

Model per a l'AIDF: guia i casos d'ús

Metodologia aplicada de l'avaluació d'impacte sobre els drets fonamentals en el disseny i desenvolupament de la IA

INDEX

Presentació	4
Part I – L'AIDF i la metodologia AIDF	6
1. Introducció.....	6
2. El paper de l'AIDF en la regulació de la IA.....	7
3. La interacció amb la normativa de protecció de dades.....	8
4. El model d'AIDF adoptat	9
4.1 La fase de planificació i determinació de l'abast	10
4.2 Fase de recopilació de dades i anàlisi de riscos	11
4.2.1 Variables clau per a l'avaluació d'impacte.....	12
4.2.2 Variables i construcció de l'índex d'impacte	13
4.3 La fase de gestió de riscos.....	14
5. La plantilla del model AIDF.....	15
Qüestionari de planificació i abast	15
Matrius de risc.....	17
Taula 1. Probabilitat	17
Taula 2. Exposició	17
Taula 3. Probabilitat que es produeixi un perjudici (<i>likelihood</i>).....	17
Taula 4. Gravetat del perjudici	18
Taula 5. Esforç per superar els perjudicis i revertir els efectes adversos.....	18
Taula 6. Gravetat (<i>severity</i>)	18
Taula 1A. Recollida de dades i anàlisi de riscos.....	19
Taula 7. Impacte general del risc	19
Taula 2A. Gestió de riscos (I)	19
Taula 3A. Gestió de riscos (II).....	20
6. Del model a la pràctica: els casos d'ús.....	20
Referències	22
Part II – Casos d'ús.....	23
Cas d'ús 1: Plataforma d'anàlisi avançada de l'aprenentatge	23
1. El context	23
2. El projecte	26
3. L'AIDF	28
4. Comentaris.....	42
Cas d'ús 2: Una eina per a la gestió dels recursos humans.....	44
1. El context	44
2. El projecte	44
3. L'AIDF	46
4. Comentaris.....	58
Cas d'ús 3: Una eina d'imatges mèdiques impulsada per IA per a la detecció del càncer.....	59
1. El context	59
2. El projecte	59
3. L'AIDF	60
4. Comentaris.....	71
Cas d'ús 4: ATENEA, IA al servei de les persones grans.....	72
1. El context	72
2. El projecte	72
3. L'AIDF	74
4. Comentaris.....	95

Coordinadors:

Alessandro Mantelero i Joana Marí

Autors:

Alessandro Mantelero (Universitat Politècnica de Torí)
Cristina Guzmán (Universitat Politècnica de Catalunya – BarcelonaTech (UPC))
Esther Garcia (CaixaBank, S.A)
Ruben Ortiz (Universitat de Barcelona)
M. Ascensión Moro (Ajuntament de Sant Feliu de Llobregat)

Participants:

Albert Portugal (Consorti de Serveis Universitaris de Catalunya)
Albert Serra (Autoritat Catalana de Protecció de Dades)
Alessandro Mantelero (Universitat Politècnica de Torí)
Cristina Guzmán (Universitat Politècnica de Catalunya – BarcelonaTech (UPC))
Esther Garcia (CaixaBank, S.A)
Joana Judas (Departament de Recerca i Universitats. Generalitat de Catalunya)
Joana Marí (Autoritat Catalana de Protecció de Dades)
Jordi Escolar (Agència de Qualitat Universitària)
M. Ascensión Moro (Ajuntament de Sant Feliu de Llobregat)
Marc Vives (Universitat Pompeu Fabra)
Maria José Campo (Fundació TIC Salut i Social)
Mariona Perramon (Consorti de Serveis Universitaris de Catalunya)
Olga Rierola (Autoritat Catalana de Protecció de Dades)
Patricia Lozano (Universitat Oberta de Catalunya)
Ruben Ortiz (Universitat de Barcelona)
Sara Hernández (Fundació TIC Salut i Social)



Barcelona, 2025

El contingut d'aquest document és propietat de l'Autoritat Catalana de Protecció de Dades i està subjecte a la llicència Creative Commons BY-NC-ND, <https://creativecommons.org/licenses/by-nc-nd/4.0/>.

L'autoria de l'obra es reconeixrà mitjançant la inclusió de la menció següent:

Obra titularitat de l'Autoritat Catalana de Protecció de Dades.

Llicència sota la llicència CC BY-NC-ND.

Avís: Quan es reutilitza o distribueix l'obra, s'han d'indicar clarament els termes de la llicència d'aquesta obra.

Avís legal: Les opinions expressades en aquest document són responsabilitat dels seus autors i no reflecteixen necessàriament l'opinió oficial de l'APDCAT. L'Autoritat Catalana de Protecció de Dades i, els autors i els membres del grup de treball no es fan responsables de les possibles conseqüències de les actuacions de les persones físiques o jurídiques arran de qualsevol informació continguda en aquest document.

Presentació

En el seu Pla estratègic 2023-2028, l'Autoritat Catalana de Protecció de Dades (APDCAT) va fixar-se com a pilar estratègic la promoció del desenvolupament de formació adreçada a les institucions. I una de les línies d'acció clau per aconseguir-ho era "impulsar la creació i potenciar la comunitat d'aprenentatge de delegats de protecció de dades."

Doncs bé, el 15 de juliol del 2023 va arrencar la primera [Xarxa de delegats i delegades de protecció de dades](#), pionera a Catalunya i a l'Estat espanyol. La plataforma va néixer amb l'objectiu de contribuir a estendre la cultura de la privacitat a Catalunya, de la mà dels delegats/delegades de protecció de dades (DPD), que vetllen per garantir el compliment de la normativa de protecció de dades, promoure la cooperació i col·laboració entre DPD i compartir coneixements i expertesa.

Aquesta plataforma aglutina els DPD de les més de 1.700 entitats que formen part de l'àmbit d'actuació de l'APDCAT, que comprèn administracions públiques com la Generalitat de Catalunya, ajuntaments, universitats públiques i privades i col·legis professionals, entre altres. Així mateix, està oberta als DPD de les entitats que presten serveis al sector públic català com a encarregades del tractament i, també, als DPD en actiu d'organitzacions amb seu a Catalunya.

La Xarxa, que actualment ja reuneix un important nombre de DPD de Catalunya, persegueix els objectius principals següents:

- Ser una institució més propera als DPD i a les organitzacions.
- Constituir un espai d'intercanvi d'idees, experiències i coneixements.
- Potenciar la figura del DPD, com a peça clau del compliment.
- Identificar i fomentar les millors pràctiques.
- Crear i difondre un model de relació i cooperació, compatible amb una supervisió útil i eficaç.
- Construir un entorn de serveis d'interacció i col·laboració obert, de generació de recursos, aprenentatge i gestió del coneixement.
- Crear un espai referent a Catalunya i Europa per als DPD.

Fruit de la creació d'aquest espai, des de l'apartat Àgora de la Xarxa es va proposar la creació d'un grup de treball anomenat "Orientacions metodològiques. Avaluacions d'impacte. Drets i llibertats." El grup el van coordinar el doctor Alessandro Mantelero¹ i la senyora Joana Marí², durant els mesos de febrer a desembre de l'any 2024, i està format pels membres de la Xarxa següents:

- Albert Portugal (Consorci de Serveis Universitaris de Catalunya)
- Albert Serra (Autoritat Catalana de Protecció de Dades)
- Cristina Guzmán (Universitat Politècnica de Catalunya – BarcelonaTech (UPC))
- Esther Garcia (Caixabank, SA)
- Joana Judas (Departament de Recerca i Universitats. Generalitat de Catalunya)
- Jordi Escolar (Agència de Qualitat Universitària)
- M. Ascensión Moro (Ajuntament de Sant Feliu de Llobregat)
- Marc Vives (Universitat Pompeu Fabra)
- Maria José Campo (Fundació TIC Salut i Social)
- Mariona Perramon (Consorci de Serveis Universitaris de Catalunya)
- Olga Rierola (Autoritat Catalana de Protecció de Dades)

¹ Professor titular de Dret Privat i Dret i Tecnologia a la Politècnica de Torí, Càtedra Jean Monnet de Societats Digitals Mediterrànies i Dret.

² Delegada de protecció de dades i responsable de Projectes Estratègics de l'Autoritat Catalana de Protecció de Dades.

- Patricia Lozano (Universitat Oberta de Catalunya)
- Ruben Ortiz (Universitat de Barcelona)
- Sara Hernández (Fundació TIC Salut i Social)

L'objectiu d'aquest equip de treball era desenvolupar una nova metodologia per a les avaluacions d'impacte en els drets fonamentals, quan s'utilitzen sistemes d'intel·ligència artificial, tenint en compte el marc normatiu europeu establert pel Reglament (UE) 2024/1689 del Parlament Europeu i del Consell, de 13 de juny de 2024 (RIA), pel que s'estableixen normes harmonitzades en matèria d'intel·ligència artificial i pel qual es modifiquen els reglaments (CE) núm. 300/2008, (UE) núm. 167/2013, (UE) núm. 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 i les directives 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828. Aquesta metodologia s'havia de diferenciar de les avaluacions d'impacte en matèria de protecció de dades que exigeix el Reglament (UE) núm. 2016/679 del Parlament Europeu i del Consell, de 27 d'abril de 2016, relatiu a la protecció de les persones físiques pel que fa al tractament de dades personals i a la lliure circulació d'aquestes dades i pel qual es deroga la Directiva 95/46 / CE (RGPD).

Fruit de la tasca d'aquest grup s'ha elaborat el document que presentem, que s'estructura en dues parts: a la primera es descriu la metodologia; i a la segona, s'aplica a casos concrets. Amb aquest treball es busca facilitar una eina útil i pràctica per a les entitats que dissenyen, desenvolupen o utilitzen sistemes o models d'intel·ligència artificial i, en particular, per a les persones encarregades d'executar les avaluacions d'impacte sobre els drets fonamentals.

En definitiva, la prevenció de la vulneració dels drets i llibertats fonamentals és una tasca comuna on l'Autoritat Catalana de Protecció de Dades hi ha de jugar un paper clau.



Meritxell Borràs i Solé
Directora
Autoritat Catalana de Protecció de Dades

Part I – L'AIDF i la metodologia AIDF

1. Introducció

Aquesta publicació presenta els resultats d'un projecte liderat per l'Autoritat Catalana de Protecció de Dades (APDCAT) amb l'objectiu de proporcionar als operadors d'IA, tant proveïdors com responsables de desplegament, una eina eficaç per desenvolupar solucions d'IA fiables i centrades en l'ésser humà³. Des d'aquest punt de vista, el disseny i desenvolupament de solucions d'IA que respectin els drets humans/fonamentals és el nucli d'una IA veritablement centrada en l'ésser humà, com demostren el Reglament d'IA (RIA) i altres iniciatives nacionals i internacionals com ara el Conveni marc del Consell d'Europa sobre intel·ligència artificial, drets humans, democràcia i estat de dret.

A la llum de l'anterior, el RIA estableix com un dels seus principals objectius garantir "un elevado nivel de protección de la salud, la seguridad y los derechos fundamentales consagrados en la Carta" (art. 1 RIA). I, en línia amb l'enfocament basat en el risc adoptat pel legislador de la UE, l'avaluació de l'impacte de la IA s'inclou en els drets fonamentals⁴, en tots els procediments de gestió de riscos establerts pel Reglament: des de l'avaluació de la conformitat fins a l'avaluació d'impacte específica sobre els drets fonamentals d'acord amb l'article 27 del RIA, inclosa una disposició específica per als models d'IA d'ús general amb risc sistèmic, **l'impacte sobre els drets fonamentals s'ha de tenir en compte en el disseny, desenvolupament i desplegament dels sistemes i models d'IA.**

En aquest context, **les disposicions sobre com dur a terme aquesta avaluació en el RIA, fins i tot en el Conveni marc del Consell d'Europa, només proporcionen una orientació limitada als qui l'han d'executar.** D'altra banda, els models proposats i el debat inicial en la literatura mostren diverses mancances des del punt de vista metodològic [Mantelero, 2024]. Per tant, en consonància amb un enfocament empíric del Dret, per provar la viabilitat i l'eficàcia dels models per dur a terme l'avaluació d'impacte sobre els drets fonamentals en el context de la IA **cal passar de l'elaboració abstracta, en el Dret i en el debat jurídic, a aplicacions concretes.**

El projecte català és la primera iniciativa basada en la implementació concreta d'una metodologia AIDF, aplicada a casos reals i fonamentada en una interacció activa amb entitats públiques i privades que apliquen solucions d'IA en els seus negocis i activitats. **Els resultats d'aquest enfocament empíric són crucials per a la implementació efectiva del RIA,** ja que demostren que és possible racionalitzar els requisits del Reglament i traduir-los, eficaçment, en un procés d'anàlisi i gestió de riscos que sigui coherent tant amb la teoria general del risc com amb el marc dels drets fonamentals.

A més, **l'evidència empírica proporcionada per aquesta publicació pot contribuir al debat internacional i de la UE** sobre el model per a l'avaluació d'impacte sobre els drets humans/fonamentals, proporcionant evidència sobre qüestions crucials com ara: (i) les variables rellevants que s'han de considerar; (ii) la metodologia per avaluar-les i crear índexs de risc; (iii) el rol que poden tenir els qüestionaris estàndard en aquest exercici i les seves limitacions; iv) el paper de l'*expert-based evaluation* en aquesta anàlisi.

El model AIDF utilitzat en els nostres casos d'ús (Mantelero, 2024; Mantelero-Exposito, 2021), així com els mateixos casos d'ús, es posen a disposició del públic per contribuir al debat internacional i de la UE

³ Vegeu també l'Agència dels Drets Fonamentals de la Unió 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.").

