



**Politecnico
di Torino**

Metodología de evaluación del impacto sobre los derechos fundamentales en IA

Universidad de Murcia
29 de julio de 2026

Alessandro Mantelero

Càtedra Jean Monnet de Societats Digitals Mediterrànies i Dret
Professor titular de Dret Privat i Dret i Tecnologia | Politècnica de Torí

I. Las raíces de la FRIA



- Los Principios Rectores sobre Empresas y Derechos Humanos de Naciones Unidas
- Diferencia entre directrices, metodologías y modelos
- Diferencia entre el análisis de impacto algorítmico (AIA), la evaluación ética y la evaluación de estrategias para el desarrollo de la IA

-
- Diferencias entre los modelos
 - De gestión de proyectos (incluidos los riesgos de mala gestión)
 - De gestión de riesgos (Artt. 9, 17, 43, RIA) — CEN-CENELEC
 - De FRIA (Artt. 27, RIA) — APDCAT/AI Office
 - De protección de datos para la IA — CNIL
 - De evaluación de impacto ético, social y jurídico — FRAIA
 - De evaluación ética

-
- Los que requiere el RIA
 - Proveedores de SIA: sistema de gestión del riego (incluida la FRIA) y de la calidad [Evaluación de la conformidad]
 - Responsables del despliegue: FRIA

 - Diferencias entre los modelos FRIA
 - Modelos de concienciación y modelos de evaluación del riesgo
 - Modelos lineales con diseño Why-What-How y modelos de evaluación de riesgo circulares

II. Errores de método



- Los enfoques puramente matemáticos son incompatibles con la lógica de los derechos fundamentales
 - La FRIA no es una herramienta de auditoría, sino de diseño
 - Diseño basado en los derechos fundamentales
 - Rendición de cuentas (herramienta para proveedores, responsables del despliegue y autoridades de supervisión)
 - Necesidad de compatibilidad con la perspectiva jurídica de las autoridades de supervisión y los tribunales

Matematización de variables clave (p. ej., Bertaina et al. 2025)

■ Combinaciones opacas de parámetros clave

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

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

III. 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

Cuestionario de 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? ¿Cuáles son las principales características del sistema? ¿En qué países se ofrecerá? ¿Qué tipos de datos se tratan (personales, no personales, categorías especiales)? 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? Identificación de los responsables: ¿qué personas/entidades están implicadas en el diseño, desarrollo y despliegue de los sistemas de IA? ¿Cuál es su papel?
Sección B Contexto de derechos fundamentales	¿Qué derechos fundamentales se ven potencialmente afectados por el uso del sistema de IA? ¿Qué instrumentos jurídicos internacionales/regionales de protección de los derechos humanos/fundamentales se han aplicado a nivel operativo? ¿Cuáles son los tribunales u organismos más relevantes en materia de derechos fundamentales en el contexto de uso? ¿Cuáles son las decisiones y disposiciones más relevantes en materia de derechos humanos/fundamentales?

■ 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 (p. ej., 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

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

- 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: la complejidad de abordar y revertir los efectos adversos (esto también pone de relieve la dimensión contextual y las interacciones)

-
- El uso de matrices para construir índices de riesgo
 - Relativamente fácil de usar y explicar
 - Transparencia y rendición de cuentas
 - No existe un índice compuesto que combine todos los impactos potenciales sobre los derechos
 - Los factores que pueden excluir el riesgo (p. ej., el carácter obligatorio de determinadas características con impacto) se consideran después de la FRIA
 - Los factores que pueden justificar el riesgo (balance a favor de intereses prevalentes) afectan a la aceptabilidad, pero no a la evaluación del nivel de impacto

■ Gestión de riesgos

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

IV. 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

ATENEA, la IA al servicio de las personas mayores (proyecto evaluado en fase piloto)

- El proyecto está dirigido específicamente a personas mayores de 65 años que viven solas y que son más vulnerables a causa de la brecha digital
- Responde a diferentes casos de uso, como:
 - Llamar a familiares
 - Llamar a emergencias (112)
 - Citas con los servicios sociales
 - Citas de atención primaria
 - Recordatorios de agenda y rutas de transporte

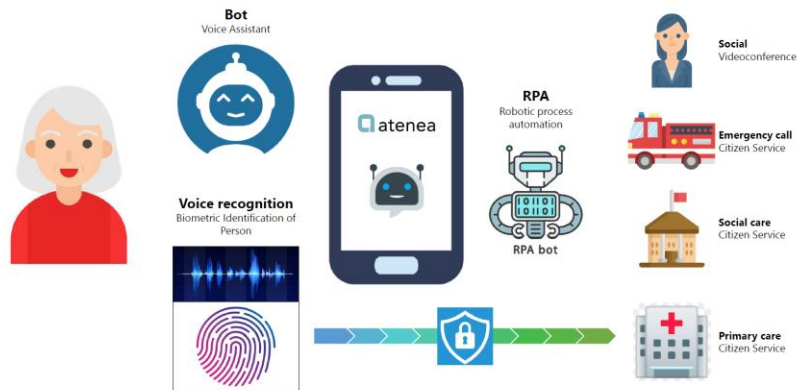


Tabla 1A. Recogida de datos y análisis de riesgos

Derechos/libertades potencialmente afectados	Descripción del impacto	Probabilidad			Gravedad		
		Probabilidad de resultados adversos	Exposición	Probabilidad	Gravedad del impacto	Esfuerzo	Gravedad
No discriminación	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). Se pueden producir discriminaciones cuando las personas usuarias no hablen correctamente el idioma del sistema (CAT/ES).	[Alto] Posibles problemas en la voz y el habla de las personas usuarias pueden generar que queden excluidas.	[Medio] Habrà un número limitado de personas con problemas de voz o de idioma.	[Medio]	[Muy alto] Impacto negativo en la igualdad de acceso a la asistencia que ofrece la solución y en la calidad del servicio.	[Alto] Sería necesario adaptar el asistente de voz para integrar las casuísticas de problemas en el habla o idioma. 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.	[Muy alto]

No discriminación	[Media]	[Muy Alta]	[Alto]	- 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.
-------------------	---------	------------	--------	--

Tabla 3A. Gestión de riesgos (II)

Derecho/libertad afectados	Probabilidad (residual)	Gravedad (residual)	Impacto residual
Protección de datos/Privacidad	[Baja] 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.	[Media] 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.	[Medio]
No discriminación	[Baja] 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).	[Media] 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).	[Medio]

Alessandro Mantelero
alessandro.mantelero@polito.it

Jean Monnet Chair in
Mediterranean Digital Societies
and Law



https://www.dpdenxarxa.cat/pluginfile.php/2468/mod_folder/content/0/FRIA_es_2.pdf