

Modelo para la EIDF: guía y casos de uso

Metodología aplicada de la evaluación de impacto sobre los derechos fundamentales en el diseño y desarrollo de la IA

ÍNDICE

Presentación	4
Parte I – La EIDF y la metodología EIDF	6
1. Introducción	6
2. El papel de la EIDF en la regulación de la IA	7
3. La interacción con la normativa de protección de datos	9
4. El modelo de EIDF adoptado	9
4.1 La fase de planificación y determinación del alcance	11
4.2 Fase de recopilación de datos y análisis de riesgos	11
4.2.1 Variables clave para la evaluación de impacto	12
4.2.2 Variables y construcción del índice de impacto	14
4.3 La fase de gestión de riesgos	15
5. La plantilla del modelo EIDF	16
Cuestionario de planificación y alcance	16
Matrices de riesgo	18
Tabla 1. Probabilidad	18
Tabla 2. Exposición	18
Tabla 3. Probabilidad de que se produzca un perjuicio (<i>likelihood</i>)	18
Tabla 4. Gravedad del perjuicio	19
Tabla 5. Esfuerzo para superar los perjuicios y revertir los efectos adversos	19
Tabla 6. Gravedad (<i>severity</i>)	19
Tabla 1A. Recogida de datos y análisis de riesgos	20
Tabla 7. Impacto general del riesgo	20
Tabla 2A. Gestión de riesgos (I)	20
Tabla 3A. Gestión de riesgos (II)	21
6. Del modelo a la práctica: los casos de uso	21
Referencias	23
Part II – Casos de uso	24
Caso de uso 1: Plataforma de analítica avanzada del aprendizaje	24
1. El contexto	24
2. El proyecto	27
3. La EIDF	30
4. Comentarios	45
Caso de uso 2: Una herramienta para la gestión de los recursos humanos	47
1. El contexto	47
2. El proyecto	48
3. La EIDF	49
4. Comentarios	61
Caso de uso 3: Una herramienta de imágenes médicas impulsada por IA para la detección del cáncer	62
1. El contexto	62
2. El proyecto	62
3. La EIDF	63
4. Comentarios	75
Caso de uso 4: ATENEA, IA al servicio de las personas mayores	76
1. El contexto	76
2. El proyecto	76
3. La EIDF	78
4. Comentarios	100

Coordinadores: Alessandro Mantelero y Joana Marí

Autores:

Alessandro Mantelero (Universidad Politécnica de Turín)
Cristina Guzmán (Universidad Politécnica de Cataluña – BarcelonaTech (UPC))
Esther Garcia (CaixaBank, S.A)
Ruben Ortiz (Universidad de Barcelona)
M. Ascensión Moro (Ayuntamiento de Sant Feliu de Llobregat)

Participantes:

Albert Portugal (Consortio de Servicios Universitarios de Cataluña)
Albert Serra (Autoridad Catalana de Protección de Datos)
Alessandro Mantelero (Universidad Politécnica de Turín)
Cristina Guzmán (Universidad Politécnica de Cataluña – BarcelonaTech (UPC))
Esther Garcia (CaixaBank, S.A)
Joana Judas (Departamento de Investigación y Universidades. Generalidad de Cataluña)
Joana Marí (Autoridad Catalana de Protección de Datos)
Jordi Escolar (Agencia de Calidad Universitaria)
M. Ascensión Moro (Ayuntamiento de Sant Feliu de Llobregat)
Marc Vives (Universidad Pompeu Fabra)
María José Campo (Fundación TIC Salud y Social)
Mariona Perramon (Consortio de Servicios Universitarios de Cataluña)
Olga Rierola (Autoridad Catalana de Protección de Datos)
Patricia Lozano (Universidad Abierta de Cataluña)
Ruben Ortiz (Universidad de Barcelona)
Sara Hernández (Fundación TIC Salud y Social)



Barcelona, 2025

El contenido de este documento es titularidad de la Autoridad Catalana de Protección de Datos y está sujeto a la licencia Creative Commons BY-NC-ND, <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

Se reconocerá la autoría de la obra mediante la inclusión de la siguiente mención:

Obra titularidad de la Autoridad Catalana de Protección de Datos.

Licenciado bajo la licencia CC BY-NC-ND.

Aviso: Al reutilizar o distribuir la obra, los términos de la licencia para esta obra deben estar claramente establecidos.

Aviso legal: Las opiniones expresadas en este documento son responsabilidad de sus autores y no reflejan necesariamente la opinión oficial de la APDCAT. La Autoridad Catalana de Protección de Datos, los autores y los miembros del grupo de trabajo no se hacen responsables de las posibles consecuencias causadas a las personas físicas o jurídicas como consecuencia de cualquier información contenida en este documento.

Presentación

En su Plan estratégico 2023-2028, la Autoridad Catalana de Protección de Datos (APDCAT) se fijó como pilar estratégico el objetivo de promover el desarrollo de formación dirigida a las instituciones. Y una de las líneas de acción clave para conseguirlo era "impulsar la creación y potenciar la comunidad de aprendizaje de delegados de protección de datos".

Así pues, el 15 de julio de 2023 arrancó la primera Red de delegados y delegadas de protección de datos (disponible en el enlace dpdenxarxa.cat), pionera en Cataluña y en el Estado español. La plataforma nació con el objetivo de contribuir a extender la cultura de la privacidad en Cataluña, de la mano de los delegados/delegadas de protección de datos (DPD), que velan por garantizar el cumplimiento de la normativa de protección de datos, promover la cooperación y colaboración entre DPD y compartir conocimientos y pericia.

Esta plataforma aglutina a los DPD de las más de 1.700 entidades que forman parte del ámbito de actuación de la APDCAT, que comprende administraciones públicas como la Generalidad de Cataluña, ayuntamientos, universidades públicas y privadas y colegios profesionales, entre otros. También está abierta a los DPD de las entidades públicas y privadas que prestan servicios en el sector público catalán como encargadas del tratamiento, así como a los DPD en activo de organizaciones con sede en Cataluña.

La Red, que actualmente cuenta con un elevado número de DPD de Cataluña, persigue los siguientes objetivos principales:

- Ser una institución más cercana a los DPD y a las organizaciones.
- Constituir un espacio de intercambio de ideas, experiencias y conocimientos.
- Potenciar la figura del DPD, como pieza clave del cumplimiento.
- Identificar y fomentar las mejores prácticas.
- Crear y difundir un modelo de relación y cooperación, compatible con una supervisión útil y eficaz.
- Construir un entorno de servicios de interacción y colaboración abierto de generación de recursos, de aprendizaje y gestión del conocimiento.
- Crear un espacio referente en Cataluña y Europa para los DPD.

Fruto de la creación de este espacio, desde el apartado Ágora de la Red se propuso la creación de un grupo de trabajo llamado "Orientaciones metodológicas. Evaluaciones de impacto. Derechos y libertades". Este grupo fue coordinado por el doctor Alessandro Mantelero y la Sra. Joana Marí, durante los meses de febrero a diciembre del año 2024, y está formado por los siguientes miembros de la Red:

- Albert Portugal (Consortio de Servicios Universitarios de Cataluña)
- Albert Serra (Autoridad Catalana de Protección de Datos)
- Cristina Guzmán (Universidad Politécnica de Cataluña – BarcelonaTech)
- Esther García (Caixabank, S.A.)
- Joana Judas (Departamento de Investigación y Universidades. Generalidad de Cataluña)
- Jordi Escolar (Agencia de Calidad Universitaria)
- M. Ascensión Moro (Ayuntamiento de Sant Feliu de Llobregat)
- Marc Vives (Universidad Pompeu Fabra)
- María José Campo (Fundación TIC Salud y Social)
- Mariona Perramon (Consortio de Servicios Universitarios de Cataluña)

- Olga Rierola (Autoridad Catalana de Protección de Datos)
- Patricia Lozano (Universidad Abierta de Cataluña)
- Ruben Ortiz (Universidad de Barcelona)
- Sara Hernández (Fundación TIC Salud y Social)

El objetivo de este equipo de trabajo era desarrollar una nueva metodología para las evaluaciones de impacto en los derechos fundamentales, cuando se utilizan sistemas de inteligencia artificial, teniendo en cuenta el marco normativo europeo establecido por el Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por lo que se modifican los Reglamentos (CE) nº 300/2008, (UE) nº 167/2013, (UE) nº 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE) 2016/797 y (UE) 2020/1828, Reglamento de Inteligencia Artificial (en adelante, RIA). Esta metodología debía diferenciarse de las evaluaciones de impacto en materia de protección de datos que exige el Reglamento (UE) núm. 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas con respecto al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46 / CE (Reglamento general de protección de datos).

Fruto del trabajo de este grupo se ha elaborado el documento que presentamos, que se estructura en dos partes: en la primera se describe la metodología y en la segunda se aplica a casos concretos. Con este trabajo, se pretende facilitar una herramienta útil y práctica para las entidades que diseñan, desarrollan o utilizan sistemas o modelos de inteligencia artificial y, en particular, para las personas encargadas de realizar las evaluaciones de impacto sobre los derechos fundamentales.

En definitiva, la prevención de la vulneración de los derechos y libertades fundamentales es una tarea común donde la Autoridad Catalana de Protección de Datos debe jugar un papel clave.



Meritxell Borràs i Solé
Directora
Autoridad Catalana de Protección de Datos

Parte I – La EIDF y la metodología EIDF

1. Introducción

Esta publicación presenta los resultados de un proyecto liderado por la Autoridad Catalana de Protección de Datos (APDCAT), con el objetivo de proporcionar a los operadores de IA, tanto proveedores como responsables de despliegue, una herramienta eficaz para desarrollar soluciones de IA fiables y centradas en el ser humano. Desde este punto de vista, el diseño y desarrollo de soluciones de IA que respeten los derechos fundamentales/humanos es el núcleo de una IA verdaderamente centrada en el ser humano,¹ como demuestran el Reglamento de IA (RIA) y otras iniciativas nacionales e internacionales, como el Convenio Marco del Consejo de Europa sobre Inteligencia Artificial, Derechos Humanos, Democracia y Estado de Derecho.

A la luz de lo anterior, el RIA establece como uno de sus principales objetivos garantizar "un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta" (art. 1 RIA) y, en línea con el enfoque basado en el riesgo adoptado por el legislador de la UE, la evaluación del impacto de la IA en los derechos fundamentales² se incluye en todos los procedimientos de gestión de riesgos establecidos por el Reglamento: desde la evaluación de la conformidad hasta la evaluación de impacto específica sobre los derechos fundamentales con arreglo al artículo 27 del RIA, incluyendo una disposición específica para los modelos de IA de uso general con riesgo sistémico, **el impacto sobre los derechos fundamentales debe tenerse en cuenta en el diseño, el desarrollo y el despliegue de los sistemas y modelos de IA.**

En este contexto, **las disposiciones sobre cómo llevar a cabo esta evaluación en el RIA, incluso en el Convenio Marco del Consejo de Europa, solo proporcionan una orientación limitada a quienes deben llevar a cabo esta evaluación.** Por otro lado, los modelos propuestos y el debate inicial en la literatura muestran varias carencias desde el punto de vista metodológico [Mantelero, 2024]. Por lo tanto, en consonancia con un enfoque empírico del Derecho, con el fin de probar la viabilidad y la eficacia de los modelos para llevar a cabo la evaluación de impacto sobre los derechos fundamentales en el contexto de la IA **es necesario pasar de la elaboración abstracta, en el Derecho y en el debate jurídico, a aplicaciones concretas.**

El proyecto catalán es la primera iniciativa basada en la implementación concreta de una metodología EIDF, aplicada a casos reales y fundamentada en una interacción activa con entidades públicas y privadas que aplican soluciones de IA en sus negocios y actividades. **Los resultados de este enfoque empírico son cruciales para la implementación efectiva del RIA,** ya que demuestran que es posible racionalizar los requisitos del Reglamento y traducirlos eficazmente en un proceso de análisis y gestión de riesgos, que sea coherente tanto con la teoría general del riesgo como con el marco de los derechos fundamentales.

Además, **la evidencia empírica proporcionada por esta publicación puede contribuir al debate internacional y de la UE** sobre el modelo para la evaluación de impacto sobre los derechos fundamentales/humanos, proporcionando evidencia sobre cuestiones cruciales como: (i) las variables relevantes que deben considerarse; (ii) la metodología para evaluarlas y crear índices de riesgo; (iii) el rol que pueden jugar los cuestionarios estándar en este ejercicio y sus limitaciones; y iv) el papel de la expert-based evaluation en este análisis.

¹ Véase también la Agencia de los Derechos Fundamentales de la Unión Europea, <https://fra.europa.eu/en/content/what-are-fundamental-rights> ("the term fundamental rights is used in a constitutional context whereas the term 'human rights' is used in international law. The two terms refer largely to the same substance as can be seen, for instance, by the many similarities between the Charter of Fundamental Rights of the EU and the European Convention on Human Rights.").

² Aquí y en las páginas siguientes, las referencias a los derechos fundamentales (o simplemente derechos) incluyen tanto los derechos fundamentales como las libertades consagradas en la Carta de los Derechos Fundamentales de la Unión Europea.

El modelo EIDF utilizado en nuestros casos de uso (Mantelero, 2024; Mantelero-Exposito, 2021), así como los propios casos de uso, se ponen a disposición del público con el fin de contribuir al debate internacional y de la UE sobre la protección de los derechos fundamentales en el contexto de la IA. También, **para servir de fuente de inspiración para las numerosas entidades de toda Europa y de países no pertenecientes a la UE que desean adoptar un enfoque de la IA** basado en los derechos fundamentales, pero no disponen de un modelo de referencia probado ni de casos concretos con los que comparar sus experiencias.

Con este objetivo en mente, en las siguientes secciones se discutirá brevemente el papel de la EIDF en la regulación de la IA, su interacción con la regulación de protección de datos, el modelo utilizado en los casos de uso, los criterios de selección de casos y las áreas cubiertas.

2. El papel de la EIDF en la regulación de la IA

La última primavera de la IA, en los últimos años, ha llevado a que las aplicaciones de aprendizaje automático se utilicen en una variedad de escenarios operativos tanto en el sector público como en el privado. Uno de los principales usos de la IA y fuente de mayores desafíos se relaciona con el papel que juegan estas aplicaciones en la toma de decisiones. Este aspecto es aún más crítico en el sector público dado el desequilibrio de poder que en muchos casos caracteriza la relación entre las personas físicas y los poderes públicos.

Este cambio de las decisiones basadas en humanos a las decisiones basadas en IA, ya sea totalmente automatizadas o con la IA apoyando a los responsables en la toma de decisiones, plantea varias preocupaciones sobre la precisión de la lógica de dichos sistemas de IA, la forma en que representan a la sociedad y la interacción humano-máquina. Si bien los desarrollos recientes en IA están aportando mejoras significativas en muchos sectores, es importante conocer y gestionar los posibles efectos secundarios de esta tecnología.³

Los riesgos asociados a las tecnologías de IA pueden tener un impacto negativo en la sociedad debido a **cuestiones de seguridad y derechos fundamentales**. Si bien los problemas de seguridad, aunque complicados y relacionados con las limitaciones inherentes a los modelos actuales de IA (por ejemplo, las llamadas alucinaciones), pueden abordarse fácilmente desde una perspectiva regulatoria utilizando prácticas y herramientas establecidas (por ejemplo, estandarización, procedimientos de certificación, auditoría, etc.), resulta mucho más complicado abordar los problemas de derechos fundamentales.

Los componentes intrínsecos, extrínsecos y contextuales de un sistema de IA pueden variar con respecto al entorno sociotécnico en el que se utiliza. **La forma en que se ha diseñado la IA** (elementos intrínsecos), en términos de conjuntos de datos de entrenamiento, ajustes, etc., puede afectar la manera en que representa los aspectos sociales relevantes para la toma de decisiones con riesgos potenciales, por ejemplo, en términos de discriminación en grupos subrepresentados, respecto a la elegibilidad, en los sistemas de IA utilizados en ayudas sociales. **La forma en que los sistemas de IA interactúan con otras tecnologías** (elementos extrínsecos), en función de la naturaleza y el rendimiento de estas últimas, puede limitar la eficacia del sistema de IA y, por tanto, afectar negativamente a los derechos conexos, como en el caso de una herramienta de imagen médica impulsada por IA que detecta mal el cáncer debido a la baja calidad del dispositivo médico utilizado para generar las imágenes, poniendo en peligro el derecho a la salud.

³ Para una categorización de los riesgos relacionados con la IA, véase también Naciones Unidas, AI Advisory Board 2024; Slattery et al. 2024.

Por último, unos sistemas que funcionen correctamente y estén bien integrados con otros dispositivos, pueden tener diferentes repercusiones en los derechos fundamentales en función del **contexto de uso**. El mismo sistema de videovigilancia basado en IA puede considerarse una medida proporcionada, – a pesar de su impacto en la privacidad individual y colectiva y la protección de datos, así como en la libertad de reunión en el caso de manifestaciones y protestas, – en presencia de altos índices de criminalidad en algunas zonas, e inapropiada en ausencia de este interés público preponderante.

A partir de este breve análisis de la interacción entre el desarrollo y el despliegue de la IA y los derechos fundamentales, queda clara *la raison d'être* de la EIDF en la regulación de la IA, así como los elementos básicos que deben tenerse en cuenta para llevar a cabo una evaluación adecuada del impacto potencial de la IA en los derechos. Las dimensiones intrínseca y extrínseca de los sistemas de IA, el contexto de uso, los derechos potencialmente afectados, la necesidad de un test de ponderación, los individuos y grupos afectados, las posibles medidas de prevención/mitigación, son las diferentes áreas de la EIDF a desarrollar para hacer frente a las cuestiones mencionadas.

En este sentido, el artículo 27.1 del RIA sobre EIDF considera (i) el contexto de uso y las categorías de actores expuestos al riesgo (art. 27.1.a), (b) y (c)), (ii) el posible perjuicio a los derechos fundamentales (art. 27.1.d)), y (iii) las medidas de prevención/mitigación que deben adoptarse (art. 27.i.e) y (f)). Esta división se refleja en las **tres fases de la metodología EIDF** adoptada en este estudio: (i) planificación, definición del alcance e identificación de riesgos; (ii) análisis de riesgos; y (iii) mitigación y gestión de riesgos.

La primera fase (planificación, definición del alcance e identificación de riesgos) incluye la descripción del sistema de IA y su contexto de uso, dados los riesgos intrínsecos (relacionados con el propio sistema) y los riesgos extrínsecos (relacionados con la interacción entre el sistema y el entorno sociotécnico en el que se implementa). Esto incluye una descripción del proceso en el que se utilizará el sistema de IA de alto riesgo (art. 27 (1) (a)), así como el período de tiempo en el que se utilizará cada sistema y la frecuencia asociada (art. 27(1) (b)). Una vez que se han definido estos elementos, es posible hacer una identificación inicial de las áreas de impacto del sistema de IA en términos de las categorías de individuos y grupos interesados (art. 27(1) (c)) y los derechos que pueden estar en juego.

Esta fase también sirve como evaluación preliminar para excluir de la EIDF aquellos casos en los que esté claro que no existe riesgo de perjuicio para las personas afectadas.

Si por el contrario, se identifica un daño potencial, éste debe examinarse en la **segunda fase (análisis de riesgos)**, que debe ir más allá de una identificación general de las posibles áreas de impacto y estimar el nivel de impacto para cada derecho o libertad.

Hay varias razones por las que la estimación individualizada del nivel de impacto es esencial. En primer lugar, es una característica de todas las evaluaciones de impacto, desde las medioambientales hasta las de ciberseguridad: no puede haber una correcta evaluación de riesgos sin una estimación de riesgos. En segundo lugar, la estimación es la base de la **tercera fase, que es la definición de las medidas de prevención/mitigación de riesgos (mitigación y gestión de riesgos)**: si no se ha estimado el impacto, no es posible identificar la idoneidad y eficacia de las medidas para eliminar/reducir el impacto inicial. En tercer lugar, la estimación es, por lo tanto, funcional para la aplicación del principio clave de la rendición de cuentas: sólo definiendo el nivel de riesgo antes y después de la adopción de las medidas de prevención/mitigación es posible demostrar que el riesgo se ha abordado de manera específica y eficaz. Por todas estas razones, las letras d) y f) del apartado 1 del artículo 27 desempeñan un papel central y constituyen la base para el desarrollo de la metodología de evaluación del riesgo.

3. La interacción con la normativa de protección de datos

La interacción entre el RIA y el RGPD en relación con la EIDF es doble. Por un lado, el **artículo 35 (1) y (7) del RGPD** ya incluye una evaluación de los riesgos para los derechos de los interesados [Grupo de Trabajo del artículo 29 sobre Protección de Datos 2017, 6], pero en la práctica de las EIPD esta parte se ha implementado escasamente. Por otro lado, el RIA hace hincapié en la evaluación del impacto en los derechos individuales, independientemente del uso de datos personales en el desarrollo y la implementación de la IA, pero hay una falta de orientación en la implementación de la EIDF.