⁴ Aquí i a les pàgines següents, les referències als drets fonamentals (o simplement drets) inclouen tant els drets fonamentals com les llibertats consagrades a la Carta dels Drets Fonamentals de la Unió Europea.

sobre la protecció dels drets fonamentals en el context de la IA. També, **per servir de font d'inspiració per a les nombroses entitats de tot Europa i de països no pertanyents a la UE que volen adoptar un enfocament de la IA** basat en els drets fonamentals, però no disposen d'un model de referència provat ni de casos concrets amb els quals comparar les seves experiències.

Amb aquest objectiu a la ment, a les seccions següents es discutirà breument el paper de l'AIDF en la regulació de la IA, la seva interacció amb la regulació de protecció de dades, el model utilitzat en els casos d'ús, els criteris de selecció de casos i les àrees cobertes.

2. El paper de l'AIDF en la regulació de la IA

L'última primavera de la IA, en els últims anys, ha portat que les aplicacions d'aprenentatge automàtic s'utilitzin en una varietat d'escenaris operatius tant en el sector públic com en el privat. Un dels principals usos de la IA i font dels desafiaments més grans es relaciona amb el paper que juguen aquestes aplicacions en la presa de decisions. Aquest aspecte encara és més crític en el sector públic, atès el desequilibri de poder que en molts casos caracteritza la relació entre les persones físiques i els poders públics.

Aquest canvi de les decisions basades en humans a les decisions basades en IA, ja sigui totalment automatitzades o amb la IA donant suport als responsables en la presa de decisions, planteja diverses preocupacions sobre la precisió de la lògica d'aquests sistemes d'IA, la forma en què representen la societat i la interacció humà-màquina. Si bé els desenvolupaments recents en IA estan aportant millores significatives en molts sectors, és important conèixer i gestionar els possibles efectes secundaris d'aquesta tecnologia.⁵

Els riscos associats a les tecnologies d'IA poden tenir un impacte negatiu en la societat, a causa **de qüestions de seguretat i drets fonamentals**. Si bé els problemes de seguretat, tot i que complicats i relacionats amb les limitacions inherents als models actuals d'IA (com ara les anomenades al·lucinacions), es poden abordar fàcilment des d'una perspectiva reguladora utilitzant pràctiques i eines establertes (per exemple, estandardització, procediments de certificació, auditoria, etc.), abordar els problemes de drets fonamentals és molt més complicat.

Els components intrínsecs, extrínsecs i contextuals d'un sistema d'IA poden variar respecte a l'entorn sociotècnic en el qual s'utilitza. **La forma en què s'ha dissenyat la IA** (elements intrínsecs), en termes de conjunts de dades d'entrenament, ajustos, etc., pot afectar la manera com representa els aspectes socials rellevants per a la presa de decisions amb riscos potencials, per exemple, de discriminació en grups subrepresentats, respecte de l'elegibilitat, en els sistemes d'IA utilitzats en ajudes socials. **La forma en què els sistemes d'IA interactuen amb altres tecnologies** (elements extrínsecs), en funció de la naturalesa i el rendiment d'aquestes últimes, pot limitar l'eficàcia dels sistemes d'IA i, per tant, afectar negativament els drets connexos, com en el cas d'una eina d'imatge mèdica impulsada per IA que detecta malament el càncer, a causa de la baixa qualitat del dispositiu mèdic utilitzat per generar les imatges, i posa en perill el dret a la salut.

Finalment, uns sistemes que funcionin correctament i estiguin ben integrats amb altres dispositius poden tenir diferents repercussions en els drets fonamentals, segons el **context d'ús**. El mateix sistema de videovigilància basat en IA es pot considerar una mesura proporcionada –malgrat el seu impacte en la privacitat individual i col·lectiva i la protecció de dades, així com en la llibertat de reunió en el cas de

⁵ Per a una categorització dels riscos relacionats amb la IA, vegeu també Nacions Unides, AI Advisory Board 2024; Slattery et al. 2024.

manifestacions i protestes– en presència d'alts índexs de criminalitat en algunes zones, i inapropiada en absència d'aquest interès públic preponderant.

A partir d'aquesta breu anàlisi de la interacció entre el desenvolupament i el desplegament de la IA i els drets fonamentals, queda clara *la raison d'être* de l'AIDF en la regulació de la IA, així com els elements bàsics que s'han de tenir en compte per dur a terme una avaluació adequada de l'impacte potencial de la IA en els drets. Les dimensions intrínseca i extrínseca dels sistemes d'IA, el context d'ús, els drets potencialment afectats, la necessitat d'un test de ponderació, els individus i grups afectats i les possibles mesures de prevenció/mitigació, són les diferents àrees de l'AIDF a desenvolupar, per fer front a les qüestions esmentades.

En aquest sentit, l'article 27.1 del RIA sobre AIDF considera (i) el context d'ús i les categories d'actors exposats al risc (art. 27.1.a, b i c); (ii) el possible perjudici en els drets fonamentals (art. 27.1.d); i (iii) les mesures de prevenció/mitigació que s'han d'adoptar (art. 27.i.e i f). Aquesta divisió es reflecteix en les **tres fases de la metodologia AIDF** adoptada en aquest estudi: (i) planificació, definició de l'abast i identificació de riscos, (ii) anàlisi de riscos i (iii) mitigació i gestió de riscos.

La primera fase (planificació, definició de l'abast i identificació de riscos) inclou la descripció del sistema d'IA i el seu context d'ús, atesos els riscos intrínsecs potencials (relacionats amb el mateix sistema) i els riscos extrínsecs (relacionats amb la interacció entre el sistema i l'entorn sociotècnic en el qual s'implementa). Això inclou una descripció del procés en el qual s'emprarà el sistema d'IA d'alt risc (art. 27.1.a), així com el període de temps en què s'utilitzarà cada sistema i la freqüència associada (art. 27.1.b). Un cop s'han definit aquests elements, és possible fer una identificació inicial de les àrees d'impacte del sistema d'IA en termes de les categories d'individus i grups interessats (art. 27.1.c) i els drets que poden estar en joc.

Aquesta fase també serveix com a avaluació preliminar, per excloure de l'AIDF els casos en què estigui clar que no hi ha risc de perjudici per a les persones afectades.

Si per contra, s'identifica un dany potencial, aquest dany s'ha d'examinar en la **segona fase (anàlisi de riscos)**, que ha d'anar més enllà d'una identificació general de les possibles àrees d'impacte i estimar el nivell d'impacte per a cada dret o llibertat.

Hi ha diverses raons per les quals l'estimació individualitzada del nivell d'impacte és essencial. En primer lloc, és una característica de totes les avaluacions d'impacte, des de les mediambientals fins a les de ciberseguretat: no hi pot haver una correcta avaluació de riscos sense una estimació de riscos. En segon lloc, l'estimació és la base de la **tercera fase, que és la definició de les mesures de prevenció/mitigació de riscos (mitigació i gestió de riscos)**: si no s'ha estimat l'impacte, no es pot identificar la idoneïtat i eficàcia de les mesures per eliminar/reduir l'impacte inicial. En tercer lloc, l'estimació és, per tant, funcional per a l'aplicació del principi clau de la rendició de comptes: només definint el nivell de risc abans i després d'adoptar les mesures de prevenció/mitigació és possible demostrar que el risc s'ha abordat de manera específica i eficaç. Per totes aquestes raons, les lletres *d* i *f* de l'apartat 1 de l'article 27 exerceixen un paper central i constitueixen la base per desenvolupar la metodologia d'avaluació del risc.

3. La interacció amb la normativa de protecció de dades

La interacció entre el RIA i l'RGPD en relació amb l'AIDF és doble. D'una banda, l'**article 35, apartats 1 i 7 de l'RGPD** ja inclou una avaluació dels riscos per als drets dels interessats [article 29 grup de treball sobre protecció de dades 2017, 6], però en la pràctica de les AIPD aquesta part s'ha implementat escassament. D'altra banda, el RIA fa èmfasi en l'avaluació de l'impacte en els drets individuals, independentment de l'ús de dades personals en el desenvolupament i la implementació de la IA, però hi ha una manca d'orientació en la implementació de l'AIDF.

L'estret vincle entre l'AIPD [APDCAT 2024] i l'AIDF també és evident en **l'article 27, apartat 4, del RIA**, que estableix 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 mateixa línia, de conformitat amb **l'article 26, apartat 9, del RIA**, els responsables de desplegament d'IA han d'utilitzar la informació facilitada en virtut de l'article 13 del RIA (transparència i comunicació d'informació als responsables de desplegament) per complir la seva obligació de dur a terme una AIPD, d'acord amb l'article 35 de l'RGPD.

En aquest context, si els responsables del desplegament de la IA no compleixen les obligacions de l'AIDF en virtut del RIA, o si les autoritats competents no apliquen adequadament aquestes obligacions, **en el futur les autoritats de protecció de dades poden exercir un paper actiu en l'aplicació de l'AIDF dels sistemes d'IA a través de les disposicions de l'article 35 de l'RGPD**, en la mesura que l'RGPD sigui aplicable (aquest és el cas en moltes situacions en què la IA afecti persones i grups, atesa l'àmplia definició de dades personals i tractament de dades a l'RGPD i el paper de les dades en el desenvolupament, la implementació i l'ús de la IA).

Tenint en compte tots aquests diferents aspectes de la interacció entre l'AIDF al RIA i l'article 35 de l'RGPD, i atesa també l'experiència de les autoritats de protecció de dades en el tractament de qüestions de drets fonamentals [Mantelero-Esposito 2021], **és adequat i necessari que aquestes autoritats exerceixin un paper actiu a l'hora de proporcionar orientació sobre l'AIDF en el context dels sistemes d'IA basats en dades personals.**

4. El model d'AIDF adoptat

Abans d'examinar la metodologia per dur a terme l'AIDF en l'àmbit de la IA, cal assenyalar que es tracta d'una avaluació realitzada per experts que ha de ser coherent tant amb els enfocaments metodològics comunament utilitzats en l'avaluació i gestió de riscos com amb la teoria jurídica dels drets fonamentals. Així, analitzar el potencial impacte de la IA en els drets individuals i col·lectius és un exercici complex que requereix diferents tipus d'experiència, que combina els drets fonamentals, el disseny de la IA i la comprensió del context social. Igual que amb l'AIPD, l'AIDF no és, per tant, un exercici que pugui ser realitzat per persones inexpertes en la matèria.

El model AIDF ha de ser coherent amb les metodologies d'avaluació i gestió de riscos. En aquest sentit, des de la perspectiva dels proveïdors i els responsables de desplegament d'IA, l'AIDF no és una tasca independent, sinó que forma part d'un conjunt integrat d'avaluacions que aquestes entitats han de dur a terme per complir les obligacions legals.

Per exemple, limitant l'atenció a les qüestions relacionades amb les dades, que són el nucli dels sistemes d'IA, els instruments jurídics nacionals i de la UE exigeixen avaluacions de riscos de protecció i seguretat de les dades.

A més, els sistemes d'IA s'utilitzen sovint com a component d'altres tecnologies (per exemple, en la prevenció i detecció d'accidents), que han de complir normes tècniques que inclouen els elements d'avaluació de riscos. Davant d'aquest escenari, l'AIDF ha d'estar en consonància amb les

metodologies comunes d'avaluació de riscos;⁶ no només en termes de la solidesa científica de l'enfocament metodològic, sinó també per garantir la coherència i la plena interoperabilitat entre els diferents components de l'estratègia general de gestió de riscos dels proveïdors i responsables de desplegament d'IA, en totes les seves activitats.

De la mateixa manera, el model AIDF ha de ser coherent amb la teoria i la pràctica jurídiques dels drets fonamentals. Referent a això, per exemple, no és possible definir l'impacte global d'un sistema d'IA en els drets fonamentals, com en el cas de la ciberseguretat, perquè els drets fonamentals s'han de considerar individualment i no es poden avaluar de forma acumulativa, ni es poden compensar impactes diferents (per exemple, un sistema d'IA amb un impacte baix en la protecció de dades i un alt impacte en la llibertat d'expressió no té un impacte mitjà global com a conseqüència d'una compensació entre aquests diferents nivells d'impacte).

D'acord amb aquests supòsits, els tres blocs principals del model AIDF són:

- (i) Una **fase de planificació i determinació de l'abast**, centrada en les principals característiques del producte/servei i en el context en què se situarà.
- (ii) Una **fase de recopilació de dades i anàlisi de riscos**, en la qual s'identificaran els riscos potencials i se n'avaluarà el seu possible impacte en els drets fonamentals.
- (iii) Una **fase de gestió de riscos**, en la qual s'adopten, proven i supervisen mesures adequades per prevenir o mitigar aquests riscos i comprovar-ne l'efectivitat.

En termes d'estructura, de conformitat amb l'article 27 del RIA i les metodologies d'avaluació de riscos, l'AIDF és una **avaluació contextual** centrada en la solució d'IA específica que s'està implementant, i no una avaluació tecnològica centrada en les tecnologies d'IA en general i els seus diversos usos potencials: examina una aplicació d'IA específica i el seu context d'ús. A més, l'AIDF també es caracteritza per un **enfocament ex ante**, la qual cosa la converteix en una eina per a un disseny d'IA orientat als drets fonamentals, que adopta l'enfocament des del disseny ja conegut en la protecció de dades.

Finalment, l'AIDF té una **estructura iterativa circular** com totes les avaluacions de riscos de situacions que poden evolucionar amb el temps, no és una avaluació prèvia puntual. Les principals fases de la gestió de riscos (planificació/abast, anàlisi de riscos, prevenció/mitigació de riscos), es repeteixen en aquesta estructura iterativa circular ja que els canvis tecnològics, socials i contextuals afecten alguns dels elements rellevants d'una avaluació prèvia (art. 27.2 RIA).

