de empresa (B2C o B2B), así como la naturaleza de las soluciones desarrolladas, todos o algunos de los principios éticos serán especialmente aplicables, y por lo tanto la elección de herramientas técnicas y la dinámica del cuestionario interactivo dependerá de ello. Igualmente, el sistema de gobernanza recomendado, así como los roles implicados y la dinámica de escalado dependerá de la estructura humana de la organización. Una lista de consejos prácticos para las empresas adaptadoras de IA en el sector industrial está disponible en la sección final del documento.

Todo ello, teniendo en cuenta el momento específico que el contexto nacional y europeo suponen para el sector industrial español en 2022, las necesidades de regulación y auditoría explicadas en la sección 2.5, así como la ventana de oportunidad para una implementación proactiva de una IA ética y responsable, tal como se describe en este artículo²⁶ de Linux Foundation:

Notas

26. The Role of Open Source for Accountable AI.

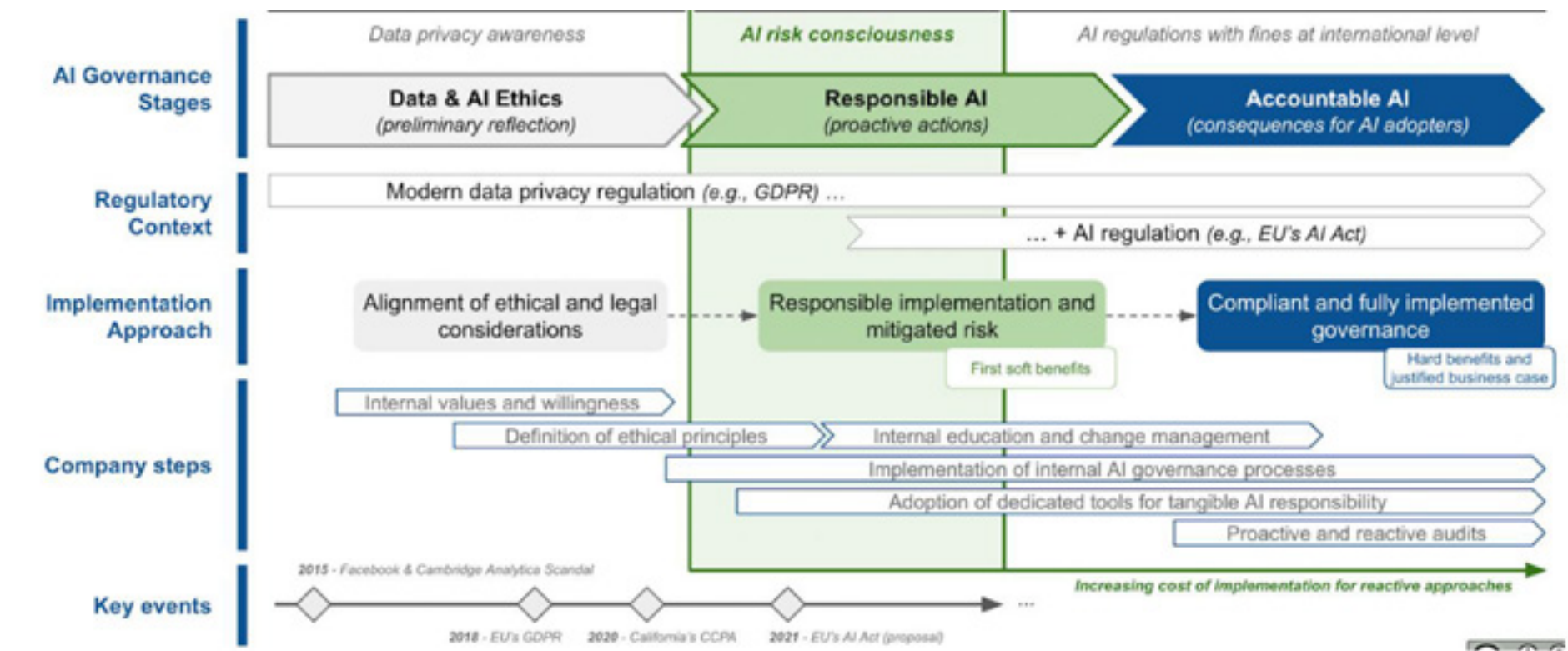


Figura 15 -Roadmap de gobernanza de IA y periodo 2021-23+



5.2. Herramienta cuestionario interactivo y medida de impacto

Para facilitar la implementación de los principios éticos en el uso y desarrollo de la IA, se pondrá a disposición pública una herramienta de autoevaluación interactiva. Esta permitirá a los usuarios, tras responder a una serie de preguntas, medir su rendimiento en cada una de las dimensiones éticas y recibir feedback sobre cómo podrían mejorarlo. Las imágenes que se muestran a continuación proceden del cuestionario de IA ética de OdiselA, desarrollado de manera conjunta con Telefónica y la GSMA. Tenga presente que la herramienta está sujeta a cambios y que la versión final puede ser diferente.

Como paso previo a responder a las cuestiones de contenido ético, además de insertar información general sobre el sistema de IA, es necesario responder a tres preguntas que definirán el nivel de riesgo del proyecto (bajo, medio, alto). Dichas preguntas tienen que ver con la severidad de un posible efecto adverso no deseado, la escala (alcance) del daño potencial y la probabilidad de que este ocurra. Así, el nivel de riesgo determina el número de preguntas; de modo que mientras menor

sea el riesgo, a menos cuestiones habrá que responder.



Figura 16 -Evaluación del riesgo en herramienta cuestionario

Las preguntas de la herramienta están estructuradas en torno a los siguientes principios: (1) Transparencia y Explicabilidad; (2) Supervisión y Autonomía Humana; (3) Justicia; (4) Privacidad; (5) Seguridad y Robustez; (6) Rendición de Cuentas; e (7) Impacto Ambiental. Estos principios se ajustan a las dimensiones éticas consideradas en el apartado anterior, a excepción de la dimensión de Derechos Humanos y Futuro del Trabajo, que se cubrirá en versiones posteriores. En función de su respuesta a las preguntas, el usuario recibirá un comentario específico. Por

ejemplo, en caso de responder negativamente a una pregunta, se indicarán acciones de mejora que deben ejecutarse para cumplir con el requerimiento ético correspondiente. Además, debajo de cada pregunta podrán insertarse comentarios como el porqué de la respuesta, el estado de la acción de mejora propuesta, o cualquier otra información que facilite la evaluación del sistema y su seguimiento.



Figura 17 -Evaluación del riesgo en herramienta cuestionario

Prólogo

Introducción

Estado de la cuestión
de la ética en IA


Análisis de impacto ético
y social en la industria
española

**Toolkit de
implementación de una
IA ética en el sector
industrial**

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



Los procesos organizativos existentes pueden aprovecharse para ayudar a aplicar las políticas éticas de la IA

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos



Finalmente, tras responder a las preguntas éticas, el usuario podrá visualizar tablas de recuento y gráficos sobre sus respuestas al cuestionario, así como generar un informe que resuma los principales resultados de la evaluación. Esto será especialmente útil para identificar áreas de mejora y también a efectos de registrar el estado ético del sistema a lo largo del tiempo. Sobre esto último, el cuestionario está concebido para ser completado tantas veces como sea necesario, como por ejemplo, durante cada fase principal del ciclo de vida del proyecto o cada vez que se apliquen cambios significativos al sistema (e.g., aumento del número de usuarios afectados por la actividad del sistema).



Figura 18 -Medida de impacto en herramienta cuestionario

Prólogo

Introducción

Estado de la cuestión de la ética en IA

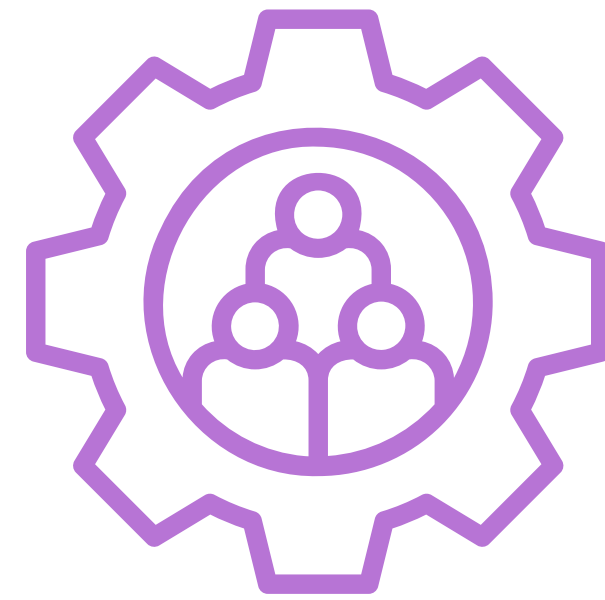
Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos



5.3. Sistema de gobernanza de la IA en empresa

5.3.1. ¿Qué es la gobernanza de la IA?

La gobernanza de la IA se refiere a la idea de que los sistemas que utilizan algoritmos para hacer inferencias y tomar decisiones requieren políticas que rijan su diseño, desarrollo y despliegue. Estas políticas deben basarse en principios éticos como los previamente mencionados.

La acción de crear políticas y garantizar su cumplimiento en toda la organización requiere una sólida estructura de gobierno con funciones y responsabilidades claras.

5.3.2. Crear una estructura adecuada

Una estructura de gobierno eficaz debe tener en cuenta la cultura organizativa, el tamaño, la estrategia y la madurez de la IA. A la hora de decidir la estructura de gobernanza de la IA es importante involucrar a las áreas relevantes de la organización. Muchas de las habilidades clave es probable que ya existan dentro de una organización. Por ejemplo:

- Los ejecutivos de la alta dirección deben participar como patrocinadores del proyecto, ya que a menudo tendrán responsabilidad final del ciclo de desarrollo del producto de IA de desarrollo de productos de IA.

- Los equipos legales y reguladores están bien versados en garantizar el cumplimiento.
- Los equipos de producto entienden cómo funcionan los sistemas desde una perspectiva técnica.
- Los profesionales de la responsabilidad social corporativa (RSC) están acostumbrados a considerar las implicaciones sociales de la actividad empresarial.

La participación de diversos actores multidisciplinares en el proceso de creación de una estructura de gobierno ayudará a aportar una diversidad de perspectivas y su posterior internalización por la organización.

Los procesos organizativos existentes pueden aprovecharse para ayudar a aplicar las políticas éticas de la IA. Por ejemplo, la mayoría de las empresas que recogen y manejan datos tienen sistemas de gobernanza establecidos para garantizar las mejores prácticas y el cumplimiento de la legislación sobre de datos. Tiene sentido inspirarse o incluso basarse en ellos a la hora de decidir cuál es la mejor manera de gobernar la IA.

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

5.3.3. Ejemplo de modelo de gobernanza: escalada de cuestiones éticas

Esta sección explora un enfoque para crear un modelo de gobernanza basado en la experiencia de las empresas. Este modelo, elegido por su flexibilidad, implica una estructura de toma de decisiones e información que puede ayudar a integrar la responsabilidad de los sistemas de IA en la cultura organizativa y mejorar la agilidad eliminando los cuellos de botella.

No cabe duda de que este modelo de gobernanza requerirá una inversión para tener éxito. Esto incluye reforzar la conciencia ética de la IA en toda la organización; formar a los gestores de productos de IA; asignar tiempo para documentar, evaluar y auditar los procesos; crear nuevas funciones o responsabilidades; comprar o la creación de herramientas técnicas y de otro tipo,



y el fomento de la colaboración entre departamentos. La introducción de nuevos principios éticos en una organización requiere un programa de gestión del cambio para que se conviertan en algo habitual.

La estructura de gobernanza propuesta por GSMA en su AI Ethics Playbook²⁷ tiene tres niveles. Estos niveles indican el proceso para escalar casos de riesgo o casos de uso más complejos.



