



**Politecnico
di Torino**

El impacto de la IA: nuevas perspectivas y continuidad en el enfoque basado en el riesgo y su implementación

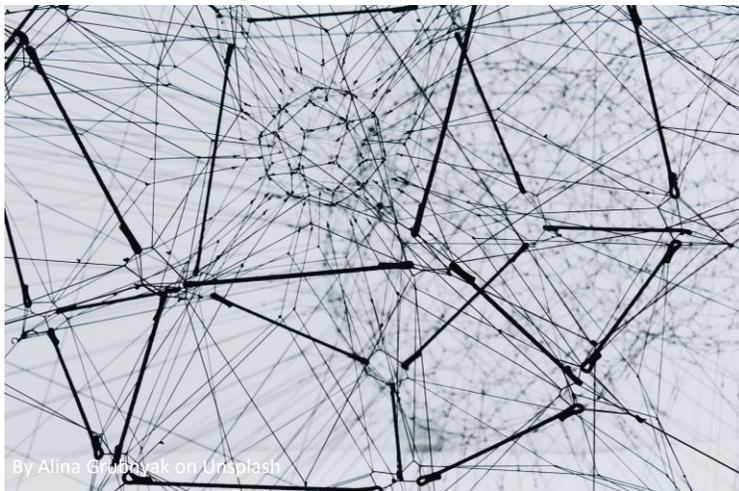
Autoridad Vasca de Protección de Datos, Donostia, 31 de mayo de 2025

Alessandro Mantelero

Jean Monnet Chair in Mediterranean Digital Societies and Law

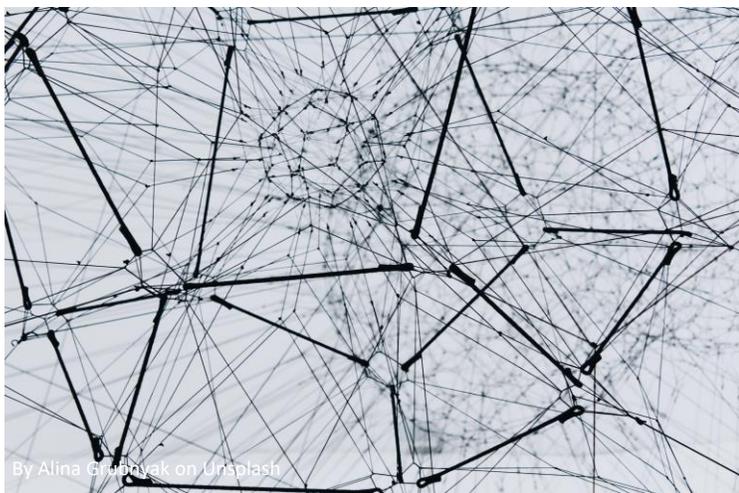
Associate Professor of Law | Polytechnic University of Turin

La sociedad moderna y el enfoque basado en el riesgo



- De las amenazas naturales a los riesgos industriales de las sociedades modernas
- De la responsabilidad civil a las obligaciones legales en materia de prevención de riesgos
 - Un mundo no estático y repercusiones más amplias
 - Las limitaciones del enfoque *ex post*
 - Detección de daños
 - Extensión del daño
 - Aplicación efectiva
 - Solidez económica

Enfoque en el riesgo y regulación digital: protección de datos



- Las raíces de la protección de datos en la prevención de riesgos
 - Desafíos críticos a los orígenes de la protección de datos
 - Enfoque procedimental y seguridad
 - Controladores internos y externos

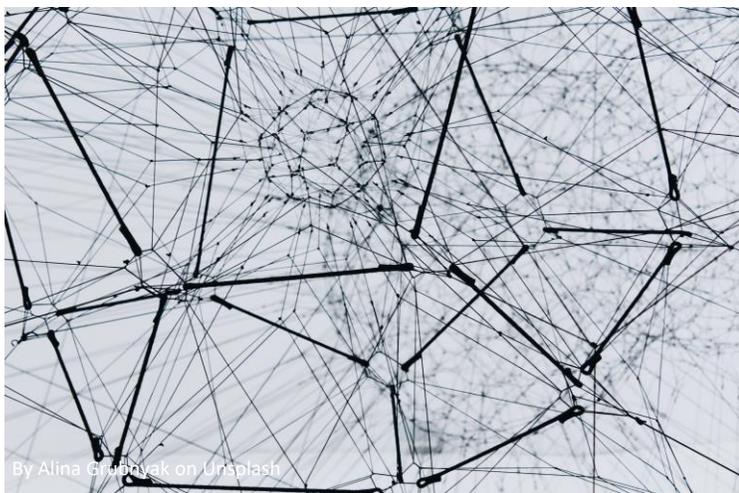
- La autoevaluación del titular de los derechos frente a la creciente complejidad tecnológica: vuelta al pasado
 - Adopción de una evaluación formalizada: PIA/DPIA
 - De la responsabilidad a la rendición de cuentas