El estrecho vínculo entre la EIPD [APDCAT 2024] y la EIDF también es evidente en el **artículo 27, apartado 4, del RIA**, que establece que "si ya se cumple cualquiera de las obligaciones establecidas en el presente artículo mediante la evaluación de impacto relativa a la protección de datos realizada con arreglo al artículo 35 del Reglamento (UE) 2016/679 o del artículo 27 de la Directiva (UE) 2016/680, la evaluación de impacto relativa a los derechos fundamentales a que se refiere el apartado 1 del presente artículo complementará dicha evaluación de impacto relativa a la protección de datos". En la misma línea, de conformidad con el **artículo 26, apartado 9, del RIA**, los responsables de despliegue de IA deben utilizar la información facilitada en virtud del artículo 13 del RIA (transparencia y comunicación de información a los responsables de despliegue) para cumplir con su obligación de llevar a cabo una EIPD de acuerdo con el artículo 35 del RGPD.

En este contexto, si los responsables de despliegue de IA no cumplen con las obligaciones de la EIDF en virtud del RIA, o si estas obligaciones no son aplicadas adecuadamente por las autoridades competentes, en el futuro **las Autoridades de Protección de Datos (APD) pueden desempeñar un papel activo en la aplicación de la EIDF de los sistemas de IA a través de las disposiciones del artículo 35 del RGPD**, en la medida en que el RGPD sea aplicable (este es el caso en muchas situaciones en las que la IA afecte a personas y grupos, dada la amplia definición de datos personales y tratamiento de datos en el RGPD y el papel de los datos en el desarrollo, la implementación y el uso de la IA).

Teniendo en cuenta todos estos diferentes aspectos de la interacción entre la EIDF en el RIA y el artículo 35 del RGPD, y también dada la experiencia de las APD en el tratamiento de cuestiones de derechos fundamentales [Mantelero-Esposito 2021], **es adecuado y necesario que estas Autoridades desempeñen un papel activo a la hora de proporcionar orientación sobre la EIDF en el contexto de los sistemas de IA basados en datos personales.**

4. El modelo de EIDF adoptado

Antes de examinar la metodología para llevar a cabo la EIDF en el ámbito de la IA, cabe señalar que se trata de una evaluación realizada por expertos que debe ser coherente tanto con los enfoques metodológicos comúnmente utilizados en la evaluación y gestión de riesgos como con la teoría jurídica de los derechos fundamentales. En este sentido, **analizar el potencial impacto de la IA en los derechos individuales y colectivos es un ejercicio complejo que requiere diferentes tipos de experiencia**, que combina los derechos fundamentales, el diseño de la IA y la comprensión del contexto social. Al igual que con la EIPD, la EIDF no es, por lo tanto, un ejercicio que pueda ser realizado por personas inexpertas en la materia.

El modelo EIDF debe ser coherente con las metodologías de evaluación y gestión de riesgos. En este sentido, desde la perspectiva de los proveedores y los responsables de despliegue de IA, la EIDF no es una tarea independiente, sino que forma parte de un conjunto integrado de evaluaciones que estas entidades deben realizar para cumplir con las obligaciones legales.

Por ejemplo, limitando la atención a las cuestiones relacionadas con los datos, que son el núcleo de los sistemas de IA, los instrumentos jurídicos nacionales y de la UE exigen evaluaciones de riesgos de protección y seguridad de los datos. Además, los sistemas de IA se utilizan a menudo como componente de otras tecnologías (por ejemplo, en la prevención y detección de accidentes), que deben cumplir normas técnicas que incluyen los elementos de evaluación de riesgos. Frente a este escenario, la EIDF debe estar en consonancia con las metodologías comunes de evaluación de riesgos⁴, no solo en términos de la solidez científica del enfoque metodológico, sino también para garantizar la coherencia y la plena interoperabilidad entre los diferentes componentes de la estrategia general de gestión de riesgos de los proveedores y responsables de despliegue de IA en todas sus actividades.

Del mismo modo, **el modelo EIDF debe ser coherente con la teoría y la práctica jurídicas de los derechos fundamentales**. A este respecto, por ejemplo, no es posible definir el impacto global de un sistema de IA en los derechos fundamentales, como en el caso de la ciberseguridad, porque los derechos fundamentales deben considerarse individualmente y **no pueden evaluarse de forma acumulativa, ni se pueden compensar impactos diferentes** (por ejemplo, un sistema de IA con un impacto bajo en la protección de datos y un alto impacto en la libertad de expresión no tiene un impacto medio global como consecuencia de una compensación entre estos diferentes niveles de impacto).

De acuerdo con estos supuestos, los **tres bloques principales del modelo EIDF** son:

- (i) Una **fase de planificación y determinación del alcance**, centrada en las principales características del producto/servicio y en el contexto en el que se situará.
- (ii) Una **fase de recopilación de datos y análisis de riesgos**, en la que se identificarán los riesgos potenciales y se evaluará su posible impacto en los derechos fundamentales.
- (iii) Una **fase de gestión de riesgos**, en la cual se adoptan, prueban y supervisan medidas adecuadas para prevenir o mitigar estos riesgos y comprobar su efectividad.

En términos de estructura, de conformidad con el artículo 27 del RIA y las metodologías de evaluación de riesgos, la EIDF es una **evaluación contextual** centrada en la solución de IA específica que se está implementando y no una evaluación tecnológica centrada en las tecnologías de IA en general y sus diversos usos potenciales: examina una aplicación de IA específica y su contexto de uso. Además, la EIDF también se caracteriza por un **enfoque ex ante**, lo que la convierte en una herramienta para un diseño de IA orientado a los derechos fundamentales, adoptando el enfoque desde el diseño ya conocido en la protección de datos.

Por último, la EIDF tiene una **estructura iterativa circular** y, como todas las evaluaciones de riesgos de situaciones que pueden evolucionar con el tiempo, no es una evaluación previa puntual, sino que tiene una estructura iterativa circular en la que se repiten las principales fases de la gestión de riesgos (planificación/alcance, análisis de riesgos, prevención/mitigación de riesgos), ya que los cambios tecnológicos, sociales y contextuales afectan a algunos de los elementos relevantes de una evaluación previa (art. 27.2 RIA).

⁴ Véase, por ejemplo, ISO, Gestión de riesgos. Directrices. ISO 31000. <https://www.iso.org/standard/65694.html>, que identifica las siguientes tres fases principales, combinadas con tres tareas complementarias (registro y presentación de informes; seguimiento y revisión; comunicación y consulta): (i) alcance, contexto y criterios; (ii) evaluación de riesgos (identificación de riesgos, análisis de riesgos, evaluación de riesgos); (iii) tratamiento de riesgos. El mismo enfoque también se puede ver en el PNUD 2024.

4.1 La fase de planificación y determinación del alcance

La fase de planificación y determinación del alcance parte del análisis de las necesidades y la descripción de las soluciones de IA que se van a desarrollar, y continúa con la consideración del escenario contextual de los derechos fundamentales (incluidos los controles ya existentes) y las áreas potencialmente afectadas. **En esta fase se examinan dos áreas principales: la dimensión inherente al sistema de IA y la dimensión contextual** (véase la sección 2 supra). Con respecto a **la dimensión contextual de las** soluciones de IA, no se limita a la identificación de los derechos potencialmente afectados y de los titulares de derechos (sin cuantificar el impacto, que es el objetivo de la siguiente fase), sino que también incluye un análisis preliminar de los elementos relevantes de la protección jurídica existente de estos derechos.

Por las razones expuestas anteriormente en relación con la variedad de usos potenciales de la IA, los contextos de uso y las personas y grupos potencialmente afectados, no es posible proporcionar un cuestionario exhaustivo que puedan utilizar los operadores de IA para abordar todas las cuestiones pertinentes de esta primera fase de la EIDF.

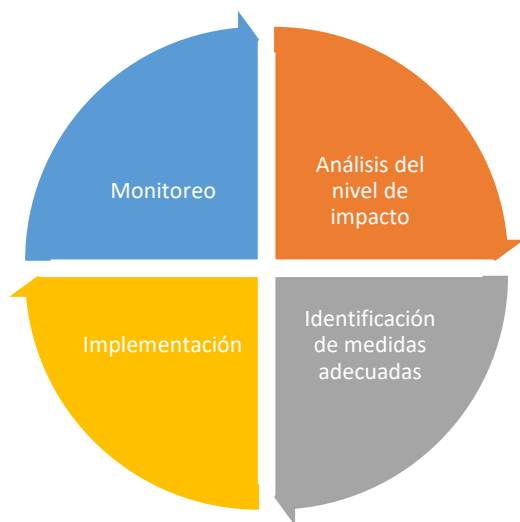
Sobre la base de la experiencia de la EIPD, es posible proporcionar **una lista no exhaustiva de posibles preguntas para la planificación y el alcance de la EIDF**, que puede ser complementada por los operadores de IA con preguntas específicas basadas en la naturaleza y el uso de las soluciones de IA que se están considerando (véase **Planificación y alcance** en la Sección 5 – La plantilla del modelo EIDF). Sin embargo, como se ha demostrado en los casos de uso realizados en Cataluña (véase la sección 6), este cuestionario es eficaz para guiar adecuadamente a los operadores de IA en la fase de planificación y determinación del alcance, cubriendo todas las áreas relevantes con preguntas que luego pueden profundizarse con preguntas adicionales centradas en el caso individual.

4.2 Fase de recopilación de datos y análisis de riesgos

Sobre la base de la información recopilada en la fase de planificación, determinación del alcance y de identificación de riesgos, será posible determinar si es necesario realizar un análisis a fondo del nivel de impacto sobre los derechos individuales. En el caso de que existan derechos potencialmente afectados, este análisis cuantificará este impacto con el fin de prevenirlo o reducirlo a través de medidas adecuadas.

Por lo tanto, el **análisis del nivel de impacto de** la solución de IA en los derechos potencialmente afectados es la primera fase del enfoque circular de la EIDF, que también incluye los tres pasos siguientes, como parte de la Fase de Gestión de Riesgos (véase la Sección 4.3): (i) la identificación de medidas adecuadas para prevenir o mitigar el riesgo, (ii) la implementación de dichas medidas, y (iii) el seguimiento del funcionamiento del sistema de IA con el fin de revisar la evaluación y las medidas adoptadas.

Evaluación de impacto (enfoque circular) [Gráfico 1]



Con respecto a la Fase de Recopilación de Datos y Análisis de Riesgos, dada la naturaleza de la evaluación, los datos se referirán a los diferentes aspectos de los derechos potencialmente en juego, incluyendo también información sobre el contexto de uso y las personas y grupos potencialmente afectados. A pesar de la variedad de estos elementos y de la especificidad de cada derecho y libertad, la fase de análisis puede basarse en parámetros comunes clave.

Estos parámetros permiten operacionalizar un concepto abstracto como el impacto en los derechos, de manera que pueda ser evaluado más fácilmente y (i) comparar el nivel de impacto sobre los diferentes derechos para priorizar la prevención/mitigación de riesgos, y (ii) comprender cómo puede cambiar el impacto sobre un derecho individual si algunos de los elementos del sistema o del contexto varían.

4.2.1 Variables clave para la evaluación de impacto

De acuerdo con la teoría del riesgo y el marco jurídico de los derechos fundamentales, el impacto sobre los derechos consta de dos dimensiones clave: la **probabilidad** de un impacto adverso y su **gravedad**.⁵ La combinación de variables relacionadas con estas dos dimensiones proporciona un **índice de riesgo** que se evalúa para cada uno de los derechos potencialmente afectados.

Para construir estos índices, es posible representar las variables relevantes en una escala de un mínimo a un máximo (suponiendo, de acuerdo con la teoría general del riesgo, que no hay riesgo cero) y utilizando variables ordinales (por ejemplo, bajo, medio, alto, muy alto). El uso del escalamiento y las variables asociadas permite comparar diferentes situaciones utilizando las mismas variables: por ejemplo los diferentes niveles de impacto sobre la no discriminación que produce un sistema de calificación crediticia cuando se introducen cambios en su algoritmo. Por lo tanto, estas variables ordinales se pueden utilizar para "medir" el impacto en una cuantificación del riesgo basada en rangos (bajo, medio, alto, muy alto).

⁵ Véase el artículo 3, apartado 2, del Reglamento de Inteligencia Artificial, que establece que por 'riesgo' se entiende "la combinación de la probabilidad de que se produzca un perjuicio y la gravedad de dicho perjuicio", y el artículo 25, apartado 1, del RGPD, que se refiere a "los riesgos de diversa probabilidad y gravedad que entraña el tratamiento para los derechos y libertades de las personas físicas [...]".

Sin embargo, la teoría de los derechos fundamentales no permite la creación de un índice compuesto, como es común en la evaluación de riesgos, donde todos los impactos potenciales se combinan para crear un índice de impacto general. Este planteamiento entra en conflicto con el enfoque jurídico de los derechos fundamentales, en el que cada derecho debe considerarse de forma independiente, en términos de su posible perjuicio, y el hecho de que uno se vea menos afectado que otro no puede dar lugar a ninguna forma de compensación.

Como muestran los casos de uso analizados en la sección 5, será posible evaluar el impacto en los diferentes derechos involucrados, pero no quiere decir que un determinado sistema de IA tenga un impacto general en los derechos que se considere bajo, medio o alto. La única interacción posible entre diferentes intereses es a través del **test** de ponderación entre los derechos en conflicto, pero esta prueba evalúa el nivel de impacto en cada derecho. El test de ponderación no se refiere al nivel de riesgo de los derechos afectados, sino a la priorización de un interés sobre otro. Por lo tanto, debe considerarse como un factor externo, que solo debe tenerse en cuenta una vez que se haya evaluado el impacto sobre los derechos individuales, y que puede influir en los resultados de la evaluación de impacto al hacer que el impacto sobre determinados derechos sea aceptable debido a un interés contrapuesto prevaleciente.⁶

Sobre la base de estas consideraciones, un modelo EIDF definirá un índice de riesgo para cada derecho potencialmente afectado utilizando dos dimensiones: probabilidad y gravedad. La **probabilidad** se entiende como una combinación de (i) la probabilidad de resultados adversos y (ii) la exposición. La primera variable se refiere a la probabilidad de que se produzcan las consecuencias adversas de un determinado riesgo, mientras que la segunda variable se refiere al grado en que las personas potencialmente en riesgo podrían verse afectadas. En cuanto a la exposición, cabe señalar que la atención se centra en las personas potencialmente afectadas por el uso del sistema de IA, no en la población en su conjunto.

La **gravedad** de las consecuencias previstas se basa en dos variables: i) la gravedad del perjuicio en el ejercicio de los derechos y libertades (gravedad),⁷ sobre la base de sus atributos, incluida la consideración de las repercusiones, la vulnerabilidad y las situaciones de dependencia específicas de cada grupo; y (ii) el esfuerzo para superarlo e invertir los efectos adversos (esfuerzo).

Tanto la probabilidad como la gravedad deben evaluarse sobre una base contextual, y la participación de las partes interesadas pertinentes puede ser de ayuda. Como es habitual en la evaluación de riesgos, la estimación de la probabilidad se basa tanto en casos anteriores, considerando situaciones comparables, como en el uso de técnicas analíticas y de simulación, basadas en posibles escenarios de uso. Los mismos enfoques se utilizan también para estimar el nivel de gravedad, pero en este caso con mayor énfasis en el análisis jurídico sobre la gravedad del perjuicio, que debe evaluarse con referencia a la jurisprudencia sobre derechos fundamentales y al marco jurídico pertinente.

Sobre la base de los valores de probabilidad y gravedad derivados de las variables anteriores, se determina un índice de riesgo, que indica el impacto global para cada uno de los derechos y libertades

⁶ Véase, por ejemplo, Parte II, Caso de uso 1, en el caso del desarrollo y uso de una plataforma de análisis de aprendizaje avanzado, en la que se considera aceptable algún impacto en los derechos de privacidad y protección de datos en vista de los beneficios para el derecho a la educación.

⁷ La **gravedad del perjuicio a un derecho fundamental/humano suele evaluarse de acuerdo con los tres elementos siguientes: (i) su intensidad, (ii) las consecuencias de la violación, y (iii) su duración**, donde la intensidad de la violación está relacionada con la importancia del interés jurídico protegido violado. Véase también Tribunal de Justicia de la Unión Europea (2022). Derecho al respeto de la vida privada y familiar, del domicilio y de la correspondencia: Guía práctica de la jurisprudencia del Tribunal Europeo de Derechos Humanos en virtud del artículo 8. https://www.echr.coe.int/documents/d/echr/guide_art_8_eng.

considerados.⁸ Vale la pena señalar que estos resultados deben combinarse con cualquier elemento que justifique una limitación de algunos derechos desde una perspectiva legal, como la naturaleza obligatoria de ciertas características que afectan a los derechos: en este caso, el riesgo potencial debe considerarse aceptable en la medida en que el sistema de IA cumpla con los requisitos legales dados.

4.2.2 Variables y construcción del índice de impacto

En muchos modelos y estándares de evaluación de impacto basados en el riesgo, los índices de riesgo se construyen utilizando matrices porque son relativamente fáciles de usar y explicar.⁹ Dado que una matriz de riesgos es un gráfico que combina dos dimensiones utilizando colores para reflejar diferentes niveles de riesgo, son útiles para evaluar índices generados por diferentes variables. Por esta razón, pueden ser utilizados en la EIDF para definir el nivel de impacto sobre cada derecho en cuestión.

La metodología aquí propuesta utiliza un índice de riesgo para cada derecho potencialmente afectado, basado en una matriz que combina dos dimensiones (probabilidad y gravedad). Cada una de estas dimensiones resulta de la combinación de dos pares de variables, también construidas mediante matrices: la probabilidad de consecuencias adversas, y la exposición, para la probabilidad; la gravedad de los perjuicios, y el esfuerzo por superarlos e invertir los efectos adversos, para la gravedad.

No existe un único modelo de matriz de riesgos que se pueda utilizar en la evaluación de riesgos; la práctica en este campo muestra una variedad de modelos. Las más comunes son las matrices 3x3, 4x4, 5x5, 5x4 y 6x4, donde los pares de números indican el número de rangos de las dos escalas que definen la dimensión considerada. Como la matriz se refiere a dos variables independientes, se pueden evaluar de acuerdo con escalas que pueden diferir en el número de rangos, por ejemplo, una escala de 6x4 donde se proporcionan seis rangos diferentes para una variable y solo cuatro para la otra.

La **matriz 4x4** puede ser la más apropiada en el contexto de la EIDF, ya que reduce el riesgo de posicionamiento medio, presta más atención a los niveles altos y muy altos de una manera coherente con el enfoque en el alto riesgo del marco regulatorio actual de la IA, y no fragmenta excesivamente la parte inferior de la escala, lo cual es menos relevante debido al enfoque antes mencionado.

En las matrices, se utilizan etiquetas descriptivas para las diferentes combinaciones de niveles en la escala de colores, como se muestra en este ejemplo de una matriz de gravedad:

⁸ Consulte la siguiente sección para conocer la metodología utilizada para combinar las diferentes variables y crear los índices.

⁹ Véanse también APDCAT 2024, 33 y 53; CNIL 2018, 23.

		Gravedad			
		Bajo (L)	Medio (M)	Alto (A)	Muy alto (MA)
Esfuerzo	Bajo (B)	B/B	B/M	B/A	B/MA
	Medio (M)	M/B	M/M	M/A	M/ MA
	Alto (A)	A/B	A/M	A/A	A/ MA
	Muy alto (MA)	MA /B	MA /M	MA /A	MA/ MA

Niveles de gravedad			
Bajo	Medio	Alto	Muy alto

4.3 La fase de gestión de riesgos

Tras el análisis de riesgos, que ha definido el nivel de impacto de la solución de IA sobre los derechos potencialmente afectados, es necesario gestionar los riesgos identificados adoptando las medidas adecuadas.¹⁰ Por lo tanto, la tercera fase de la EIDF se articula en tres etapas, de la siguiente manera:

- (i) La **identificación de medidas adecuadas** para prevenir o mitigar el riesgo, teniendo en cuenta su impacto en el nivel de riesgo de acuerdo con un análisis de escenarios específicos del contexto.
- (ii) La **implementación** de dichas medidas.
- (iii) El **seguimiento** del funcionamiento del sistema de IA con el fin de revisar la evaluación y las medidas adoptadas en caso de que los cambios tecnológicos, sociales y contextuales afecten al nivel de riesgo o a la eficacia de las medidas aplicadas.