4.1 La fase de planificació i determinació de l'abast

La fase de planificació i determinació de l'abast parteix de l'anàlisi de les necessitats i la descripció de les solucions d'IA que es desenvoluparan, i continua amb la consideració de l'escenari contextual dels drets fonamentals (inclosos els controls ja existents) i les àrees potencialment afectades. **En aquesta fase s'examinen dues àrees principals: la dimensió inherent al sistema d'IA i la dimensió contextual** (vegeu la secció 2 supra). Pel que fa a la **dimensió contextual de les solucions d'IA**, no es limita a identificar els drets potencialment afectats i els titulars de drets (sense quantificar l'impacte, que és l'objectiu de la fase següent), sinó que també inclou una anàlisi preliminar dels elements rellevants de la protecció jurídica existent d'aquests drets.

⁶ Vegeu, per exemple, ISO, Gestió de riscos. Directrius. ISO 31000. <https://www.iso.org/standard/65694.html>, que identifica les següents tres fases principals, combinades amb tres tasques complementàries (registre i presentació d'informes; seguiment i revisió; comunicació i consulta): (i) abast, context i criteris; (ii) avaluació de riscos (identificació de riscos, anàlisi de riscos, avaluació de riscos); (iii) tractament de riscos. El mateix enfocament també es pot veure al PNUD 2024.

Per les raons exposades anteriorment en relació amb la varietat d'usos potencials de la IA, els contextos d'ús i les persones i grups potencialment afectats, no és possible proporcionar un qüestionari exhaustiu que puguin utilitzar els operadors d'IA per abordar totes les qüestions pertinents d'aquesta primera fase de l'AIDF.

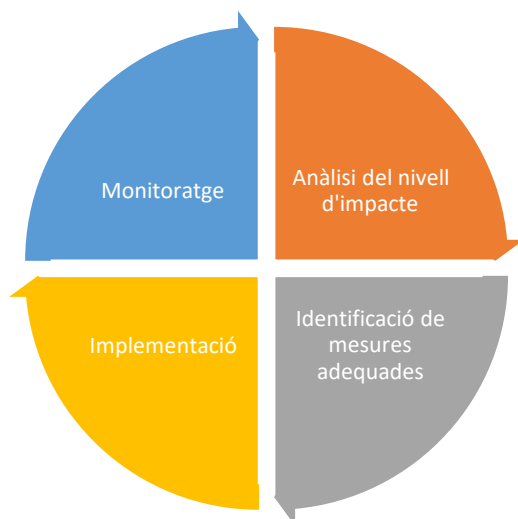
Sobre la base de l'experiència de l'AIPD, es pot facilitar **una llista no exhaustiva de possibles preguntes per a la planificació i l'abast de l'AIDF**, que els operadors d'IA podent complementar amb preguntes específiques basades en la naturalesa i l'ús de les solucions d'IA que s'estan considerant (vegeu la **Planificació i abast** a la secció 5 – La plantilla del model AIDF). Tanmateix, com s'ha demostrat en els casos d'ús realitzats a Catalunya (vegeu la secció 6), aquest qüestionari és eficaç per guiar adequadament els operadors d'IA en la fase de planificació i determinació de l'abast i cobreix totes les àrees rellevants amb preguntes que, després, es poden aprofundir amb preguntes addicionals centrades en el cas individual.

4.2 Fase de recopilació de dades i anàlisi de riscos

Sobre la base de la informació recopilada en la fase de planificació, determinació de l'abast i d'identificació de riscos, és possible determinar si cal realitzar una anàlisi a fons del nivell d'impacte sobre els drets individuals. En el cas que hi hagi drets potencialment afectats, aquesta anàlisi quantifica aquest impacte per prevenir-lo o reduir-lo amb mesures adequades.

Per tant, l'**anàlisi del nivell d'impacte** de la solució d'IA en els drets potencialment afectats és la primera fase de l'enfocament circular de l'AIDF, que també inclou els tres passos següents com a part de la fase de gestió de riscos (vegeu la secció 4.3): (i) la identificació de mesures adequades per prevenir o mitigar el risc, (ii) la implementació d'aquestes mesures i (iii) el seguiment del funcionament del sistema d'IA, per tal de revisar l'avaluació i les mesures adoptades.

Avaluació d'impacte (enfocament circular) [Gràfic 1]



Pel que fa a la fase de recopilació de dades i anàlisi de riscos, atesa la naturalesa de l'avaluació les dades es referiran als diferents aspectes dels drets potencialment en joc, inclosa també informació sobre el context d'ús i les persones i grups potencialment afectats. Malgrat la varietat d'aquests elements i de l'especificitat de cada dret i llibertat, la fase d'anàlisi es pot basar en paràmetres comuns clau.

Aquests paràmetres permeten operacionalitzar un concepte abstracte com l'impacte en els drets, de manera que també pugui ser avaluat més fàcilment, i (i) comparar el nivell d'impacte sobre els diferents drets per prioritzar la prevenció/mitigació de riscos i (ii) comprendre com pot canviar l'impacte sobre un dret individual, si alguns dels elements del sistema o del context varien.

4.2.1 Variables clau per a l'avaluació d'impacte

D'acord amb la teoria del risc i el marc jurídic dels drets fonamentals, l'impacte sobre els drets consta de dues dimensions clau: la **probabilitat** d'un impacte advers i la seva **gravetat**.⁷ La combinació de variables relacionades amb aquestes dues dimensions proporciona un **índex de risc**, que s'avalua per a cadascun dels drets potencialment afectats.

Per construir aquests índexs, és possible representar les variables rellevants en una escala d'un mínim a un màxim (suposant, d'acord amb la teoria general del risc, que no hi ha risc zero) i utilitzar variables ordinals (per exemple, baix, mitjà, alt i molt alt). L'ús de l'escalament i les variables associades permet comparar diferents situacions, utilitzant les mateixes variables: per exemple, els diferents nivells d'impacte sobre la no discriminació que produeix un sistema de qualificació creditícia, quan s'introdueixen canvis en el seu algoritme. Per tant, aquestes variables ordinals es poden utilitzar per "mesurar" l'impacte en una quantificació del risc basada en rangs (baix, mitjà, alt i molt alt).

Tanmateix, la teoria dels drets fonamentals no permet crear un índex compost, com és comú en l'avaluació de riscos, on tots els impactes potencials es combinen per crear un índex d'impacte general. Aquest plantejament entra en conflicte amb l'enfocament jurídic dels drets fonamentals, en el qual cada dret s'ha de considerar de manera independent, en termes del seu possible perjudici, i el fet que un dret es vegi menys afectat que un altre no pot donar lloc a cap forma de compensació.

Com mostren els casos d'ús analitzats a la secció 5, es pot avaluar l'impacte en els diferents drets involucrats, però això no vol dir que un determinat sistema d'IA tingui un impacte general en els drets que es consideri baix, mitjà o alt. L'única interacció possible entre diferents interessos és a través del **test** de ponderació entre els drets en conflicte, però aquesta prova avalua el nivell d'impacte en cada dret. El test de ponderació no es refereix al nivell de risc dels drets afectats, sinó a la prioritització d'un interès sobre un altre. Per tant, cal considerar-lo com un factor extern, que només s'ha de tenir en compte un cop s'hagi avaluat l'impacte sobre els drets individuals i que pot influir en els resultats de l'avaluació d'impacte, en fer que l'impacte sobre determinats drets sigui acceptable a causa d'un interès contraposat prevalent.⁸

Sobre la base d'aquestes consideracions, un model AIDF definirà un índex de risc per a cada dret potencialment afectat utilitzant dues dimensions: probabilitat i gravetat. La **probabilitat** s'entén com una combinació de (i) la probabilitat de resultats adversos i (ii) l'exposició. La primera variable es refereix a la probabilitat que es produeixin les conseqüències adverses d'un determinat risc, mentre que la segona variable es refereix al grau en què les persones potencialment en risc podrien veure's afectades. Quant a l'exposició, cal assenyalar que l'atenció se centra en les persones potencialment exposades a l'ús del sistema d'IA (la població identificada) i no en la població en el seu conjunt.

⁷ Vegeu l'article 3, apartat 2, del Reglament d'intel·ligència artificial, que estableix que per 'risc' s'entén "la combinació de la probabilitat que es produeixi un perjudici i la gravetat d'aquest perjudici", i l'article 25, apartat 1, de l'RGPD, que es refereix als "riscos de diversa probabilitat i gravetat que comporta el tractament per als drets i llibertats de les persones físiques [...]."

⁸ Vegeu, per exemple, la part II, cas d'ús 1, en el cas del desenvolupament i ús d'una plataforma d'anàlisi d'aprenentatge avançat, en la qual es considera acceptable algun impacte en els drets de privacitat i protecció de dades, en vista dels beneficis per al dret a l'educació.

La **gravetat** de les conseqüències previstes es basa en dues variables: i) la gravetat del perjudici en l'exercici dels drets i llibertats (gravetat),⁹ sobre la base dels seus atributs, inclosa la consideració de les repercussions, la vulnerabilitat i les situacions de dependència específiques de cada grup; i (ii) l'esforç per superar-lo i revertir els efectes adversos (esforç).

Tant la probabilitat com la gravetat s'han d'avaluar sobre una base contextual, i la participació de les parts interessades pertinents pot ser d'ajut. Com és habitual en l'avaluació de riscos, l'estimació de la probabilitat es basa tant en casos anteriors, considerant situacions comparables, com en l'ús de tècniques analítiques i de simulació basades en possibles escenaris d'ús. Els mateixos enfocaments s'utilitzen també per estimar el nivell de gravetat, però en aquest cas amb més èmfasi en l'anàlisi jurídica sobre la gravetat del perjudici, que s'ha d'avaluar amb referència a la jurisprudència sobre drets fonamentals i al marc jurídic pertinent.

Sobre la base dels valors de probabilitat i gravetat derivats de les variables anteriors es determina un índex de risc, que indica l'impacte global per a cadascun dels drets i llibertats considerats.¹⁰ Val la pena assenyalar que aquests resultats s'han de combinar amb qualsevol element que justifiqui una limitació d'alguns drets des d'una perspectiva legal, com la naturalesa obligatòria de certes característiques que afecten els drets: en aquest cas, el risc potencial s'ha de considerar acceptable en la mesura que el sistema d'IA compleixi amb els requisits legals donats.

4.2.2 Variables i construcció de l'índex d'impacte

En molts models i estàndards d'avaluació d'impacte basats en el risc, els índexs de risc es construeixen utilitzant matrius perquè són relativament fàcils d'usar i explicar¹¹. Atès que una matriu de riscos és un gràfic que combina dues dimensions utilitzant colors per reflectir diferents nivells de risc, són útils per avaluar índexs generats per diferents variables. Per aquesta raó, es poden utilitzar en l'AIDF per definir el nivell d'impacte sobre cada dret en qüestió.

La metodologia aquí proposada emprà un índex de risc per a cada dret potencialment afectat, basat en una matriu que combina dues dimensions (probabilitat i gravetat). Cadascuna d'aquestes dimensions resulta de la combinació de dos parells de variables, també construïdes mitjançant matrius: la probabilitat de conseqüències adverses i l'exposició, per a la probabilitat; la gravetat dels perjudicis i l'esforç per superar-los i revertir els efectes adversos, per a la gravetat.

No hi ha un únic model de matriu que es pugui utilitzar en l'avaluació de riscos; la pràctica en aquest camp mostra una varietat de models. Les més comunes són les matrius 3x3, 4x4, 5x5, 5x4 i 6x4, on els parells de números indiquen el nombre de rangs de les dues escales que defineixen la dimensió considerada. Com que la matriu es refereix a dues variables independents, es poden avaluar d'acord amb escales que poden diferir en el nombre de rangs, per exemple, una escala de 6x4 on es proporcionen sis rangs diferents per a una variable i només quatre per a l'altra.

La **matriu 4x4** pot ser la més apropiada en el context de l'AIDF, ja que redueix el risc de posicionament mitjà, presta més atenció als nivells alts i molt alts d'una manera coherent, amb l'enfocament en l'alt

⁹ La **gravetat del perjudici a un dret fonamental/humà sol avaluar-se d'acord amb els tres elements següents: (i) la seva intensitat, (ii) les conseqüències de la violació, i (iii) la seva durada**, on la intensitat de la violació està relacionada amb la importància de l'interès jurídic protegit violat. Vegeu també Tribunal de Drets Humans de la Unió Europea (2022). Dret al respecte de la vida privada i familiar, del domicili i de la correspondència: Guia pràctica de la jurisprudència del Tribunal Europeu de Drets Humans en virtut de l'article 8. https://www.echr.coe.int/documents/d/echr/guide_art_8_eng.

¹⁰ Consulteu la secció següent per conèixer la metodologia utilitzada per combinar les diferents variables i crear els índexs.

¹¹ Vegeu també APDCAT 2024, 33 i 53; CNIL 2018, 23.

risc del marc regulador actual de la IA, i no fragmenta excessivament la part inferior de l'escala, la qual cosa és menys rellevant a causa de l'enfocament abans esmentat.

En les matrius, s'utilitzen etiquetes descriptives per a les diferents combinacions de nivells en l'escala de colors, com es mostra en aquest exemple d'una matriu de gravetat:

		Gravetat			
		Baixa (L)	Mitjana (M)	Alta (A)	Molt alta (MA)
Esforç	Baix (B)	B/B	B/M	B/A	B/MA
	Mitjà (M)	M/B	M/M	M/A	M/ MA
	Alt (A)	A/B	A/M	A/A	A/ MA
	Molt alt (MA)	MA/B	MA/M	MA/A	MA/ MA

Nivells de gravetat			
Baix	Mitjà	Alt	Molt alt

4.3 La fase de gestió de riscos

Després de l'anàlisi de riscos, que ha definit el nivell d'impacte de la solució d'IA sobre els drets potencialment afectats, cal gestionar els riscos identificats adoptant les mesures adequades.¹² Per tant, la tercera fase de l'AIDF s'articula en tres etapes, de la manera següent:

- (i) La **identificació de mesures adequades** per prevenir o mitigar el risc, tenint en compte el seu impacte en el nivell de risc d'acord amb una anàlisi d'escenaris específics del context.
- (ii) La **implementació** d'aquestes mesures.
- (iii) El **seguiment** del funcionament del sistema d'IA per tal de revisar l'avaluació i les mesures adoptades en cas que els canvis tecnològics, socials i contextuals afectin el nivell de risc o l'eficàcia de les mesures preses.