Figura 19-Niveles de gobernanza, GSMA AI Playbook

5.3.4. Enfoques alternativos

Los sistemas de gobernanza de IA con niveles pueden enfocarse de dos maneras, en función de la dinámica de escalado elegida:

- 1 Modelo de permiso (top-down), en el cual el nivel ejecutivo permite a los equipos trabajar en aplicaciones específicas.
- 2 Modelo de problemas (bottom-up), que permite a los equipos de desarrollo analizar las necesidades de análisis ético de IA, y contactar con los niveles superiores cuando lo crean necesario o cuando no consiguen ponerse de acuerdo sobre la aplicación de uno o varios principios éticos.

Notas

27. GSMA Playbook.

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

**Toolkit de
implementación de una
IA ética en el sector
industrial**

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



5.4. Recursos de formación

El rol de la formación y el acompañamiento para una comprensión adecuada de la inteligencia artificial, su componente ética, y por supuesto su aplicación en el sector industrial, es especialmente relevante y permite que las empresas eduquen a su personal y les ayuden a implementar la noción de IA ética en empresa.

Existen multitud de formaciones en IA, pero hemos seleccionado las siguientes con un foco especial en ética en IA, acompañada de recursos generales IA tipo 101 de calidad, aunque pocas se enfocan solo en el sector industrial:



Programa de formación IA con enfoque ético y responsable. OdiselA dispone de un programa verticalizable al sector industrial sobre IA ética de 20 horas de duración, que ya ha sido impartido a miembros del Ministerio de Asuntos Económicos y Transformación Digital (MINECO), y de la Secretaría de Estado de Digitalización e Inteligencia Artificial (SEDIA):

- **Módulo 1** - Introducción a la IA
- **Módulo 2** - Estrategia IA para sectores industriales específicos
- **Módulo 3** - Framework para gestión de proyectos IA
- **Módulo 4** - Tecnologías aplicadas al negocio
- **Módulo 5** - Aspectos éticos de la IA
- **Módulo 6** - Aspectos normativos de la IA
- **Módulo 7** - Practica grupal para implementación de IA en empresa, incluyendo gobernanza de IA ética



Recursos en línea

- **AI Business School**

www.microsoft.com/en-us/ai/ai-business-school

- **AI School for Manufacturing**

docs.microsoft.com/en-us/learn/paths/ai-business-school-manufacturing

- **AI Development 101**

www.youtube.com/playlist?list=PLlrxD0HtieHjAGOj3FrgN7wZS9afg7X6R

- **AI Business 101**

www.youtube.com/playlist?list=PLCICBn7ebwvslHwZJ5yRDjfsDkj_Wzmx



- **AI Ethics: Global Perspectives, Global AI Ethics Consortium**

aiethicscourse.org/modules.html

- **Ethics in AI and Data Science, Linux Foundation**

training.linuxfoundation.org/training/ethics-in-ai-and-data-science-lfs112

- **Ethics of AI, University of Helsinki**

ethics-of-ai.mooc.fi

- **Programa especializado: Ethics in the Age of AI, Coursera**

es.coursera.org/specializations/ethics-in-ai

- **Data Ethics, AI and Responsible Innovation, University of Edinburgh**

www.edx.org/course/Data-Ethics-AI-and-Responsible-Innovation

- **Practical Data Ethics, fast.ai**

ethics.fast.ai/

- **Cursos en el catálogo del consorcio AI4EU**

www.ai4europe.eu/education/education-catalog?field_category_target_id=143

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

**Toolkit de
implementación de una
IA ética en el sector
industrial**

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

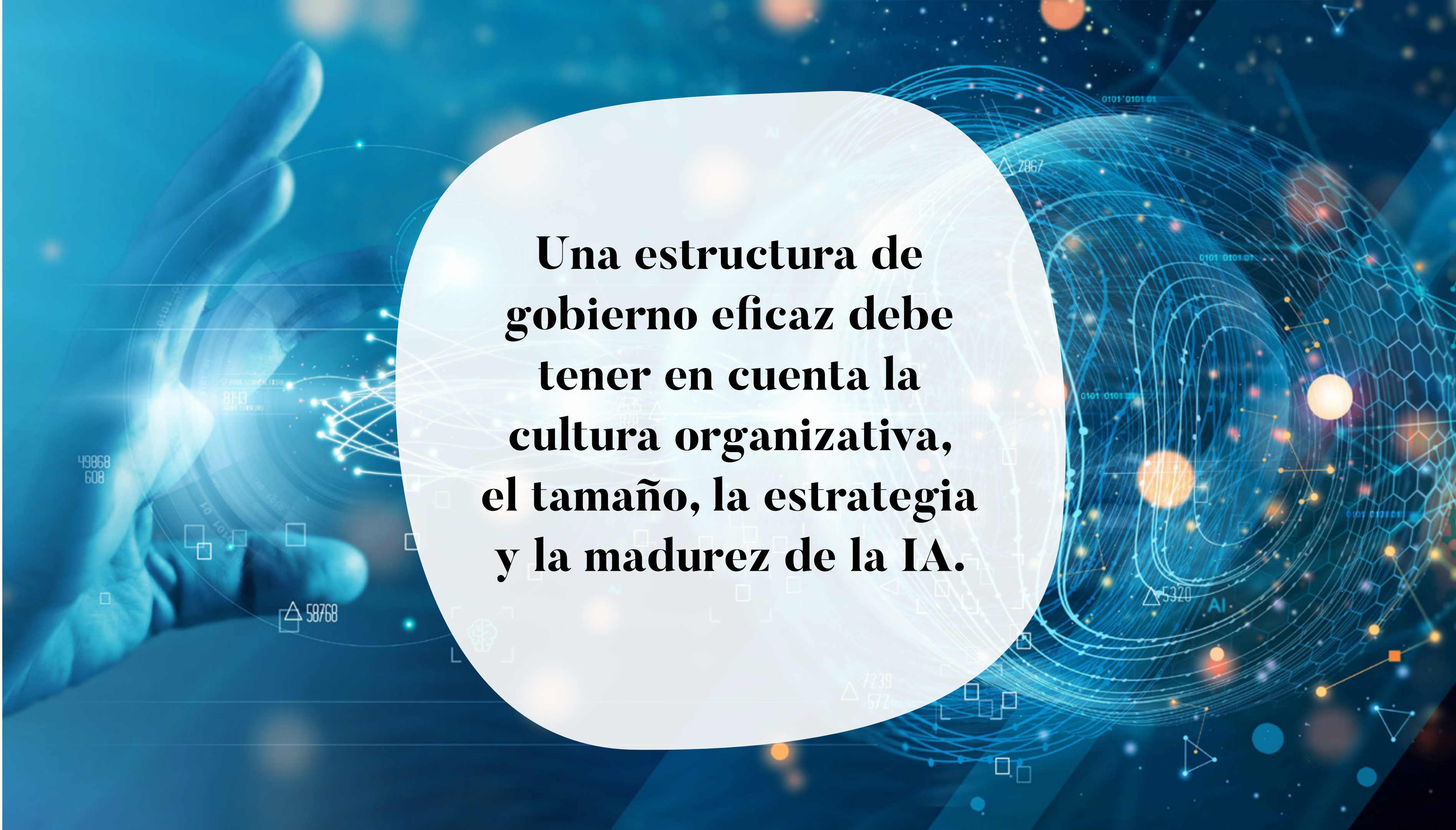
Análisis de impacto ético
y social en la industria
española

**Toolkit de
implementación de una
IA ética en el sector
industrial**

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



**Una estructura de
gobierno eficaz debe
tener en cuenta la
cultura organizativa,
el tamaño, la estrategia
y la madurez de la IA.**

Prólogo

Introducción

Estado de la cuestión de la ética en IA

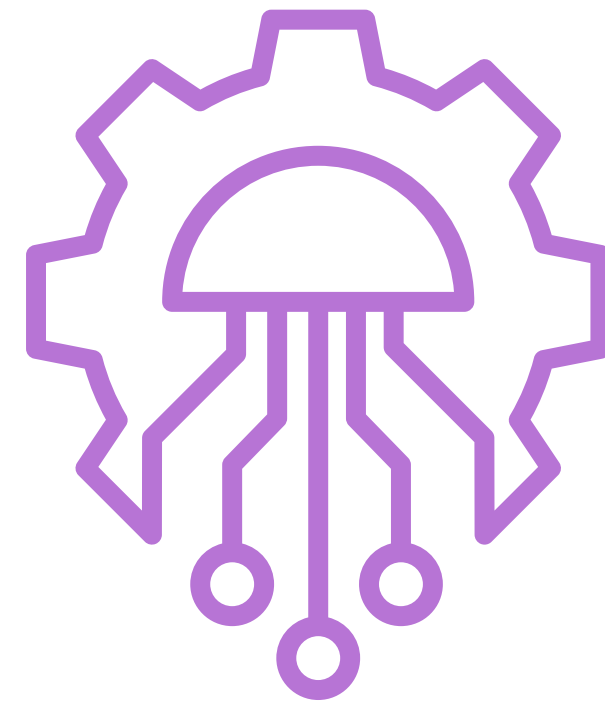
Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos



5.5. Herramientas técnicas disponibles

A continuación, se listan una serie de herramientas disponibles para cada uno de los principios éticos. Por el momento, no se trata de una recomendación (ya que dicho análisis se podrá efectuar en futuras publicaciones de IndesIA) sino de una lista de herramientas y librerías a disposición de las empresas que adopten la IA, con opciones a nivel comercial y open source.

Nota: los principios de “rendición de cuentas”, y “DDHH y futuro del trabajo” figuran como N/A (no aplicable) al no ser medibles mediante herramientas técnicas, pero mecanismos para su control pueden ser incorporados a nivel de gobernanza.



Principio	Herramienta técnica disponible	Empresa
Privacidad y Gobernanza	Smart Noise	Microsoft
	Presidio	Microsoft
	SEAL Homomorphic Encryption	Microsoft
	Confidential computing for ML	Microsoft
	AI Privacy 360 toolbox	IBM
	Smart Noise	OpenDP
Rendición de Cuentas	N/A	-
Seguridad y Robustez	Error Analysis	Microsoft
	Counterfeit	Microsoft
	Adversarial Robustness 360 Toolbox	Linux Foundation IBM
	Alibi Detect	Seldon.io

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

**Toolkit de
implementación de una
IA ética en el sector
industrial**

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



Principio	Herramienta técnica disponible	Empresa
Transparencia y Explicabilidad	InterpretML	Microsoft
	DiCE (Diverse Counterfactual Explanations)	Microsoft
	EconML	Microsoft
	Datasheets for Datasets	Microsoft
	AI Explainability 360	Linux Foundation IBM
	Uncertainty Quantification 360	IBM
	AI FactSheets 360	IBM
	Watson Openscale (Cloud Pak)	IBM
	Causal Inference 360	IBM
	Visual Data Mining and ML	SAS
	Alibi Explain	Seldon.io
	Fiddler Explainable AI	Fiddler.ai
	AyGLOO XAI	AyGLOO

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

Principio	Herramienta técnica disponible	Empresa
Justicia e Inclusión	Microsoft Fairlearn	Microsoft
	AI Fairness 360 toolkit	Linux Foundation / IBM
	AI Fairness Checklist	Microsoft
	Human-AI Experiences (HAX)	Microsoft
	Watson OpenScale (Cloud Pak)	IBM
	Causal Inference 360	IBM
	LinkedIn Fairness Toolkit (LiFT)	Linkedin
	AyGLOO XAI	AyGLOO
Supervisión Humana	GAMMA FACET	BCG GAMMA
	Deepchecks Open Source	Deepchecks
DDHH y Futuro del Trabajo	N/A	-
Sostenibilidad e Impacto Ambiental	CodeCarbon	Mila (Montreal), BCG GAMMA, Haverford College, Comet.ml



Para aquellas empresas que busquen no solo explorar sino entender cómo este tipo de soluciones funcionan en el día a día, de manera aterrizada, la iniciativa GuIA de OdiselA ha publicado un informe detallado que explora algunas de estas herramientas. Concretamente, el capítulo 3 del presente documento.