Dado que la EIDF no es una comprobación final de una solución de IA, sino **una herramienta de diseño para guiar el desarrollo y el despliegue de la IA** hacia un enfoque orientado a los derechos fundamentales, el seguimiento del funcionamiento del sistema de IA puede formar parte de la fase previa a la comercialización en la que se prueban e implementan diferentes soluciones de diseño con el fin de seleccionar la más adecuada. Por lo tanto, en consonancia con el **enfoque circular de la evaluación de riesgos** y el diseño de la IA, es posible que se lleven a cabo varias series de evaluaciones de riesgo, de aplicación de medidas de mitigación y reevaluaciones hasta que la versión final del producto/servicio de IA conlleve un nivel de **riesgo residual** satisfactorio en términos de aceptabilidad y pueda introducirse en el mercado o ponerse en servicio.

Además, **pueden producirse cambios en el escenario tecnológico y social o en el contexto específico de uso después de** que la herramienta de IA se haya introducido en el mercado o se haya puesto en servicio. Ello puede repercutir en el nivel de riesgo evaluado previamente con respecto a los derechos de que se trate, así como plantear nuevas preocupaciones con respecto a otros derechos. En tales casos, también será necesario reevaluar las soluciones de IA adoptadas y aplicar las medidas adecuadas.¹¹

¹⁰ Véase también el artículo 27, apartado 1, letra f), del Reglamento de Inteligencia Artificial.

¹¹ Véase también el artículo 27, apartado 2, del RIA.

De acuerdo con estas observaciones, la plantilla del modelo EIDF (véase la sección 5) incluye una matriz que muestra el impacto de las medidas de prevención/mitigación de riesgos adoptadas sobre los niveles de riesgo identificados en la fase de análisis de riesgos y el riesgo residual resultante.

5. La plantilla del modelo EIDF

El modelo utilizado en los casos de uso se basa en las tres fases de la EIDF discutidas en la sección anterior. Consta de varios elementos.

El primero es un cuestionario que abarca las cuatro áreas principales de la fase de planificación y determinación del alcance, en concreto, la descripción y el análisis del sistema de IA, el contexto jurídico, los controles ya establecidos y la participación de las partes interesadas (cuestionario de **planificación y alcance**).

El segundo elemento de la plantilla es un conjunto de matrices (**Tablas 3, 6 y 7**) y criterios de cuantificación de variables asociadas que se utilizan para evaluar la probabilidad y la gravedad de los posibles perjuicios a cada derecho y libertad (**Tablas 1, 2, 4 y 5, y Tabla 1A**), y para estimar el impacto global asociado.

El tercer elemento de la plantilla consiste en dos tablas, una que indica el nivel de impacto sobre cada derecho y libertad y las medidas de prevención/mitigación identificadas para reducir el riesgo (**Tabla 2A**), y otra que estima el riesgo residual resultante de la adopción de estas medidas (**Tabla 3A**).

Cuestionario de planificación y alcance

<p>Sección A</p> <p>Descripción y análisis del sistema de IA, incluidos los flujos de datos relacionados</p>	<p>¿Cuáles son los principales objetivos del sistema de IA?</p> <p>¿Cuáles son las principales características del sistema?</p> <p>¿En qué países se ofrecerá?</p> <p>¿Qué tipos de datos se tratan (personales, no personales, categorías especiales)?</p> <p>Identificación de los posibles titulares de derechos: ¿quiénes son los individuos o grupos que pueden verse afectados por los sistemas de IA? ¿Se incluyen entre ellos individuos o grupos vulnerables?</p> <p>Identificación de los responsables: ¿qué personas/entidades están implicadas en el diseño, desarrollo y despliegue de los sistemas de IA? ¿Cuál es su papel?</p>
<p>Sección B</p> <p>Contexto de derechos fundamentales</p>	<p>¿Qué derechos fundamentales se ven potencialmente afectados por el uso del sistema de IA?</p> <p>¿Qué instrumentos jurídicos internacionales/regionales de protección de los derechos humanos/fundamentales se han aplicado a nivel operativo?</p> <p>¿Cuáles son los tribunales u organismos más relevantes en materia de derechos fundamentales en el contexto de uso?</p> <p>¿Cuáles son las decisiones y disposiciones más relevantes en materia de derechos humanos/fundamentales?</p>

<p>Sección C</p> <p>Controles implementados</p>	<p>¿Qué políticas y procedimientos se han implementado para evaluar el posible impacto sobre los derechos fundamentales, incluida la participación de las partes interesadas?</p> <p>¿Se ha realizado, desarrollado y aplicado una evaluación de impacto en relación con cuestiones específicas (por ejemplo, protección de datos) o algunas características del sistema (por ejemplo, el uso de la biometría)?</p>
<p>Sección D</p> <p>Intervención de las partes interesadas y diligencia debida</p>	<p>¿Cuáles son los principales grupos o comunidades potencialmente afectados por el sistema de IA, incluido su desarrollo?</p> <p>¿Qué partes interesadas, además de los individuos o grupos que puedan verse afectados por los sistemas de IA, deben participar (por ejemplo, la sociedad civil y las organizaciones internacionales, expertos, asociaciones industriales, periodistas)?</p> <p>¿Hay otros titulares de obligaciones que deban participar, aparte del proveedor y el responsable del despliegue de IA (por ejemplo, autoridades nacionales, organismos gubernamentales)?</p> <p>¿Han participado en el proceso de evaluación los socios comerciales, incluidos los proveedores de servicios (por ejemplo, subcontratistas en sistemas de IA y conjuntos de datos)?</p> <p>¿Ha llevado a cabo el proveedor de IA una evaluación de su cadena de suministro, para identificar si las actividades de los proveedores/contratistas que participan en el desarrollo de productos/servicios pueden afectar a los derechos fundamentales?</p> <p>¿El proveedor ha promovido estándares o auditorías de derechos fundamentales, para garantizar el respeto de los derechos fundamentales entre los proveedores?</p> <p>¿El proveedor y el responsable del despliegue de la IA han comunicado públicamente las posibles repercusiones del sistema de IA en los derechos fundamentales?</p> <p>¿El proveedor y el responsable del despliegue de la IA han proporcionado formación sobre estándares de derechos fundamentales al personal encargado de la gestión y la contratación relacionadas con el sistema de IA?</p>

Matrices de riesgo

Tabla 1. Probabilidad

Bajo	El riesgo de perjuicio es improbable o altamente improbable
Medio	El riesgo puede producirse
Alto	Existe una alta probabilidad de que se produzca el riesgo
Muy alto	Es muy probable que se produzca el riesgo

Tabla 2. Exposición

Bajo	Pocos o muy pocos de la población identificada de titulares de derechos están potencialmente afectados
Medio	Parte de la población identificada está potencialmente afectada
Alto	La mayoría de la población identificada está potencialmente afectada
Muy alto	Casi toda la población identificada está potencialmente afectada

Tabla 3. Probabilidad de que se produzca un perjuicio (*likelihood*)

		Probabilidad			
		Baja	Media	Alta	Muy alta
Exposición	Bajo	B	B/M	B/A	B/MA
	Medio	M/B	M	M/A	M/MA
	Alto	A/B	A/M	A	A/MA
	Muy alto	MA/B	MA/M	MA/A	MA

Probabilidad			
Bajo	Medio	Alto	Muy alto

Tabla 4. Gravedad del perjuicio

Baja	Los individuos y grupos afectados pueden encontrar sólo perjuicios menores en el ejercicio de sus derechos y libertades.
Media	Los individuos y grupos afectados pueden encontrarse con importantes perjuicios.
Alta	Los individuos y grupos afectados pueden enfrentarse a graves perjuicios.
Muy alta	Los individuos y grupos afectados pueden encontrarse con perjuicios graves o incluso irreversibles.

Tabla 5. Esfuerzo para superar los perjuicios y revertir los efectos adversos

Bajo	Los perjuicios sufridos pueden superarse sin problemas (por ejemplo, tiempo dedicado a modificar la información, molestias, irritaciones, etc.)
Medio	Los perjuicios sufridos pueden superarse a pesar de algunas dificultades (por ejemplo, costes adicionales, miedo, incomprensión, estrés, pequeñas dolencias físicas, etc.).
Alto	Los perjuicios sufridos pueden superarse aunque con serias dificultades (por ejemplo, pérdidas económicas, daños materiales, empeoramiento de la salud, etc.).
Muy alto	Los perjuicios sufridos pueden no superarse (por ejemplo, dolencias psicológicas o físicas de larga duración, muerte, etc.).

Tabla 6. Gravedad (severity)

		Gravedad			
		Baja	Media	Alta	Muy alta
Esfuerzo	Bajo	B	B/M	B/A	B/MA
	Medio	M/B	M	M/A	M/MA
	Alto	A/B	A/M	A	A/MA
	Muy alto	MA/B	MA/M	MA/A	MA

Gravedad			
Baja	Media	Alta	Muy alta

Tabla 1A. Recogida de datos y análisis de riesgos

Derechos/ Libertades potencialmente afectados	Descripción del impacto	Probabilidad			Gravedad		
		Probabilidad de resultados adversos	Exposición	Probabilidad	Gravedad del impacto	Esfuerzo	Gravedad

Tabla 7. Impacto general del riesgo

		Gravedad			
		Baja	Media	Alta	Muy alta
Probabilidad	Baja				
	Media				
	Alta				
	Muy alta				

Riesgo global			
Bajo	Medio	Alto	Muy alto

Tabla 2A. Gestión de riesgos (I)

Derechos/libertades afectados	Probabilidad	Gravedad	Impacto gobal	Medidas para prevenir/mitigar los impactos

Tabla 3A. Gestión de riesgos (II)

Derechos/libertades afectados	Probabilidad (residual)	Gravedad (residual)	Impacto residual

6. Del modelo a la práctica: los casos de uso

El debate metodológico en curso sobre la EIDF y su implementación se ha caracterizado por ser un debate principalmente teórico y centrado en las políticas, con poca atención al análisis empírico y a la plena implementación en el mundo real de los diferentes modelos propuestos. En este contexto, el proyecto liderado por la Autoridad Catalana de Protección de Datos marca la diferencia al acercar el debate de la EIDF a la experiencia concreta de los proveedores y responsables de despliegue directamente implicados en el uso de la IA.

Este enfoque empírico basado en casos es crucial para probar la eficacia del modelo propuesto en la consecución de los objetivos políticos y de diseño de la EIDF elaborados por el legislador de la UE en el RIA. Más concretamente, los casos de uso de este proyecto han demostrado que es posible racionalizar el procedimiento de la EIDF evitando la adopción de una larga lista de control y centrándose en los elementos esenciales del impacto en los derechos fundamentales.

Los casos de uso también han demostrado que, al menos para la primera ronda de evaluación, mitigación y reevaluación, las personas con la formación adecuada pueden completar la EIDF en dos o tres reuniones breves (3 horas por reunión). Esto confirma que **la EIDF, si se formula adecuadamente, no impone una carga adicional excesiva a las entidades privadas y públicas de la UE para cumplir con el RIA.**

Por último, cada caso de uso se basó en cuatro interacciones diferentes: (i) un análisis interno inicial del caso por parte de los expertos (normalmente DPD) de las entidades que llevan a cabo la EIDF, con la objetivo de esbozar los elementos clave de la EIDF prestando especial atención a la fase de planificación y alcance; (ii) una discusión con un experto en la EIDF para revisar la evaluación inicial; (iii) un debate en grupo en el que participan los expertos de todas las entidades participantes en el proyecto; y (iv) una revisión final por parte de los expertos de la entidad que realiza la EIDF. Este proceso de cuatro pasos ilustra dos aspectos clave de la EIDF: **la importancia de una evaluación de expertos y la importancia de una evaluación basada en equipos** que involucre diferentes conocimientos para mejorar el nivel de análisis.

En cuanto a **la selección de los casos**, cabe destacar que se trata de un proyecto en curso y los casos presentados en este informe son los primeros que se han comentado y en los que se ha aplicado el modelo EIDF. Otros casos están en evaluación y se publicarán en el futuro en la web “DPD en xarxa”, (<https://www.dpdenxarxa.cat/>) y de la Autoridad Catalana de Protección de Datos (<https://www.apdcat.cat>). Por otra parte, algunos casos en los que se ha aplicado el modelo EIDF con un impacto relevante en el diseño de soluciones de IA no se han incluido en este informe por razones de confidencialidad, pero han sido útiles para que todos los participantes elaboren mejor la práctica de la plantilla de la EIDF.

En cuanto a los **ámbitos cubiertos**, los casos de uso se refieren a cuatro de los ámbitos clave del anexo III del RIA, concretamente, la educación (evaluación de los resultados del aprendizaje y predicción del abandono escolar), la gestión de los trabajadores (sistemas de apoyo a la toma de

decisiones en la gestión de recursos humanos), el acceso a la asistencia sanitaria (tratamiento del cáncer basado en imágenes médicas) y los servicios sociales (asistente de voz para personas mayores). Estos ámbitos representan las áreas en las que las soluciones de IA se utilizan cada vez más, con mayor impacto en individuos y grupos. En este sentido, la naturaleza de los casos de uso discutidos también los hará útiles para muchas otras entidades públicas y privadas de otros países interesadas en diseñar sistemas/modelos de IA que cumplan con los derechos fundamentales en estas áreas fundamentales.

De acuerdo con el objetivo de este proyecto, los casos de uso se presentan tal y como fueron desarrollados por los participantes, más que como mejores prácticas o casos estandarizados. El proyecto fue diseñado para probar la eficacia de la plantilla del modelo y la metodología asociada. Siguiendo la línea de experiencia en la elaboración de las EIPD, hemos dado libertad a los participantes para desarrollar las diferentes partes de la plantilla según su enfoque. Así, encontramos análisis más extensos que otros. Sin embargo, los elementos básicos (preguntas, matrices y metodología de evaluación) siguen siendo los mismos.

La idea principal que se quiere poner de relieve en este informe es la importancia de reflexionar sobre el ejercicio de análisis realizado con el objetivo de mostrar los resultados obtenidos, y no presentar las EIDF llevadas a cabo como casos ficticios y perfectos. La EIDF ha sido y será implementada por varios actores, en algunos casos con mayor detalle, en otros con algunas limitaciones; pero, en la medida en que contribuya a un análisis eficaz y a la prevención y mitigación del impacto en los derechos fundamentales, habrá logrado su objetivo principal.

Referencias

Todos los sitios web citados en este documento fueron consultados entre septiembre y diciembre de 2024.

APDCAT. 2024. Evaluación de impacto relativa a la protección de datos,

https://apdcat.gencat.cat/web/.content/03-documentacio/Reglament_general_de_proteccio_de_dades/documents/Guia-EIPD.pdf.

APDCAT. 2020. Inteligencia Artificial. Decisiones automatizadas en Cataluña.

https://apdcat.gencat.cat/es/documentacio/intel·ligencia_artificial/

Article 29 Data Protection Working Party. 2017. Guidelines on Data Protection Impact Assessment (DPIA) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679, WP 248 rev.01, <https://ec.europa.eu/newsroom/article29/items/611236>.

CNIL. 2018. Privacy Impact Assessment (PIA). Templates, <https://www.cnil.fr/en/privacy-impact-assessment-pia>.

European Court of Human Rights.2022. Guide on Article 8 of the European Convention on Human Rights. Right to respect for private and family life, home and correspondence,

<https://ks.echr.coe.int/web/echr-ks/article-8>.

Mantelero, A. 2024. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. 54 Computer Law & Security Review 106020, <https://doi.org/10.1016/j.clsr.2024.106020> (open access).

Mantelero, A. and Esposito, M.S. 2021. An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. 41 Computer Law & Security Review 105561, <https://doi.org/10.1016/j.clsr.2021.105561> (open access).

Slattery, P. et al. 2024. A systematic evidence review and common frame of reference for the risks from artificial intelligence. <http://doi.org/10.13140/RG.2.2.28850.00968> and <https://airisk.mit.edu/>.

United Nations, AI Advisory Board. 2024. Governing AI for Humanity.

https://www.un.org/sites/un2.un.org/files/governing_ai_for_humanity_final_report_en.pdf.

UNDP. 2024. Visual guide. Business Process for Risk Management,

https://popp.undp.org/sites/g/files/zskgke421/files/2024-05/Risk_Management_Full_Visual_Guide_13.pdf.

Part II – Casos de uso

Caso de uso 1: Plataforma de analítica avanzada del aprendizaje

1. El contexto

El caso de uso que expondremos a continuación se encuentra relacionado con uno de los grandes problemas del sistema educativo superior, que es el abandono prematuro de la educación y la formación de la población entre 18 y 24 años. Esta situación ha generado una preocupación a nivel europeo y los datos estadísticos europeos que se pueden consultar en la base de datos de Eurostat¹² lo ponen de manifiesto. Tanto es así que la prioridad nº 1 del marco estratégico para la cooperación europea, en el ámbito de la educación y de la formación con miras al Espacio Europeo de Educación (EEE) y más allá (2021-2030)¹³, es “aumentar la calidad, la equidad, la inclusión y el éxito de todos en el ámbito de la educación y la formación”. Aunque el abandono prematuro se ha reducido en la última década, sigue constituyendo un reto. Para evitar limitaciones de acceso de los jóvenes a oportunidades socioeconómicas en un futuro, se pretende prestar especial atención a los grupos que presenten un riesgo de bajo rendimiento educativo y un abandono prematuro en sus estudios.

Las instituciones de educación superior deben promover estrategias educativas que fomenten la finalización con éxito de los itinerarios de educación y formación, reduzcan los índices de abandono prematuro y afronten las causas del bajo rendimiento. Por eso, es importante discernir entre los datos de que se dispone, estructurarlos, extraer la información que nos proporcionan y utilizarla para la finalidad concreta que queremos conseguir.

- Los datos que nos pueden interesar para identificar la situación en la que se encuentra cada estudiante en cada momento mientras cursa sus estudios, y poder predecir lo que sucederá, son:
- Los datos históricos de los estudios previos de los alumnos recabados durante años anteriores.
- Los datos proporcionados por el centro de educación anterior.
- Los datos proporcionados por el propio estudiante en el momento de la matrícula.
- Los datos obtenidos durante la evolución de sus estudios.

A partir de la información que proporcionan los datos, se pueden establecer patrones de actuación del estudiantado. Además de disponer de una visión global de la situación del conjunto de estudiantes que está cursando unos mismos estudios y de su futura evolución, es posible tener, a la vez, una visión individualizada sobre la situación particular en la que se encuentra

¹² Acceso a Eurostat: <https://ec.europa.eu/eurostat>.

¹³ Resolución del Consejo relativa a un marco estratégico para la cooperación europea en el ámbito de la educación y la formación con miras al Espacio Europeo de Educación y más allá

(2021-2030): <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32021G0226%2801%29>.

cada estudiante y anticiparse a su porvenir. Esta información ayuda a promover estrategias acordes a la realidad, ofreciendo un tratamiento personalizado y atendiendo a las necesidades de cada estudiante.

El análisis de estos datos y su interpretación para llevar a cabo actuaciones de mejora y progreso en el ámbito de la educación entran dentro de lo que se conoce como “análítica del aprendizaje” (en inglés, *learning analytics*). En la primera Conferencia Internacional sobre Análisis y Conocimiento de Aprendizaje (en inglés, *Learning Analytics and Knowledge, LAK*), celebrada en 2011, se definió la analítica del aprendizaje¹⁴ como la medición, recopilación, análisis y comunicación de datos sobre los estudiantes y sus contextos, con el fin de comprender y optimizar el aprendizaje y los entornos en los que se produce.

Pese a que hace mucho tiempo que se investiga sobre la analítica del aprendizaje y que se ha utilizado tanto para predecir el éxito académico de los estudiantes como el riesgo de abandono de sus estudios, la aparición de nuevas tecnologías que proporcionan técnicas de análisis alternativas ha puesto de manifiesto la necesidad de abordar los requerimientos legales que han surgido de las recientes regulaciones, para poder hacer un uso correcto de estas.

Dada la situación actual, en la que los sistemas de inteligencia artificial (en adelante, sistemas de IA) están apareciendo de forma habitual en nuestros quehaceres cotidianos, es fundamental que las instituciones detecten y conozcan sus riesgos; así, podrán prevenirlos, minimizarlos y abordarlos

para hacer un uso óptimo de ellos, como aliados en la modernización y digitalización del sistema universitario.

Una de las dificultades detectadas, aparte de la pericia en la interpretación correcta de la información conseguida por los datos, es la ausencia de fluidez de adaptación a las nuevas tecnologías por parte de las instituciones y de su personal, en sus planteamientos y metodologías de enseñanza. Con los recursos de los que se disponen en el ámbito de la educación, es un hecho que las instituciones de educación superior tendrán que ir introduciendo sistemas de IA en los procesos en que la intervención humana se pueda encontrar limitada. Debemos ser conscientes de los cambios que se están generando, como la irrupción de la inteligencia artificial generativa¹⁵ dentro del modelo de aprendizaje del estudiantado, y los que se pueden llegar a generar.