Atès que l'AIDF no és una comprovació final d'una solució d'IA, sinó **una eina de disseny per guiar el desenvolupament i el desplegament de la IA** cap a un enfocament orientat als drets fonamentals, el seguiment del funcionament del sistema d'IA pot també formar part de la fase prèvia a la comercialització en la qual es proven i implementen diferents solucions de disseny per seleccionar la més adequada. Per tant, en consonància amb l'**enfocament circular de l'avaluació de riscos** i el disseny de la IA, és possible que es duguin a terme diverses sèries d'avaluacions de risc, d'aplicació de mesures de mitigació i reavaluacions, fins que la versió final del producte/servei d'IA comporti un nivell de **risc residual** satisfactori, en termes d'acceptabilitat, i pugui introduir-se en el mercat o posar-se en servei.

A més, **es poden produir canvis en l'escenari tecnològic i social o en el context específic d'ús, després** que l'eina d'IA s'hagi introduït en el mercat o s'hagi posat en servei. Això pot repercutir en el

¹² Vegeu també l'article 27, apartat 1, lletra f), del RIA.

nivell de risc avaluat prèviament respecte dels drets de què es tracti, així com plantejar noves preocupacions respecte d'altres drets. En aquests casos, caldrà reavaluar les solucions d'IA adoptades i prendre les mesures adequades.¹³

D'acord amb aquestes observacions, la plantilla del model AIDF (vegeu la secció 5) inclou una matriu que mostra l'impacte de les mesures de prevenció/mitigació de riscos adoptades, sobre els nivells de risc identificats en la fase d'anàlisi de riscos i el risc residual resultant.

5. La plantilla del model AIDF

El model utilitzat en els casos d'ús es basa en les tres fases de l'AIDF discutides en la secció anterior. Consta de diversos elements.

El primer és un qüestionari, que abasta les quatre àrees principals de la fase de planificació i determinació de l'abast; en concret, la descripció i l'anàlisi del sistema d'IA, el context jurídic, els controls ja establerts i la participació de les parts interessades (**qüestionari de planificació i abast**).

El segon element de la plantilla és un conjunt de matrius (**taules 3, 6 i 7**) i criteris de quantificació de variables associades que s'utilitzen per avaluar la probabilitat i la gravetat dels possibles perjudicis a cada dret i llibertat (**taules 1, 2, 4 i 5, i la taula 1A**) i per estimar l'impacte global associat.

El tercer element de la plantilla consisteix en dues taules: una que indica el nivell d'impacte sobre cada dret i llibertat i les mesures de prevenció/mitigació identificades per abordar el risc (**taula 2A**), i una altra que estima el risc residual resultant de l'adopció d'aquestes mesures (**taula 3A**).

Qüestionari de planificació i abast

<p>Secció A</p> <p>Descripció i anàlisi del sistema d'IA, inclosos els fluxos de dades relacionats</p>	<p>Quins són els principals objectius del sistema d'IA?</p> <p>Quines són les principals característiques del sistema?</p> <p>En quins països s'oferirà?</p> <p>Quins tipus de dades es tracten (personals, no personals, categories especials)?</p> <p>Identificació dels possibles titulars de drets: qui són els individus o grups que es poden veure afectats pels sistemes d'IA? S'hi inclouen individus o grups vulnerables?</p> <p>Identificació dels responsables: quines persones/entitats estan implicades en el disseny, desenvolupament i desplegament dels sistemes d'IA? Quin és el seu paper?</p>
<p>Secció B</p> <p>Context de drets fonamentals</p>	<p>Quins drets fonamentals es veuen potencialment afectats per l'ús del sistema d'IA?</p>

¹³ Vegeu també l'article 27, apartat 2, del RIA.

	<p>Quins instruments jurídics internacionals/regionals de protecció dels drets humans/fonamentals s'han aplicat a nivell operatiu?</p> <p>Quins són els tribunals o organismes més rellevants en matèria de drets fonamentals en el context d'ús?</p> <p>Quines són les decisions i disposicions més rellevants en matèria de drets humans/fonamentals?</p>
<p>Secció C</p> <p>Controls implementats</p>	<p>Quines polítiques i procediments s'han implementat per avaluar el possible impacte sobre els drets fonamentals, inclosa la participació de les parts interessades?</p> <p>S'ha efectuat, desenvolupat i aplicat una avaluació d'impacte en relació amb qüestions específiques (per exemple, protecció de dades) o algunes característiques del sistema (per exemple, l'ús de la biometria)?</p>
<p>Secció D</p> <p>Intervenció de les parts interessades i diligència deguda</p>	<p>Quins són els principals grups o comunitats potencialment afectats pel sistema d'IA, inclòs el seu desenvolupament?</p> <p>Quines parts interessades, a més dels individus o grups que poden veure's afectats pels sistemes d'IA, hi han de participar (per exemple, la societat civil i les organitzacions internacionals, experts, associacions industrials, periodistes)?</p> <p>Hi ha altres titulars d'obligacions que hi hagin de participar, a part del proveïdor i el responsable del desplegament de la IA (per exemple, autoritats nacionals, organismes governamentals)?</p> <p>Han participat en el procés d'avaluació els socis comercials, inclosos els proveïdors de serveis (per exemple, subcontractistes en sistemes d'IA i conjunts de dades)?</p> <p>El proveïdor d'IA ha dut a terme una avaluació de la seva cadena de subministrament, per identificar si les activitats dels proveïdors/contractistes que participen en el desenvolupament de productes/serveis poden afectar els drets fonamentals?</p> <p>El proveïdor ha promogut estàndards o auditories de drets fonamentals, per garantir el respecte dels drets fonamentals entre els proveïdors?</p> <p>El proveïdor i el responsable del desplegament de la IA han comunicat públicament les possibles repercussions del sistema d'IA en els drets fonamentals?</p> <p>El proveïdor i el responsable del desplegament de la IA han proporcionat formació sobre estàndards de drets fonamentals al personal encarregat de la gestió i la contractació relacionades amb el sistema d'IA?</p>

Matrius de risc

Taula 1. Probabilitat

Baixa	El risc de perjudici és improbable o altament improbable.
Mitjana	El risc es pot produir.
Alta	Hi ha una alta probabilitat que es produeixi el risc.
Molt alta	És molt probable que es produeixi el risc.

Taula 2. Exposició

Baixa	Pocs o molt pocs individus de la població identificada de titulars de drets estan potencialment afectats.
Mitjana	Part de la població identificada està potencialment afectada.
Alta	La majoria de la població identificada està potencialment afectada.
Molt alta	Gairebé tota la població identificada està potencialment afectada.

Taula 3. Probabilitat que es produeixi un perjudici (*likelihood*)

		Probabilitat			
		Baixa	Mitjana	Alta	Molt alta
Exposició	Baixa	B	B/M	B/A	B/MA
	Mitjana	M/B	M	M/A	M/MA
	Alta	A/B	A/M	A	A/MA
	Molt Alta	MA/B	MA/M	MA/A	MA

Probabilitat			
Baixa	Mitjana	Alta	Molt alta

Taula 4. Gravetat del perjudici

Baixa	Els individus i grups afectats poden trobar només perjudicis menors en l'exercici dels seus drets i llibertats.
Mitjana	Els individus i grups afectats poden trobar-se amb importants perjudicis.
Alta	Els individus i grups afectats poden enfrontar-se a greus perjudicis.
Molt alta	Els individus i grups afectats poden trobar-se amb perjudici greus o fins i tot irreversibles.

Taula 5. Esforç per superar els perjudicis i revertir els efectes adversos

Baix	Els perjudicis soferts es poden superar sense problemes (per exemple, temps dedicat a modificar la informació, molèsties, irritacions, etc.).
Mitjà	Els perjudicis soferts es poden superar malgrat algunes dificultats (per exemple, costos addicionals, por, incomprensió, estrès, petites dolències físiques, etc.).
Alt	Els perjudicis soferts es poden superar encara que amb serioses dificultats (per exemple, pèrdues econòmiques, danys materials, empitjorament de la salut, etc.).
Molt alt	Els perjudicis soferts poden no superar-se (per exemple, dolències psicològiques o físiques de llarga durada, mort, etc.).

Taula 6. Gravetat (severity)

		Gravetat			
		Baix	Mitjana	Alta	Molt alta
Esforç	Baix	B	B/M	B/A	B/MA
	Mitjà	M/B	M	M/A	M/MA
	Alt	A/B	A/M	A	A/MA
	Molt Alt	MA/B	MA/M	MA/A	MA

Gravetat			
Baix	Mitjà	Alt	Molt alt

Taula 1A. Recollida de dades i anàlisi de riscos

Drets/ Llibertats potencialment afectats	Descripció de l'impacte	Probabilitat			Gravetat		
		Probabilitat de resultats adversos	Exposició	Probabilitat	Gravetat de l'impacte	Esforç	Gravetat

Taula 7. Impacte general del risc

		Gravetat			
		Baixa	Mitjana	Alta	Molt alta
Probabilitat	Baixa				
	Mitjana				
	Alta				
	Molt Alta				

Risc global			
Baixa	Mitjana	Alta	Molt alta

Taula 2A. Gestió de riscos (I)

Drets/llibertats afectats	Probabilitat	Gravetat	Impacte global	Mesures per prevenir/mitigar els impactes

Taula 3A. Gestió de riscos (II)

Drets/llibertats afectats	Probabilitat (residual)	Gravetat (residual)	Impacte residual

6. Del model a la pràctica: els casos d'ús

El debat metodològic en curs sobre l'AIDF i la seva implementació s'ha caracteritzat per ser principalment teòric i centrat en les polítiques, amb poca atenció a l'anàlisi empírica i a la plena implementació dels diferents models proposats en el món real. En aquest context, el projecte liderat per l'Autoritat Catalana de Protecció de Dades marca la diferència, en apropar el debat de l'AIDF a l'experiència concreta dels proveïdors i responsables de desplegament directament implicats en l'ús de la IA.

Aquest enfocament empíric, basat en casos, és crucial per provar l'eficàcia del model proposat en la consecució dels objectius polítics i de disseny de l'AIDF elaborats pel legislador de la UE al RIA. Més concretament, els casos d'ús d'aquest projecte han demostrat que és possible **racionalitzar el procediment de l'AIDF, evitant adoptar d'una llarga llista de control i centrant-se en els elements essencials de l'impacte en els drets fonamentals.**

Els casos d'ús també han demostrat que, almenys per a la primera ronda d'avaluació, mitigació i reavaluació, les persones amb la formació adequada poden completar l'AIDF en dues o tres reunions breus (3 hores per reunió). Això confirma que, **si es formula adequadament, l'AIDF no imposa una càrrega addicional excessiva a les entitats privades i públiques de la UE per complir amb el RIA.**

Finalment, cada cas d'ús es va basar en quatre interaccions diferents: (i) una anàlisi interna inicial del cas per part dels experts (normalment DPD) de les entitats que duen a terme l'AIDF, amb l'objectiu d'esbossar els elements clau de l'AIDF, amb especial atenció a la fase de planificació i abast; (ii) una discussió amb un expert en l'AIDF, per revisar l'avaluació inicial; (iii) un debat en grup, en el qual participen els experts de totes les entitats participants en el projecte; i (iv) una revisió final per part dels experts de l'entitat que realitza l'AIDF. Aquest procés de quatre passos il·lustra dos aspectes clau de l'AIDF: **la importància d'una avaluació d'experts i la importància d'una avaluació basada en equips**, que involucri diferents coneixements per millorar el nivell d'anàlisi.

Pel que fa a **la selecció dels casos**, cal destacar que es tracta d'un projecte en curs i els casos presentats en aquest informe són els primers que s'han comentat i en els quals s'ha aplicat el model AIDF. Altres casos estan en avaluació i es publicaran en el futur al [web "DPD en xarxa"](#) i en el de [l'Autoritat Catalana de Protecció de Dades](#). D'altra banda, alguns casos en què s'ha aplicat el model AIDF amb un impacte rellevant en el disseny de solucions d'IA no s'han inclòs en aquest informe, per raons de confidencialitat, però han estat útils perquè tots els participants elaborin millor la pràctica de la plantilla de l'AIDF.

Pel que fa als **àmbits coberts**, els casos d'ús es refereixen a quatre dels àmbits clau de l'annex III del RIA, concretament: l'educació (avaluació dels resultats de l'aprenentatge i predicció de l'abandonament escolar); la gestió dels treballadors (sistemes de suport a la presa de decisions en la gestió de recursos humans); l'accés a l'assistència sanitària (tractament del càncer basat en imatges mèdiques); i els serveis socials (assistent de veu per a gent gran). Aquests àmbits

representen les àrees en les quals les solucions d'IA s'utilitzen cada vegada més, amb més impacte en individus i grups. En aquest sentit, la naturalesa dels casos d'ús discutits també els farà útils per a moltes altres entitats públiques i privades d'altres països, interessades a dissenyar sistemes/models d'IA que compleixin amb els drets fonamentals en aquestes àrees fonamentals.