06

CASO DE ÉXITO DE GOBERNANZA DE IA EN EMPRESA

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

**Caso de éxito de
gobernanza de IA
en empresa**

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



CASOS DE ÉXITO



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

**Caso de éxito de
gobernanza de IA
en empresa**

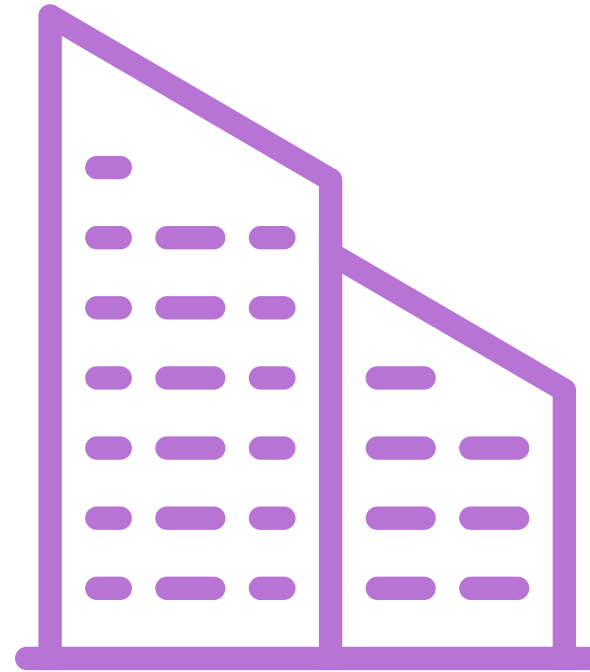
Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

Telefónica, con una historia casi centenaria, es una de las multinacionales emblemáticas que han acompañado al mercado español e internacional a lo largo de los años. Desde sus orígenes en 1920, los servicios de telecomunicaciones y las infraestructuras en las que se basan han evolucionado hasta convertirse en los ejes básicos de una innovación, sostenida por los datos. La compañía se ha adaptado proactivamente a la secuencia de tecnologías disruptivas, desde la adopción temprana de big data, hasta reconocidas iniciativas de datos e inteligencia artificial (IA) como Telefónica Kernel (antes 4ª Plataforma), y el rol pionero de la compañía en términos de gobernanza de datos e Inteligencia Artificial.

Este caso de éxito pretende contextualizar el rol de Telefónica como empresa de innovación y pionera en materia de inteligencia artificial, con un foco especial en su enfoque propio de IA responsable y la contribución al estado del arte de dicha materia, que es para la mayor parte de empresas una tarea pendiente, tanto a nivel de industria nacional e internacional, como en comparación con otros actores relevantes de la industria de telecomunicaciones.





6.1. Telefónica: Orígenes e IA responsable

La historia de Telefónica como empresa tecnológica - que va más allá de las actividades puras de telecomunicaciones - no puede ser entendida sin una estrategia de datos, como los generados por las redes, aquellos relacionados a los usuarios de los servicios, y el conocimiento generado, a modo de insights. Todos ellos han sido fuente y motor del resto de actividades de la marca a lo largo de los últimos años.

La relación entre Telefónica y los datos (tradicionalmente conocidos como big data) no es nueva. El periodo entre 2011 y 2015 supuso el paso inicial entre la empresa que hasta entonces era considerada un actor relevante de la industria telco y la transformación total para dar lugar a una empresa de innovación en datos e inteligencia artificial. La experiencia interna obtenida con las iniciativas de big data dio lugar en 2016 a la creación de una unidad de negocio para proyectos con clientes (LUCA), integrada ahora en Telefónica Tech, y la creación de un nuevo rol ejecutivo: el Chief Data Officer con Chema Alonso a la cabeza, la que en su día fue una decisión pionera del actual Presidente de la empresa, José María Álvarez-Pallete.

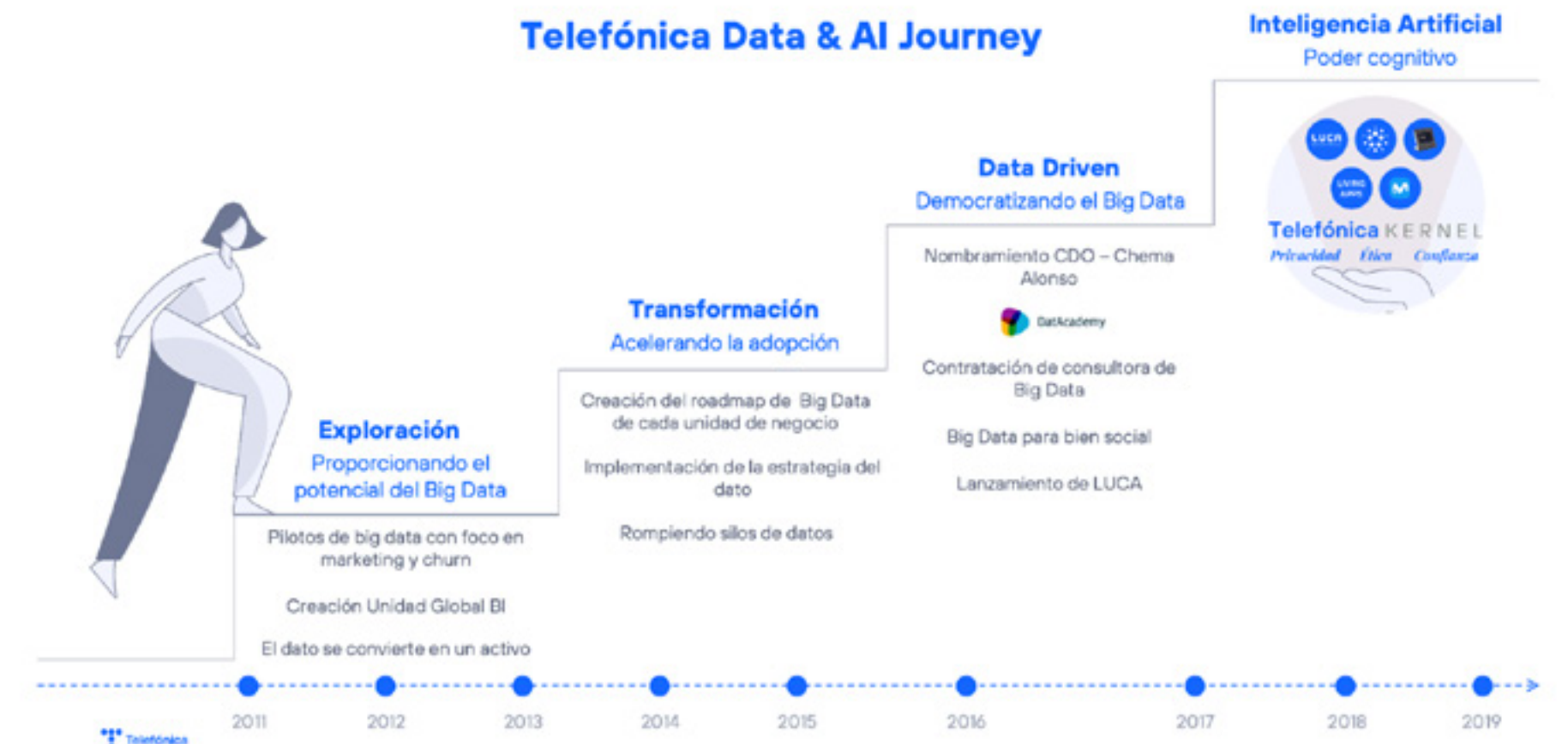


Figura 20 - Trayectoria de Telefónica entre 2011 y 2019



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

**Caso de éxito de
gobernanza de IA
en empresa**

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos

Esa nueva etapa dio lugar a la Telefónica actual, la misma que atrae la atención en el Mobile World Congress (MWC), año tras año, con presentaciones vanguardistas, y con la innovación como ADN. Esta innovación le permite aprovechar todas las oportunidades de progreso y de negocio que ofrece el ecosistema digital actual, caracterizado por su dinamismo y rapidez, y con la que ha cambiado para siempre la manera de entender los datos y la relación entre la telco y sus clientes, siendo la 4ª Plataforma su exponente máximo y todos los servicios digitales que se crean sobre la misma usando sus capacidades.

Telefonica Kernel es una capa de abstracción y digitalización que permite disponer de una visión 360 de cliente. Para conseguirlo, se apoya sobre distintas capacidades de gestión de identidad, de “apificación” (APIs) y de abstracción de datos que implementa bajo la máxima de privacidad-por-diseño. Con este ecosistema de capacidades que ofrece, se pueden desarrollar aplicaciones, productos y servicios digitales más rápido, con menos fricción y además habilita la creación de productos que incorporan inteligencia artificial para enriquecer la experiencia del cliente en Telefónica. Algunos de estos productos son: Aura, el asistente virtual con Inteligencia Artificial de Telefónica; y la aplicación Mi Movistar. Ambos ofrecen una nueva forma de relación con los clientes para gestionar su experiencia digital con la compañía. Soluciones a las que hay que sumar otras como las Living Apps, aplicaciones integradas en Movistar Plus+, que ofrecen una nueva experiencia de consumo desde la televisión o soluciones de mercado con terceros a través de Telefónica Tech, líder en transformación digital que integra, entre otros, el negocio de Big Data e IoT del Grupo Telefónica.



Prólogo

Introducción

Estado de la cuestión de la ética en IA

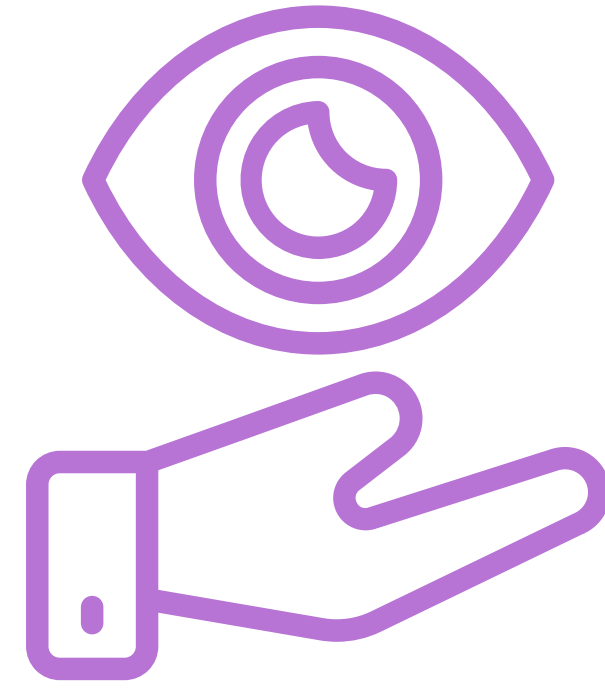
Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos



6.2. Visión y principios de la IA

Telefónica hace girar todas sus acciones sobre los Objetivos de Desarrollo Sostenible (ODS), tal y como indica en sus Principios de negocio responsable y en la Política de derechos humanos. Esto le permite asegurar la sostenibilidad del negocio además de poder aportar, de forma significativa, a un proyecto determinante para el futuro de las personas, favoreciendo el progreso social y económico a través de la digitalización a la par que generar confianza para asegurar una transición digital centrada en las personas.