De este modo, si se toma la iniciativa de incorporar los sistemas de IA que tenemos a nuestro alcance, previamente se pueden establecer herramientas adecuadas para proteger los derechos y libertades de los estudiantes desde una fase inicial y evitar, así, un perjuicio con su integración en el sistema universitario y, por ende, al estudiantado. No obstante, aún hay desafíos pendientes de abordar, como el sesgo que se puede dar en los datos o los dilemas éticos que se planteen en cada caso, así como cuestiones inherentes al desarrollo de los sistemas de IA y el impedimento o resistencia de la incorporación de los sistemas de IA en algunos ámbitos, que pueden llevar a entorpecer o ralentizar su propio crecimiento. En nuestro caso, se podría dejar el sistema universitario en la

¹⁴ <https://www.solaresearch.org/about/what-is-learning-analytics/>

¹⁵ Podéis encontrar la definición de IA generativa y cómo funciona en la *Guía para el uso de IA generativa en educación e investigación* de la UNESCO, 2024: <https://unesdoc.unesco.org/ark:/48223/pf0000389227>.

obsolescencia y lograr un efecto contrario al deseado, pudiendo generar un retroceso en la enseñanza y en la formación de los jóvenes.

Desde las fases embrionarias en las que se plantea el uso de sistemas de IA, hay que estudiar y reflexionar sobre los diferentes planteamientos y casuísticas con las que nos podemos encontrar, para aplicar todas las cautelas necesarias e instaurar las medidas técnicas y organizativas para construir modelos sólidos que garanticen la protección de los jóvenes, respetando los derechos fundamentales y los valores sociales.

El rápido aumento de los usos de los sistemas de IA en la educación está transformando la manera de enseñar y aprender y tiene repercusiones directas en las instituciones, en su personal y en el estudiantado. Esta transformación va aparejada a la necesidad de capacitar al personal de nuevas competencias, para desenvolverse en este nuevo panorama tecnológico en constante evolución y que, por consiguiente, se encuentre facultado¹⁶ para utilizar la información que le proporciona el sistema de IA.

El uso de la inteligencia artificial en la educación está planteando cuestiones fundamentales. Por eso, el Reglamento europeo de Inteligencia Artificial¹⁷ (RIA) ha determinado como sistemas de IA de alto riesgo “los sistemas de inteligencia artificial destinados a ser utilizados para evaluar los resultados del aprendizaje, también cuando dichos resultados se utilicen para orientar

el proceso de aprendizaje de las personas físicas en centros educativos” (apartado 3.b del anexo III).

Si bien hay sectores reticentes al uso de la inteligencia artificial en sus procesos, por lo que puede conllevar, es incuestionable que un buen uso en la educación parece ser enriquecedor con lo que respecta la aproximación de los estudios a las nuevas generaciones de estudiantes.

En defecto de integrar en la institución educativa una aplicación propia, en el mercado ya se dispone de plataformas de analítica de aprendizaje que recopilan la información obtenida de los datos proporcionados por el propio estudiante que se encuentran en el sistema de la institución (en inglés, *Student Information System, SIS*) y de los datos generados en los sistemas de gestión del aprendizaje (en inglés, *Learning Management System, LMS*), para identificar o predecir a los estudiantes en riesgo de abandono. Por ejemplo:

- *Assessment and Learning in Knowledge Spaces*¹⁸ (ALEKS)
- DreamBox¹⁹
- Carnegie Learning²⁰
- Smart Sparrow²¹
- The IntelliBoard²²

¹⁶ Recomendable la lectura de la publicación *AI competency framework for teachers* de la UNESCO, 2024 (actualmente versión disponible solo en inglés): <https://unesdoc.unesco.org/ark:/48223/pf0000391104>.

¹⁷ Reglamento (UE) 2024/1689 del Parlamento Europeo y del Consejo, de 13 de junio de 2024, por el que se establecen normas armonizadas en materia de inteligencia artificial y por el que se modifican los Reglamentos (CE) n° 300/2008, (UE) n° 167/2013, (UE) n° 168/2013, (UE) 2018/858, (UE) 2018/1139 y (UE) 2019/2144 y las Directivas 2014/90/UE, (UE)

2016/797 y (UE) 2020/1828 (Reglamento de Inteligencia Artificial): https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=OJ%3AL_202401689.

¹⁸ <https://www.aleks.com>

¹⁹ <https://www.dreambox.com>

²⁰ <https://www.carnegielearning.com>

²¹ <https://www.smartsparrow.com>

²² Casos reales de aplicación de la plataforma The IntelliBoard: <https://intelliboard.net/customers/>

En lo que respecta a los esfuerzos para reducir la tasa de abandono prematuro de la educación y la formación, prever los diferentes escenarios posibles facilita el establecimiento de políticas y medidas más efectivas, ya que se pueden centrar en la raíz del problema. La puesta a disposición del personal de las instituciones de mecanismos apropiados, y personalizados a las necesidades de cada estudiante, puede implicar un empoderamiento de los jóvenes a continuar con su crecimiento personal en el ámbito educativo, que luego se trasladará a su ámbito profesional. Por este motivo, teniendo en cuenta el potencial que tienen los sistemas de IA de transformar la noción actual que tenemos de la educación, el caso de uso que se ha llevado a cabo se ha centrado en el marco de la analítica del aprendizaje con el uso de sistemas de IA de alto riesgo, con un algoritmo predictivo.

Como podréis observar en los siguientes apartados, el caso de uso planteado y analizado ha dejado al margen el uso de algoritmos de decisión automatizada (ADA). Para conocer casos de uso de los ADA, recomiendo el informe *Inteligencia artificial. Decisiones Automatizadas en Cataluña*²³ elaborado por la Autoridad Catalana de Protección de Datos (APDCAT).

2. El proyecto

El proyecto tiene por objetivo diseñar y desarrollar un nuevo ecosistema de analítica del aprendizaje para el sistema educativo superior, con la creación de una plataforma de analítica avanzada del aprendizaje con uso de un sistema de IA, que permita evaluar los resultados del aprendizaje y predecir

el riesgo de abandono del estudiantado. En concreto, la plataforma servirá para:

- Gestionar los datos relacionados con los procesos de enseñanza y aprendizaje.
- Elaborar paneles de seguimiento (en inglés, *dashboards*), que monitoricen, analicen y muestren de manera visual tanto los resultados de todos los estudiantes en su conjunto como los resultados individualizados de cada estudiante, de acuerdo a los indicadores previamente definidos.
- Diseñar y modular la metodología empleada.
- Analizar el rendimiento académico de los estudiantes.
- Detectar a los estudiantes en riesgo de abandono de sus estudios.

Los datos que se usan en el marco de la plataforma tienen como objetivo detectar de manera temprana, rápida y eficaz el riesgo de abandono de los estudios que se cursan. Asimismo, la finalidad última del uso de los datos es mejorar el proceso de aprendizaje, con la tarea de orientación psicopedagógica que se ha de proporcionar al estudiantado. No debemos olvidar que, en la formación académica que recibe, los estudiantes tienen derecho a tutorías y asesoramiento por parte del personal de la institución para obtener orientaciones adecuadas a su proceso de aprendizaje.

Para poder proporcionar información personalizada y oportuna sobre su aprendizaje, es necesario disponer de los siguientes datos:

1. Datos históricos agregados generados en los sistemas de gestión del aprendizaje (LMS), así como los recabados de las bases de

²³ Acceso al informe *Inteligencia artificial. Decisiones Automatizadas en Cataluña*:
https://apdcat.gencat.cat/web/contenut/03-documentacio/inteligencia_artificial/documents/Informe-IA-es.pdf

datos internas de la institución, sobre el estudiantado de años anteriores que superó o abandonó los estudios de enseñanza superior.

2. Datos que provienen directamente de las instituciones educativas en las que el/la estudiante ha estado antes de acceder al sistema educativo superior.
3. Datos que solicita la institución del sistema educativo superior durante el proceso de matrícula y que proporciona el propio estudiante (SIS).
4. Datos obtenidos del comportamiento del estudiante durante la evolución de sus estudios (información referente al expediente académico).
5. Datos relacionados con circunstancias personales a lo largo de los estudios (por ejemplo, si compatibiliza estudios con trabajo, si obtiene una beca, si hace movilidad a otra institución, etc.).

Para proporcionar un mejor contexto al algoritmo predictivo y así obtener resultados más fiables, hay que caracterizar al estudiantado de acuerdo con los datos que se dispongan en cada momento. Antes del comienzo de los estudios, es necesario agrupar a los estudiantes de acuerdo con la información proporcionada por la institución educativa relacionada con la prueba de acceso al sistema educativo superior, así como por el mismo estudiante en el momento de formalizar la matrícula. Este perfilado inicial ofrecerá una primera visión de cómo puede comportarse el/la estudiante probabilísticamente. Según los datos que se vayan obteniendo, la clasificación del estudiantado podrá ir variando acorde con las métricas e indicadores definidos provenientes de los datos históricos. Con el fin de garantizar el cumplimiento de los requisitos en materia de datos del RIA, los conjuntos de datos usados para entrenar el sistema de IA para realizar las clasificaciones del estudiantado y la obtención del perfilado, son datos anonimizados y sin trazabilidad alguna.

Aunque en el siguiente apartado nos adentraremos en la evaluación concreta realizada del caso de uso, a continuación menciono algunos de los diferentes aspectos encontrados que se han considerado, para hacer más comprensible el contexto del caso de uso analizado.

Si bien durante la evaluación del proyecto se ha detectado la necesidad de incorporar determinadas categorías especiales de datos personales, como podrían ser los datos relativos a la salud para incluirlos en la información del estudiantado con necesidades de soporte educativo (NSE), en el caso de uso no se tuvo en cuenta la recopilación de estos datos.

En cuanto a los paneles de seguimiento, en un principio se contempló la posibilidad de disponer de diferentes paneles adaptados visualmente a la información seleccionada y relevante para el sujeto que podía acceder. Se distinguieron tres colores diferentes, según la situación del estudiantado en cada momento: color verde, indicando que el/la estudiante no abandonará sus estudios con una probabilidad superior al 80%; color amarillo, indicando que la probabilidad de abandono se sitúa entre el 20% y el 60%; y color rojo, indicando que el/la estudiante tiene una probabilidad de abandono de sus estudios superior al 50%. No obstante lo anterior, una vez realizada la evaluación se introdujeron los siguientes cambios:

- Se determinó que el acceso a la totalidad de la información de los paneles debía estar limitado a los tutores, para que fueran ellos los que tuvieran conocimiento de la predicción del riesgo de abandono de los estudiantes y actuaran al respecto, conforme a sus competencias y según lo establecido en la normativa de aplicación al sistema de educación superior.
- Se acotó la información que aparecía en los paneles del profesorado, limitando su acceso a los datos agregados de la evolución del estudiantado en su conjunto dentro de sus clases. Sin la posibilidad de acceder a la información individualizada de

cada estudiante, se evita el efecto inverso a lo que se pretende conseguir, si inconscientemente desde un inicio se etiqueta a ese grupo reducido de estudiantes que tienen un rendimiento bajo.

- Se eliminó la visualización de los paneles por parte del estudiantado, a efectos de cuidar de su salud mental y emocional y no provocar situaciones de agobio o ansiedad por el hecho de ver un color determinado indicativo de su rendimiento académico.

La evaluación, también puso de manifiesto el hecho que hay información de sucesos que influyen en el rendimiento del estudiantado (por ejemplo: la muerte de una persona allegada) que el sistema no tiene en cuenta. Por lo tanto, es importante tratar el riesgo potencial de abandono de los estudios caso por caso y recopilar la información relevante relacionada que debe tenerse en cuenta, no solo a nivel individual, sino también para considerar su inclusión en las variables que alimentan el sistema de IA predictiva.

Para complementar la evaluación del impacto sobre los derechos fundamentales en materia de inteligencia artificial (EIDF), se realizó una evaluación de impacto en materia de protección de datos (EIPD). Aunque no se incorpora a este documento, esta evaluación conllevó que se realizasen cambios desde la perspectiva de protección de datos, en aplicación de los principios establecidos por el Reglamento general de protección de datos (RGPD)²⁴. Fueron cambios adicionales que enriquecieron la evaluación final.

²⁴ Reglamento (UE) 2016/679 del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en lo que respecta al tratamiento de datos personales y a la libre circulación de estos datos y por el que se deroga la Directiva 95/46/CE

(Reglamento general de protección de datos - RGPD): <https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=celex%3A32016R0679>.

3. La EIDF

Planificación y alcance

Sección A Descripción y análisis del sistema de IA, incluidos los flujos de datos relacionados	¿Cuáles son los principales objetivos del sistema de IA?	<ul style="list-style-type: none"> a) Proporcionar indicadores de rendimiento académico. b) Mostrar de forma predictiva la probabilidad de abandono de los estudios superiores. c) Contribuir a la mejora del proceso de aprendizaje. <p>Con la información proporcionada por el sistema de IA, las instituciones de educación superior pueden establecer políticas para reducir la tasa de abandono prematura, así como ayudar a los tutores con la tarea de orientación psicopedagógica que deben prestar al estudiantado.</p>
	¿Cuáles son las principales características del sistema?	<p>Predecir situaciones en las cuales se puede actuar para mejorar el proceso de aprendizaje.</p>
	¿En qué países se ofrecerá?	<p>España, pero se puede hacer extensible a otras instituciones extranjeras de educación superior (dentro o fuera de la Unión Europea), con las cuales se formalicen acuerdos para impartir estudios conjuntos.</p>

	<p>¿Qué tipos de datos se tratan (personales, no personales, categorías especiales)?</p>	<ul style="list-style-type: none"> • Datos históricos agregados sobre el estudiantado de años anteriores. • Datos personales y no personales que provienen directamente de las instituciones educativas en las que el/la estudiante ha estado de forma previa al acceso al sistema educativo superior. • Datos que solicita la institución del sistema educativo superior durante el proceso de matrícula y que proporciona el mismo estudiante. • Datos obtenidos del comportamiento del estudiante durante la evolución de sus estudios (información referente al expediente académico). • Datos relacionados con circunstancias personales a lo largo de los estudios. •
	<p>Identificación de los posibles titulares de derechos: ¿quiénes son los individuos o grupos que pueden verse afectados por los sistemas de IA? ¿Se incluyen entre ellos individuos o grupos vulnerables?</p>	<p>Estudiantado.</p>
	<p>Identificación de los responsables: ¿qué personas/entidades están implicadas en el diseño, desarrollo y despliegue de los sistemas de IA? ¿Cuál es su papel?</p>	<p>Los responsables del diseño, suministro y desarrollo del sistema de IA son las propias instituciones de educación superior que quieran implementarla.</p> <p>En la gestión de los datos participará el personal de la institución.</p> <p>Asimismo, el tratamiento de la información proporcionada por los datos en los paneles de seguimiento se visualizará de manera diferente según el perfil del usuario: (i) estudiantado, (ii) profesorado, (iii) tutor/a y (iv) gestor/a.</p>

<p>Sección B</p> <p>Contexto de derechos fundamentales</p>	<p>¿Qué derechos fundamentales se ven potencialmente afectados por el uso del sistema de IA?</p>	<p>Analizados todos los derechos individuales, civiles, políticos, económicos y sociales establecidos en la Carta de los Derechos Fundamentales de la Unión Europea²⁵, se ha concluido que se ven potencialmente afectados los siguientes derechos:</p> <p><input checked="" type="checkbox"/> Dignidad humana (artículo 1)</p> <p>Justificación: falta de visión completa; pérdida de autonomía personal, si la IA toma decisiones y ofrece soluciones por cuenta de las personas sin intervención humana; presión excesiva hacia el/la estudiante e impacto en la visión de sí mismo.</p> <p><input checked="" type="checkbox"/> Respeto de la vida privada y familiar (artículo 7)</p> <p>Justificación: invasión de la intimidad a causa de la monitorización constante del rendimiento académico; afectación a la privacidad familiar; afectación a la “privacidad decisional”.</p> <p><input checked="" type="checkbox"/> Protección de datos de carácter personal (artículo 8)</p> <p>Justificación: implicación de perfilado o valoración de sujetos, datos a gran escala y uso de nuevas tecnologías.</p> <p><input checked="" type="checkbox"/> No discriminación (artículo 21)</p> <p>Justificación: el uso de algoritmos basados en datos históricos y patrones que contemplen sesgos discriminatorios del pasado puede perpetuar y amplificar perjuicios; no tener en cuenta determinados factores o variables que podrían incidir; evaluación basada en los resultados de la predicción.</p>
---	---	---

²⁵ Carta de los Derechos Fundamentales de la Unión Europea: <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:12012P/TXT&from=ES>.

		Si bien se tuvo en cuenta el derecho a la educación (artículo 14), no se encuentra afectado porque el sistema de IA no limita el derecho de acceso a la enseñanza.
	¿Qué instrumentos jurídicos internacionales/regionales de protección de los derechos humanos/fundamentales se han aplicado a nivel operativo?	La normativa relativa a protección de datos personales y la normativa relativa al sistema universitario, así como la relacionada con los colectivos afectados (por ejemplo, Estatuto del Estudiante Universitario).
	¿Cuáles son los tribunales u organismos más relevantes en materia de derechos fundamentales en el contexto de uso?	Las autoridades de control en materia de protección de datos del país/región donde se desarrolle y utilice el sistema de IA, así como los tribunales competentes de ese país/región.
	¿Cuáles son las decisiones y disposiciones más relevantes en materia de derechos humanos/fundamentales?	No aplicable (N/A).
Sección C Controles implementados	¿Qué políticas y procedimientos se han implementado para evaluar el posible impacto sobre los derechos fundamentales, incluida la participación de las partes interesadas?	No aplicable (N/A).
	¿Se ha realizado, desarrollado y aplicado una evaluación de impacto en relación con	Se ha realizado una evaluación de impacto en materia de protección de datos personales (EIPD).

	cuestiones específicas (por ejemplo, protección de datos) o algunas características del sistema (por ejemplo, el uso de la biometría)?	
Sección D Intervención de las partes interesadas y diligencia debida	¿Cuáles son los principales grupos o comunidades potencialmente afectados por el sistema de IA, incluido su desarrollo?	Estudiantado.
	¿Qué partes interesadas, además de los individuos o grupos que puedan verse afectados por los sistemas de IA, deben participar (por ejemplo, la sociedad civil y las organizaciones internacionales, expertos, asociaciones industriales, periodistas)?	Familias y profesorado.
	¿Hay otros titulares de obligaciones que deban participar, aparte del proveedor y el responsable del despliegue de IA (por ejemplo, autoridades nacionales, organismos gubernamentales)?	Autoridad de control en protección de datos, Departamento de Educación/Universidades; Agencia Española de Supervisión de la IA; Comisión de IA de la Generalitat de Catalunya. Unidad/órgano competente de la Universidad.
	¿Han participado en el proceso de evaluación los socios comerciales, incluidos los proveedores de servicios (por ejemplo, subcontratistas en sistemas de IA y conjuntos de datos)?	No
	¿Ha llevado a cabo el proveedor de IA una evaluación de su cadena de suministro, para identificar si las actividades de los proveedores/contratistas que participan en el desarrollo de productos/servicios pueden afectar a los derechos fundamentales?	No aplicable (N/A).

	¿El proveedor ha promovido estándares o auditorías de derechos fundamentales, para garantizar el respeto de los derechos fundamentales entre los proveedores?	
	¿El proveedor y el responsable del despliegue de la IA han comunicado públicamente las posibles repercusiones del sistema de IA en los derechos fundamentales?	No es obligatoria su publicación, pero es recomendable indicar/añadir un resumen sobre el análisis realizado (tanto de la EIPD como la evaluación en derechos fundamentales). Principio de transparencia; confianza en el sistema de IA.
	¿El proveedor y el responsable del despliegue de la IA han proporcionado formación sobre estándares de derechos fundamentales al personal encargado de la gestión y la contratación relacionadas con el sistema de IA?	No aplicable (N/A).