D'acord amb l'objectiu d'aquest projecte, els casos d'ús es presenten tal com els van desenvolupar els participants, més que com a millors pràctiques o casos estandarditzats. El projecte es va dissenyar per provar l'eficàcia de la plantilla del model i la metodologia associada. Seguint la línia de l'experiència amb les AIPD, hem donat llibertat als participants per desenvolupar les diferents parts de la plantilla segons el seu enfocament. Així, hi trobem anàlisis més extenses que d'altres. Tanmateix, els elements bàsics (les preguntes, les matrius i la metodologia d'avaluació) continuen essent els mateixos.

La idea principal que es vol posar en relleu en aquest informe és la importància de reflexionar sobre l'exercici d'anàlisi realitzat amb l'objectiu de mostrar els resultats obtinguts, i no presentar les AIDF dutes a terme com a casos ficticis i perfectes. L'AIDF ha estat i serà implementada per diversos actors, en alguns casos amb més detall, en d'altres amb algunes limitacions; però, en la mesura que contribueixi a una anàlisi eficaç i a la prevenció i mitigació de l'impacte en els drets fonamentals, haurà assolit el seu objectiu principal.

Referències

Tots els llocs web citats en aquest document van ser consultats entre setembre i desembre de 2024.

APDCAT. 2024. Avaluació d'impacte relativa a la protecció de dades, https://apdcat.gencat.cat/web/.content/03-documentacio/Reglament_general_de_proteccio_de_dades/documents/Guia-Practica-avaluacio-impacte-proteccio-de-dades-2019.pdf.

APDCAT. 2020. Intel·ligència Artificial. Decisions Automatitzades a Catalunya, https://apdcat.gencat.cat/web/.content/03-documentacio/intelligencia_artificial/documents/INFORME-INTELLIGENCIA-ARTIFICIAL-FINAL-WEB-OK.pdf.

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 d'ús

Cas d'ús 1: Plataforma d'analítica avançada de l'aprenentatge

1. El context

El cas d'ús que exposarem a continuació està relacionat amb un dels grans problemes del sistema educatiu superior, que és l'abandonament prematur de l'educació i la formació de la població entre 18 i 24 anys. Aquesta situació ha generat una preocupació en l'àmbit europeu i les dades estadístiques europees, que es poden consultar a la base de dades d'Eurostat,¹⁴ ho posen de manifest. Per això, la prioritat núm. 1 del marc estratègic per a la cooperació europea en l'àmbit de l'educació i de la formació, amb vista a l'Espai Europeu d'Educació (EEE) i més enllà (2021-2030)¹⁵, és “augmentar la qualitat, l'equitat, la inclusió i l'èxit de tots en l'àmbit de l'educació i la formació.” Tot i que l'abandonament prematur s'ha reduït en la darrera dècada, continua constituint un repte. Per evitar que els joves pateixin limitacions d'accés a oportunitats socioeconòmiques en un futur, es pretén prestar atenció especial als grups que presenten un risc de baix rendiment educatiu i un abandonament prematur en els seus estudis.

Les institucions d'educació superior han de promoure estratègies educatives que fomentin que els itineraris d'educació i formació finalitzin

amb èxit, redueixin els índexs d'abandonament prematur i afrontin les causes del baix rendiment. Per això, és important discernir entre les dades de què es disposa, estructurar-les i extreure la informació que ens proporcionen per utilitzar-la per a la finalitat concreta que volem aconseguir.

Les dades que ens poden interessar per identificar la situació en què es troba cada estudiant en cada moment mentre cursa els seus estudis, i poder predir el que succeirà, són:

- Les dades històriques dels estudis previs dels alumnes de recollides durant anys anteriors.
- Les dades proporcionades pel centre d'educació anterior.
- Les dades proporcionades pel mateix estudiant en el moment de la matrícula.
- Les dades obtingudes durant l'evolució dels estudis.

A partir de la informació que proporcionen les dades, es poden establir patrons d'actuació dels estudiants. També permet disposar d'una visió global de la situació del conjunt de l'estudiantat que cursa uns mateixos estudis i de la seva futura evolució i tenir, alhora, una visió individualitzada sobre la situació particular de cada estudiant i anticipar-se al seu futur. Aquesta informació ajuda a promoure estratègies d'acord amb la realitat, oferint un tractament personalitzat i atenent les necessitats de cada estudiant.

¹⁴Accés a Eurostat: <https://ec.europa.eu/eurostat>.

¹⁵Resolució del Consell relativa a un marc estratègic per a la cooperació europea en l'àmbit de l'educació i la formació amb vista a l'Espai Europeu d'Educació i més enllà (2021-2030): <https://eur-lex.europa.eu/legal-content/CA/TXT/?uri=celex%3A32021G0226%2801%29>.

L'anàlisi d'aquestes dades i la seva interpretació per dur a terme actuacions de millora i progrés en l'àmbit de l'educació entren dins del que es coneix com a "analítica de l'aprenentatge" (en anglès, *learning analytics*). A la primera Conferència Internacional sobre Anàlisi i Coneixement d'Aprenentatge (en anglès, *Learning Analytics and Knowledge, LAK*), celebrada el 2011, es va definir l'analítica de l'aprenentatge¹⁶ com el mesurament, la recopilació, l'anàlisi i la comunicació de dades sobre els estudiants i els seus contextos, per comprendre i optimitzar l'aprenentatge i els entorns en què es produeix.

Tot i que fa molt de temps que s'investiga sobre l'analítica de l'aprenentatge, i que s'ha utilitzat tant per predir l'èxit acadèmic dels estudiants com el risc d'abandonament dels seus estudis, l'aparició de noves tecnologies que proporcionen tècniques d'anàlisi alternatives ha posat de manifest la necessitat d'abordar els requeriments legals que han sorgit de les regulacions recents, a fi de poder fer-ne un ús correcte.

Atesa la situació actual, en què els sistemes d'intel·ligència artificial (d'ara endavant, sistemes d'IA) estan apareixent habitualment en les nostres tasques quotidianes, és fonamental que les institucions en detectin i coneguin els riscos; així, es podran prevenir, minimitzar i abordar, per fer-ne un ús òptim com a aliats en la modernització i digitalització del sistema universitari.

Una de les dificultats detectades, a banda de l'habilitat en la interpretació correcta de la informació resultant, és la manca de fluïdesa d'adaptació a les noves tecnologies per part de les institucions i del personal que hi treballa, en els plantejaments i les metodologies d'ensenyament. Amb els recursos de què es disposen en l'àmbit de l'educació, és evident que les

institucions d'educació superior han d'anar introduint sistemes d'IA en els processos en què la intervenció humana es pot trobar limitada. Cal ser conscients dels canvis que s'estan generant, com ho ha fet la irrupció de la intel·ligència artificial generativa¹⁷ dins del model d'aprenentatge de l'estudiantat, i els que es poden arribar a generar. D'aquesta manera, si es pren la iniciativa d'incorporar els sistemes d'IA que tenim al nostre abast, es poden establir prèviament eines adequades per protegir els drets i les llibertats des d'una fase inicial, i evitar que la seva integració generi un perjudici al sistema universitari i, consegüentment, a l'estudiantat. Tot i això, encara hi ha desafiaments pendents d'abordar, com el possible biaix en les dades o els dilemes ètics que es plantegin en cada cas, així com qüestions inherents al desenvolupament dels sistemes d'IA i l'impediment i la resistència a la incorporació d'aquests sistemes en alguns àmbits, que poden portar a entorpir o alentir-ne el creixement. En el nostre cas, deixar el sistema universitari en l'obsolescència i aconseguir un efecte contrari al desitjat podria generar una reculada en l'ensenyament i en la formació dels joves.

Des de les fases embrionàries en què es planteja l'ús de sistemes d'IA, cal estudiar i reflexionar sobre els diferents plantejaments i casuístiques amb què ens podem trobar, per aplicar totes les cauteles necessàries i instaurar les mesures tècniques i organitzatives per construir models sòlids, que garanteixin la protecció dels joves i respectin els drets fonamentals i els valors socials.

El ràpid augment dels usos dels sistemes d'IA en l'educació està transformant la manera d'ensenyar i d'aprendre, cosa que té repercussions directes en les institucions, en el seu personal i en l'estudiantat mateix.

¹⁶ <https://www.solaresearch.org/about/what-is-learning-analytics/>.

¹⁷ Podeu trobar la definició d'IA generativa i com funciona a la *Guia per a l'ús d'IA generativa en educació i investigació* de la UNESCO, 2024: <https://unesdoc.unesco.org/ark:/48223/pf0000389227>.

Aquesta transformació va aparellada amb la necessitat de capacitar el personal amb noves competències, a fi que es pugui desenvolupar en aquest panorama tecnològic en constant evolució i, en conseqüència, estigui facultat¹⁸ per utilitzar la informació que li proporciona el sistema d'IA.

L'ús de la intel·ligència artificial en l'educació està plantejant qüestions fonamentals. Per això, el mateix Reglament europeu d'intel·ligència artificial¹⁹ (RIA) ha determinat com a sistemes d'IA d'alt risc “els sistemes d'intel·ligència artificial destinats a ser utilitzats per avaluar els resultats de l'aprenentatge, també quan aquests resultats s'utilitzen per orientar el procés d'aprenentatge de les persones físiques en centres educatius” (apartat 3.b de l'annex III).

Si bé hi ha sectors reticents a emprar la intel·ligència artificial en els seus processos, pel que pot comportar, és inqüestionable que fer-ne un bon ús en l'educació pot ser enriquidor per a l'aproximació dels estudis a les noves generacions d'estudiants.

En lloc d'integrar a la institució educativa una aplicació pròpia, al mercat ja hi ha plataformes d'anàlisi d'aprenentatge que recopilen les dades que el mateix estudiant ha proporcionat al sistema de la institució (en anglès, *Student Information System, SIS*), o les dades generades als sistemes de gestió de l'aprenentatge (en anglès, *Learning Management System, LMS*), per identificar i predir els estudiants en risc d'abandonament. Per exemple:

- *Assessment and Learning in Knowledge Spaces*²⁰ (ALEKS)
- DreamBox²¹
- Carnegie Learning²²
- Smart Sparrow²³
- The IntelliBoard²⁴

Pel que fa als esforços per reduir la taxa d'abandonament prematur de l'educació i la formació, preveure els diferents escenaris possibles facilita que es puguin establir polítiques i mesures més efectives, que es poden centrar en l'arrel del problema. Posar a disposició del personal de les institucions mecanismes apropiats i personalitzats a les necessitats de cada estudiant pot empoderar els joves a continuar amb el seu creixement personal en l'àmbit educatiu, que després es traslladarà al seu àmbit professional. Per això, tenint en compte el potencial dels sistemes d'IA per transformar la noció actual de l'educació, el cas d'ús que s'ha preparat s'ha centrat en el marc de l'anàlisi de l'aprenentatge amb l'ús de sistemes d'IA d'alt risc amb un algorisme predictiu.

Com podreu observar en els apartats següents, el cas ha deixat al marge l'ús d'algorismes de decisió automatitzada (ADA). Per conèixer aquests casos d'ús, recomano l'informe *Intel·ligència artificial. Decisions automatitzades a Catalunya*²⁵, elaborat per l'Autoritat Catalana de Protecció de Dades (APDCAT).

¹⁸Recomanable la lectura de la publicació *AI competency framework for teachers* de la UNESCO, 2024 (actualmente, versió disponible només en anglès):

<https://unesdoc.unesco.org/ark:/48223/pf0000391104>.

¹⁹Reglament (UE) 2024/1689 del Parlament Europeu i del Consell, de 13 de juny de 2024, pel qual s'estableixen normes harmonitzades en matèria d'intel·ligència artificial i pel qual es modifiquen els reglaments (CE) núm. 300/2008 (UE) núm. 167/2013, (UE) núm. 168/2013, (UE) 2018/858, (UE) 2018/1139 i (UE) 2019/2144 i les directives 2014/90/UE, (UE) 2016/797 i (UE) 2020/1828 (Reglament d'intel·ligència artificial): https://eur-lex.europa.eu/legal-content/ES/TXT/?uri=OJ%3AL_202401689.

²⁰ <https://www.aleks.com>.

²¹ <https://www.dreambox.com>.

²² <https://www.carnegielearning.com>.

²³ <https://www.smartsparrow.com>.

²⁴Casos reals d'aplicació de la plataforma The IntelliBoard: <https://intelliboard.net/customers/>.

²⁵Accés a l'informe *Intel·ligència artificial. Decisions automatitzades a Catalunya*:

https://apdcat.gencat.cat/web/.content/03-documentacio/intelligencia_artificial/documents/Informe-IA-es.pdf.

2. El projecte

El projecte té per objectiu dissenyar i desenvolupar un nou ecosistema d'analítica de l'aprenentatge per al sistema educatiu superior, amb la creació d'una plataforma d'analítica avançada de l'aprenentatge amb ús d'un sistema d'IA que permeti avaluar els resultats de l'aprenentatge i predir el risc d'abandonament de l'estudiantat. En concret, la plataforma servirà per a:

- Gestionar les dades relacionades amb els processos d'ensenyament i aprenentatge.
- Elaborar panells de seguiment (en anglès, *dashboards*), que monitoritzin, analitzin i mostrin de manera visual tant els resultats del conjunt de tots els estudiants com els resultats individualitzats de cada estudiant, d'acord amb els indicadors prèviament definits.
- Dissenyar i modular la metodologia emprada.
- Analitzar el rendiment acadèmic dels estudiants.
- Detectar els estudiants en risc d'abandonament dels estudis.

Les dades que es fan servir en el marc de la plataforma tenen com a objectiu detectar de manera primerenca, ràpida i eficaç el risc d'abandonament dels estudis de l'estudiantat. Així mateix, la darrera finalitat de l'ús de les dades és millorar el procés d'aprenentatge amb la tasca d'orientació psicopedagògica que s'ha de proporcionar als estudiants; no hem d'oblidar que tenen dret a tutories i assessorament per part del personal de la institució, per obtenir orientacions adequades al procés d'aprenentatge.