La premisa principal es que la tecnología debe contribuir a crear una sociedad más inclusiva y ofrecer mejores oportunidades para todos, sin dejar a nadie atrás y la inteligencia artificial puede contribuir a estos objetivos. Con el fin de guiar a la empresa y sus empleados en su aplicación de la Inteligencia Artificial y el big data en todas las líneas de negocio, Telefónica publicó en 2018 sus «Principios de IA»:



Figura 21 - Principios de la inteligencia artificial según Telefónica



Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

El motivo de la creación de dichos principios fue justamente señalar la importancia de la ética para la inteligencia artificial, la cual – según Telefónica – no puede estar completa si no es acompañada de dicha visión. Cada uno de los cinco principios tiene un razonamiento específico:



Principio	Razonamiento
Justicia	Las aplicaciones de la tecnología de IA deben dar resultados justos, sin impactos discriminatorios en relación con la raza, el origen étnico, la religión, el género, la orientación sexual, la discapacidad o cualquier otra condición personal.
Transparencia y explicabilidad	El objetivo es que los usuarios sepan que están interactuando con un sistema de IA, qué datos suyos se usan y para qué. Telefónica se asegurará de comprender la lógica que hay detrás de las decisiones del sistema.
Centrada en las personas	La IA debe estar al servicio de la sociedad y generar beneficios tangibles para las personas, cuyos derechos humanos no pueden verse vulnerados. Además, la compañía se ha propuesto ayudar a cumplir los Objetivos de Desarrollo Sostenible (ODS) con la IA.
Privacidad y seguridad desde el diseño	Explicabilidad e interpretabilidad de los algoritmos, incluye el modelo y los datos usados.
Trabajando con socios y terceros	Las Políticas de Privacidad y Seguridad de la Compañía cobran en estos Principios especial relevancia para preservar los datos tanto personales como anónimos y agregados.

Para mayor información sobre los Principios de IA Responsable en Telefónica:

- **Octubre 2018 - Telefónica AI Principles**

iethicscourse.org/modules.html

- **Infografía sobre los Principios de IA Responsable en Telefónica**

www.telefonica.com/es/wp-content/uploads/sites/4/2021/03/infografia-ia-esp.pdf

Estos principios de IA ética se enmarcan dentro de los principios de Diseño Responsable de Telefónica:



Figura 22 - Principios de diseño responsable

La presentación de la alianza estratégica entre Telefónica y Microsoft para impulsar la inteligencia artificial y crear nuevos servicios, con los máximos responsables de ambas empresas presentes en el MWC'19, ilustró un paso más en la etapa de madurez IA de Telefónica, caracterizada como Cognitive Intelligence o inteligencia cognitiva por la empresa. Desde entonces, la adopción de la Inteligencia Artificial ha continuado como una tendencia natural, y ha habido aportaciones muy significativas en el ámbito de la ética y la IA responsable.

En los dos últimos años, hemos visto un auge general de los conceptos de IA responsable, los cuales partieron de una realidad teórica que ha ido evolucionando hacia prácticas más aterrizadas. Telefónica no ha sido una excepción, y ha trabajado en la definición de nuevas prácticas de gobernanza y nuevos roles relacionados, los cuales vamos a explorar en la sección 3.



6.3. Factor diferenciador: Gobernanza IA interna

La gobernanza de inteligencia artificial y el enfoque propio de Telefónica son motivos suficientes para este caso de éxito que nos ocupa. Durante los últimos tres años, la empresa ha desarrollado un sistema de IA responsable con tres niveles, en los que cada persona tiene la capacidad para analizar posibles riesgos potenciales.

Los puntos principales de este sistema de gobernanza de Inteligencia Artificial Responsable son:

- **Los principios IA como punto de partida para cualquier implementación que incluya inteligencia artificial y/o datos masivos.**
- **La introducción del modelo de gobernanza como parte clave de los procesos oficiales de desarrollo de productos y servicios.**
- **La creación de un cuestionario autoevaluación con el objetivo de que los responsables de producto, junto a los equipos de desarrollo, lo completen durante la fase de diseño de los productos y servicios que utilizan Inteligencia Artificial y a posteriori en sucesivas fases durante todo el ciclo de vida del producto.**
- **La definición de un proceso de escalado entre niveles para garantizar el análisis detallado.**
- **La figura de los RAI Champions, como parte del proceso de gobierno de la IA Ética, los cuales están cerca tanto de las unidades de negocio como de los equipos de desarrolladores.**

6.3.1. Rol del Responsible AI (RAI) Champion

La figura del Champion de IA responsable forma parte de la nueva granularidad de roles relacionados a las iniciativas de inteligencia artificial, más allá de los casos típicos de científicos e ingenieros de datos. El *RAI Champion* es más un rol que un puesto específico dentro de una organización, por lo general ocupado por profesionales con perfiles muy diversos, desde científicos de datos, responsables de privacidad, hasta la responsabilidad social de la empresa, con mayor o menor conocimiento técnico, pero que traen un punto de vista diferente a la hora de analizar el impacto ético de un proyecto de Inteligencia Artificial.

Este rol lo utilizan también algunas empresas que están muy avanzadas en la implementación de la IA responsable, pero Telefónica ha adaptado dicho perfil a los procesos y necesidades internas. La premisa principal es que se necesita un *RAI Champion* siempre que la unidad de negocio use o planea utilizar inteligencia artificial o sistemas de datos masivos en sus soluciones. Si el nivel de adopción es intensivo, puede haber más de un *Champion*.



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

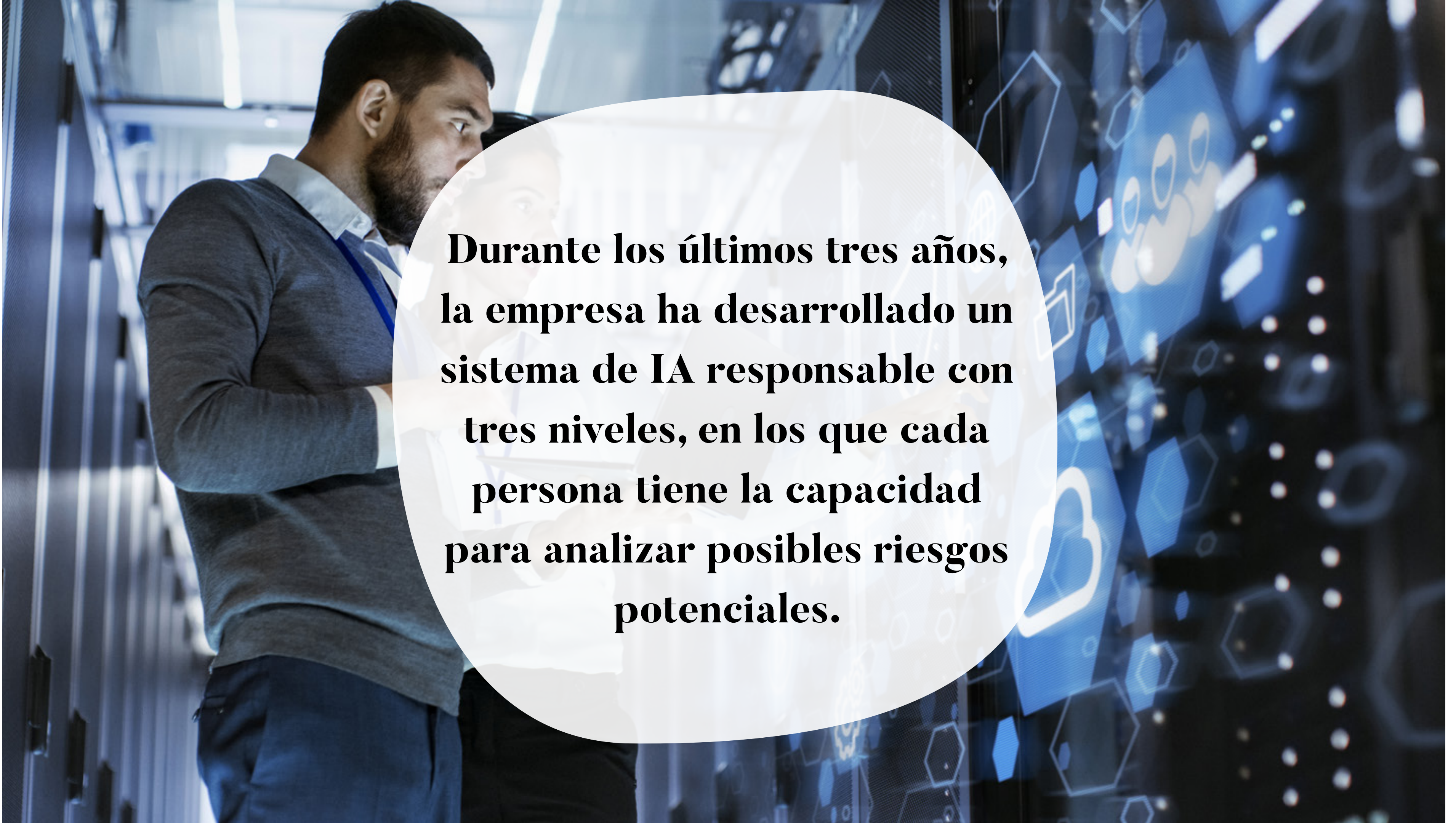
Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

**Caso de éxito de
gobernanza de IA
en empresa**

Caso de uso aplicado para la
implementación de IA ética en la
industria

Conclusiones
y consejos prácticos



**Durante los últimos tres años,
la empresa ha desarrollado un
sistema de IA responsable con
tres niveles, en los que cada
persona tiene la capacidad
para analizar posibles riesgos
potenciales.**

Las funciones de un RAI Champion incluyen:

- Informar sobre la importancia de aplicar los principios de Inteligencia Artificial.
- Formar a sus áreas de influencia sobre cómo aplicar dichos principios de Inteligencia Artificial y el modelo de gobernanza.
- Guiar y escalar, y ser el punto de referencia para dudas y cuestiones éticas, así como escalar al grupo de expertos cuando sea necesario.
- Coordinar con diferentes áreas como Privacidad, Seguridad, Chief Data Office, Responsabilidad Social Corporativa, Legal, etc.
- Conectar y fomentar la construcción de una comunidad de expertos en Inteligencia Artificial IA ética.
- Gestionar el cambio, para que las consideraciones éticas sean parte del business as usual de los equipos.