-
- Compatibilidad, no alternativa, entre riesgo y protección de derechos

 - Diferentes enfoques del riesgo
 - Enfoque basado en los riesgos/beneficios (regulación eficiente)
 - Enfoque basado en el daño
 - Enfoque basado en los derechos

-
- Diferentes tipos de intereses en juego y aceptabilidad del riesgo
 - Solo daños materiales/contractuales
 - Derechos fundamentales
 - Perjuicio directo
 - Perjuicio indirecto (normas de procedimiento del RGPD)

El impacto más amplio de la IA



- De la protección de datos a los derechos fundamentales
- La experiencia de las autoridades de protección de datos y los límites de la perspectiva de la protección de datos
- Retos
 - ‘Vernaculisation’
 - Falta de experiencia
 - Arquitectura confusa de las autoridades supervisoras
- Gestión de riesgos: la diferencia entre la evaluación previa y las decisiones judiciales

- Regulación de la IA en la UE: ventajas y límites
 - Ser los primeros en actuar y el dilema de Collingridge
 - Atención a los derechos fundamentales

 - Lógica predominante centrada en la seguridad de los productos
 - Calidad de la redacción y efectos negativos de los compromisos alcanzados en el trílogo (por ejemplo, el artículo 27 del RIA)
 - Amplio papel de la CE y de la Oficina de IA
 - Agenda política y necesidades normativas

-
- Necesidad de coherencia sistémica en la aplicación del RIA (regulación en materia de ciberseguridad, protección de datos y otras regulaciones sectoriales basadas en el riesgo)

 - Coherencia con la teoría y la forma de actuar en materia de riesgos
 - Probabilidad y gravedad
 - Niveles de riesgo
 - Evaluación basada en el contexto
 - Enfoque circular

- Coherencia con la lógica y el marco de los derechos fundamentales
 - Evaluación independiente de cada derecho
 - Uso de escalas (bajo, medio, alto y muy alto) para definir las variables de riesgo, tal y como se hace en la jurisprudencia

El modelo FRIA



- Las raíces de la FRIA
 - Los Principios Rectores sobre las empresas y los derechos humanos de las Naciones Unidas
 - La práctica de la EIDH (HRIA)
 - Herramientas de política industrial frente a instrumentos normativos

- Diferencia entre la evaluación del impacto algorítmico (AIA), la evaluación ética y la evaluación de las estrategias nacionales de IA

-
- Diferencias entre modelos
 - Modelos de concienciación y modelos de evaluación de riesgos
 - Modelos lineales (diseño Why-What-How design) y modelos circulares
 - Modelos matemáticos y el enfoque jurídico de la FRIA

Errores de método:
matematizar el FRIA
o utilizar la lógica
formal



Combinaciones matemáticas de elementos heterogéneos, incluidos los derechos fundamentales (por ejemplo, Bertaina et al. 2025)

- Media del riesgo total, que incluye características heterogéneas (tipología del algoritmo, propiedad y control, proceso de toma de decisiones y supervisión humana, datos de entrada, rendimiento, transparencia, explicabilidad, equidad, seguimiento y mantenimiento)
- Impacto conjunto en los derechos fundamentales
- Ponderación arbitraria de los riesgos generales (30 %) y del impacto conjunto en los derechos fundamentales (70 %)

Do the third parties provide the relevant information to perform this assessment? Not applicable (e.g., AI system has been developed internally, third parties' contribution is non-relevant for FRAIIA) (risk level 1); Yes (choose a risk level between 2 and 3); No (risk level 4).