Per poder proporcionar a l'estudiantat informació personalitzada i oportuna sobre el seu aprenentatge, cal disposar de les dades següents:

1. Dades històriques agregades generades als sistemes de gestió de l'aprenentatge (LMS), així com els recollits de les bases de dades internes de la institució sobre l'estudiantat d'anys anteriors que va superar o abandonar els estudis d'ensenyament superior.
2. Dades que provenen directament de les institucions educatives on l'estudiant ha estat abans d'accedir al sistema educatiu superior.

3. Dades que sol·licita la institució del sistema educatiu superior durant el procés de matrícula, i que proporciona el mateix estudiant (SIS).
4. Dades obtingudes del comportament de l'estudiant durant l'evolució dels seus estudis (informació sobre l'expedient acadèmic).
5. Dades relacionades amb circumstàncies personals al llarg dels estudis (per exemple, si compatibilitza estudis amb feina, si obté una beca, si hi ha hagut mobilitat a una altra institució, etc.).

Per proporcionar un context millor a l'algorisme predictiu i així obtenir resultats més fiables, cal caracteritzar l'estudiantat d'acord amb les dades de què es disposi en cada moment. Abans que comencin els estudis, cal agrupar els estudiants d'acord amb la informació proporcionada per la institució educativa prèvia sobre a la prova d'accés al sistema educatiu superior, i pel mateix estudiant al moment de formalitzar la matrícula. Aquest perfilat inicial oferirà una primera visió de com pot comportar-se l'estudiant probabilísticament. Segons les dades que es vagin obtenint, la classificació de l'estudiantat podrà anar variant d'acord amb les mètriques i els indicadors definits provinents de les dades històriques. Amb la finalitat de garantir el compliment dels requisits en matèria de dades del RIA, els conjunts de dades utilitzats per a entrenar el sistema d'IA per a realitzar les classificacions dels estudiants, així com les dades incorporades per obtenir el perfilat, són dades anonimitzades i sense cap traçabilitat

Tot i que a l'apartat següent ens endinsarem en l'avaluació concreta del cas d'ús, a continuació esmento alguns dels diferents aspectes trobats que s'han considerat, per fer més comprensible el context del cas analitzat.

Si bé durant l'avaluació del projecte s'ha detectat la necessitat d'incorporar determinades categories especials de dades personals, com ara les relatives a la salut per incloure la informació de l'estudiantat amb necessitats de suport educatiu (NSE), en el cas d'ús no es va tenir en compte la recollida d'aquestes dades.

Pel que fa als panells de seguiment, al principi es van preveure diferents panells adaptats visualment a la informació seleccionada i rellevant per al subjecte que podia accedir-hi, amb tres colors diferents segons la situació en cada moment: color verd; per indicar que l'estudiant no abandonarà els estudis amb una probabilitat superior al 80 %; color groc, per indicar que la probabilitat d'abandonament se situa entre el 20 % i el 60 %; i color vermell, per indicar que l'estudiant té una probabilitat d'abandonament superior al 50 %). Això no obstant, un cop feta l'avaluació es van introduir els canvis següents:

- Es va determinar que l'accés a la totalitat de la informació dels panells havia d'estar limitat als tutors. Així, serien ells els que tindrien coneixement de la predicció del risc d'abandonament dels estudiants i hi actuarien, d'acord amb les seves competències i segons l'establert a la normativa d'aplicació al sistema d'educació superior.
- Es va delimitar la informació que apareixia als panells del professorat, limitant l'accés a les dades agregades de l'evolució de l'estudiantat en el seu conjunt dins de les classes. Sense la possibilitat d'accedir a la informació individualitzada de cada estudiant, s'evita l'efecte invers al que es pretén aconseguir, quan inconscientment des d'un inici es podria etiquetar el grup reduït d'estudiants que tenen un rendiment baix.
- Es va eliminar la possibilitat que l'estudiantat visualitzés els panells, per tenir cura de la seva salut mental i emocional i no provocar situacions d'angoixa o ansietat pel fet de veure un color determinat indicatiu del rendiment acadèmic.

L'avaluació, també va posar de manifest que hi ha informació d'esdeveniments que influeixen en el rendiment de l'estudiantat (com ara la mort d'una persona propera) que el sistema no té en compte. Per tant, és important tractar el risc potencial d'abandonament dels estudis cas per cas i recopilar la informació rellevant relacionada que s'ha de tenir en compte, no només a nivell individual, sinó també per considerar la seva inclusió en les variables que alimenten el sistema d'IA predictiva.

Per complementar l'avaluació de l'impacte sobre els drets fonamentals en matèria d'intel·ligència artificial (AIDF), es va efectuar una avaluació d'impacte en matèria de protecció de dades (AIPD). Tot i que aquesta darrera avaluació no s'incorpora en aquest document, això va comportar que es fessin canvis des de la perspectiva de protecció de dades, en aplicació als principis establerts pel Reglament general de protecció de dades (RGPD).²⁶ Van ser canvis addicionals que van enriquir l'avaluació final.

²⁶Reglament (UE) 2016/679 del Parlament Europeu i del Consell, de 27 d'abril de 2016, relatiu a la protecció de les persones físiques pel que fa al tractament de dades personals i a la lliure circulació d'aquestes dades i pel qual es deroga la Directiva 95/46/CE (Reglament

general de protecció de dades - RGPD): <https://eur-lex.europa.eu/legal-content/CA/TXT/?uri=celex%3A32016R0679>.

3. L'AIDF

Planificació i abast

Secció A Descripció i anàlisi del sistema d'IA, inclosos els fluxos de dades relacionats	Quins són els principals objectius del sistema d'IA?	<ul style="list-style-type: none"> a) Proporcionar indicadors de rendiment acadèmic. b) Mostrar de manera predictiva la probabilitat d'abandonament dels estudis superiors. c) Contribuir a millorar el procés d'aprenentatge. <p>Amb la informació proporcionada pel sistema d'IA, les institucions d'educació superior poden establir polítiques per reduir la taxa d'abandonament prematur, així com ajudar els tutors en la tasca d'orientació psicopedagògica que han de prestar a l'estudiantat.</p>
	Quines són les principals característiques del sistema?	<p>La predicció de situacions en les quals es pot actuar per millorar el procés d'aprenentatge.</p>
	En quins països s'oferirà?	<p>A Espanya, però es pot fer extensible a altres institucions estrangeres d'educació superior (dins o fora de la Unió Europea), amb les quals es formalitzin acords per impartir estudis conjunts.</p>

	<p>Quins tipus de dades es tracten (personals, no personals, categories especials)?</p>	<ul style="list-style-type: none"> ▪ Dades històriques agregades sobre l'estudiantat d'anys anteriors. ▪ Dades que provenen directament de les institucions educatives on l'estudiant ha estudiat abans d'accedir al sistema educatiu superior. ▪ Dades que sol·licita la institució del sistema educatiu superior durant el procés de matrícula, que proporciona el mateix estudiant. ▪ Dades obtingudes del comportament de l'estudiant durant l'evolució dels seus estudis (informació referent a l'expedient acadèmic). ▪ Dades relacionades amb circumstàncies personals al llarg dels estudis.
	<p>Identificació dels possibles titulars de drets: qui són els individus o grups que es poden veure afectats pels sistemes d'IA? S'hi inclouen individus o grups vulnerables?</p>	<p>Estudiantat.</p>

	<p>Identificació dels responsables: quines persones/entitats estan implicades en el disseny, desenvolupament i desplegament dels sistemes d'IA? Quin és el seu paper?</p>	<p>Les mateixes institucions d'educació superior que vulguin implementar-la.</p> <p>En la gestió de les dades hi participarà el personal de la institució.</p> <p>El tractament de la informació proporcionada per les dades als panells de seguiment es visualitzarà de manera diferent segons el perfil de l'usuari: (i) estudiantat, (ii) professorat, (iii) tutor/a i (iv) gestor/a.</p>
<p>Secció B Context de drets fonamentals</p>	<p>Quins drets fonamentals es veuen potencialment afectats per l'ús del sistema d'IA?</p>	<p>Analitzats tots els drets individuals, civils, polítics, econòmics i socials establerts a la Carta dels Drets Fonamentals de la Unió Europea²⁷, s'ha conclòs que es veuen potencialment afectats els drets següents:</p> <p><input checked="" type="checkbox"/> Dignitat humana (article 1)</p> <p>Justificació: manca de visió completa; pèrdua d'autonomia personal, si la IA pren decisions i ofereix solucions per compte de les persones sense intervenció humana; pressió excessiva cap a l'estudiantat i impacte en la visió de si mateix.</p> <p><input checked="" type="checkbox"/> Respecte de la vida privada i familiar (article 7)</p> <p>Justificació: invasió de la intimitat a causa de la monitorització constant del rendiment acadèmic; afectació a la privadesa familiar; afectació a la 'privadesa decisonal'.</p>

²⁷Carta dels Drets Fonamentals de la Unió Europea: <https://eur-lex.europa.eu/legal-content/ES/TXT/HTML/?uri=CELEX:12012P/TXT&from=ES>.

		<p><input checked="" type="checkbox"/> Protecció de dades de caràcter personal (article 8)</p> <p>Justificació: implicació del perfilat o valoració de subjectes, dades a gran escala i ús de noves tecnologies.</p> <p><input checked="" type="checkbox"/> No discriminació (article 21)</p> <p>Justificació: l'ús d'algorismes basats en dades històriques i patrons que prevegin biaixos discriminatoris del passat pot perpetuar i amplificar perjudicis; no tenir en compte determinats factors o variables que hi podrien incidir; avaluació basada en els resultats de la predicció.</p> <p>Tot i que es va tenir en compte el dret a l'educació (article 14), no està afectat perquè el sistema d'IA no limita el dret d'accés a l'ensenyament.</p>
	<p>Quins instruments jurídics internacionals/regionals de protecció dels drets humans/fonamentals s'han aplicat a nivell operatiu?</p>	<p>La normativa sobre protecció de dades personals i la relativa al sistema universitari, així com la relacionada amb els col·lectius afectats (per exemple, l'Estatut de l'estudiant universitari).</p>
	<p>Quins són els tribunals o organismes més rellevants en matèria de drets fonamentals en el context d'ús?</p>	<p>Les autoritats de control en matèria de protecció de dades del país/regió on es desenvolupi i utilitzi el sistema d'IA, així com els tribunals competents d'aquest país/regió.</p>

	Quines són les decisions i disposicions més rellevants en matèria de drets humans/fonamentals?	No aplicable (N/A).
Secció C Controls implementats	Quines polítiques i procediments s'han implementat per avaluar el possible impacte sobre els drets fonamentals, inclosa la participació de les parts interessades?	No aplicable (N/A).
	S'ha efectuat, desenvolupat i aplicat una avaluació d'impacte en relació amb qüestions específiques (per exemple, protecció de dades) o algunes característiques del sistema (per exemple, l'ús de la biometria)?	S'ha realitzat una avaluació d'impacte en matèria de protecció de dades personals (AIPD).
Secció D Intervenció de les parts interessades i diligència deguda	Quins són els principals grups o comunitats potencialment afectats pel sistema d'IA, inclòs el seu desenvolupament?	Estudiantat.
	Quines parts interessades, a més dels individus o grups que poden veure's afectats pels sistemes d'IA, hi han de participar (per exemple, la societat civil i les organitzacions internacionals, experts, associacions industrials, periodistes)?	Famílies i professorat.

	<p>Hi ha altres titulars d'obligacions que hi hagin de participar, a part del proveïdor i el responsable del desplegament de la IA (per exemple, autoritats nacionals, organismes governamentals)?</p>	<p>Autoritat de control en protecció de dades, Departament d'Educació/Universitats.</p> <p>Unitat/òrgan competent de la Universitat.</p>
	<p>Han participat en el procés d'avaluació els socis comercials, inclosos els proveïdors de serveis (per exemple, subcontractistes en sistemes d'IA i conjunts de dades)?</p>	<p>No</p>
	<p>El proveïdor d'IA ha dut a terme una avaluació de la seva cadena de subministrament, per identificar si les activitats dels proveïdors/contractistes que participen en el desenvolupament de productes/serveis poden afectar els drets fonamentals?</p> <p>El proveïdor ha promogut estàndards o auditories de drets fonamentals, per garantir el respecte dels drets fonamentals entre els proveïdors?</p>	<p>No aplicable (N/A)</p>
	<p>El proveïdor i el responsable del desplegament de la IA han comunicat públicament les possibles repercussions del sistema d'IA en els drets fonamentals?</p>	<p>La publicació no n'és obligatòria, però és recomanable indicar/afegir un resum sobre l'anàlisi realitzada (tant de l'AIPD com l'avaluació en drets fonamentals).</p>

	<p>El proveïdor i el responsable del desplegament de la IA han proporcionat formació sobre estàndards de drets fonamentals al personal encarregat de la gestió i la contractació relacionades amb el sistema d'IA?</p>	<p>No aplicable (N/A)</p>
--	---	---------------------------

Matrius de risc

Taula 1. Probabilitat

Baix	El risc de perjudici és improbable o altament improbable
Mitjà	El risc es pot materialitzar
Alt	Hi ha una alta probabilitat que el risc es materialitzi
Molt alt	És molt probable que el risc es materialitzi

Taula 2. Exposició

Baix	Pocs o molt pocs titulars de drets de població identificada estan potencialment afectats
Mitjà	Part de la població identificada està potencialment afectada
Alt	La majoria de la població identificada està potencialment afectada
Molt alt	Gairebé tota la població identificada està potencialment afectada

Taula 3. Probabilitat que es produeixi un perjudici (*likelihood*)

		Probabilitat			
		Baixa	Mitjana	Alta	Molt alta
Exposició	Baixa	B	B/M	B/A	B/MA
	Mitjana	M/B	M	M/A	M/MA
	Alta	A/B	A/M	A	A/MA
	Molt alta	MA/B	MA/M	MA/H	MA

Probabilitat			
Baixa	Mitjana	Alta	Molt alta

Taula 4. Gravetat del perjudici

Baixa	Els individus i grups afectats poden patir només perjudicis menors en l'exercici dels seus drets i llibertats
Mitjana	Els individus i grups afectats poden patir perjudicis importants
Alta	Els individus i els grups afectats poden enfrontar-se a perjudicis greus
Molt alta	Els individus i grups afectats poden patir perjudicis greus o fins i tot irreversibles

Taula 5. Esforç per superar el perjudici i revertir els efectes adversos

Baix	Els perjudicis soferts es poden superar sense problemes (per exemple, temps dedicat a modificar la informació, molèsties, irritacions, etc.)
Mitjà	Els perjudicis soferts es poden superar malgrat algunes dificultats (per exemple, costos addicionals, por, incomprensió, estrès, petites malalties físiques, etc.).
Alt	Els perjudicis soferts es poden superar, tot i que amb dificultats serioses (per exemple, pèrdues econòmiques, danys materials, empitjorament de la salut, etc.).
Molt alt	Els perjudicis patits poden no superar-se (per exemple, malalties psicològiques o físiques de llarga durada, mort, etc.).