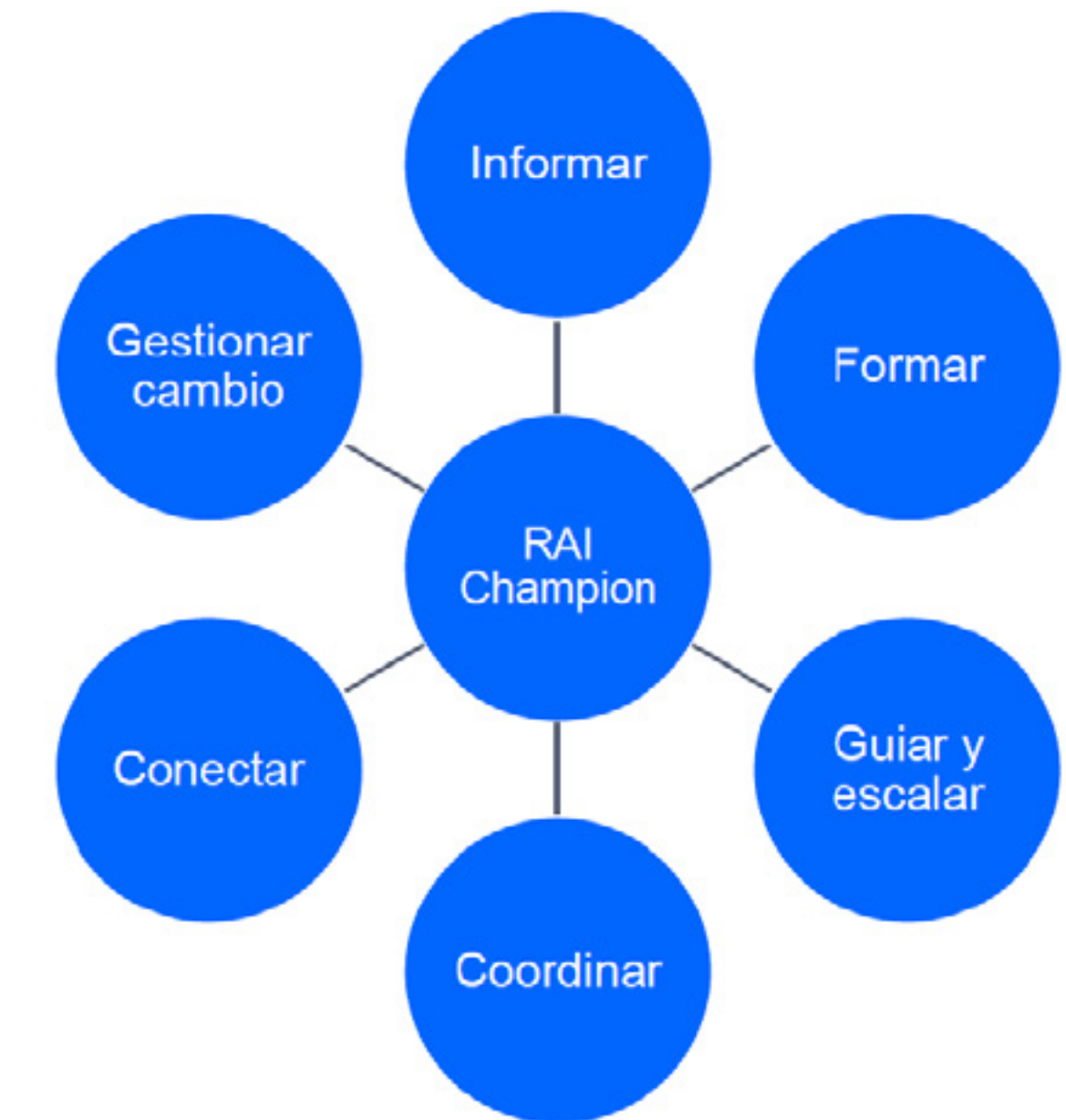
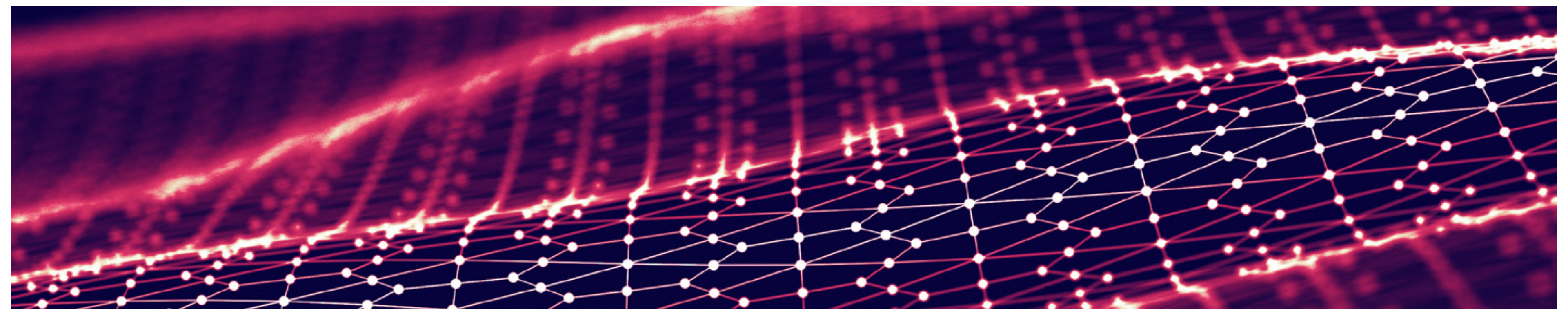


Figura 23 - Responsabilidades del Responsable AI Champion



Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos



6.3.2. Sistema de escalado en tres niveles

Tal y como hemos comentado anteriormente, el escalado forma parte del alcance de actividades del RAI Champion. Dicho escalado no se corresponde a la jerarquía tradicional de empresa, sino que Telefónica ha creado una estructura dedicada a la IA ética y responsable. No hay muchas empresas que hayan alcanzado dicho nivel de madurez interna en el mundo, y en el caso de Telefónica incluye no sólo los RAI Champions de cada área (definidos como el nivel 1 del modelo), sino un equipo core RAI (nivel 2), que coordina a todos los RAI Champions y otros actores como CDO (Chief Data Office), RSC (Responsabilidad Social Corporativa) o DPO (Data Privacy Office), y la Oficina de Negocio Responsable (nivel 3). Concretamente, la relación entre niveles y los factores de escalado son:

- **Nivel 1 a 2:** Si hay dudas en torno a las preguntas o resultados del cuestionario (detalles en la sección 4), y dichas dudas no pueden ser resueltas por el RAI Champion o su comunidad de expertos, entonces el caso se escala al nivel del equipo core RAI para consulta.

- **Nivel 2 a 3:** En el caso extraordinario de que el equipo core RAI y su red ampliada de expertos multidisciplinares no puedan encontrar una solución, se procederá al nivel final de escalado, en el que la Oficina de Negocio Responsable podrá intervenir y solucionar el caso.

Dicho sistema promueve la responsabilidad individual, y una habilidad de escalado bajo demanda, en vez de un sistema tradicional de aprobación top-down. Se parte de la base de que todo el mundo quiere implementar una IA ética y responsable, y para ello se ponen a disposición una serie de herramientas.



Para más información sobre el modelo de gobierno:
[Junio 2021 - Telefónica's Approach to the Responsible Use of AI](#)

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

6.4. Implementación: De los principios éticos a la acción

6.4.1. Formación como herramienta de conocimiento y concienciación

Siendo la Inteligencia Artificial y la ética en inteligencia artificial campos de conocimiento relativamente nuevos para muchas personas, Telefónica ha dado el peso necesario a la formación interna para democratizar el acceso a dicho conocimiento, y para que cada uno de los empleados entiendan la importancia absoluta de la ética en la Inteligencia Artificial. Hoy en día, este programa de formación ha formado a más de 3.500 empleados de la compañía, y el número sigue creciendo.

Los aspectos cubiertos en dicha formación incluyen los fundamentos de la inteligencia artificial y los datos masivos, los principios de IA Responsable de Telefónica y su implementación en función del tipo de proyecto, el sistema de gobernanza y los roles en el día a día, las herramientas disponibles para implementar una IA ética y responsable, y por supuesto ejemplos prácticos para aterrizar los conceptos aprendidos.

Este programa, a la altura de cualquier iniciativa de formación de calidad en IA ética y responsable, permite desarrollar el conocimiento y un entendimiento común de los principios, el argot, y el potencial realista de la inteligencia artificial y los datos masivos.



Figura 24 - Formación interna de IA y ética

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

6.4.2. Cuestionario de evaluación como punto de partida del análisis ético

Después de la formación, los empleados tienen la información y el conocimiento necesario para aplicar los principios en sus proyectos. Con el fin de facilitar el análisis, Telefónica ha creado un cuestionario dinámico que debe ser integrado en el diseño de productos y servicios que utilizan inteligencia artificial.

Este cuestionario está disponible a través de un acceso en línea y contiene una serie de preguntas relevantes sobre cada uno de los principios IA de Telefónica. Dichas preguntas aportan una reflexión en función de las respuestas introducidas y recomendaciones para garantizar la ética y la responsabilidad entre las fases de diseño e implementación. Este cuestionario es completado por los responsables del producto o servicio a analizar.

Entre otras, las cuestiones se centran en:

- Los sets de datos y sus variables, con foco especial en aquellas que puedan ser sensibles (por ejemplo, el género o la nacionalidad de una persona, o la presencia de grupos minoritarios).
- La correlación entre variables, como parte del ejercicio de preparación de los datos que permite reducir datos innecesarios que puedan introducir sesgos en los modelos.
- La explicabilidad de los sistemas y el posible impacto en la vida de las personas.

- Las cuestiones de privacidad de datos y el cumplimiento de la regulación.
- La relación con sistemas de terceras partes, y su alineamiento con los principios IA de Telefónica.

Por último, y tal como hemos visto en la sección 3, el RAI Champion correspondiente solventa dudas, propone medidas, e incluso escala el caso si es necesario.



6.4.3. Gobernanza IA como parte del proceso de desarrollo de productos

El sistema de gobernanza IA responsable pretende privilegiar un enfoque: “Responsible AI first”, en el que todo tipo de stakeholders (desde los más técnicos, hasta las unidades de negocio y los niveles ejecutivos) formen parte del análisis y la reflexión para garantizar una implementación ética y responsable de la IA en los productos y servicios. Este sistema es una realidad del día a día en Telefónica, y alinea a toda la empresa en torno a los procesos y principios previamente definidos.

Telefónica utiliza los procesos ya existentes de desarrollo de productos y servicios, para integrar como una fase más del desarrollo, esta evaluación de los principios de IA Ética siempre que el producto o servicio incorpore inteligencia artificial o datos masivos.



Modelo de gobernanza: IA ética en el desarrollo de productos y servicios



Telefónica

Figura 25 - Modelo de gobernanza de IA ética

Este sistema, compartido y adoptado internamente, en conjunto con el resto de las herramientas y acciones formativas previamente explicadas, permite tener un enfoque estandarizado y validado que garantice la implementación eficaz del sistema en el día a día y que tenga en consideración la ética en todas las fases del desarrollo del producto, partiendo del diseño. Esta madurez de pensamiento y organización ha permitido a Telefónica y sus empleados desarrollar una experiencia muy valiosa, con aprendizajes interesantes que van a ser presentados a continuación.

6.5. Aprendizajes y recomendaciones de mejores prácticas

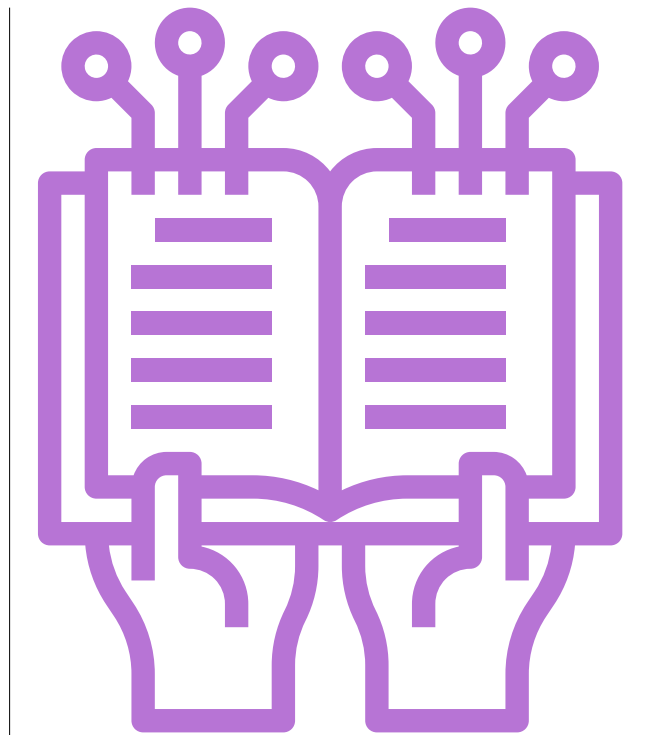
Si consideramos el camino de Telefónica para convertirse en una empresa de datos e inteligencia artificial desde 2011 (sección 1), así como el nivel de innovación y soluciones presentadas a lo largo de los años, es justo pensar que Telefónica ha desarrollado un conocimiento y una opinión formada en torno a las mejores prácticas para facilitar una adopción ética y responsable de la Inteligencia Artificial. Pocos ejemplos nacionales pueden ilustrar mejor este tipo de recorrido y sus aprendizajes.