- Incoherencia: la transparencia debería reducir la puntuación de riesgo, no aumentarla

Table 8

The list of possible values for class severity, class effort and duration and class probability of occurrence, together with an explanation of the value meaning.¹³

Value	Class Severity	Class Effort and Duration	Class Probability of Occurrence
Neutral (-)	no impact at all is expected for the considered FR, and thus the Severity/Effort and Duration of Classes is not even evaluated		
Irrelevant (0)	the Severity/Effort and Duration is evaluated, and the result is that no impact or an irrelevant impact is expected		Probability of Occurrence < 1%.
Low (1)	a low impact is expected		1% < Probability of Occurrence < 35%. ⁷
Medium (2)	a medium impact is expected		35% < Probability of Occurrence < 70%. ⁷
High (3)	a high impact is expected		Probability of Occurrence > 70%. ⁷

Matematización de variables clave (por ejemplo, la probabilidad en Bertaina et al. 2025)

- Dependencia de series históricas en lugar de un enfoque basado en rangos (definición arbitraria de series relevantes)
- Combinaciones opacas de parámetros clave (tablas 8 y 9)

Table 9

The mapping table that defines the value for class risk score given a combination of class severity, class effort and duration and class probability of occurrence.

Class Risk Score	Corresponding combinations of: Class severity – class effort and duration – class probability of occurrence
0	1-0-1; 0-3-1; 0-2-1; 0-1-1; 0-0-1; 0-3-2; 0-2-2; 0-1-2; 0-0-2; 0-3-3; 0-2-3; 0-1-3; 0-0-3; 3-3-0; 3-2-0; 3-1-0; 3-0-0; 2-3-0; 2-2-0; 2-1-0; 2-0-0; 1-3-0; 1-2-0; 1-1-0; 1-0-0; 0-0-0
1	3-0-1; 2-1-1; 2-0-1; 2-0-2; 1-2-1; 1-1-1; 1-1-2; 1-0-2; 1-0-3
2	3-3-1; 3-2-1; 3-1-1; 3-1-2; 3-0-2; 3-0-3; 2-3-1; 2-2-1; 2-2-2; 2-1-2; 2-1-3; 2-0-3; 1-3-1; 1-3-2; 1-2-2; 1-2-3; 1-1-3
3	3-3-2; 3-2-2; 3-3-3; 3-2-3; 3-1-3; 2-3-2; 2-3-3; 2-2-3; 1-3-3
-	Whenever Class Severity is Neutral (and as a consequence Class Effort and Duration is Neutral), independently of Class Probability of Occurrence

Let \mathcal{W}_1 be the initial set of empirical information consisting of $\{R(a), I(a), B(a), S(a), N(a)\}$. Where $R(a)$ represents the premise that the AI is intended to be used for recruitment; $I(a)$ that the AI is designed for conducting interviews; $B(a)$ that the AI uses biometrics; $S(a)$ that the AI is used for screening application materials and online activity; and $N(a)$ that the AI employs NLP.³⁰

And let \mathcal{D}_1 be the initial set of defaults consisting of $\{\delta_1, \delta_2, \delta_3, \delta_4\}$ (with no priority relation between them, i.e., $< = \emptyset$). Where $\delta_1 = \{R(x) \rightarrow HR(x)\}$ (the default that if the AI is intended for recruitment purposes, then it is a high-risk system)³¹; $\delta_2 = \{HR(x) \rightarrow FRIA(x)\}$ (the default that if the AI is classified as high-risk, then a FRIA must be implemented before deployment)³²; $\delta_3 = \{APP(y) \rightarrow ND(y)\}$ (the default that anyone applying for a job has the right to

non-discrimination)³³; and $\delta_4 = \{APP(y) \rightarrow PR(y)\}$ (the default that anyone applying for a job has the right to privacy).³⁴

In this scenario, given that $\delta_1 - \delta_4$ are all binding, Cybercruitment is rationally committed to accepting, on the one hand, that since the AI system is intended to be used for recruitment purposes, then it is high-risk ($HR(a)$) and so a FRIA must be conducted prior to deployment ($FRIA(a)$); and on the other, that if someone b submits a job application, their fundamental rights of non-discrimination ($ND(b)$) and privacy ($PR(b)$) must be protected throughout the implementation of the AI system.

Now, in order to determine whether the AI poses any serious risk to these fundamental rights, Cybercruitment can move then to the next scenario.

Uso de la lógica formal abstracta (por ejemplo, García-Godínez, Miguel. 2025, <https://doi.org/10.1007/s43681-025-00761-1>)

S2. Risk Identification

Let $\mathcal{W}_2 = \{\mathcal{W}_1\}$, $\mathcal{D}_2 = \{\mathcal{D}_1 \cup (\delta_5, \delta_6)\}$, and $< = \emptyset$. Where $\{\mathcal{W}_1\}$ and $\{\mathcal{D}_1\}$ are as before, and $\delta_5 = \{B(x) \rightarrow \neg ND(y)\}$ (if the AI uses inferential biometric technology, then it undermines the right to non-discrimination) and $\delta_6 = \{N(x) \rightarrow \neg PR(y)\}$ (if the AI uses NLP to score application materials, then it affects right to privacy).