Matrices de riesgo

Tabla 1. Probabilidad

Baja	El riesgo de perjuicio es improbable o altamente improbable
Media	El riesgo puede producirse
Alta	Existe una alta probabilidad de que se produzca el riesgo
Muy alta	Es muy probable que se produzca el riesgo

Tabla 2. Exposición

Baja	Pocos o muy pocos de la población identificada de titulares de derechos están potencialmente afectados
Media	Parte de la población identificada está potencialmente afectada
Alta	La mayoría de la población identificada está potencialmente afectada
Muy alta	Casi toda la población identificada está potencialmente afectada

Tabla 3. Probabilidad de que se produzca un perjuicio (*likelihood*)

		Probabilidad			
		Baja	Media	Alta	Muy alta
Exposición	Baja	B	B/M	B/A	B/MA
	Media	M/B	M	M/A	M/MA
	Alta	A/B	A/M	A	A/MA
	Muy alta	MA/B	MA/M	MA/H	MA

Probabilidad			
Baja	Media	Alta	Muy alta

Tabla 4. Gravedad del perjuicio

Baja	Los individuos y grupos afectados pueden encontrar sólo perjuicios menores en el ejercicio de sus derechos y libertades
Media	Los individuos y grupos afectados pueden encontrarse con importantes perjuicios
Alta	Los individuos y grupos afectados pueden enfrentarse a graves perjuicios
Muy alta	Los individuos y grupos afectados pueden encontrarse con perjuicios graves o incluso irreversibles

Tabla 5. Esfuerzo para superar el perjuicio y revertir los efectos adversos

Bajo	Los perjuicios sufridos pueden superarse sin problemas (por ejemplo, tiempo dedicado a modificar la información, molestias, irritaciones, etc.)
Medio	Los perjuicios sufridos pueden superarse a pesar de algunas dificultades (por ejemplo, costes adicionales, miedo, incomprensión, estrés, pequeñas dolencias físicas, etc.).
Alto	Los perjuicios sufridos pueden superarse aunque con serias dificultades (por ejemplo, pérdidas económicas, daños materiales, empeoramiento de la salud, etc.).
Muy alto	Los perjuicios sufridos pueden no superarse (por ejemplo, dolencias psicológicas o físicas de larga duración, muerte, etc.).

Tabla 6. Gravedad (severity)

		Gravedad			
		Baja	Media	Alta	Muy alta
Esfuerzo	Bajo	B	B/M	B/A	B/MA
	Medio	M/B	M	M/A	M/MA
	Alto	A/B	A/M	A	A/MA
	Muy alto	MA/B	MA/M	MA/A	MA

Gravedad			
Baja	Media	Alta	Muy alta

Tabla 1A. Recogida de datos y análisis de riesgos

Derechos/libertades potencialmente afectados	Descripción del impacto	Probabilidad			Gravedad		
		Probabilidad de resultados adversos	Exposición	Probabilidad	Gravedad	Esfuerzo	Gravedad
Dignidad humana	El algoritmo tiene en cuenta unos parámetros generados por datos históricos recabados dentro de un	[Alta] Existe una alta probabilidad de	[Baja] La exposición es baja, ya que	[Media]	[Media] Aunque se trate de colectivos	[Medio] Los perjuicios sufridos pueden	[Media]

	<p>contexto socioeconómico concreto, pero no todos los que podrían tener una afectación directa en el rendimiento académico actual (por ejemplo: las mejoras docentes/adaptaciones solicitadas; las personas que no se identifican con un género determinado, acceso a nuevas tecnologías, etc.).</p>	<p>que se produzca el riesgo, porque la información proporcionada con los datos actuales no recaba toda la información para englobar a todas las categorías de grupos que se podrían ver afectados.</p>	<p>afecta a un número limitado de casos de falta de información.</p>		<p>reducidos los que se puedan ver afectados por no tener en cuenta determinados parámetros, el estudiantado afectado podría encontrarse con importantes perjuicios. El algoritmo podría indicar una predisposición determinada a que ese colectivo reducido hará unas actuaciones concretas, las cuales no reflejan la realidad de su situación.</p>	<p>superarse a pesar de algunas dificultades. En el caso del estudiantado que solicita mejoras docentes, se informa de forma previa al profesorado y a los tutores. Por lo que respecta a otros parámetros que no se tienen en cuenta, pero que pueden afectar a colectivos minoritarios, pueden tenerse en cuenta para su incorporación en el sistema de IA para mejorarlo.</p>	
--	---	---	--	--	---	--	--

<p>Respeto de la vida privada y familiar</p>	<p>Seguimiento constante del rendimiento académico; afectación a la privacidad familiar; afectación a la “privacidad decisional”.</p>	<p>[Alta]</p> <p>Existe una alta probabilidad de que se produzca el riesgo. Aunque los datos por separado son necesarios dentro de los diferentes procedimientos que lleva a cabo la institución, disponer de cierta información recopilada de forma conjunta para una finalidad concreta conlleva una afectación sobre el control de dicha información.</p>	<p>[Muy alta]</p> <p>Dado que afectaría a todo el estudiantado.</p>	<p>[Muy alta]</p>	<p>[Baja]</p> <p>El estudiantado afectado puede encontrar perjuicios menores en el ejercicio de sus derechos y libertades, dado que la información se encuentra en el marco de las funciones de la institución de llevar a cabo la enseñanza.</p>	<p>[Bajo]</p> <p>El personal de las instituciones de educación superior tiene unos deberes y obligaciones en el marco de sus funciones para salvaguardar los derechos del estudiantado. Asimismo, la institución debe formar a su personal en la materia, para que conozca la normativa de aplicación y pueda actuar frente a las diferentes situaciones en las que se pueda encontrar.</p>	<p>[Baja]</p>
---	---	--	---	-------------------	---	---	---------------

<p>Protección de datos personales</p>	<p>El algoritmo recoge datos a gran escala y hace uso de nuevas tecnologías.</p> <p>Además, el algoritmo implica perfilado o valoración de sujetos, para poder hacer el análisis predictivo de abandono.</p>	<p>[Media]</p> <p>Existe el riesgo de que se elaboren perfiles y predicciones inexactos. Por lo tanto, es necesario realizar una evaluación de impacto en materia de protección de datos.</p>	<p>[Muy alta]</p> <p>Dado que afectaría a todo el estudiantado.</p>	<p>[Alta]</p>	<p>[Media]</p> <p>La elaboración de perfiles incorrectos repercute negativamente en la representación correcta del rendimiento de los estudiantes y de los resultados esperados. Se ha considerado si este es un caso en el que los estudiantes también tienen derecho a no ser perfilados.</p>	<p>[Medio]</p> <p>En aplicación del RGPD, deben establecerse las medidas organizativas adecuadas y proteger los derechos de los estudiantes en relación con el tratamiento de datos, pues la forma en que se generan y utilizan los perfiles puede requerir algunos cambios en el diseño de los sistemas de IA (por ejemplo, su puesta a punto) y en su uso.</p>	<p>[Media]</p>
--	--	---	---	---------------	---	--	----------------

<p>No discriminación</p>	<p>Teniendo en cuenta que el algoritmo compara datos históricos, obtiene datos de otras instituciones y recoge datos de la matrícula, puede haber patrones que contemplen sesgos discriminatorios del pasado que pueden perpetuarse y amplificar perjuicios; no tener en cuenta determinados factores o variables que podrían incidir; evaluación basada en los resultados de la predicción.</p>	<p>[Media]</p> <p>Dado el peso de las variables en la consideración del modelo predictivo del riesgo de abandono del modelo predictivo</p>	<p>[Muy alta]</p> <p>Dado que el impacto afectaría potencialmente a todos los estudiantes a los que se les aplicaría el algoritmo.</p>	<p>[Alta]</p>	<p>[Media]</p> <p>La clasificación del estudiantado puede estar sesgada y proporcionar información errónea sobre los primeros indicadores de riesgo de abandono, lo que da lugar a un trato desigual injustificado.</p>	<p>[Medio]</p> <p>La clasificación del estudiantado no es estática, por lo que los datos iniciales no los encasillarán en una clasificación determinada, sino que irá cambiando según su evolución en los estudios.</p>	<p>[Media]</p>
---------------------------------	--	--	--	---------------	---	---	----------------

Tabla 7. Impacto general del riesgo

		Gravedad			
		Baja	Media	Alta	Muy alta
Probabilidad	Baja				
	Media				
	Alta				
	Muy alta				

Tabla 3A. Gestión de riesgos (I)

Derecho/libertad afectados	Probabilidad	Gravedad	Impacto global	Medidas para prevenir/mitigar los impactos
Dignidad humana	[Media]	[Media]	[Media]	<ul style="list-style-type: none"> • Utilizar el modelo predictivo como herramienta de apoyo a la toma de decisiones y no como herramienta de toma de decisiones automatizadas; uso limitado de los resultados proporcionados per el sistema de IA. • No hacer visible para el estudiantado los índices del riesgo de abandono. • Poner a disposición del personal de la institución directrices de uso del sistema de IA (política de uso).
Respeto de la vida privada y familiar	[Muy alta]	[Baja]	[Media]	<ul style="list-style-type: none"> • Diseñar el modelo predictivo de forma que garantice en todo momento que se tiene el control de los datos. • Limitar el acceso al perfil individual, sin que se puedan visualizar los perfiles de los otros estudiantes. • Que la herramienta de predicción no tenga en cuenta las interlocuciones ni comunicaciones del estudiante con el profesorado o entre sí. • Que la herramienta sea utilizada como apoyo al establecimiento de medidas y no como herramienta de toma de decisiones automatizadas.
Protección de datos	[Alta]	[Media]	[Media]	<ul style="list-style-type: none"> • Restringir el acceso a los datos: acceso total a los tutores y sólo datos agregados a los profesores.
No discriminación	[Alto]	[Media]	[Media]	<ul style="list-style-type: none"> • Revisar periódicamente que los datos introducidos en las bases de datos no generan perfiles discriminatorios. • Hacer el cálculo inicial para la construcción de perfiles de forma periódica a partir de la incorporación de nuevos datos en la base de datos, para que los nuevos comportamientos puedan revisar los potenciales sesgos. • Revisar de forma periódica que el modelo de predicción no es discriminatorio y se diseña un algoritmo sensible a la discriminación y al potencial sesgo.

Tabla 3A. Gestión de riesgos (II)

Derecho/libertad afectados	Probabilidad (residual)	Gravedad (residual)	Impacto residual
Dignidad humana	[Media]	[Baja]	[Bajo]
Respecte de la vida privada i familiar	[Alta]	[Baja]	[Medio]
Protección de datos	[Media]	[Media]	[Medio]
No discriminación	[Media]	[Media]	[Medio]

4. Comentarios

Formar parte de este grupo de trabajo propuesto por la Autoridad Catalana de Protección de Datos (APDCAT) ha dado la posibilidad de trabajar con personas pertenecientes a otros sectores. Esto ha constituido una gran mejora en el análisis del caso de uso expuesto, dado que se han aportado diferentes perspectivas que han evidenciado distintas sensibilidades respecto a la evaluación de cada uno de los derechos fundamentales.

Realizar la evaluación de impacto relativa a los derechos fundamentales para los sistemas de IA de alto riesgo, nos muestra la importancia de adoptar un planteamiento amplio en el momento de establecer la identificación de los riesgos, que abarque todos y cada uno de los derechos fundamentales, así como las medidas mitigadoras de los riesgos identificados. Uno de los aspectos más difíciles ha sido concretar los riesgos residuales, ya que la determinación del riesgo resultante tras las medidas

de mitigación previstas hace que el análisis de escenarios en el ámbito de los derechos fundamentales no sea fácil.

El enfoque del análisis desde la óptica de un delegado de protección de datos (DPD) pone de manifiesto la prioridad de salvaguardar un derecho (el de la protección de datos personales) que se encuentra ampliamente regulado en comparación con otros derechos fundamentales, así como la necesidad de abordarlo conjuntamente con otros derechos que están íntimamente relacionados, para cumplir con las obligaciones derivadas de todo el marco regulatorio.

Asimismo, se ha puesto de manifiesto lo siguiente:

- La necesidad de formar a todos los agentes implicados en materia de derechos fundamentales.
- Evitar los listados de cumplimiento (*checklists*) relacionados con los derechos fundamentales, dado que no se profundiza en los diferentes aspectos que los integran.

- Dar a conocer la repercusión de los sistemas de IA en el ámbito de la educación.
- Comprender la definición de sistemas de IA proporcionada por el RIA (artículo 3.1) y la repercusión de los sistemas de IA considerados de alto riesgo del anexo III.
- Establecer la evaluación por defecto y desde el diseño.
- Concienciar a los responsables del despliegue de los sistemas de IA de alto riesgo de la necesidad de llevar a cabo y disponer de la evaluación de impacto relativa a los derechos fundamentales.
- Adaptar las evaluaciones al contexto en que nos encontremos en cada momento.
- Aunque la evaluación de impacto relativa a los derechos fundamentales para los sistemas de IA de alto riesgo no tiene que ir de la mano de la EIPD, es recomendable realizarla.

Caso de uso 2: Una herramienta para la gestión de los recursos humanos

1. El contexto

Este caso de uso se enmarca en el proceso de selección de personas que tiene implementado la organización y, por tanto, se guía por los principios de actuación de los que esta se ha dotado. Lo lleva a cabo el departamento de recursos humanos de la entidad, directamente o a través de proveedores.

En concreto, la organización ha definido diferentes palancas en su plan director de gestión de personas, para i) Impulsar una cultura de equipo ilusionante, comprometida con el nuevo proyecto, colaborativa y ágil, y al mismo tiempo promover un liderazgo cercano, motivador, no jerárquico, con capacidades transformadoras; ii) Impulsar nuevas formas de trabajo, con respeto por la diversidad, la igualdad de oportunidades, la inclusión y la no discriminación, e incorporando la sostenibilidad en los procesos de recursos humanos; iii) Transformar la gestión del modelo de desarrollo de las personas: más proactivo en la capacitación de los equipos, con foco en habilidades críticas; iv) Desarrollar una propuesta de valor única y diferencial para el empleado; y v) Evolucionar hacia una cultura *data driven* de la función de personas, a través de la optimización de la estructura de datos y de la aplicación de inteligencia artificial y nuevas tecnologías, para facilitar el análisis de la información y tomar decisiones basadas en datos en materia de personas.

En el marco de esta última palanca, se propuso, analizó y finalmente se implementó el caso de uso que seguidamente exponemos.

Antes de entrar en el detalle, merece la pena señalar que la entidad es una organización madura en los esquemas de cumplimiento de protección de datos y seguridad de la información, y avanzada en relación con la inteligencia artificial y su gobernanza, en la medida que ya había adoptado e implementado, entre otras, las medidas siguientes:

1. Creación y aplicación de metodologías internas para el desarrollo e implementación de sistemas de inteligencia artificial, que incluye los 144 controles establecidos por la AEPD en su guía de auditorías de tratamientos que incluyen inteligencia artificial. Ello permite que los sistemas de IA desarrollados en interno cumplan **por defecto y desde el diseño** con multitud de controles, tales como su inventario, asociación al tratamiento de datos al que dan servicio, evaluación de su necesidad y proporcionalidad, valoración de calidad del dato (lo que incluye, entre otros, el análisis y mitigación de los posibles sesgos), de su explicabilidad, transparencia y robustez, tanto en el sentido del Reglamento de inteligencia artificial como del Reglamento general de protección de datos. También, medidas en el ámbito de la validación y verificación de la bondad del sistema.

2. Análisis de estos casos de uso antes de su implementación en el marco de la evaluación de impacto en protección de datos, con ampliación de su objeto, por parte de los equipos de asesoría jurídica innovación y privacidad, asesoría jurídico-laboral, sistemas (CDO –Chief data officer-, donde se incluye desde una perspectiva técnica el equipo de IA responsable) y seguridad de la información (CISO –Chief Information Security Officer). Todo ello permite realizar un segundo control sobre la bondad de estos sistemas.

3. Evaluación y sanción, si procede, de estas iniciativas por los comités corporativos correspondientes.

2. El proyecto

Desarrollo y aplicación de un sistema de inteligencia artificial (4 modelos de *machine learning*), basado en la experiencia previa de la organización en el marco de la selección de personal para la cobertura de ciertas vacantes. En concreto, el sistema realiza una tarea muy concreta y limitada: la predicción de un dato adicional para cada empleado/a, que consiste en la probabilidad de que se adecue a esa vacante en función de datos de la relación laboral y las características del centro de destino. El resultado se integra como un dato más en el marco del proceso de selección de personal, que en todo caso es gestionado y liderado por los técnicos especializados de recursos humanos quienes, junto con el resto de información disponible y de acuerdo con los procesos internos de la entidad, podrán utilizar esta información en sus funciones de selección.

En este sentido, y para mejor entendimiento, ante una vacante específica el sistema permitirá a los técnicos de recursos humanos visualizar a los empleados de la entidad ordenados en base a esa probabilidad de adecuación a la vacante, que pueden o no haber solicitado. Será, en todo caso, el personal especializado de recursos humanos quien decida utilizar o no esta información adicional.

La finalidad del sistema es, por tanto, ayudar a la eficiencia del proceso de selección de personal y mejorarla, poniendo a disposición del personal del departamento de recursos humanos información sistematizada que, de otra manera, deberían compilar y estructurar de manera manual. En ningún caso el sistema adopta decisión alguna.

3. La EIDF

Planificación y alcance

Sección A Descripción y análisis del sistema de IA, incluidos los flujos de datos relacionados	¿Cuáles son los principales objetivos del sistema de IA?	Mejorar el proceso de selección de ciertas vacantes, poniendo a disposición del personal especializado del departamento de recursos humanos información adicional sobre la adecuación de un candidato a una determinada vacante, en base a criterios objetivos. En particular: <ol style="list-style-type: none"> i. Eficiencia y ahorro de tiempo del personal de selección, a la hora de verificar ciertos requisitos objetivos relevantes para la cobertura de una vacante. ii. Objetivación del proceso. iii. Promoción de la proactividad de la organización en el ofrecimiento de la vacante a candidatos/as que no la han solicitado.
	¿Cuáles son las principales características del sistema?	Se configura como una herramienta de soporte que realiza una tarea complementaria y limitada, consistente en la generación de un dato adicional para cada empleado/a en el marco del proceso de selección de ciertas vacantes. Permite que los técnicos de recursos humanos puedan visualizar al personal de la entidad ordenado en base a su adecuación para una vacante determinada.
	¿En qué países se ofrecerá?	España
	¿Qué tipos de datos se tratan (personales, no personales, categorías especiales)?	<ul style="list-style-type: none"> • Datos relativos a la actividad laboral (nivel y grupo, funciones asignadas, datos de productividad, calidad y cumplimiento). • Características del centro de destino (tamaño y tipo de centro).

	Identificación de los posibles titulares de derechos: ¿quiénes son los individuos o grupos que pueden verse afectados por los sistemas de IA? ¿Se incluyen entre ellos individuos o grupos vulnerables?	<p>Personal de la organización.</p>
	Identificación de los responsables: ¿qué personas/entidades están implicadas en el diseño, desarrollo y despliegue de los sistemas de IA? ¿Cuál es su papel?	<p>El departamento de recursos humanos, asesoría jurídica en general (incluyendo asesoría jurídico-laboral), DPD, sistemas (CDO), seguridad de la información (CISO). El primero ha desarrollado y utilizado la herramienta. Los segundos son equipos evaluadores que han acompañado en el desarrollo y aplicación del sistema, estableciendo e implementando, en su caso, los controles necesarios que excedían del ámbito del desarrollo.</p>
Sección B Contexto de derechos fundamentales	¿Qué derechos fundamentales se ven potencialmente afectados por el uso del sistema de IA?	<input checked="" type="checkbox"/> Protección de datos personales. <input checked="" type="checkbox"/> No discriminación. <input checked="" type="checkbox"/> Igualdad entre hombres y mujeres. <input checked="" type="checkbox"/> Derecho a la información y a la consulta de los trabajadores en la empresa.
	¿Qué instrumentos jurídicos internacionales/regionales de protección de los derechos humanos/fundamentales se han aplicado a nivel operativo?	<p>La normativa de protección de datos personales, la de relaciones laborales (entre otros, Estatuto de los trabajadores y “Guía práctica y herramienta sobre la obligación empresarial de información sobre el uso de algoritmos en el ámbito laboral”, del Ministerio de Trabajo) y el Reglamento de inteligencia artificial.</p>

	<p>¿Cuáles son los tribunales u organismos más relevantes en materia de derechos fundamentales en el contexto de uso?</p>	<p>Autoridades de Protección de Datos Ministerio de Trabajo</p>
	<p>¿Cuáles son las decisiones y disposiciones más relevantes en materia de derechos humanos/fundamentales?</p>	<p>Acervo comunitario, tanto en materia del derecho fundamental a la protección de datos como del derecho a la igualdad y no discriminación.</p>
<p>Sección C Controles implementados</p>	<p>¿Qué políticas y procedimientos se han implementado para evaluar el posible impacto sobre los derechos fundamentales, incluida la participación de las partes interesadas?</p> <p>¿Se ha realizado, desarrollado y aplicado una evaluación de impacto en relación con cuestiones específicas (por ejemplo, protección de datos) o algunas características del sistema (por ejemplo, el uso de la biometría)?</p>	<p>El sistema de IA se ha desarrollado en base a metodología interna de desarrollo, que incorpora por defecto y desde el diseño los controles establecidos por la AEPD en su Guía de requisitos para auditorías de tratamientos que incluyan IA (tales como el inventario y registro del sistema, el control de los datos y del sesgo, la intervención humana, la explicabilidad, la validación y verificación del sistema, etc.).</p> <p>Adicionalmente, se ha realizado una evaluación de impacto en protección de datos que incluye ya el pronunciamiento sobre la afectación al resto de derechos fundamentales, en base a la metodología propia de la organización. En concreto, la EIPD ha contemplado:</p> <ol style="list-style-type: none"> i. El análisis del cumplimiento del tratamiento de datos de la normativa de protección de datos (lo que incluye las obligaciones jurídicas, las de seguridad y las relativas al propio sistema de IA). ii. El análisis de los potenciales daños y perjuicios materiales o inmateriales que podrían causarse, así como en su caso los controles identificados y las medidas mitigadoras establecidas.