6.5.1 Aprendizajes clave de Telefónica

Desde las áreas de Chief Data Office y Chief Responsibility Office nos comparten sus reflexiones finales:

- Los principios de IA son un excelente punto de partida para unificar el enfoque de innovación de inteligencia artificial responsable y ética en cualquier empresa. Cualquier compañía que comience su andadura en el mundo de los datos e inteligencia artificial debe adoptar unos principios, se recomienda crear un set propio de principios, en base a los objetivos de la empresa, contexto de negocio, valores y visión de la compañía, y el uso o aplicación específico de la inteligencia artificial.

- Definir y adoptar un modelo de gobernanza de IA responsable permite aterrizar los conceptos, y facilita la implementación en el día a día de la empresa. Introducir dicha gobernanza en los procesos oficiales de desarrollo de productos es clave, y una metodología de éxito probada durante los últimos años por Telefónica.
- Formar a los empleados es clave para garantizar la comprensión de los conceptos y del modelo. Se deben privilegiar formaciones híbridas que cubran aspectos técnicos y de negocio, y que sean accesibles a cualquier empleado de la empresa, más allá de su rol, como adopters potenciales de la inteligencia artificial y los datos masivos en sus productos y servicios.
- Crear la figura de los RAI Champions y sus comunidades de conocimiento permiten tener un punto de referencia para la adopción correcta de una inteligencia artificial responsable y ética. Dichos Champions son embajadores de la visión y los principios de Inteligencia Artificial, y suelen adoptar ese rol como parte de sus responsabilidades, sin que ello suponga una dedicación exclusiva. Las ganas de continuar aprendiendo y ayudando a los equipos de desarrollo y de negocio son la motivación clave para este tipo de profesionales



6.5.2 Beneficios de la IA ética y responsable

Desde Telefónica conocen la importancia que tienen tanto sus grupos de interés (clientes, proveedores, empleados y la sociedad en su conjunto), como sus accionistas e inversores. Estos son cada vez más conscientes de la importancia del desarrollo de una tecnología ética y responsable, así como de los crecientes requerimientos y exigencias a tener en cuenta.

Por todo lo anterior, Telefónica fue una de las primeras empresas en el mundo en definir un marco de principios éticos de Inteligencia Artificial, que permitiera, además de hacer las cosas bien, adelantarse a los posibles riesgos de cumplimiento, así como amenazas, que pudieran surgir, reforzando así especialmente sus sistemas de control y prevención de riesgos.

Compañías como Telefónica que cuenten con un sólido sistema de control de sus riesgos en materia de IA, tendrán un especial atractivo e interés para accionistas e inversores que desde hace años recalcan la importancia de los aspectos ESG (Environmental, Social and Governance) en la estrategia de las compañías.



07

CASO DE USO APLICADO PARA LA IMPLEMENTACIÓN DE IA ÉTICA EN LA INDUSTRIA

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

**Caso de uso aplicado
para la implementación
de IA ética en la industria**

Conclusiones
y consejos prácticos



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

**Caso de uso aplicado
para la implementación
de IA ética en la industria**

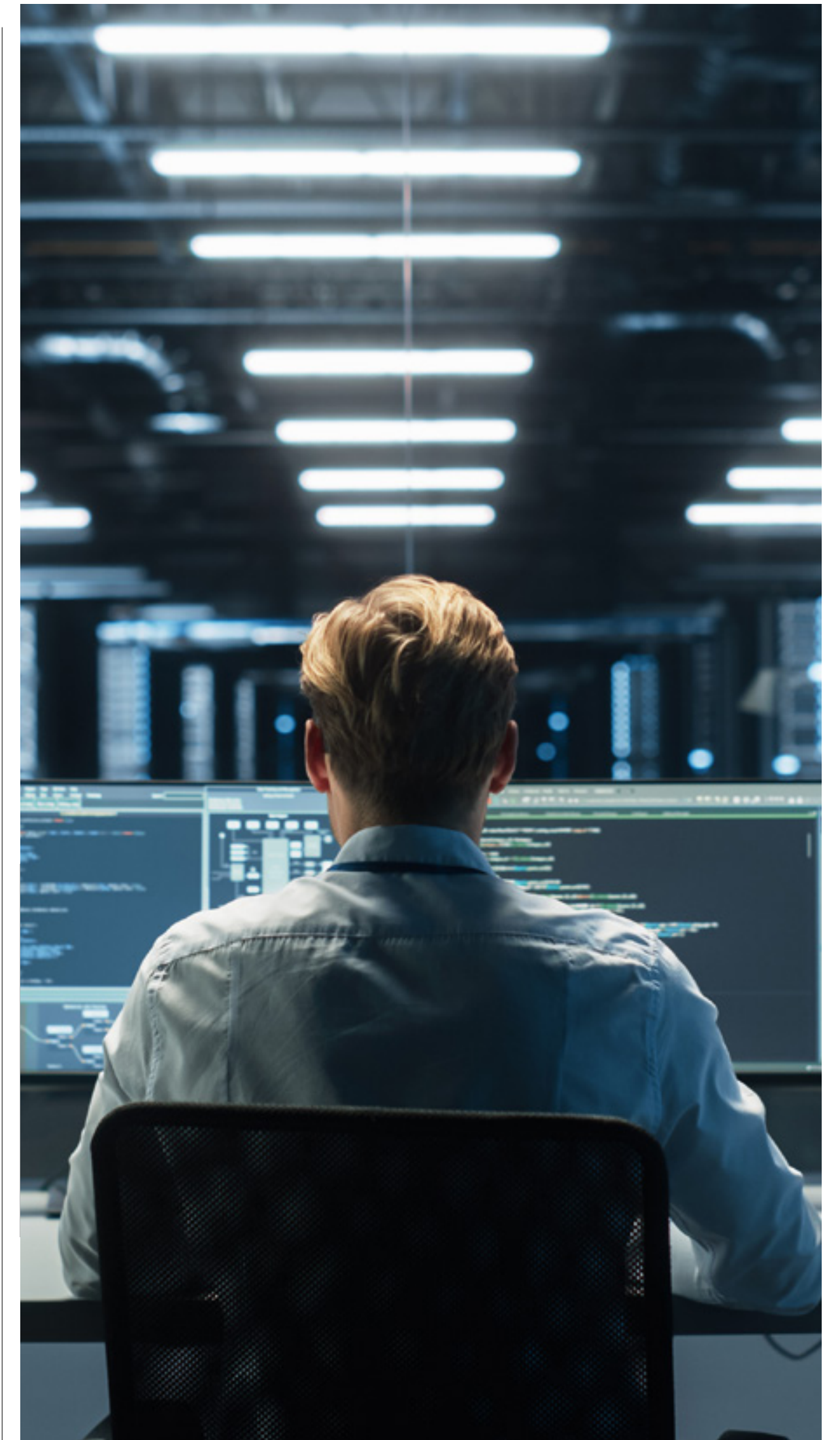
Conclusiones
y consejos prácticos



7.1. Contexto del caso y consideraciones iniciales

La disponibilidad de recursos y modelos de referencia no implica que la evaluación del riesgo potencial de un caso de uso IA específico sea sencillo. Como de costumbre, la complejidad está en los detalles, y la mejor manera de adoptar una IA ética y responsable es comenzar aplicando lo que sabemos a proyectos específicos.

Dentro del análisis de casos de uso IA de IndesIA en el capítulo 3, varios proyectos hacen referencia - directa o indirectamente - a sistemas de mantenimiento predictivo basados en tecnologías IA. Si tomamos este ejemplo como un caso ilustrativo, hay varios elementos a considerar:



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

**Caso de uso aplicado
para la implementación
de IA ética en la industria**

Conclusiones
y consejos prácticos

- **¿Qué tipo de equipamiento estamos analizando?** No es lo mismo tratar de predecir el mantenimiento para un elemento crítico (con consecuencias y peligros potenciales para las personas), que para otros elementos que - en el peor de los casos - no impactan la seguridad física de las mismas.
- **¿Qué tipo de datos necesitamos?** Varios casos de IndesIA incluyen inspección visual basada en fotos y videos (datos no estructurados), que requieren de visión artificial para detectar objetos y patrones tales como defectos, colores, etc. Este tipo de datos suelen obtenerse a través de cámaras fijas, o drones de inspección. Otra opción son los datos de sistemas IoT, que permiten capturar métricas en forma de datos relativamente estructurados (en forma de tabla simple, o ficheros de marcado JSON, etc.).
- **¿Cuál es la decisión automatizada que se está buscando?** Aunque esto está relacionado más con los fundamentos de IA que con la parte ética, es importante saber qué tipo de decisión IA estamos buscando. Por ejemplo, detección de objetos específicos, defectos, o patrones como podrían ser deterioro o corrosión en objetos metálicos. De manera complementaria, un enfoque de clasificación binaria (casos buenos y casos malos) también es factible. El componente importante para una IA ética es analizar el riesgo potencial de una mala decisión, es decir, el impacto de falsos positivos y falsos negativos, y cómo establecer un plan de mitigación para dichos casos.
- **¿Cómo de automatizado puede estar el proceso en el conjunto de actividades de la empresa?** Si consideramos la inspección y el

mantenimiento predictivo como algo muy importante a nivel de seguridad, o para reducir gastos por mantenimiento reactivo, es posible que no se quiera automatizar 100% el proceso, y se prefiera diseñar un sistema IA que genere recomendaciones y alertas que serán después interpretadas por una persona experta en la materia.

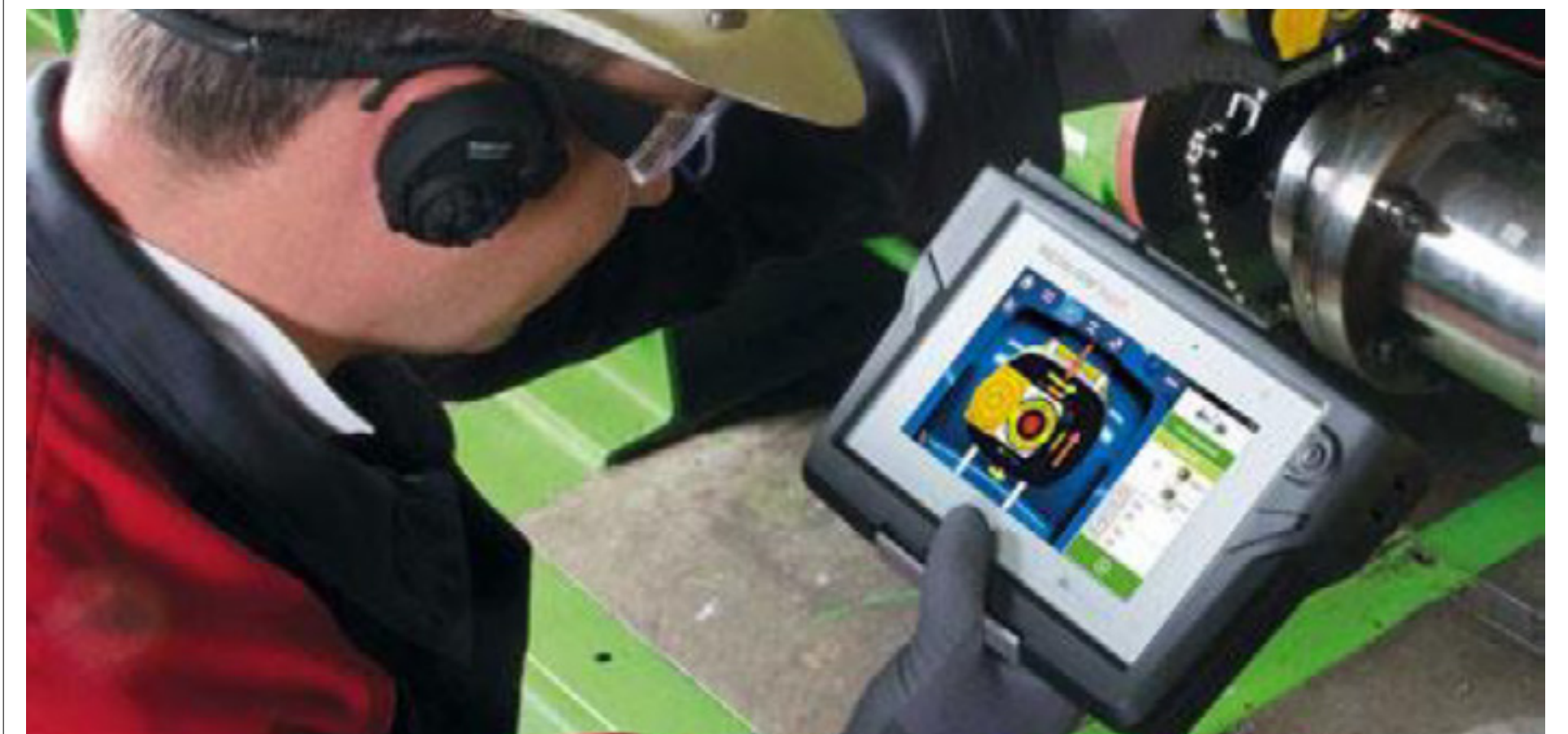


Figura 26 - Ejemplos de mantenimiento preventivo con personas y drones

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

7.2. Elementos de análisis

Esta lista contiene los principales factores de análisis para evaluar un caso de mantenimiento predictivo en el contexto del sector industrial español:



Principios aplicables

En este caso, hay varios principios relevantes (lo cual no implica que el resto de principios no sean importantes). Concretamente:

Seguridad / safety

Dependiendo del tipo de activo que se analice, este componente es especialmente relevante. Si hacemos mantenimiento predictivo para un sistema crítico, debemos garantizar "safety" y evitar cualquier accidente fatal directa o indirectamente relacionado con el sistema IA.