The rationale behind δ_5 is that biometric systems may disproportionately disadvantage certain demographic groups due to biases in data and algorithmic design [56]. And the justification behind δ_6 is that NLP may expose individuals to invasive and potentially prejudiced assessments [57]. So, in the specific situation where b is evaluated through the AI, the proper scenario would then yield both $\neg ND(b)$ and $\neg PR(b)$.

However, this result conflicts with the consequence that obtains from S1. In particular, with the normative conclusions that follow from δ_3 and δ_4 , viz., that the applicant's right to non-discrimination and privacy must be protected (i.e., $ND(b)$ and $PR(b)$). To solve this conflicting situation, Cybercruitment can move next to the risk mitigation scenario.

Los elementos de la FRIA



- Fase de planificación y delimitación del alcance (cuestionario de evaluación)
 - Centrarse en las principales características del producto/servicio y en el contexto en el que se utilizará
- Fase de recopilación de datos y análisis de riesgos
 - Identificar los riesgos potenciales y estimar los posibles impactos en los derechos fundamentales
- Fase de gestión de riesgos
 - Adoptar las medidas adecuadas para prevenir o mitigar los riesgos identificados, comprobar la eficacia de estas medidas y realizar un seguimiento continuo cuando sea necesario

- **El cuestionario de evaluación**

- Contextualizar el sistema de IA y su uso
- Identificar las categorías que podrían verse afectadas
- La FRIA no consiste en marcar casillas, sino que es una herramienta de diseño

- Las secciones del cuestionario
 - Sección A - Descripción y análisis del sistema de IA, incluidos los flujos de datos relacionados
 - Sección B - Contexto de derechos fundamentales
 - Sección C - Controles implementados
 - Sección D - Intervención de las partes interesadas y diligencia debida

Análisis y gestión de riesgos

- Evaluación del impacto
 - Uso de escalas (nivel mínimo/máximo de impacto)
 - Uso de variables ordinales (por ejemplo, bajo, medio, etc.)
 - Las escalas y las variables facilitan la evaluación comparativa, la transparencia y la rendición de cuentas
- Adopción de medidas adecuadas de prevención y mitigación de riesgos
- Aplicación de dichas medidas
- Seguimiento

-
- Componentes de riesgo (índice de riesgo) para cada derecho que pueda verse afectado
 - Probabilidad de impacto adverso (basada tanto en casos anteriores, analizando situaciones comparables, como en el uso de técnicas analíticas y de simulación, basadas en posibles escenarios de uso)
 - Probabilidad de impactos perjudiciales
 - Exposición

-
- Gravedad (mayor énfasis en el análisis jurídico relativo a la gravedad del perjuicio: jurisprudencia sobre derechos humanos/fundamentales, marco jurídico)
 - Gravedad: gravedad del perjuicio en el ejercicio de los derechos y libertades (gravedad), teniendo en cuenta el impacto específico en determinados grupos, la vulnerabilidad y las situaciones de dependencia
 - Esfuerzo: el esfuerzo por superar los prejuicios y revertir los efectos adversos (esto también pone de relieve la dimensión contextual y las interacciones)

		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

Bajo	Medio	Alto	Muy alto

		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
Probabilidad	Baja				
	Media				
	Alta				
	Muy alta				

- No existe un índice compuesto que combine todos los impactos potenciales sobre los derechos
- Factores que pueden excluir el riesgo (por ejemplo, el carácter obligatorio de determinadas características con impacto)
- El uso de matrices para construir índices de riesgo
 - Relativamente fácil de usar y explicar
 - Transparencia y rendición de cuentas (desarrollo de la IA centrado en el ser humano)
 - Coherencia con la lógica jurídica de los tribunales

■ Gestión de riesgos

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

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

- Identificar las medidas adecuadas y evaluar el riesgo residual
 - Si bien deben mitigarse todos los impactos, los diferentes niveles de impacto sobre los distintos derechos pueden ayudarnos a comprender dónde intervenir primero y dónde es necesario revisar más a fondo el diseño de la IA
- Una evaluación dinámica y circular

Un modelo
científicamente
validado y probado en
aplicaciones reales.