		<p>iii. La afectación a los derechos fundamentales, cuyo elenco se ha configurado tanto con los establecidos en la Carta de derechos fundamentales de la UE como de la Constitución Española.</p> <p>Tanto las metodologías señaladas como la propia EIPD han sido evaluadas por el comité corporativo de la entidad, a propuesta de un equipo evaluador integrado por miembros de AJ, DPD, CISO y CDO.</p> <p>Adicionalmente, se ha contemplado el correspondiente circuito de información a la representación de los trabajadores.</p>
Sección D Intervención de las partes interesadas y diligencia debida	¿Cuáles son los principales grupos o comunidades potencialmente afectados por el sistema de IA, incluido su desarrollo?	Personal de la entidad.
	¿Qué partes interesadas, además de los individuos o grupos que puedan verse afectados por los sistemas de IA, deben participar (por ejemplo, la sociedad civil y las organizaciones internacionales, expertos, asociaciones industriales, periodistas)?	Equipos evaluadores establecidos por la entidad (DPD, CDO y CISO), además de asesoría jurídica incluyendo la laboral.
	¿Hay otros titulares de obligaciones que deban participar, aparte del proveedor y el responsable del despliegue de IA (por ejemplo, autoridades nacionales, organismos gubernamentales)?	No
	¿Han participado en el proceso de evaluación los socios comerciales, incluidos los proveedores	No

	de servicios (por ejemplo, subcontratistas en sistemas de IA y conjuntos de datos)?	
	<p>¿Ha llevado a cabo el proveedor de IA una evaluación de su cadena de suministro, para identificar si las actividades de los proveedores/contratistas que participan en el desarrollo de productos/servicios pueden afectar a los derechos fundamentales?</p> <p>¿El proveedor ha promovido estándares o auditorías de derechos fundamentales, para garantizar el respeto de los derechos fundamentales entre los proveedores?</p>	N/A
	¿El proveedor y el responsable del despliegue de la IA han comunicado públicamente las posibles repercusiones del sistema de IA en los derechos fundamentales?	La organización ha comunicado a su personal y la representación de los trabajadores el uso del sistema de IA, así como su finalidad, lógica y consecuencias de acuerdo con lo previsto tanto en la normativa de protección de datos como en la normativa laboral.
	¿El proveedor y el responsable del despliegue de la IA han proporcionado formación sobre estándares de derechos fundamentales al personal encargado de la gestión y la contratación relacionadas con el sistema de IA?	N/A

Matrices de riesgo

Tabla 1. Probabilidad

Baja	El riesgo de perjuicio es improbable o altamente improbable
Media	El riesgo puede producirse
Alta	Existe una alta probabilidad de que se produzca el riesgo
Muy alta	Es muy probable que se produzca el riesgo

Tabla 2. Exposición

Baja	Pocos o muy pocos titulares de derechos de la población identificada están potencialmente afectados
Media	Parte de la población identificada está potencialmente afectada
Alta	La mayoría de la población identificada está potencialmente afectada
Muy alta	Casi toda la población identificada está potencialmente afectada

Tabla 3. Probabilidad de que se produzca un perjuicio (*likelihood*)

		Probabilidad			
		Baja	Media	Alta	Muy alta
Exposición	Baja	B	B/M	B/A	B/MA
	Media	M/B	M	M/A	M/MA
	Alta	A/B	A/M	A	A/MA
	Muy alta	MA/B	MA/M	MA/H	MA

Probabilidad			
Baja	Media	Alta	Muy alta

Tabla 4. Gravedad del perjuicio

Baja	Los individuos y grupos afectados pueden encontrar sólo perjuicios menores en el ejercicio de sus derechos y libertades
Media	Los individuos y grupos afectados pueden encontrarse con importantes perjuicios
Alta	Los individuos y grupos afectados pueden enfrentarse a graves perjuicios
Muy alta	Los individuos y grupos afectados pueden encontrarse con perjuicios graves o incluso irreversibles

Tabla 5. Esfuerzo para superar el perjuicio y revertir los efectos adversos

Bajo	Los perjuicios sufridos pueden superarse sin problemas (por ejemplo, tiempo dedicado a modificar la información, molestias, irritaciones, etc.)
Medio	Los perjuicios sufridos pueden superarse a pesar de algunas dificultades (por ejemplo, costes adicionales, miedo, incomprensión, estrés, pequeñas dolencias físicas, etc.)
Alto	Los perjuicios sufridos pueden superarse aunque con serias dificultades (por ejemplo, pérdidas económicas, daños materiales, empeoramiento de la salud, etc.)
Muy alto	Los perjuicios sufridos pueden no superarse (por ejemplo, dolencias psicológicas o físicas de larga duración, muerte, etc.)

Tabla 6. Gravedad (severity)

		Gravedad			
		Baja	Media	Alta	Muy alta
Esfuerzo	Bajo	B	B/M	B/A	B/MA
	Medio	M/B	M	M/A	M/MA
	Alto	A/B	A/M	A	A/MA
	Muy alto	MA/B	MA/M	MA/A	MA

Gravedad			
Baja	Media	Alta	Muy alta

Tabla 1A. Recogida de datos y análisis de riesgos

Derechos/libertades potencialmente afectados	Descripción del impacto	Probabilidad			Gravedad		
		Probabilidad de resultados adversos	Exposición	Probabilidad	Gravedad	Esfuerzo	Gravedad
Protección de datos	El algoritmo requiere la utilización de datos de la relación laboral del personal de la organización. Por tanto, cualquier uso que infringiese la normativa de protección de datos podría producir una afectación a este derecho	<p>[Baja]</p> <p>Ya que tanto la creación del sistema como su uso han sido sometidos una evaluación de impacto en protección de datos</p>	<p>[Muy alta]</p> <p>El impacto afecta potencialmente a todas las personas a las que se aplica el algoritmo</p>	[Media]	<p>[Baja]</p> <p>Aunque el tratamiento de datos se realiza en el marco de la selección de personal, el sistema realiza una tarea muy limitada con datos limitados al contexto laboral y objetivos</p>	<p>[Medio]</p> <p>Si el tratamiento fuese no conforme, se podrían adoptar medidas para cumplir los deberes de transparencia e información, interrumpir el tratamiento en su caso e, incluso, suprimir los datos generados por el sistema</p>	[Baja]

<p>No discriminación e igualdad entre hombres y mujeres</p>	<p>El algoritmo se entrena con datos históricos, por lo que de haber sesgos y perpetuarse podrían darse situaciones de discriminación, que podrían incluir y afectar a la igualdad entre hombres y mujeres</p>	<p>[Baja]</p> <p>Se ha verificado la ausencia de sesgos, tanto en la creación como en el uso del modelo, incluyendo controles para corregir los sesgos históricos que contenían los datos de entrenamiento</p>	<p>[Media]</p> <p>El impacto potencial afectaría a parte del personal de la entidad</p>	<p>[Baja]</p>	<p>[Media]</p> <p>El sistema provee de un dato adicional a disposición de los técnicos especializados de recursos humanos que, en todo caso, lideran la selección del personal. El sistema no adopta decisiones</p>	<p>[Alto]</p> <p>Si bien el potencial perjuicio, consistente en un resultado de adecuación incorrecto, podría corregirse en el propio proceso de gestión de la vacante a través la actuación del técnico de recursos humanos o a posteriori, se califica como alto para el caso que fuese resuelto una vez adjudicada la vacante</p>	<p>[Media]</p>
<p>Derecho a la información y a la consulta de los trabajadores</p>	<p>El sistema se incluye en el marco de la gestión de la selección de personal, de la</p>	<p>[Baja]</p> <p>Tanto en la creación como en el uso del sistema</p>	<p>[Muy alta]</p> <p>El impacto afecta potencialmente</p>	<p>[Media]</p>	<p>[Baja]</p> <p>El potencial perjuicio de esta afectación sería</p>	<p>[Bajo]</p> <p>La mitigación de este</p>	<p>[Baja]</p>

	entidad por lo que un incumplimiento de la normativa laboral, en particular en relación con las obligaciones de información a los trabajadores o a sus representantes, podría afectar este derecho	han participado, entre otros, los equipos de asesoría jurídica en general, incluyendo la laboral. Los procedimientos establecidos garantizan los deberes de información tanto de los empleados como de sus representantes	a todas las personas a las que se aplica el algoritmo		la falta de información preceptiva a la representación laboral	potencial perjuicio pasaría por la subsanación de la información deficiente	
--	--	---	---	--	--	---	--

Tabla 7. Impacto general del riesgo

		Gravedad [derecho/libertad afectados]			
		Baja	Media	Alta	Muy alta
Probabilidad	Baja				
	Media				
	Alta				
	Muy alta				

Tabla 2A. Gestión de riesgos

Derecho/libertad afectados	Probabilidad	Gravedad	Impacto global	Medidas para prevenir/mitigar los impactos
Protección de datos	[Media]	[Baja]	[Bajo]	<ul style="list-style-type: none"> N/A. Medidas adoptadas tanto en el desarrollo del algoritmo como antes de su uso (tales como la EIPD, facilitar información, etc.).
No discriminación e igualdad de hombres y mujeres	[Baja]	[Media]	[Bajo]	<ul style="list-style-type: none"> Formación a los técnicos de recursos humanos, para evitar la excesiva confianza en el resultado del algoritmo.
Derecho a la información y a la consulta de los trabajadores	[Media]	[Baja]	[Bajo]	<ul style="list-style-type: none"> N/A. Medidas adoptadas tanto en el desarrollo del algoritmo como antes de su uso. Se ha facilitado la información preceptiva tanto al personal de la entidad como a la representación laboral, siguiendo los modelos establecidos por el Ministerio de Trabajo.

4. Comentarios

Tal y como se ha expuesto, la organización cuenta por defecto y desde el diseño con estructuras, procedimientos y controles que se centran en varias de las cuestiones cubiertas por el RIA. Esto ha permitido que la EIDF sobre el caso de uso y su impacto en los derechos fundamentales arroje un nivel de riesgo de impacto bajo.. Esas estructuras, procedimientos y controles permiten que el sistema se desarrolle en un *framework* controlado y que lleva al *data scientist* a incluir en el desarrollo e implementación del sistema de IA, por defecto, ciertas medidas que mitigan los riesgos que a priori ha identificado la entidad.

Adicionalmente, para el caso que no se hayan identificado riesgos o medidas a priori, la participación de los distintos equipos evaluadores y su acompañamiento en el desarrollo del sistema permite que se identifiquen y mitiguen en el momento del desarrollo/implementación.

Caso de uso 3: Una herramienta de imágenes médicas impulsada por IA para la detección del cáncer

1. El contexto

En Europa se está produciendo un avance importante en el desarrollo de herramientas de IA que usan imágenes de cáncer. Ya en 2012, un equipo investigador de la Universidad de Maastricht propuso el concepto de “radiomics”, que hace referencia al método de extracción de un gran número de características de las imágenes médicas, utilizando algoritmos de caracterización de datos. El incremento del desarrollo de sistemas de IA destinados a usar imágenes médicas para tratar el cáncer se puede ilustrar consultando el número de publicaciones sobre “radiomics AI” en el portal PubMed: 42 resultados en 2019; 99 resultados en 2020; 165 resultados en 2021; 235 resultados en 2022; 309 resultados en 2023; y 338 resultados en 2024. Por tanto, estamos ante un tipo de sistema de inteligencia artificial que cada vez será más común, no sólo en el mundo académico sino también en el mundo asistencial.

Además, hay tipos de cáncer comunes en el mundo en que los pacientes reciben un alto grado de sobretatamiento y un alto grado de efectos considerables que se podrían evitar. En consecuencia, el uso de un sistema de IA que analizara las imágenes médicas de los pacientes permitiría a los profesionales sanitarios disponer de una herramienta de soporte para predecir la respuesta a la terapia y, por tanto, poder ajustar la terapia para que fuera lo más eficiente posible. Es decir, permitiría alcanzar el objetivo marcado con el mínimo tratamiento. Además, estos sistemas de IA también

proporcionarían tanto a los profesionales sanitarios como a los pacientes una predicción de la evolución del paciente durante los próximos años.

2. El proyecto

El proyecto se estructura en una primera fase que consiste en el desarrollo de un sistema de IA basado en imágenes médicas, que es entrenado con la información de 5.000 pacientes provenientes de diez países de Europa. Por tanto, el conjunto de datos de entrenamiento es multicéntrico.

Además, se prevé una segunda fase del proyecto que implicará la validación del sistema de IA en ocho centros sanitarios de todo el mundo, fuera de Europa. El objetivo de esta segunda fase es que el sistema de IA sea probado en un centro sanitario ubicado en Asia, otro en África y otro en América del Sur.

3. La EIDF

Planificación y alcance

Sección A Descripción y análisis del sistema de IA, incluidos los flujos de datos relacionados	¿Cuáles son los principales objetivos del sistema de IA?	Mejorar el tratamiento de los pacientes con cáncer X mediante la predicción de: a) Respuesta del paciente al tratamiento. b) Efectos secundarios (toxicidad y sensibilidad). c) Previsiones para los próximos cinco años.
	¿Cuáles son las principales características del sistema?	Reconocimiento de imágenes basado en la IA. Uso de imágenes médicas para predecir la respuesta de un paciente a un determinado tratamiento y para ayudar a los profesionales sanitarios a determinar su aplicación en casos concretos, así como el nivel de uso una vez que se aplique.
	¿En qué países se ofrecerá?	Distribución mundial.
	¿Qué tipos de datos se tratan (personales, no personales, categorías especiales)?	<ul style="list-style-type: none"> • Datos demográficos (sexo, edad y país). • Características del cáncer (tipo de cáncer y zona afectada). • Estadio del cáncer y subtipo molecular. • Información relativa al tratamiento previo. • Régimen de tratamiento (esquema y duración). • Informe de patología (post-tratamiento).
	Identificación de los posibles titulares de derechos: ¿quiénes son los individuos o grupos	Personas entre 18 y 85 años.

	<p>que pueden verse afectados por los sistemas de IA? ¿Se incluyen entre ellos individuos o grupos vulnerables?</p>	<p>Dado que todas las personas implicadas están afectadas por el cáncer, deben considerarse vulnerables, debido a sus condiciones de salud y a las relaciones entre estas condiciones y la finalidad del sistema de IA.</p>
	<p>Identificación de los responsables: ¿qué personas/entidades están implicadas en el diseño, desarrollo y despliegue de los sistemas de IA? ¿Cuál es su papel?</p>	<p>En el diseño participan hospitales y centros de investigación; estos últimos sólo en el diseño del sistema de IA y los primeros también en el tratamiento sanitario relacionado.</p>
<p>Sección B Contexto de derechos fundamentales</p>	<p>¿Qué derechos fundamentales se ven potencialmente afectados por el uso del sistema de IA?</p>	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> Derecho a la intimidad (incluida la protección de datos). <input checked="" type="checkbox"/> Libertad frente a la discriminación. <input checked="" type="checkbox"/> Derecho a un nivel de vida adecuado (incluido el derecho a la salud física y mental).
	<p>¿Qué instrumentos jurídicos internacionales/regionales de protección de los derechos humanos/fundamentales se han aplicado a nivel operativo?</p>	<p>Declaración universal de los derechos humanos, Carta de derechos fundamentales de la UE, normativas de protección de datos aplicable.</p>
	<p>¿Cuáles son los tribunales u organismos más relevantes en materia de derechos fundamentales en el contexto?</p>	<p>Autoridades de protección de datos del país/región donde se desarrollen y utilicen los sistemas de IA, y órganos jurisdiccionales, Tribunal de Justicia de la Unión Europea y Tribunal Europeo de Derechos Humanos.</p>

	¿Cuáles son las decisiones y disposiciones más relevantes en materia de derechos humanos/fundamentales?	No aplicable
Sección C Controles implementados	¿Qué políticas y procedimientos se han implementado para evaluar el posible impacto sobre los derechos fundamentales, incluida la participación de las partes interesadas?	Se ha previsto la constitución de un comité de ética específico para el proyecto.
	¿Se ha realizado, desarrollado y aplicado una evaluación de impacto en relación con cuestiones específicas (por ejemplo, protección de datos) o algunas características del sistema (por ejemplo, el uso de la biometría)?	Se debe realizar una evaluación de impacto sobre la protección de datos.
Sección D Intervención de las partes interesadas y diligencia debida	¿Cuáles son los principales grupos o comunidades potencialmente afectados por el sistema de IA, incluido su desarrollo?	Pacientes con cáncer X [anonimizados].
	¿Qué partes interesadas, además de los individuos o grupos que puedan verse afectados por los sistemas de IA, deben participar (por ejemplo, la sociedad civil y las organizaciones internacionales, expertos, asociaciones industriales, periodistas)?	Asociaciones de pacientes con cáncer.
	¿Hay otros titulares de obligaciones que deban participar, aparte del proveedor y el responsable del despliegue de IA (por ejemplo, autoridades nacionales, organismos gubernamentales)?	Autoridad de protección de datos, Departamento de Sanidad local, comité ético de investigación científica, autoridad de supervisión de IA.

	<p>¿Han participado en el proceso de evaluación los socios comerciales, incluidos los proveedores de servicios (por ejemplo, subcontratistas en sistemas de IA y conjuntos de datos)?</p>	<p>No</p>
	<p>¿Ha llevado a cabo el proveedor de IA una evaluación de su cadena de suministro, para identificar si las actividades de los proveedores/contratistas que participan en el desarrollo de productos/servicios pueden afectar a los derechos fundamentales?</p> <p>¿El proveedor ha promovido estándares o auditorías de derechos fundamentales, para garantizar el respeto de los derechos fundamentales entre los proveedores?</p>	<p>N/A</p>
	<p>¿El proveedor y el responsable del despliegue de la IA han comunicado públicamente las posibles repercusiones del sistema de IA en los derechos fundamentales?</p>	<p>No</p>
	<p>¿El proveedor y el responsable del despliegue de la IA han proporcionado formación sobre estándares de derechos fundamentales al personal encargado de la gestión y la contratación relacionadas con el sistema de IA?</p>	<p>N/A</p>

Matrices de riesgo

Tabla 1. Probabilidad

Baja	El riesgo de perjuicio es improbable o altamente improbable
Media	El riesgo puede producirse
Alta	Existe una alta probabilidad de que se produzca el riesgo
Muy alta	Es muy probable que se produzca el riesgo

Tabla 2. Exposición

Baja	Pocos o muy pocos de la población identificada de titulares de derechos están potencialmente afectados
Media	Parte de la población identificada está potencialmente afectada
Alta	La mayoría de la población identificada está potencialmente afectada
Muy alta	Casi toda la población identificada está potencialmente afectada

Tabla 3. Probabilidad de que se produzca un perjuicio (*likelihood*)

		Probabilidad			
		Baja	Media	Alta	Muy alta
Exposición	Baja	B	B/M	B/A	B/MA
	Media	M/B	M	M/A	M/MA
	Alta	A/B	A/M	A	A/MA
	Muy alta	MA/B	MA/M	MA/H	MA

Probabilidad			
Baja	Media	Alta	Muy alta

Tabla 4. Gravedad del perjuicio

Baja	Los individuos y grupos afectados pueden encontrar sólo perjuicios menores en el ejercicio de sus derechos y libertades
Media	Los individuos y grupos afectados pueden encontrarse con importantes perjuicios
Alta	Los individuos y grupos afectados pueden enfrentarse a graves perjuicios
Muy alta	Los individuos y grupos afectados pueden encontrarse con perjuicios graves o incluso irreversibles

Tabla 5. Esfuerzo para superar el perjuicio y revertir los efectos adversos

Bajo	Los perjuicios sufridos pueden superarse sin problemas (por ejemplo, tiempo dedicado a modificar la información, molestias, irritaciones, etc.)
Medio	Los perjuicios sufridos pueden superarse a pesar de algunas dificultades (por ejemplo, costes adicionales, miedo, incomprensión, estrés, pequeñas dolencias físicas, etc.)
Alto	Los perjuicios sufridos pueden superarse aunque con serias dificultades (por ejemplo, pérdidas económicas, daños materiales, empeoramiento de la salud, etc.)
Muy alto	Los perjuicios sufridos pueden no superarse (por ejemplo, dolencias psicológicas o físicas de larga duración, muerte, etc.)