Rendición de cuentas

Si asumimos la criticidad previamente descrita, la rendición de cuentas cobrará especial relevancia. ¿Qué sucede si se produce un accidente y se prueba que la causa raíz está relacionada al sistema? ¿Quién toma la responsabilidad a nivel corporativo o

individual? Este tipo de cuestiones no tienen una respuesta única, pero el sistema de gobernanza interno debe servir para definir un plan de acción en los casos.

Supervision Humana

Partiendo de la base de que ningún sistema IA es perfecto, y las predicciones suelen incluir un % de error más o menos alto dependiendo del caso, es importante definir de manera explícita el grado de autonomía del sistema de IA. En caso de que se trate de un componente de seguridad (safety, alto riesgo), la supervisión humana es importante - HITL (human-in-the-loop) para reducir el riesgo; todas las decisiones pasan por una persona. En caso de menos criticidad, una autonomía de monitorización puede ser adecuada - HOTL (human-on-the-loop); la máquina ejecuta las decisiones, pero están supervisadas por una

persona que puede intervenir y cambiar la decisión.

Explicabilidad

También en casos de alto riesgo, la predicción del sistema debería estar acompañada de información que explique por qué el sistema IA ha llegado a su conclusión.



Prólogo

Introducción

Estado de la cuestión de la ética en IA

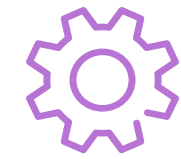
Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos



Medida de impacto potencial

Dicho impacto puede ser evaluado a través de la herramienta cuestionario interactivo mencionada en la sección 4.2.

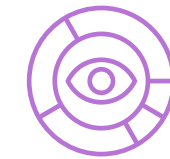


Beneficio vs. riesgo

Como en cualquier otro sistema, una evaluación objetivo del riesgo potencial en comparación del beneficio para la empresa es necesario. ¿Por qué hacemos esto? ¿Tiene sentido desde un punto de vista operacional e incluso económico? Si la respuesta es positiva, se puede cuantificar:

- XX% reducción del gasto por intervención manual
- YY% aumento de la precisión en la toma de decisión en comparación con el proceso actual

Esta información se pretende captar en la primera parte del cuestionario mencionado en la sección 4.2.



Consideraciones de supervisión humana

En este caso, si asumimos que un humano debe supervisar los resultados del sistema IA, hay que definir:

Tipos de información, notificaciones, y alertas a compartir con el humano en función del tipo de decisión y niveles de confianza del modelo IA.

Tipo de interfaz para la colaboración entre el sistema IA y el humano. En caso de operaciones de mantenimiento en el terreno, integrar los datos relevantes en una tablet o teléfono inteligente pueden facilitar la labor del operario.

El factor humano como “persona” de la solución. Es decir, el usuario potencial de dicha solución - el operador - puede y debe estar involucrado en el proceso de diseño. Cómo utilizar la solución IA, qué tipo de información necesita, por qué las alertas y predicciones son importantes y la guía fundamental para detectar casos de mantenimiento prioritarios, etc.

Adicionalmente, la evaluación del operador para aceptar o rechazar las recomendaciones del sistema IA pueden ser trazadas y utilizadas posteriormente para re-educar al sistema IA (como una fuente de etiquetado para nuevas iteraciones de aprendizaje supervisado).



Regulaciones aplicables y auditoría

Tal como hemos visto en la sección 2.5, las regulaciones a nivel de IA son un work in progress a nivel español y europeo, pero los pilotos ya han comenzado y se espera que el periodo 2022-23 sea clave para la consolidación del AI Act. Otras regulaciones específicas asociadas al subsector en concreto deben ser analizadas. Por ejemplo, si hablamos de inspección de barcos, el departamento legal de la empresa podrá verificar otras regulaciones de seguridad y su cumplimiento, más allá del enfoque puramente centrado en la IA.

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

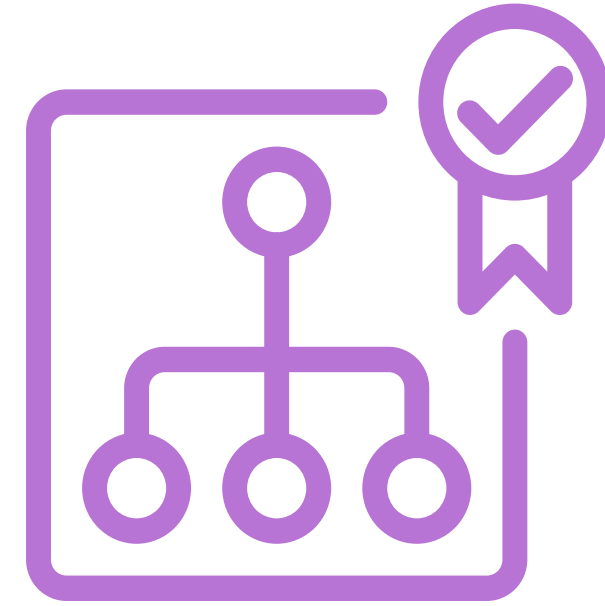
Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

**Caso de uso aplicado
para la implementación
de IA ética en la industria**

Conclusiones
y consejos prácticos



7.3. Relación con la gobernanza IA

Tal como hemos visto en el caso de Telefónica (capítulo 5), el sistema de gobernanza incluye las personas y los procesos necesarios para analizar las consideraciones anteriores a tiempo, y en la secuencia correcta.

Para el caso de mantenimiento predictivo basado en tecnologías IA, algunos pasos que podrían facilitar la tarea son:

- **Antes de implementar la solución**

No es lo mismo tratar de predecir el mantenimiento para un elemento crítico (con consecuencias y peligros potenciales para las personas), que para otros elementos que - en el peor de los casos - no impactan la seguridad física de las mismas.

- **Durante la implementación del proyecto**

El equipo IA puede añadir herramientas dedicadas para la evaluación de los principios de IA aplicables al mantenimiento preventivo (aquellos explicados en la sección 6.2)

- **Durante y al finalizar la implementación IA**

Se pueden definir los sistemas de evaluación del rendimiento de la solución IA, así como los planes de mitigación si el resultado del sistema IA provoca una situación crítica.

- La evaluación continua y las auditorías (voluntarias y obligatorias) cobrarán especial relevancia una vez que la regulación AI Act europea esté vigente.

En este caso, la designación de un RAI Champion a nivel de unidad de negocio o proyecto que pueda coordinar y facilitar la discusión ética con el resto de grupos involucrados va a ayudar a dar la importancia y prioridad a este tipo de análisis.

Si imaginamos este caso como el primer caso de IA en una empresa, tomarlo como un ejemplo piloto para implementar la dinámica de IA ética y responsable permitirá replicar los pasos y marcos de evaluación establecidos para otros proyectos.



08

CONCLUSIONES

Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

**Conclusiones
y consejos prácticos**



Tras lo analizado en el presente documento, se puede concluir que si bien las discusiones éticas en torno a la inteligencia artificial están presentes en varios ámbitos y estratos de la sociedad, su aplicación al sector industrial es un “work in progress” para la mayor parte de las empresas nacionales e internacionales.

Las iniciativas que sostienen dichas discusiones ya sean públicas o privadas, ayudan no sólo a la reflexión e implementación de sistemas que garantizan una aplicación responsable de la inteligencia artificial en las empresas, con base en modelos de referencia, casos de éxito, y transferencia de conocimiento. En ese sentido, la misión de IndesIA contribuye a tal fin y este documento no es más que un primer paso para ayudar a las empresas españolas del sector industrial a seguir madurando su adopción y modelos de gobernanza IA.

La visión de IndesIA y OdiselA: La ética en IA comienza una fase tangible en la que las empresas pueden adoptar de manera proactiva principios éticos y enfoques responsables. La ventana de oportunidad es maravillosa, antes de la puesta en marcha de las normativas de IA, e IndesIA y OdiselA proponen una serie de puntos que pueden facilitar las iniciativas éticas en IA de las empresas españolas, los cuales pueden y deben ser adaptados al contexto específico de cada organización.

Nuestros consejos prácticos para aquellas empresas que quieran implementar una IA con enfoque ético y responsable en el sector industrial, en 6 simples puntos:



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

**Conclusiones
y consejos prácticos**

- **Tomar este documento de IndesIA y OdiselA como punto de partida.** El toolkit del capítulo 4 es material de base para la discusión interna en las empresas. La manera de adaptarlo al contexto interno dependerá de factores tales como la madurez general de IA, el nivel de transformación digital, la naturaleza de las soluciones desarrolladas, etc.
- **Incentivar la discusión interna para elegir la lista de principios éticos de la empresa.** La elección o adopción de principios éticos tiene una componente muy específica a la cultura y los valores de cada empresa. Sin embargo, en 2022 se comienza a ver una cierta convergencia en cuanto a tipos de principios, que son llamados de diferentes maneras pero que suelen referirse a cuestiones similares (transparencia, explicabilidad, etc.). Por eso mismo, recomendamos la lista de principios éticos compartida en el capítulo 3.1.
- **Utilizar la herramienta cuestionario para entender los niveles de riesgo ligados a las actividades IA de la empresa, así como el impacto para los principios previamente seleccionados.**
 - Especial atención merece la distinción entre casos de uso IA orientados a personas o a máquinas/procesos, ya que el impacto en términos de importancia por cada uno de los principios va a variar en función de ello. Por ejemplo, aquellos que afecten a personas suelen centrarse en temas de transparencia y explicabilidad, mientras que los de máquinas/procesos tienen un impacto en safety (seguridad física).



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

**Conclusiones
y consejos prácticos**

- **Elegir a las personas y departamentos involucrados en la iniciativa ética.** El modelo de gobernanza IA, tanto en los aspectos de roles involucrados como en la dinámica de escalado de casos y riesgos, dependerá mayormente de la estructura humana y consideraciones tales como la jerarquía y los grupos definidos dentro de la empresa. Es decir, los puntos elaborados en el capítulo 4.3 deberán ser adaptados al potencial realista.

- Independientemente de la estructura humana de la empresa, que puede tener o no un departamento dedicado para datos e IA, es importante encontrar y designar a aquellos “RAI champions” (como en el caso de Telefónica) que puedan liderar los primeros pasos de la gobernanza.