- Mantelero, A. 2024. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: roots, legal obligations and key elements for a model template. *Computer Law & Security Review*. 54, <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. *Computer Law & Security Rev.* 41, <https://doi.org/10.1016/j.clsr.2021.105561> (open access)



- Desde los modelos teóricos hasta su validación concreta
 - Una plataforma de analítica avanzada del aprendizaje
 - Una herramienta para la gestión de los recursos humanos
 - Una herramienta de imágenes médicas impulsada por IA para la detección del cáncer
 - ATENEA, IA al servicio de las personas mayores

Una herramienta de imagen médica basada en IA para la detección del cáncer

- Primera fase: desarrollo de un sistema de IA basado en imágenes médicas (5000 pacientes de diez países europeos)
- Segunda fase: validación del sistema de IA en ocho centros sanitarios de todo el mundo fuera de Europa (un centro sanitario en Asia, uno en África y uno en Sudamérica)

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.

	que pueden verse afectados por los sistemas de IA? ¿Se incluyen entre ellos individuos o grupos vulnerables?	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.
	Identificación de los responsables: ¿qué personas/entidades están implicadas en el diseño, desarrollo y despliegue de los sistemas de IA? ¿Cuál es su papel?	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.
Sección B Contexto de derechos fundamentales	¿Qué derechos fundamentales se ven potencialmente afectados por el uso del sistema de IA?	<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).
	¿Qué instrumentos jurídicos internacionales/regionales de protección de los derechos humanos/fundamentales se han aplicado a nivel operativo?	Declaración universal de los derechos humanos, Carta de derechos fundamentales de la UE, normativas de protección de datos aplicable.
	¿Cuáles son los tribunales u organismos más relevantes en materia de derechos fundamentales en el contexto?	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.

	¿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>	No
	<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
	<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>	No
	<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>	N/A

Tabla 1A. Recogida de datos y análisis de riesgos

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/A	MA

Probabilidad			
Baja	Media	Alta	Muy alta

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

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 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.

Derecho a la salud física y mental	<p>[Alta]</p> <p>Cuando se utiliza en pacientes europeos.</p> <p>[Muy alta]</p> <p>Cuando se utiliza en pacientes no europeos.</p>	<p>[Alta]</p> <p>Cuando el control posterior del cáncer puede corregir el error del sistema</p> <p>[Muy alta]</p> <p>Cuando el control posterior del cáncer no puede corregir el error del sistema.</p>	<p>[Alto]</p> <p>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]</p> <p>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]	[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.

Integrar FRIA y DPIA



- Cuestiones jurídicas y operativas
 - Artículo 27(5) de la Ley de IA
 - DPIA/FRIA/CA: solapamientos y flujos de datos necesarios
 - Evitar la duplicación y la incoherencia

- Retos principales
 - Variedad de modelos de DPIA y su compatibilidad con la FRIA
 - Identificar una estructura principal común
 - Aplicar un procedimiento eficaz de intercambio de datos y metodologías comunes de evaluación de riesgos

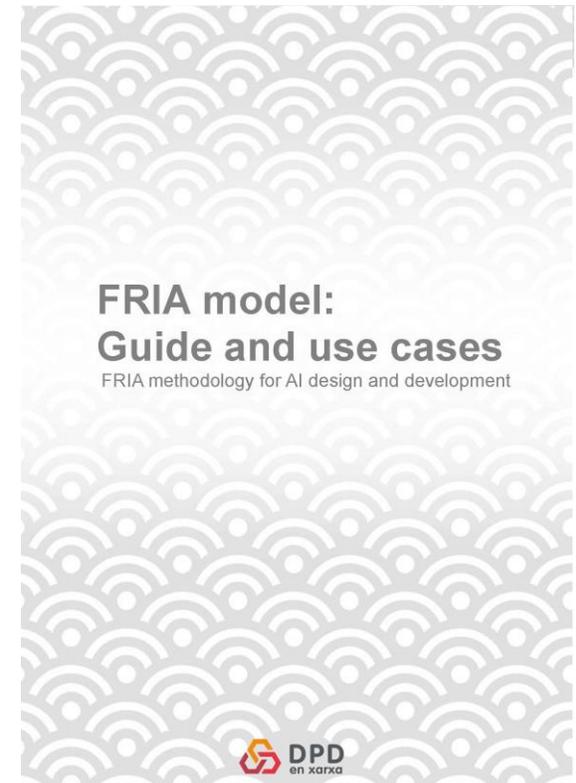


Alessandro Mantelero
alessandro.mantelero@polito.it



Mantelero, A. 2024. The Fundamental Rights Impact Assessment (FRIA) in the AI Act: roots, legal obligations and key elements for a model template. Computer Law & Security Review. 54, <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. Computer Law & Security Rev. 41, <https://doi.org/10.1016/j.clsr.2021.105561> (open access)



https://apdcat.gencat.cat/en/documentacio/intelligencia_artificial/