Tabla 6. Gravedad (severity)

		Gravedad			
		Baja	Media	Alta	Muy alta
Esfuerzo	Bajo	B	B/M	B/A	B/MA
	Medio	M/B	M	M/A	M/MA
	Alto	A/B	A/M	A	A/MA
	Muy alto	MA/B	MA/M	MA/A	MA

Gravedad			
Baja	Media	Alta	Muy alta

Tabla 1A. Recogida de datos y análisis de riesgos

Derechos/libertades potencialmente afectados	Descripción del impacto	Probabilidad			Gravedad		
		Probabilidad de resultados adversos	Exposición	Probabilidad	Gravedad	Esfuerzo	Gravedad
Protección de datos/Privacidad	El desarrollo del sistema de IA se basa en el uso de categorías especiales de datos personales y otra información personal de los pacientes. Cualquier operación de tratamiento que no cumpla la normativa aplicable en materia de protección de datos personales podría afectar a este derecho.	[Baja] El proyecto está sujeto a evaluaciones específicas de impacto ético y de protección de datos.	[Muy alta] El impacto afecta potencialmente a todas las personas a las que se aplica el algoritmo.	[Media]	[Media] El tratamiento ilegal de datos sanitarios relacionados con el cáncer y el uso ilícito de esta información pueden ser invasivos y afectar a la intimidad de las personas.	[Medio] La recogida y el tratamiento ilícitos de datos pueden detectarse y detenerse, con la supresión de la información que se haya recogido ilegalmente.	[Media]
No discriminación	El algoritmo se entrenó con datos de centros sanitarios europeos, por lo que es posible que se produzcan	[Alta] La etnia puede provocar algunas diferencias en las imágenes médicas, que	[Muy alta] Todas las personas del grupo pertinente (grupo étnico)	[Muy alta]	[Muy alta] Impacto negativo en la igualdad de acceso a la asistencia sanitaria y en la	[Alto] Sería necesario adaptar o incluso volver a entrenar el	[Muy alta]

	discriminaciones cuando se utilice en los tres centros sanitarios no pertenecientes a la UE.	pueden afectar a la precisión del diagnóstico.	a las que se aplica el algoritmo.		calidad del tratamiento oncológico recibido.	algoritmo con datos que eviten la discriminación.	
Derecho a la salud física y mental	El funcionamiento incorrecto del algoritmo puede dar lugar a un tratamiento sanitario ineficaz y perjudicial para el paciente, con el consiguiente perjuicio para el derecho a la salud.	[Media] Cuando se utiliza en pacientes europeos. [Alta] Cuando se utiliza en pacientes no europeos.	[Muy alta] El impacto afecta potencialmente a todas las personas a las que se aplica el algoritmo.	[Alta] Cuando se utiliza en pacientes europeos. [Muy alta] Cuando se utiliza en pacientes no europeos.	[Muy alta] El funcionamiento incorrecto del algoritmo puede dar lugar a un tratamiento sanitario ineficaz y perjudicial para el paciente.	[Medio] Patologías en las que un seguimiento posterior puede corregir el error del sistema. [Alto] Patología en la que el control posterior no puede corregir el error del sistema.	[Alta] Cuando el control posterior del cáncer puede corregir el error del sistema [Muy alta] Cuando el control posterior del cáncer no puede corregir el error del sistema.

Tabla 7. Impacto general del riesgo

		Gravedad			Muy alta
		Baja	Media	Alta	
Probabilidad	Baja				
	Media				
	Alta				
	Muy alta				

Tabla 2A. Gestión de riesgos (I)

Derecho/libertad afectados	Probabilidad	Gravedad	Impacto global	Medidas para prevenir/mitigar los impactos
Protección de datos/Privacidad	[Media]	[Media]	[Medio]	<ul style="list-style-type: none"> Publicar información sobre los procedimientos utilizados para obtener y procesar los datos originales utilizados para entrenar el sistema de IA.
No discriminación	[Muy alta]	[Muy alta]	[Muy alto]	<ul style="list-style-type: none"> Ampliar el conjunto de datos de entrenamiento, evitando una baja representación de los grupos relevantes.

<p>Derecho a la salud física y mental</p>	<p>[Alta] Cuando se utiliza en pacientes europeos.</p> <p>[Muy alta] Cuando se utiliza en pacientes no europeos.</p>	<p>[Alta] Cuando el control posterior del cáncer puede corregir el error del sistema</p> <p>[Muy alta] Cuando el control posterior del cáncer no puede corregir el error del sistema.</p>	<p>[Alto] Sólo cuando se utiliza en pacientes europeos y cuando la patología es tal que el seguimiento posterior puede corregir el error del sistema.</p> <p>[Muy alto] Para los otros tres escenarios.</p>	<ul style="list-style-type: none"> • Informar a los profesionales sanitarios de las limitaciones de la herramienta. Por ejemplo, indicando el tipo de errores. • Diferenciar entre patologías con o sin evolución rápida. • Informar a los profesionales sanitarios de que debe tenerse en cuenta la tasa de error. del equipo de diagnóstico por imagen utilizado.
---	--	---	---	--

Tabla 3A. Gestión de riesgos (II)

Derecho/libertad afectados	Probabilidad (residual)	Gravedad (residual)	Impacto residual
Protección de datos/Privacidad	[Media] Ya que la probabilidad se ha reducido a baja.	[Baja] Ya que el esfuerzo se ha reducido a bajo.	[Medio]
No discriminación	[Media] Ya que la probabilidad se ha reducido a baja.	[Media] Ya que el esfuerzo se ha reducido a bajo.	[Medio]
Derecho a la salud física y mental	[Media] Cuando se utiliza en pacientes europeos, ya que la probabilidad se ha reducido a baja. [Alta] Cuando se utiliza en pacientes no europeos, ya que la probabilidad se ha reducido a media.	[Media] Si se trata de una patología en la que un seguimiento posterior puede corregir el error del sistema, ya que el esfuerzo se ha reducido a bajo. [Alta] Si se trata de una patología en la que el seguimiento posterior no puede corregir el error del sistema, ya que el esfuerzo se ha reducido a medio.	[Medio] Sólo cuando se utilice en pacientes europeos y cuando la patología sea tal que el seguimiento posterior pueda corregir el error del sistema. [Alto] Para los tres escenarios restantes.

4. Comentarios

Las principales dificultades encontradas en el uso de la metodología en el caso de uso planteado se describen a continuación.

Esta metodología requiere, como mínimo, identificar los derechos fundamentales y libertades sobre los que impactará el sistema de inteligencia artificial; por tanto, se requiere tener un conocimiento experto sobre el contenido esencial de cada uno de los derechos fundamentales y libertades que son objeto de examen. Se puede partir de la base que los delegados de protección de datos tienen dicho conocimiento en relación con el derecho a la protección de datos personales, así como con el derecho a la intimidad personal y familiar. Sin embargo, un delegado de protección de datos no necesariamente debe tener este conocimiento experto en relación con el resto de derechos fundamentales y libertades. En consecuencia, la primera dificultad se encuentra en que se requiere disponer de un conocimiento ciertamente profundo sobre cada uno de los derechos fundamentales y libertades, para poder identificar en primer lugar cuáles serán afectados. Una vez identificados, siempre se puede solicitar asesoramiento a personas expertas en su contenido esencial.

La dificultad anterior se vuelve más compleja cuando se prevé utilizar el sistema de inteligencia artificial fuera de la Unión Europea. Si bien se puede llegar a definir un marco común en relación con el contenido de los derechos fundamentales y libertades dentro de la Unión Europea, teniendo en cuenta la Carta de derechos fundamentales de la UE y los pronunciamientos del TJUE sin llegar al detalle de las diferencias establecidas por los órganos jurisdiccionales nacionales, difícilmente se puede hablar de un marco común cuando se examina el contenido de los derechos fundamentales y libertades en todo el mundo. La aplicación de esta metodología en la UE y fuera de la UE con el mismo nivel de detalle requeriría un análisis de derecho comparado, que difícilmente podría ser asumido por la gran

mayoría de organizaciones dados los recursos humanos, temporales y económicos que requeriría. Por lo tanto, sería útil disponer de una guía con el contenido mínimo de cada derecho fundamental y libertad a nivel mundial, o bien por regiones o tradiciones jurídicas, que permitiera usar la presente metodología sin requerir recursos desproporcionados en relación con su fin; es decir, disponer de un análisis *ex ante* que facilite el diseño del sistema de inteligencia artificial.

La última dificultad encontrada ha consistido en identificar a las personas que deben intervenir en la ejecución de la evaluación de impacto, tanto por sus conocimientos como por su papel en el desarrollo y ejecución de este tipo de inteligencia artificial. Por ejemplo, en identificar qué personas expertas en el desarrollo del sistema de inteligencia artificial con fines sanitarios debían participar, sin que su participación implicara revelar información del proyecto a la “competencia”.

Caso de uso 4: ATENEA, IA al servicio de las personas mayores

1. El contexto

Las administraciones públicas, en general, se encuentran inmersas en un proceso de transformación digital, con la idea de reformar los servicios públicos aprovechando las ventajas que aporta la evolución exponencial de las tecnologías. Pero estas políticas de transformación digital se deben formular e implementar generando un impacto positivo en términos de inclusión social, combinando el impulso de la digitalización con las políticas sociales para minimizar al máximo la brecha digital que se genera, inevitablemente, en estos procesos de cambio tan disruptivos. Se trata, pues, de poner la tecnología al servicio de las personas, con el objetivo de mejorar las relaciones con la ciudadanía, la atención social, haciendo frente a las desigualdades que origina la creciente digitalización, garantizando la igualdad de oportunidades y, en general, mejorar las condiciones de vida de la ciudadanía.

Es en este contexto, y bajo estas premisas y compromisos sociales, que se ha impulsado el proyecto ATENEA (en fase de prueba piloto) que tiene como finalidad contribuir a la transformación digital de los territorios, poniendo en el centro a la ciudadanía más vulnerable, con los objetivos específicos de reducir la brecha digital existente y la soledad no deseada, aumentar la seguridad de las personas, especialmente dentro del hogar, fomentando la inclusión, el bienestar, la salud y, en definitiva, la calidad de vida de las personas. El proyecto va dirigido específicamente a personas mayores de 65 años, que viven solas y que sufren con más incidencia las vulnerabilidades originadas por la brecha digital.

2. El proyecto

ATENEA es un proyecto basado en el desarrollo de un sistema de IA generativa (redes neuronales), asistente de voz, biometría por voz y *robotic process automation*, combinadas con tecnologías maduras como *data analytics*, *cloud computing* y *smartphone*. Mediante el reconocimiento biométrico de la voz permite dar respuesta a las peticiones de las personas mayores usuarias en relación con distintos casos de uso: llamada y videollamada a un familiar, llamada a emergencias (112), llamada al ayuntamiento de referencia y/o reserva automática de cita previa con los servicios sociales, reserva de cita previa con el centro de atención primaria de referencia, recordatorios de agenda, indicaciones de rutas de transporte y, en un futuro, compra on-line, gestiones bancarias y gestiones de suministros. La solución ATENEA no requiere habilidades digitales ni manipulación por parte del usuario, la identificación es biométrica con el fin de garantizar el uso individual exclusivo y la seguridad.

Con todo, ATENEA es una inteligencia artificial en forma de *tablet*, sin botones, sin pantallas táctiles, que funciona sólo con una interacción tan sencilla como la voz. A una persona mayor, probablemente en situación de dependencia, le da respuestas ágiles a sus necesidades básicas. Esta inteligencia artificial permite realizar con una conversación las gestiones diarias, como conocer el extracto bancario, pedir cita previa al médico, hacer una videoconferencia con un familiar, llamar a emergencias o a servicios sociales.

ATENEA ha sido diseñada con la cocreación de las personas mayores, de los cuidadores profesionales y del entorno de la persona usuaria, estableciendo un vínculo de confianza y apoyo en caso de necesidad.

La iniciativa está liderada por una alianza estratégica entre entidades tecnológicas y sociales que se encargan, respectivamente, del diseño de la solución, el contacto con los usuarios y el despliegue de integradores

sociodigitales sobre el terreno, los servicios de cloud, la dotación de dispositivos (tablets), la tecnología del habla, la automatización robótica de procesos, la provisión de comunicaciones, la seguridad, la evaluación de resultados e impacto y la garantía de los derechos de las personas usuarias. Se trata de una colaboración público-privada. La fase actual del proyecto está subvencionada por el Departamento de Derechos Sociales del Gobierno de Cataluña en el marco del Plan de recuperación, transformación y resiliencia, financiado por la Unión Europea, Next Generation EU. Esta experiencia tiene el apoyo de distintos ayuntamientos y administraciones públicas que ponen su territorio como prueba piloto para el testeo de la solución y se encargan de encontrar a los usuarios potenciales.

3. La EIDF

Planificación y alcance

Sección A Descripción y análisis del sistema de IA, incluidos los flujos de datos relacionados	¿Cuáles son los principales objetivos del sistema de IA?	Dar respuesta automatizada a las peticiones de atención realizadas mediante la voz de las personas usuarias: <ul style="list-style-type: none"> • Llamada y videollamada a un familiar. • Llamada a emergencias (112). • Llamada al ayuntamiento de referencia y/o reserva automática de cita previa con los servicios sociales. • Reserva de cita previa con el Centro de Atención Primaria de referencia • Recordatorios de agenda • Información de rutas de transporte • Recordatorios de medicación • Agenda de la ciudad • En desarrollo: compra en supermercado on-line, televisión y radio a la carta
	¿Cuáles son las principales características del sistema?	Un asistente de voz basado en inteligencia artificial que interactúa con los usuarios mediante comandos de voz, proporcionando información, ejecutando tareas y ofreciendo servicios en tiempo real. Su diseño combina tecnologías de reconocimiento de voz, procesamiento del lenguaje natural (NLP) y síntesis de voz para entender, procesar y responder a las solicitudes de los usuarios de manera natural. Las peticiones por voz que realizan las personas usuarias son procesadas mediante un sistema de automatización robótica de procesos, que es una tecnología que permite automatizar tareas repetitivas, basadas en reglas y que generalmente son realizadas por humanos, utilizando robots de software o bots. Estos bots imitan y ejecutan

		acciones humanas en entornos digitales, como hacer clics, mover archivos, completar formularios, copiar y pegar información, o procesar datos de forma masiva. En suma, ATENEA permite hacer las gestiones diarias con una conversación con la IA.
	¿En qué países se ofrecerá?	España
	¿Qué tipos de datos se tratan (personales, no personales, categorías especiales)?	<ul style="list-style-type: none"> • Datos demográficos (edad, ciudad, situación familiar, país) • Información personal, sin incluir los datos relacionados que las personas usuarias tratan con los servicios accedidos a través de la solución. • Configuración del dispositivo • Idioma de los diálogos (CAT/SPA) • Información de acceso cita previa médico de familia • Teléfono de cita previa servicios sociales municipales/OAC • Teléfono de emergencias/teleasistencia • Teléfono y nivel de parentesco de familias • Teléfono de los integradores sociotecnológicos asignados. • Información relacionada recordatorios y alertas personalizadas.
	Identificación de los posibles titulares de derechos: ¿quiénes son los individuos o grupos que pueden verse afectados por los sistemas de IA? ¿Se incluyen entre ellos individuos o grupos vulnerables?	<p>Las personas usuarias, mayores de 65 años, con un nivel cognitivo suficiente para relacionarse con el dispositivo.</p> <p>Dado que todas las personas implicadas están afectadas por la brecha digital y la soledad en el hogar, deben considerarse vulnerables, debido a sus</p>

		condiciones socio-demográficas y a las relaciones entre estas condiciones y la finalidad del sistema de IA.
	<p>Identificación de los responsables: ¿qué personas/entidades están implicadas en el diseño, desarrollo y despliegue de los sistemas de IA? ¿Cuál es su papel?</p>	<p>La iniciativa está liderada por una alianza estratégica entre entidades tecnológicas y sociales que se encargan, respectivamente, del diseño de la solución, el contacto con los usuarios y el despliegue de integradores sociodigitales sobre el terreno, servicios de cloud, dotación de dispositivos (tablets), tecnología del habla, la automatización robótica de procesos, provisión de comunicaciones, seguridad, evaluación de resultados e impacto y garantía de los derechos de las personas usuarias. Se trata de una colaboración público-privada. La fase actual del proyecto está subvencionada por el Departamento de Derechos Sociales del Gobierno de Cataluña en el marco del Plan de recuperación, transformación y resiliencia, financiado por la Unión Europea, Next Generation EU. Esta experiencia tiene el apoyo de distintos ayuntamientos y Administraciones públicas que ponen su territorio como prueba piloto para el testeo de la solución y se encargan de encontrar a los usuarios potenciales.</p>
<p>Sección B Contexto de derechos fundamentales</p>	<p>¿Qué derechos fundamentales se ven potencialmente afectados por el uso del sistema de IA?</p>	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> Derecho a la intimidad (incluida la protección de datos) <input checked="" type="checkbox"/> Derecho a la privacidad <input checked="" type="checkbox"/> Libertad frente a la discriminación <input checked="" type="checkbox"/> Derecho al cuidado de la salud <input checked="" type="checkbox"/> Derecho a la asistencia social <input checked="" type="checkbox"/> Derecho de acceso a los servicios de interés económico general <input checked="" type="checkbox"/> Derecho de acceso a los servicios de interés general

	<p>¿Qué instrumentos jurídicos internacionales/regionales de protección de los derechos humanos/fundamentales se han aplicado a nivel operativo?</p>	<p>El proyecto se inició antes de la publicación del Reglamento de Inteligencia Artificial que, también será aplicable a esta solución. En su momento, la normativa aplicable fue el Reglamento (UE) 2016/679, del Parlamento Europeo y del Consejo, de 27 de abril de 2016, relativo a la protección de las personas físicas en cuanto al tratamiento de datos personales y a la libre circulación de estos datos (Reglamento general de protección de datos - RGPD) y la Ley Orgánica 3/2018, de 5 de diciembre, de protección de datos personales y garantía de los derechos digitales (LOPD-GDD) y normativa complementaria</p>
	<p>¿Cuáles son los tribunales u organismos más relevantes en materia de derechos fundamentales en el contexto de uso?</p>	<p>En España, la protección de los derechos fundamentales se garantiza a través de un marco normativo nacional e instituciones judiciales y no judiciales. Además, España está vinculada a los sistemas de protección internacionales y regionales, especialmente al Sistema Europeo de Derechos Fundamentales.</p> <p>A continuación, se presentan los tribunales y organismos más relevantes en materia de derechos fundamentales en España:</p> <ul style="list-style-type: none"> • Tribunal Constitucional • Tribunal Supremo • Audiencia Nacional • Juzgados y Tribunales ordinarios • Defensor del Pueblo e instituciones análogas en las regiones • Agencias autonómicas de defensa de los derechos • Fiscalía General del Estado • Autoridades de control en materia de protección de datos del País/Región

	<p>¿Cuáles son las decisiones y disposiciones más relevantes en materia de derechos humanos/fundamentales?</p>	<p>Tribunal Constitucional:</p> <ul style="list-style-type: none">• Sentencia 135/2024: esta sentencia abordó la vulneración del derecho a la tutela judicial efectiva, al aplicar preceptos legales que habían sido declarados inconstitucionales previamente.• Sentencia 113/2021: en este fallo, el Tribunal reconoció la vulneración del derecho a la tutela judicial efectiva en relación con la protección a la familia y los menores, enfatizando la necesidad de una motivación reforzada en casos que afectan a derechos fundamentales sustantivos.• Sentencia 58/2018: esta resolución trató sobre la protección de datos personales y la libertad de expresión, estableciendo criterios sobre el equilibrio entre ambos derechos. <p>Disposiciones legislativas:</p> <ul style="list-style-type: none">• Ley Orgánica 3/2007, de 22 de marzo, para la igualdad efectiva de mujeres y hombres: establece medidas para eliminar la discriminación por razón de género.• Ley 15/2022, de 12 de julio, integral para la igualdad de trato y la no discriminación: refuerza el marco legal contra cualquier forma de discriminación.• Ley Orgánica 1/2004, de 28 de diciembre, de Medidas de Protección Integral contra la Violencia de Género: aborda la prevención, protección y sanción de la violencia contra las mujeres.
--	---	--