- **Planear recursos dedicados para la comprensión e implementación de los marcos regulatorios y de estandarización** (por ejemplo, AI Act o las nuevas normas ISO, de la misma manera que se hizo en su día con GDPR). A la hora de analizar el nivel de riesgo de un sistema IA, además de las discusiones multidisciplinares entre las partes involucradas en la empresa, los niveles de riesgo del AI Act (explicados en el capítulo 2.5) son la referencia de cara al cumplimiento de la futura regulación europea. Adicionalmente, y tal como hemos visto en el caso de uso del capítulo 6 para mantenimiento predictivo, se puede analizar la criticidad del proyecto IA más allá de un nivel específico del AI Act. Por ejemplo, muchos de los casos industriales están en el mismo nivel de riesgo alto debido a la componente de seguridad “safety”, pero la criticidad de este depende de la importancia de los componentes industriales en sí. Es decir, no es lo mismo - a nivel de riesgo para las personas - un mantenimiento predictivo de una vía de tren, que de un equipo secundario para la limpieza de un barco.

Prólogo

Introducción

Estado de la cuestión de la ética en IA

Análisis de impacto ético y social en la industria española

Toolkit de implementación de una IA ética en el sector industrial

Caso de éxito de gobernanza de IA en empresa

Caso de uso aplicado para la implementación de IA ética en la industria

Conclusiones y consejos prácticos

- **Diseñar y facilitar un programa de formación específico a las IA y a su componente ética.** El rol de la educación en ética IA (y la IA en general) es fundamental para desarrollar conocimiento y concienciación dentro de las empresas. No hay mucha formación adaptada al contexto industrial, pero los ejemplos de formaciones propuestas en el capítulo 4.4 son un buen punto de partida para la mayoría de las empresas que adopten IA. Estos recursos pueden ser complementarios a otros desarrollados en el futuro por asociaciones como IndesIA.

- En el caso de empresas muy noveles en el ámbito de datos e IA, este punto de formación debe ser el primero, para que las personas y departamentos involucrados puedan formarse antes de comenzar las discusiones. Para otros casos, las actividades de formación pueden hacerse en paralelo con el resto de los puntos mencionados.

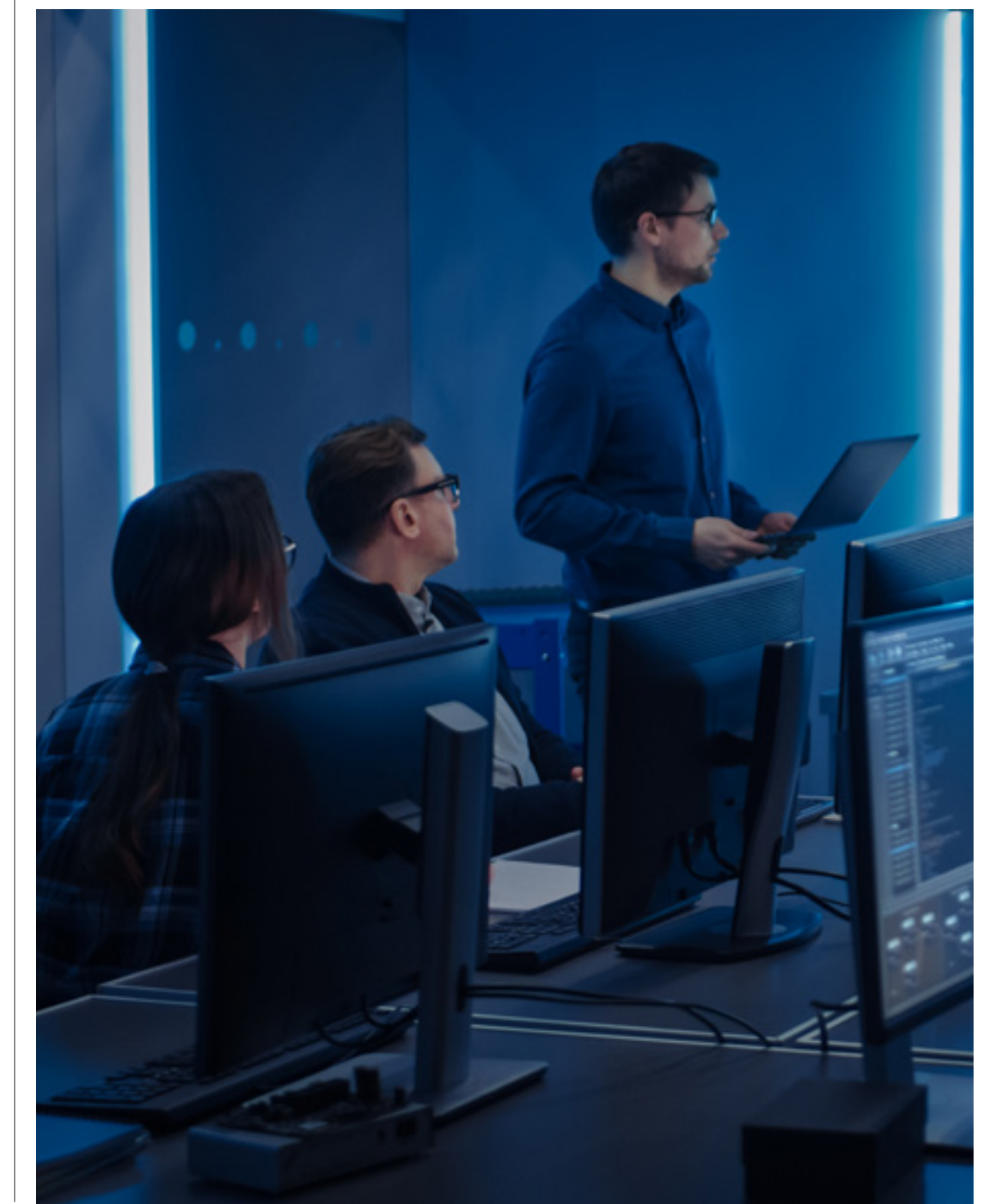
- **Establecer un entorno de pruebas** con los equipos de datos e IA para que puedan probar, evaluar, y seleccionar herramientas de IA responsable. La elección de dichas herramientas técnicas depende del

conocimiento previo y la proactividad de los equipos técnicos y del liderazgo. La lista propuesta en el capítulo 4.5 contiene herramientas de valor probado, pero este ecosistema está evolucionando cada año. Es responsabilidad de cada empresa y equipo de IA seguir explorando soluciones a incorporar dentro de sus "MLOps".

- En el caso de los proyectos IA, suele ser una buena práctica involucrar a los perfiles técnicos de los equipos (científicos, ingenieros, etc.) para que sean ellos los que exploren, prueben, y elijan las herramientas. Esto facilita la adopción a largo plazo, y la lista propuesta incluye varias herramientas gratuitas y de código abierto, lo cual puede facilitar la fase de exploración sin la necesidad de invertir en software.

Como se puede ver, no hay una fórmula mágica pero sí unos building blocks que pueden servir para comenzar el viaje hacia una IA responsable en la empresa. La misión - importante y apasionante a partes iguales - recién comienza. Ya no sólo se trata de sacar provecho del potencial de la inteligencia artificial, sino de hacerlo de una manera justa, ética y responsable.

Además, factores como la creciente inversión en IA a nivel nacional, el componente cultural (con una preocupación generalizada por las causas "justas"), o el enfoque ético de la Estrategia Nacional de Inteligencia Artificial de la SEDIA, hacen que los próximos años sean decisivos para las empresas españolas.



Prólogo

Introducción

Estado de la cuestión
de la ética en IA

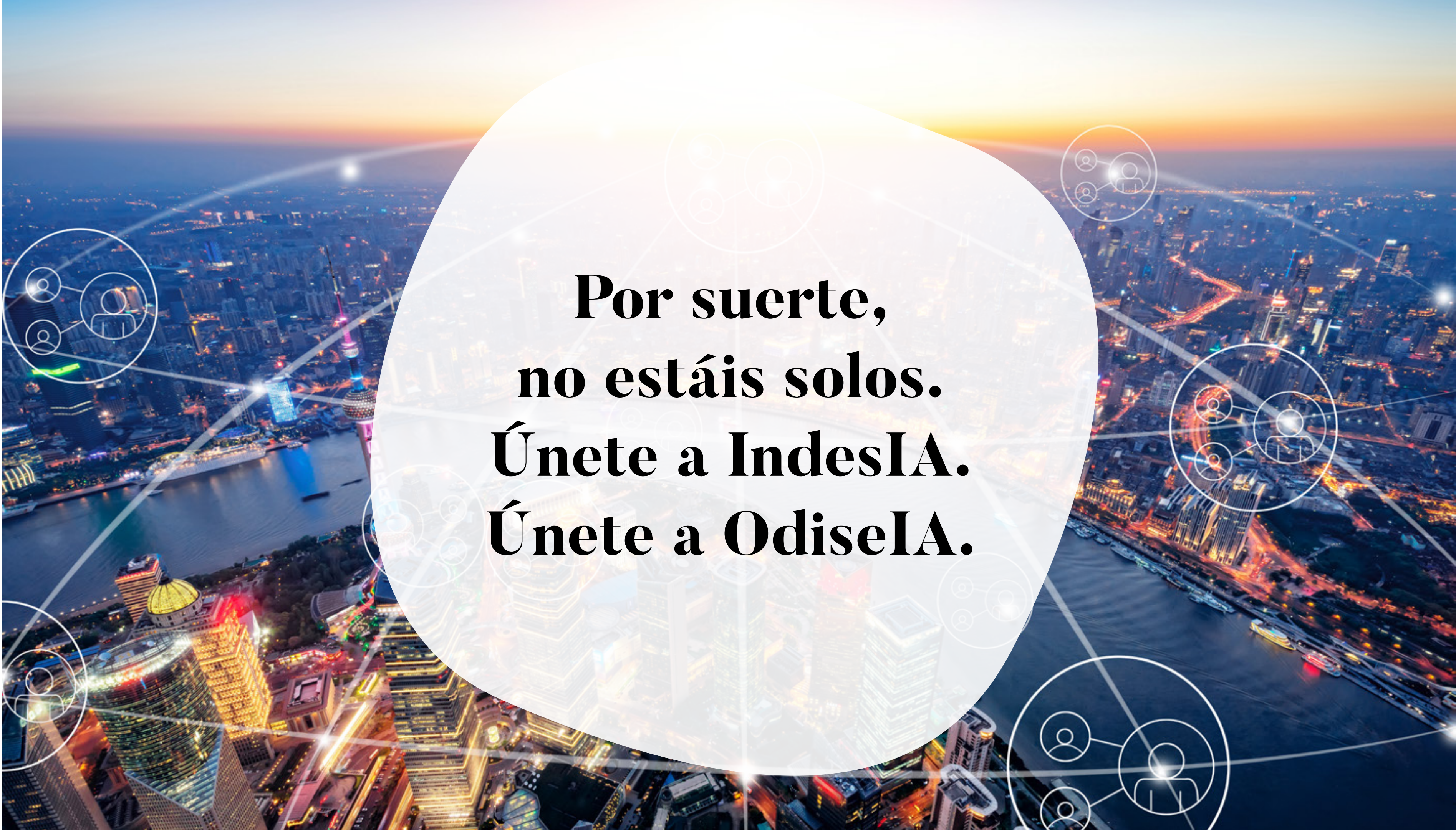
Análisis de impacto ético
y social en la industria
española

Toolkit de implementación de
una IA ética en el sector
industrial

Caso de éxito de gobernanza
de IA en empresa

Caso de uso aplicado para la
implementación de IA ética en la
industria

**Conclusiones
y consejos prácticos**



**Por suerte,
no estáis solos.
Únete a IndesIA.
Únete a OdiseIA.**

INDESIA