Taula 6. Gravetat (severity)

		Gravetat			
		Baixa	Mitjana	Alta	Molt alta
Esforç	Baix	B	B/M	B/A	B/MA
	Mitjà	M/B	M	M/A	M/MA
	Alt	A/B	A/M	A	A/MA
	Molt alt	MA/B	MA/M	MA/A	MA

Gravetat			
Baix	Mitjà	Alt	Molt alt

Taula 1A. Recollida de dades i anàlisi de riscos

Drets/llibertats potencialment afectats	Descripció de l'impacte	Probabilitat			Gravetat		
		Probabilitat de resultats adversos	Exposició	Probabilitat	Gravetat	Probabilitat de resultats adversos	Exposició
Dignitat humana	L'algorisme té en compte uns paràmetres generats per dades històriques recollides dins d'un context socioeconòmic concret, però no tots els que podrien tenir una afectació directa en el rendiment acadèmic actual (per exemple, les millores docents/adaptacions sol·licitades; les persones que no s'identifiquen amb un gènere determinat; accés a noves tecnologies, etc.)	[Alta] Hi ha una alta probabilitat que es produeixi el risc, perquè la informació proporcionada amb les dades actuals no recull tota la informació per englobar totes les categories de grups que es podrien veure afectats	[Baixa] L'exposició és baixa, ja que afecta a nombre limitat de casos de falta d'informació.	[Mitjana]	[Mitjana] Encara que els col·lectius que es puguin veure afectats siguin reduïts, pel fet de no tenir en compte determinats paràmetres l'estudiantat afectat podria patir importants perjudicis. L'algorisme podria indicar una predisposició determinada al fet que aquest col·lectiu reduït tindrà unes actuacions concretes, que	Dignitat humana	[Medium]

					no reflecteixen la realitat de la seva situació		
Respecte a la vida privada i familiar	Seguiment constant del rendiment acadèmic; impacte en la privadesa familiar; impacte en la "privadesa decisional".	[Alta] Hi ha una alta probabilitat que es produeixi el risc. Tot i que les dades per separat són necessàries dins dels diferents procediments que duu a terme la institució, disposar de certa informació recopilada de manera conjunta per a una finalitat concreta comporta una afectació sobre el control d'aquesta informació	[Molt alta] Atès que afectaria tot l'estudiantat	[Molt alta]	[Baixa] L'estudiantat afectat patir perjudicis menors en l'exercici dels seus drets i les seves llibertats, atès que la informació es troba en el marc de les funcions de la institució educativa	[Baix] El personal de les institucions d'educació superior té uns deures i unes obligacions en el marc de les seves funcions, per salvaguardar els drets de l'estudiantat. Així mateix, la institució ha de formar el seu personal en la matèria, a fi que conegui la normativa aplicable i pugui actuar davant de les diferents situacions en què es pugui trobar	[Baixa]

<p>Protecció de dades personals</p>	<p>L'algorisme recull dades a gran escala i fa ús de noves tecnologies.</p> <p>A més, implica perfilat o valoració de subjectes per poder fer l'anàlisi predictiva d'abandonament</p>	<p>[Mitjana]</p> <p>Existeix el risc que s'elaborin perfils i prediccions inexactes. Per tant, cal fer una avaluació d'impacte en matèria de protecció de dades</p>	<p>[Molt alta]</p> <p>Atès que afectaria tot l'estudiantat</p>	<p>[Alta]</p>	<p>[Mitjana]</p> <p>L'elaboració de perfils incorrectes repercuteix negativament en la representació correcta del rendiment dels estudiants i dels resultats esperats. S'ha considerat si aquest és un cas en el qual els estudiants també tenen dret a no ser perfilats.</p>	<p>[Mitjà]</p> <p>En aplicació de l'RGPD, s'han d'establir les mesures organitzatives adequades i protegir els drets dels estudiants en relació amb el tractament de dades, perquè la forma en què es generen i utilitzen els perfils pot requerir alguns canvis en el disseny dels sistemes d'IA (per exemple la seva posada a punt) i en el seu ús.</p>	<p>[Mitjana]</p>
<p>No discriminació</p>	<p>Tenint en compte que l'algorisme compara dades històriques, n'obté d'altres institucions i recull dades de la matrícula, hi</p>	<p>[Mitjana]</p> <p>Atès el pes de les variables en la consideració del model</p>	<p>[Molt alta]</p> <p>Atès que l'impacte afectaria potencialment</p>	<p>[Alta]</p>	<p>[Mitjana]</p> <p>La classificació de l'estudiantat pot ser esbiaixada i</p>	<p>[Mitjà]</p> <p>La classificació de l'estudiantat no és estàtica, per la qual</p>	<p>[Mitjana]</p>

	pot haver patrons que comportin biaixos discriminatoris del passat que es poden perpetuar i amplificar perjudicis; no tenir en compte determinats factors o variables que podrien incidir; avaluació basada en els resultats de la predicció	predictiu del risc d'abandonament del model predictiu	tots els estudiants als quals s'aplicaria l'algorisme		proporcionar informació errònia sobre els primers indicadors de risc d'abandonament, la qual cosa dona lloc a un tracte desigual injustificat.	cosa les dades inicials no les encasellen en una classificació determinada, sinó que anirà canviant segons la seva evolució en els estudis	
--	--	---	---	--	--	--	--

Taula 7. Impacte general del risc

		Gravetat			
		Baixa	Mitjana	Alta	Molt alta
Probabilitat	Baixa				
	Mitjana				
	Alta				
	Molt alta				

Taula 2A. Gestió de riscos (I)

Dret/llibertat afectats	Probabilitat	Gravetat	Impacte global	Mesures per prevenir/mitigar els impactes
Dignitat humana	[Mitjana]	[Mitjana]	[Mitjà]	<ul style="list-style-type: none"> Utilitzar el model predictiu com a eina de suport a la presa de decisions i no com a eina de presa de decisions automatitzades; ús limitat dels resultats proporcionats pel sistema d'IA. No fer visible per als estudiants els índexs del risc d'abandonament. Posar a disposició del personal de la institució directrius d'ús del sistema d'IA (política d'ús).
Respecte de la vida privada i familiar	[Molt alta]	[Baixa]	[Mitjà]	<ul style="list-style-type: none"> Dissenyar el model predictiu de manera que garanteixi en tot el moment que es té el control de les dades. Limitar l'accés al perfil individual, sense visualitzar els perfils dels altres estudiants. Que l'eina de predicció no tingui en compte les interlocucions ni les comunicacions de l'estudiant amb el professorat o entre si. Que l'eina sigui utilitzada com a eina de suport a l'establiment de mesures i no com a eina de presa de decisions automatitzades.
Protecció de dades	[Alta]	[Mitja]	[Mitja]	<ul style="list-style-type: none"> Restringir l'accés a les dades: accés total als tutors i només dades agregades als professors.
No discriminació	[Alt]	[Mitja]	[Mitja]	<ul style="list-style-type: none"> Revisar periòdicament que les dades introduïdes a les bases de dades no generen perfils discriminatoris. Revisar el càlcul inicial per la construcció de perfils periòdicament a partir de la incorporació de noves dades a la base de dades, perquè els nous comportaments puguin revisar els potencials biaixos. Revisar periòdicament que el model de predicció no sigui discriminatori i que es dissenya un algorisme sensible a la discriminació i al potencial biaix.

Taula 3A. Gestió de riscos (II)

Dret/llibertat afectats	Probabilitat (residual)	Gravetat (residual)	Impacte residual
Dignitat humana	[Mitjana]	[Baixa]	[Baix]
Respecte de la vida privada i familiar	[Alta]	[Baixa]	[Mitjà]
Protecció de dades	[Mitjana]	[Mitjana]	[Mitjà]
No discriminació	[Mitjana]	[Mitjana]	[Mitjà]

4. Comentaris

Formar part d'aquest grup de treball proposat per l'Autoritat Catalana de Protecció de Dades (APDCAT) ha donat la possibilitat de treballar amb persones pertanyents a altres sectors. Això ha significat una gran millora en l'anàlisi del cas d'ús exposat, atès que s'han aportat diferents perspectives que han evidenciat diverses sensibilitats pel que fa a l'avaluació de cadascun dels drets fonamentals.

Realitzar l'avaluació d'impacte relativa als drets fonamentals per als sistemes d'IA d'alt risc, ens mostra la importància d'adoptar un plantejament ampli en el moment d'establir la identificació dels riscos, que abasti tots i cada un dels drets fonamentals, així com les mesures mitgadores dels riscos identificats. Un dels aspectes més difícils ha sigut concretar els riscos residuals, ja que la determinació del risc resultant tres les mesures de mitigació previstes fa que l'anàlisi d'escenaris en l'àmbit dels drets fonamentals no sigui fàcil.

L'enfocament de l'anàlisi des de l'òptica d'un delegat de protecció de dades (DPD) posa de manifest la prioritat d'aplicació d'un dret (el de la protecció de dades personals) que està àmpliament regulat en comparació amb altres drets fonamentals. També, la necessitat d'abordar-lo conjuntament amb altres drets que estan íntimament relacionats, per complir amb les obligacions derivades de tot el marc regulador.

Així mateix, s'ha posat de manifest que cal:

- Formar tots els agents implicats en matèria de drets fonamentals.
- Evitar els llistats de compliment (*checklists*) relacionats amb els drets fonamentals, atès que no s'aprofundeix en els diferents aspectes que els integren.
- Donar a conèixer la repercussió dels sistemes d'IA en l'àmbit de l'educació.
- Comprendre la definició de sistemes d'IA proporcionada pel RIA (article 3.1) i la repercussió dels sistemes d'IA considerats d'alt risc de l'annex III.
- Establir l'avaluació per defecte i des del disseny.

- Conscienciar els responsables del desplegament dels sistemes d'IA de risc alt de la necessitat de dur a terme i disposar de l'avaluació d'impacte relativa als drets fonamentals.
- Adaptar les avaluacions al context actual en què ens trobem en cada moment.
- Tot i que l'avaluació d'impacte relativa als drets fonamentals per als sistemes d'IA d'alt risc no ha d'anar de la mà de l'AIPD, es recomanable efectuar-la.

Cas d'ús 2: Una eina per a la gestió dels recursos humans

1. El context

Aquest cas d'ús s'emmarca en el procés de selecció de persones que té implementat l'organització i, per tant, es guia pels principis d'actuació de què s'ha dotat. El duu a terme el departament de recursos humans de l'entitat, directament o mitjançant proveïdors.

En concret, l'organització ha definit diferents palanques en el seu pla director de gestió de persones: i) Impulsar una cultura d'equip il·lusionant, compromesa amb el nou projecte, col·laborativa i àgil, i alhora promoure un lideratge proper, motivador, no jeràrquic, amb capacitats transformadores; ii) Impulsar noves formes de treball, amb respecte per la diversitat, la igualtat d'oportunitats, la inclusió i la no-discriminació, i incorporar la sostenibilitat en els processos de recursos humans; iii) Transformar la gestió del model de desenvolupament de les persones: més proactiu en la capacitat dels equips, amb focus en habilitats crítiques; iv) Desenvolupar una proposta de valor única i diferencial per a l'empleat; i v) Evolucionar cap a una cultura *data driven* de la funció de persones, mitjançant l'optimització de l'estructura de dades i l'aplicació d'intel·ligència artificial i noves tecnologies, per facilitar l'anàlisi de la informació i prendre decisions basades en dades en matèria de persones.

En el marc d'aquesta darrera palanca, es va proposar, analitzar i finalment implementar el cas que exposem tot seguit.

Abans d'entrar en el detall, escau assenyalar que l'entitat és una organització madura respecte dels esquemes de compliment de protecció de dades i seguretat de la informació, i avançada en relació amb la

intel·ligència artificial i la seva governança, atès que ja havia adoptat i implementat, entre d'altres, les mesures següents:

1. Creació i aplicació de metodologies internes per desenvolupar i implementar sistemes d'intel·ligència artificial, que inclou els 144 controls establerts per l'AEPD a la guia d'auditories de tractaments que inclouen intel·ligència artificial. Això permet que els sistemes d'IA desenvolupats en intern compleixin per defecte i des del disseny una multitud de controls, com ara l'inventari, associació al tractament de dades a què donen servei, avaluació de la necessitat i proporcionalitat, valoració de qualitat de la dada (que inclou l'anàlisi i mitigació dels possibles biaixos), de la seva explicabilitat, transparència i robustesa, tant en el sentit del Reglament d'intel·ligència artificial com del Reglament general de protecció de dades. També, mesures en l'àmbit de la validació i la verificació de la bondat del sistema.
2. Anàlisi d'aquests casos d'ús abans de la seva implementació en el marc de l'avaluació d'impacte en protecció de dades, amb ampliació del seu objecte, per part dels equips d'assessoria jurídica innovació i privadesa, assessoria jurídica-laboral, sistemes (CDO -*chief data Officer* en anglès-, on s'inclou des d'una perspectiva tècnica l'equip d'IA responsable) i la seguretat de la informació (CISO, *chief information security officer* en anglès). Tot això permet fer un segon control sobre la bondat d'aquests sistemes.
3. Avaluació d'aquestes iniciatives pels comitès corporatius corresponents i sanció, si escau.