Sección C Controles implementados	¿Qué políticas y procedimientos se han implementado para evaluar el posible impacto sobre los derechos fundamentales, incluida la participación de las partes interesadas?	<p>ATENEA es un proyecto de colaboración voluntaria, por parte de las personas usuarias para mejorar su bienestar y calidad de vida. Es un proyecto que desde el inicio las personas usuarias participaron en el desarrollo del plan funcional del proyecto.</p> <p>De hecho, para poder participar en el proyecto, las personas usuarias tuvieron que dar su consentimiento informado, ejerciendo el derecho a su autonomía. Para ejercer este derecho la persona fue informada verbalmente y por escrito. Posteriormente, este consentimiento informado inicial fue validado por las propias personas usuarias y se les volvió a informar de los riesgos y beneficios del uso de este sistema.</p>
	¿Se ha realizado, desarrollado y aplicado una evaluación de impacto en relación con cuestiones específicas (por ejemplo, protección de datos) o algunas características del sistema (por ejemplo, el uso de la biometría)?	<p>Se ha realizado la evaluación de impacto en protección de datos personales (EIPD) y cumplimiento de la normativa vigente en materia de seguridad y protección de datos personales. Así se recoge en los convenios de colaboración con los territorios piloto.</p>
Sección D Intervención de las partes interesadas y diligencia debida	¿Cuáles son los principales grupos o comunidades potencialmente afectados por el sistema de IA, incluido su desarrollo?	<p>Personas mayores de 65 años que viven solas.</p>
	¿Qué partes interesadas, además de los individuos o grupos que puedan verse afectados por los sistemas de IA, deben participar (por ejemplo, la sociedad civil y las organizaciones internacionales, expertos, asociaciones industriales, periodistas)?	<p>Administraciones públicas, empresas privadas y entidades del tercer sector.</p>

	<p>¿Hay otros titulares de obligaciones que deban participar, aparte del proveedor y el responsable del despliegue de IA (por ejemplo, autoridades nacionales, organismos gubernamentales)?</p>	<p>Autoridad de protección de datos, Administraciones públicas locales y autonómicas, comité ético de investigación científica de una Universidad pública, autoridad de supervisión de IA.</p>
	<p>¿Han participado en el proceso de evaluación los socios comerciales, incluidos los proveedores de servicios (por ejemplo, subcontratistas en sistemas de IA y conjuntos de datos)?</p>	<p>En el proceso de evaluación han participado los partners del proyecto constituidos como una agrupación de empresas sin entidad jurídica.</p> <p>Desde el primer momento, todas las partes que han participado en el proyecto (se trata de una colaboración público-privada), han tenido muy presente la necesidad de velar por el cumplimiento normativo, así como por la garantía de los derechos fundamentales de la ciudadanía que pudieran verse afectados por la puesta en marcha de este proyecto. Por este motivo, para la realización del proyecto ATENEA se contó con la participación de un comité externo (Universidad pública de Cataluña) que valoró dicho proyecto dándole la acreditación necesaria para poder llevarlo a cabo.</p> <p>Por otro lado, en los convenios firmados con las Administraciones públicas participantes, se prevé como obligación la realización de evaluaciones de impacto, el seguimiento continuo, la formación, la transparencia y la rendición de cuentas de este proyecto.</p>
	<p>¿Ha llevado a cabo el proveedor de IA una evaluación de su cadena de suministro, para identificar si las actividades de los proveedores/contratistas que participan en el desarrollo de productos/servicios pueden afectar a los derechos fundamentales?</p>	<p>No (no es un requisito legal)</p>

	¿El proveedor ha promovido estándares o auditorías de derechos fundamentales, para garantizar el respeto de los derechos fundamentales entre los proveedores?	
	¿El proveedor y el responsable del despliegue de la IA han comunicado públicamente las posibles repercusiones del sistema de IA en los derechos fundamentales?	No (no es un requisito legal)
	¿El proveedor y el responsable del despliegue de la IA han proporcionado formación sobre estándares de derechos fundamentales al personal encargado de la gestión y la contratación relacionadas con el sistema de IA?	Solo en materia de protección de datos

Matrices de riesgo

Tabla 1. Probabilidad

Baja	El riesgo de perjuicio es improbable o altamente improbable
Media	El riesgo puede producirse
Alta	Existe una alta probabilidad de que se produzca el riesgo
Muy alta	Es muy probable que se produzca el riesgo

Tabla 2. Exposición

Baja	Pocos o muy pocos de la población identificada de titulares de derechos están potencialmente afectados
Media	Parte de la población identificada está potencialmente afectada
Alta	La mayoría de la población identificada está potencialmente afectada
Muy alta	Casi toda la población identificada está potencialmente afectada

Tabla 3. Probabilidad de que se produzca un perjuicio (*likelihood*)

		Probabilidad			
		Baja	Media	Alta	Muy alta
Exposición	Baja	B	B/M	B/A	B/MA
	Media	M/B	M	M/A	M/MA
	Alta	A/B	A/M	A	A/MA
	Muy alta	MA/B	MA/M	MA/H	MA

Probabilidad			
Baja	Media	Alta	Muy alta

Tabla 4. Gravedad del perjuicio

Baja	Los individuos y grupos afectados pueden encontrar sólo perjuicios menores en el ejercicio de sus derechos y libertades.
Media	Los individuos y grupos afectados pueden encontrarse con importantes perjuicios.
Alta	Los individuos y grupos afectados pueden enfrentarse a graves perjuicios.
Muy alta	Los individuos y grupos afectados pueden encontrarse con perjuicios graves o incluso irreversibles.

Tabla 5. Esfuerzo para superar el perjuicio y revertir los efectos adversos

Bajo	Los perjuicios sufridos pueden superarse sin problemas (por ejemplo, tiempo dedicado a modificar la información, molestias, irritaciones, etc.)
Medio	Los perjuicios sufridos pueden superarse a pesar de algunas dificultades (por ejemplo, costes adicionales, miedo, incomprensión, estrés, pequeñas dolencias físicas, etc.).
Alto	Los perjuicios sufridos pueden superarse aunque con serias dificultades (por ejemplo, pérdidas económicas, daños materiales, empeoramiento de la salud, etc.).
Muy alto	Los perjuicios sufridos pueden no superarse (por ejemplo, dolencias psicológicas o físicas de larga duración, muerte, etc.).

Tabla 6. Gravedad (severity)

		Gravedad			
		Baja	Media	Alta	Muy alta
Esfuerzo	Bajo	B	B/M	B/A	B/MA
	Medio	M/B	M	M/A	M/MA
	Alto	A/B	A/M	A	A/MA
	Muy alto	MA/B	MA/M	MA/A	MA

Gravedad			
Baja	Media	Alta	Muy alta

Tabla 1A. Recogida de datos y análisis de riesgos

Derechos/libertades potencialmente afectados	Descripción del impacto	Probabilidad			Gravedad		
		Probabilidad de resultados adversos	Exposición	Probabilidad	Gravedad del impacto	Esfuerzo	Gravedad
Derecho a la protección de datos	El desarrollo del sistema de IA se basa en el uso del reconocimiento de la	[Medio] El proyecto está sujeto a evaluaciones	[Muy alto] El impacto afecta potencialmen	[Alta]	[Alto] Los riesgos asociados con el uso de la	[Medio] La recogida y el tratamiento ilícitos de datos	[Medio]

	<p>voz. En este sentido, la IA es usada para elaborar/enviar las solicitudes del usuario (de voz a texto) a otras tecnologías robóticas no basadas en IA que realizarán la tarea solicitada</p> <p>ATENEA escucha todas las conversaciones a la espera del comando wake up word “Hola ATENEA”. Este proceso de escucha se mantiene únicamente en el dispositivo (Tablet), no se procesa ni se almacena en cloud.</p> <p>Solo el comando “Hola ATENEA” activa el servicio y su comunicación al cloud, donde se procesa y almacena la</p>	<p>específicas de impacto ético y de protección de datos.</p> <p>Las personas usuarias deben tener un nivel cognitivo adecuado para prestar su consentimiento informado a la utilización de este sistema.</p> <p>El proyecto ha sido evaluado por un comité de ética reconocido de una universidad pública de Cataluña y el sistema cumple con las medidas de seguridad establecidas en las normativas técnicas vigentes en España (ENS, normas UNE...) según el informe elaborado por la Agencia de</p>	<p>te a todos los individuos a las que se aplica el algoritmo.</p>		<p>biometría son limitados debido a la naturaleza opcional de la identificación biométrica; La identificación biométrica puede ser una solución proporcionada para las personas mayores con movilidad reducida. Existe el riesgo de que las interacciones de voz se utilicen para la elaboración de perfiles.</p> <p>El principal riesgo es el tratamiento ilícito de datos, incluidas categorías especiales de</p>	<p>pueden detectarse y detenerse, con la supresión de la información que se haya recogido ilegalmente.</p>	
--	---	--	--	--	---	--	--

	<p>conversación para entrenar el algoritmo.</p> <p>Todas las conversaciones del usuario relacionadas con un servicio específico accedido a través de ATENEA, por ejemplo, llamadas de emergencia u otras llamadas, se mantienen privadas.</p> <p>La biometría por voz sólo se utiliza para identificar a la persona, y se activa en base a la decisión de la propia persona usuaria.</p> <p>La identificación se puede basar en biometría de voz o en la tradicional combinación de usuario y contraseña.</p> <p>Cualquier operación de tratamiento que no cumpla la normativa aplicable en materia</p>	<p>Ciberseguridad de Cataluña</p>			<p>datos relacionados con la situación vulnerable de los usuarios, que puede ser invasivo y afectar a la privacidad de las personas.</p>		
--	---	-----------------------------------	--	--	--	--	--

	de protección de datos personales podría afectar a este derecho.						
No discriminación	<p>El mecanismo de IA funciona con tecnología del habla, por lo que es posible que se produzcan discriminaciones cuando el asistente de voz no entienda al hablante por algún problema de comunicación (por ejemplo, impedimento o deterioro del habla).</p> <p>Se pueden producir discriminaciones cuando las personas usuarias no hablen correctamente el idioma del sistema (CAT/ES).</p>	<p>[Alto]</p> <p>Posibles problemas en la voz y el habla de las personas usuarias pueden generar que queden excluidas.</p>	<p>[Medio]</p> <p>Habrà un número limitado de personas con problemas de voz o de idioma.</p>	[Medio]	<p>[Muy alto]</p> <p>Impacto negativo en la igualdad de acceso a la asistencia que ofrece la solución y en la calidad del servicio.</p>	<p>[Alto]</p> <p>Sería necesario adaptar el asistente de voz para integrar las casuísticas de problemas en el habla o idioma.</p> <p>Las personas usuarias deben disponer de cierto nivel cognitivo. Se podría excluir también para el uso las personas que tengan algún tipo de problema en el habla o en el idioma.</p>	[Muy alto]
Derecho a la salud	Un funcionamiento inadecuado del sistema puede provocar que no se	<p>[Bajo]</p> <p>El proyecto está sujeto a</p>	<p>[Muy alto]</p> <p>El impacto afecta</p>	[Bajo]	<p>[Alto]</p> <p>No poder acceder cuando</p>	<p>[Medio]</p> <p>Este sistema no es el único al</p>	[Medio]

	garantice adecuadamente este derecho	evaluaciones continuas y a un seguimiento específico para detectar posibles errores en el funcionamiento. Solo en algunos casos el funcionamiento inadecuado puede afectar significativamente al derecho a la salud (p. ej., llamada de emergencia).	potencialmente a todas las personas a las que se aplica el algoritmo.		se necesita a los servicios de salud puede causar un grave perjuicio a usuarios que potencialmente están más expuestas a situaciones de riesgo	que pueden recurrir las personas en caso de necesidad, ya que existen otros canales de acceso a sistemas sanitarios (así como a otros servicios), teniendo en cuenta que los potenciales usuarios deben tener un nivel cognitivo determinado	
Derecho a la asistencia social	Un funcionamiento inadecuado del sistema puede provocar que no se garantice adecuadamente este derecho	[Bajo] El proyecto está sujeto a evaluaciones continuas y a un seguimiento específico para detectar posibles errores en el funcionamiento.	[Muy alto] El impacto afecta potencialmente a todas las personas a las que se aplica el algoritmo.	[Bajo]	[Alto] No poder acceder cuando se necesita a los servicios de asistencia social puede causar un grave perjuicio para estas personas que	[Bajo] Este sistema no es la única solución a la que pueden recurrir las personas, ya que existen otros canales de acceso a servicios de	[Medio]

		Solo en algunos casos de uso el funcionamiento inadecuado puede afectar a este derecho.			potencialmente están más expuestas a situaciones de riesgo.	asistencia social, teniendo en cuenta que los potenciales usuarios deben tener un grado cognitivo determinado. Además, estas personas cuentan con una persona de contacto de los servicios sociales que hace un seguimiento de su situación, y no se trata de servicios de emergencia.	
Derecho de acceso a los servicios de interés económico general	Este derecho se podría ver vulnerado si no todas las personas pueden tener acceso a este sistema por falta de recursos de las Administraciones públicas	[Medio] El proyecto se encuentra en fase piloto con financiación provenientes de los fondos Next Generation.	[Medio] El impacto afecta potencialmente a todas las personas que no pueden beneficiarse	[Medio]	[Medio] No poder acceder a este sistema bien por falta de información o de medios, puede vulnerar la igualdad de	[Medio] Se trataría de poner los recursos suficientes para que todas las personas susceptibles de disponer de este	[Medio]

			todavía de este proyecto, por falta de recursos o de información		oportunidades y la cohesión social del territorio.	sistema puedan acceder a él	
--	--	--	--	--	--	-----------------------------	--

Tabla 7. Impacto general del riesgo

		Gravedad			
		Baja	Media	Alta	Muy alta
Probabilidad	Baja				
	Media				
	Alta				
	Muy alta				

Tabla 2A. Gestión de riesgos (I)

Derecho/libertad afectados	Probabilidad	Gravedad	Impacto global	Medidas para prevenir/mitigar los impactos
Derecho a la protección de datos	[Alta]	[Media]	[Alto]	<ul style="list-style-type: none"> • Publicación de la información sobre los procedimientos utilizados para obtener y procesar los datos originales utilizados para entrenar el sistema de IA (transparencia algorítmica y registro público de estos algoritmos) • La identificación biométrica mediante la voz es opcional y se basa en el consentimiento previo. Esta identificación no se utilizará para otra finalidad, incluida la identificación compartida con terceros. • Exclusión de cualquier actividad de profiling basada en el uso de los servicios accesibles a través de ATENEA. • La implementación y mantenimiento de un sistema de gestión de riesgos y de la seguridad durante todo el ciclo de vida del sistema. • Garantizar la calidad de los datos de entrenamiento, validación y prueba mediante prácticas adecuadas de gobernanza y gestión de datos. • <i>Accountability</i>: elaborar y mantener actualizada la documentación técnica para el sistema.

				<ul style="list-style-type: none"> • Permitir una supervisión humana efectiva durante su uso, con personal capacitado. • Garantizar el ejercicio de los derechos de las personas usuarias.
No discriminación	[Media]	[Muy Alta]	[Alto]	<ul style="list-style-type: none"> • Entrenar al asistente de voz para evitar algunas de las situaciones de discriminación descritas. • Restringir las personas usuarias potenciales para evitar estas situaciones. • Implantación y mantenimiento de un sistema de gestión de riesgos y de la seguridad durante todo el ciclo de vida del sistema. • Garantizar la calidad de los datos de entrenamiento, validación y prueba mediante prácticas adecuadas de gobernanza y gestión de datos. • <i>Accountability</i>: elaborar y mantener actualizada la documentación técnica para el sistema. • Permitir una supervisión humana efectiva durante su uso, con personal capacitado y pruebas periódicas de eficacia del sistema en los casos críticos. • Garantizar el ejercicio de los derechos de las personas usuarias.
Cuidado de la salud	[Baja]	[Media]	[Medio]	<ul style="list-style-type: none"> • Garantizar la asistencia técnica adecuada y mejorar la calidad técnica del sistema.

				<ul style="list-style-type: none"> • Informar de todos los canales que tienen a su disposición estas personas para acceder al cuidado de la salud.
Asistencia social	[Baja]	[Media]	[Medio]	<ul style="list-style-type: none"> • Garantizar la asistencia técnica adecuada y mejorar la calidad técnica del sistema. • Informar de todos los canales que tienen a su disposición estas personas para acceder a la asistencia social.
Acceso a los servicios de interés económico general	[Media]	[Media]	[Medio]	<ul style="list-style-type: none"> • Garantizar los recursos y medios suficientes para que todas las personas usuarias potenciales tengan acceso a este servicio.

Tabla 3A. Gestión de riesgos (II)

Derecho/libertad afectados	Probabilidad (residual)	Gravedad (residual)	Impacto residual
Protección de datos/Privacidad	<p>[Baja]</p> <p>La probabilidad de que se produzca un impacto sobre este derecho (probabilidad de resultados adversos) ha disminuido a bajo, ya que las medidas adoptadas han limitado la posibilidad de consecuencias negativas.</p>	<p>[Media]</p> <p>El nivel de gravedad del impacto ha disminuido a medio, ya que las medidas adoptadas han restringido el ámbito de las consecuencias negativas que podrían producirse. El nivel de gravedad sigue siendo medio, pero de un nivel más bajo que el inicial.</p>	[Medio]
No discriminación	<p>[Baja]</p> <p>La probabilidad se ha reducido a baja porque la probabilidad de efectos adversos se ha reducido a media, debido a la disminución de los casos posibles de funcionamiento inadecuado, y la exposición se ha reducido a baja, debido a la reducción del número de personas afectadas (se han identificado algunas limitaciones de uso relacionadas con el conocimiento de idiomas).</p>	<p>[Media]</p> <p>El nivel de gravedad se ha reducido de muy alto a medio debido a la implementación de medidas complementarias/alternativas para acceder a los servicios (reduciendo la gravedad del impacto de muy alto a alto) y el esfuerzo menor requerido para reaccionar debido a las medidas ya implementadas (reduciendo el esfuerzo de alto a medio).</p>	[Medio]

Cuidado de la salud	[Baja]	[Media] El nivel de gravedad ha disminuido, debido a las medidas complementarias/alternativas para acceder a los servicios (reducción de la gravedad del impacto de muy alto a alto) y al reducido esfuerzo necesario para reaccionar debido a las medidas ya implementadas (reducción de esfuerzo de medio a bajo). La gravedad sigue siendo media, pero de un nivel más bajo que el inicial.	[Medio]
Asistencia social	[Baja]	[Media] La gravedad ha disminuido, debido a las medidas complementarias/alternativas para acceder a los servicios (reducción de la gravedad del impacto de muy alto a alto) y al reducido esfuerzo necesario para reaccionar debido a las medidas ya implementadas (reducción de esfuerzo de medio a bajo). La gravedad sigue siendo media, pero de un nivel más bajo que el inicial.	[Medio]
Acceso a los servicios de interés económico general	[Media]	[Media] Se mantiene el nivel de gravedad ya que garantizar este derecho depende de los recursos que los gobiernos destinen a ello.	[Medio]

4. Comentarios

La aparición de nuevas tecnologías no solo está causando una revolución en la manera en que trabajamos y prestamos servicios públicos, sino también en la necesidad de repensar, desde la etapa de diseño, el impacto de estas aplicaciones en los derechos fundamentales de los ciudadanos. De hecho, este cambio de paradigma ya se hizo evidente con la entrada en vigor del Reglamento General de Protección de Datos (RGPD) y ahora está cobrando impulso.

Una metodología para llevar a cabo una evaluación de impacto sobre los derechos fundamentales (EIDF) es una herramienta esencial para identificar los riesgos y las medidas organizativas y técnicas a aplicar, con un enfoque transversal que debe integrar en los equipos de trabajo multidisciplinares responsables de la implementación de la IA en las administraciones públicas. Con esta finalidad, como ya se ha observado con la aplicación del RGPD, es absolutamente necesario que, desde el primer momento, los delegados de protección de datos se impliquen en las estrategias y proyectos de transformación digital a implementar en cada organización. De esta manera, estos riesgos pueden ser analizados desde el principio y por defecto, y se pueden tomar las medidas técnicas y organizativas necesarias.

Por otro lado, también es recomendable revisar las descripciones de los puestos de trabajo para incluir este tipo de competencias y responsabilidades, invertir en formación y desarrollar perfiles profesionales adecuados. Además, en este momento de transición hacia “la administración inteligente” (basada en el uso de la IA), se debería desarrollar algún tipo de guía práctica como instrucción interna, de manera que todos en las organizaciones sean conscientes de estos riesgos, incluidas las directrices sobre las mejores prácticas en el uso de la IA en la actividad administrativa.