2. El projecte

Desenvolupament i aplicació d'un sistema d'intel·ligència artificial (4 models d'aprenentatge automàtic), basat en l'experiència prèvia de l'organització en la selecció de personal per cobrir determinades vacants. En concret, el sistema fa una tasca molt concreta i limitada, consistent a predir una dada

addicional per a cada empleat/ada: la probabilitat que s'adeqüi a aquesta vacant, d'acord amb dades de la relació laboral i les característiques del centre de destinació.

El resultat que dona el sistema s'integra com una dada més en el procés de selecció de personal, que sempre està gestionat i liderat pels tècnics especialitzats de recursos humans que, juntament amb la resta d'informació disponible i d'acord amb els processos interns de l'entitat, podran utilitzar aquesta informació a l'hora de desenvolupar les funcions de selecció.

Perquè s'entengui millor, davant d'una vacant específica, el sistema permetrà als tècnics de recursos humans visualitzar els empleats/ades de l'entitat, ordenats segons aquesta probabilitat d'adequació a la vacant, l'hagin sol·licitat o no. A partir d'aquí, serà el personal especialitzat de recursos humans qui decidirà si utilitza aquesta informació addicional o no.

Per tant, la finalitat del sistema és ajudar i millorar l'eficiència del procés de selecció de personal, posant a disposició del departament de recursos humans una informació sistematitzada que, altrament, caldria compilar i estructurar manualment. El sistema no pren mai cap decisió.

3. L'AIDF

Planificació i abast

Secció A Descripció i anàlisi del sistema d'IA, inclosos els fluxos de dades relacionats	Quins són els principals objectius del sistema d'IA?	Millorar el procés de selecció de determinades vacants, posant a disposició del personal especialitzat del departament de recursos humans informació addicional sobre l'adequació d'un candidat a una determinada vacant, sobre la base de criteris objectius. <ol style="list-style-type: none"> I. Eficiència i estalvi de temps del personal de selecció, a l'hora de verificar certs requisits objectius rellevants per cobrir una vacant. II. Objectivització del procés. III. Promoció de la proactivitat de l'organització en l'ofertament de la vacant a candidats/ates que no l'han sol·licitada.
	Quines són les principals característiques del sistema?	Es configura com una eina de suport que duu a terme una tasca complementària i limitada, consistent a generar una dada addicional per a cada empleat/ada en el procés de selecció de determinades vacants. Permet que els tècnics de recursos humans visualitzin el personal de l'entitat ordenat d'acord amb la seva adequació a una vacant concreta.
	En quins països s'oferirà?	Espanya
	Quins tipus de dades es tracten (personals, no personals, categories especials)?	<ul style="list-style-type: none"> · Dades relatives a l'activitat laboral (nivell i grup, funcions assignades, dades de productivitat, qualitat i compliment). · Característiques del centre de destinació (mida i tipus de centre).

	<p>Identificació dels possibles titulars de drets: qui són els individus o grups que es poden veure afectats pels sistemes d'IA? S'hi inclouen individus o grups vulnerables?</p>	<p>Personal de l'organització.</p>
	<p>Identificació dels responsables: quines persones/entitats estan implicades en el disseny, desenvolupament i desplegament dels sistemes d'IA? Quin és el seu paper?</p>	<p>El departament de recursos humans, assessoria jurídica en general (inclosa l'assessoria jurídic-laboral), DPD sistemes (CDO), seguretat de la informació (CISO). El primer ha desenvolupat i utilitzat l'eina. Els segons són equips avaluadors que han acompanyat en el desenvolupament i l'aplicació del sistema, establint i implementant, si escau, els controls necessaris que excedien l'àmbit del desenvolupament.</p>
<p>Secció B</p> <p>Context de drets fonamentals</p>	<p>Quins drets fonamentals es veuen potencialment afectats per l'ús del sistema d'IA?</p>	<p><input checked="" type="checkbox"/> Protecció de dades personals.</p> <p><input checked="" type="checkbox"/> No discriminació.</p> <p><input checked="" type="checkbox"/> Igualtat entre homes i dones.</p> <p><input checked="" type="checkbox"/> Dret a la informació i a la consulta dels treballadors a l'empresa.</p>
	<p>Quins instruments jurídics internacionals/regionals de protecció dels drets humans/fonamentals s'han aplicat a nivell operatiu?</p>	<p>La normativa de protecció de dades personals, la de relacions laborals (entre d'altres, l'Estatut dels treballadors i la 'Guia pràctica i eina sobre l'obligació empresarial d'informació sobre l'ús d'algorismes en l'àmbit laboral', del Ministeri de Treball, així com el Reglament d'intel·ligència artificial).</p>

	<p>Quins són els tribunals o organismes més rellevants en matèria de drets fonamentals en el context d'ús?</p>	<p>Autoritats de Protecció de Dades Ministeri de Treball</p>
	<p>Quines són les decisions i disposicions més rellevants en matèria de drets humans/fonamentals?</p>	<p>Cabal comunitari, tant en matèria del dret fonamental a la protecció de dades com del dret a la igualtat i no discriminació.</p>
<p>Secció C Controls implementats</p>	<p>Quines polítiques i procediments s'han implementat per avaluar el possible impacte sobre els drets fonamentals, inclosa la participació de les parts interessades?</p> <p>S'ha efectuat, desenvolupat i aplicat una avaluació d'impacte en relació amb qüestions específiques (per exemple, protecció de dades) o algunes característiques del sistema (per exemple, l'ús de la biometria)?</p>	<p>El sistema d'IA s'ha desenvolupat sobre la base de metodologia interna de desenvolupament, que incorpora per defecte i des del disseny els controls establerts per l'AEPD a la seva guia de requisits per a auditories de tractaments que incloguin IA (com ara l'inventari i el registre del sistema, el control de les dades i del biaix, la intervenció humana, l'explicabilitat, la validació i verificació del sistema, etc.).</p> <p>Adicionalment, s'ha fet una avaluació d'impacte sobre la protecció de dades, que ja inclou el pronunciament sobre l'afectació a la resta de drets fonamentals en base a la metodologia pròpia de l'organització. En concret, l'AIPD ha previst:</p> <ol style="list-style-type: none"> I. L'anàlisi sobre el compliment del tractament de dades de la normativa de protecció de dades (cosa que inclou les obligacions jurídiques, de seguretat i les relatives al mateix sistema d'IA). II. L'anàlisi dels potencials danys i perjudicis materials o immaterials que es podrien causar, així com si escauen els controls identificats i les mesures mitigadores establertes. III. L'afectació als drets fonamentals, la relació dels quals s'ha configurat tant amb els establerts a la Carta de drets fonamentals de la UE com de la Constitució espanyola.

		<p>Tant les metodologies assenyalades com la mateixa AIPD han estat avaluades pel comitè corporatiu de l'entitat, a proposta d'un equip avaluador integrat per membres de l'AJ, DPD, CISO i CDO.</p> <p>Adicionalment, s'ha previst el circuit d'informació corresponent a la representació dels treballadors.</p>
Secció D Intervenció de les parts interessades i diligència deguda	Quins són els principals grups o comunitats potencialment afectats pel sistema d'IA, inclòs el seu desenvolupament?	Personal de l'entitat.
	Quines parts interessades, a més dels individus o grups que poden veure's afectats pels sistemes d'IA, hi han de participar (per exemple, la societat civil i les organitzacions internacionals, experts, associacions industrials, periodistes)?	Equips avaluadors establerts per l'entitat (DPD, CDO i CISO), a més de l'assessoria jurídica inclosa la laboral.
	Hi ha altres titulars d'obligacions que hi hagin de participar, a part del proveïdor i el responsable del desplegament de la IA (per exemple, autoritats nacionals, organismes governamentals)?	No
	Han participat en el procés d'avaluació els socis comercials, inclosos els proveïdors de serveis (per exemple, subcontractistes en sistemes d'IA i conjunts de dades)?	No

	<p>El proveïdor d'IA ha dut a terme una avaluació de la seva cadena de subministrament, per identificar si les activitats dels proveïdors/contractistes que participen en el desenvolupament de productes/serveis poden afectar els drets fonamentals?</p> <p>El proveïdor ha promogut estàndards o auditories de drets fonamentals, per garantir el respecte dels drets fonamentals entre els proveïdors?</p>	<p>N/A</p>
	<p>El proveïdor i el responsable del desplegament de la IA han comunicat públicament les possibles repercussions del sistema d'IA en els drets fonamentals?</p>	<p>L'organització ha comunicat al seu personal i a la representació dels treballadors l'ús del sistema d'IA. També la seva finalitat, lògica i conseqüències, d'acord amb el que preveuen tant la normativa de protecció de dades com la normativa laboral.</p>
	<p>El proveïdor i el responsable del desplegament de la IA han proporcionat formació sobre estàndards de drets fonamentals al personal encarregat de la gestió i la contractació relacionades amb el sistema d'IA?</p>	<p>N/A</p>

Matrius de risc

Taula 1. Probabilitat

Baixa	El risc de perjudici és improbable o altament improbable
Mitjana	El risc es pot materialitzar
Alta	Hi ha una alta probabilitat que el risc es materialitzi
Molt alta	És molt probable que el risc es materialitzi

Taula 2. Exposició

Baixa	Pocs o molt pocs titulars de drets de població identificada estan potencialment afectats
Mitjana	Part de la població identificada està potencialment afectada
Alta	La majoria de la població identificada està potencialment afectada
Molt alta	Gairebé tota la població identificada està potencialment afectada

Taula 3. Probabilitat que es produeixi un perjudici (*likelihood*)

		Probabilitat			
		Baixa	Mitjana	Alta	Molt alta
Exposició	Baixa	B	B/M	B/A	B/MA
	Mitjana	M/B	M	M/A	M/MA
	Alta	A/B	A/M	A	A/MA
	Molt alta	MA/B	MA/M	MA/H	MA

Probabilitat			
Baixa	Mitjana	Alta	Molt alta

Taula 4. Gravetat del perjudici

Baixa	Els individus i grups afectats poden patir només perjudicis menors en l'exercici dels seus drets i llibertats
Mitjana	Els individus i grups afectats poden patir perjudicis importants
Alta	Els individus i els grups afectats poden enfrontar-se a perjudicis greus
Molt alta	Els individus i grups afectats poden patir perjudicis greus o fins i tot irreversibles

Taula 5. Esforç per superar el perjudici i revertir els efectes adversos

Baix	Els perjudicis soferts es poden superar sense problemes (per exemple, temps dedicat a modificar la informació, molèsties, irritacions, etc.)
Mitjà	Els perjudicis soferts es poden superar malgrat algunes dificultats (per exemple, costos addicionals, por, incomprensió, estrès, petites malalties físiques, etc.).
Alt	Els perjudicis soferts es poden superar, tot i que amb dificultats serioses (per exemple, pèrdues econòmiques, danys materials, empitjorament de la salut, etc.).
Molt alt	Els perjudicis patits poden no superar-se (per exemple, malalties psicològiques o físiques de llarga durada, mort, etc.).

Taula 6. Gravetat (severity)

		Gravetat			
		Baixa	Mitjana	Alta	Molt alta
Esforç	Baix	B	B/M	B/A	B/MA
	Mitjà	M/B	M	M/A	M/MA
	Alt	A/B	A/M	A	A/MA
	Molt alt	MA/B	MA/M	MA/A	MA

Gravetat			
Baix	Mitjà	Alt	Molt alt

Taula 1A. Recollida de dades i anàlisi de riscos

Drets/llibertats potencialment afectats	Descripció de l'impacte	Probabilitat			Gravetat		
		Probabilitat de resultats adversos	Exposició	Probabilitat	Gravetat	Esforç	Gravetat
Protecció de dades	L'algorisme requereix utilitzar dades de la relació laboral del personal de l'organització. Per tant, qualsevol ús que infringeixi la normativa de protecció de dades podria afectar aquest dret	[Baixa] Ja que tant la creació com l'ús del sistema s'han sotmès a una avaluació d'impacte sobre la protecció de dades	[Molt alta] L'impacte afecta potencialment totes les persones a qui s'aplica l'algorisme	[Mitjana]	[Baixa] Tot i que el tractament de dades es realitza en el marc de la selecció de personal, el sistema fa una tasca molt limitada amb dades circumscrites al context laboral objectiu	[Mitjà] Si el tractament fos no conforme, es podrien adoptar mesures per complir els deures de transparència i informació, interrompre el tractament, si escau, i fins i tot suprimir les dades generades pel sistema	[Baixa]
No discriminació i igualtat entre homes i dones	L'algoritme s'entrena amb dades històriques. Per tant, si hi haguessin biaixos i	[Baixa] S'ha verificat l'absència de	[Mitjana] L'impacte potencial	[Baixa]	[Mitjana] Ja que el sistema proveeix d'una	[Alt] El perjudici potencial d'un	[Mitjana]

	<p>es perpetuessin, es podrien produir situacions de discriminació que podrien afectar la igualtat entre homes i dones</p>	<p>biaixos en el sistema i el seu ús, inclosos controls per corregir els biaixos històrics que contenen les dades d'entrenament</p>	<p>afectaria part del personal de l'entitat</p>		<p>dada addicional els tècnics especialitzats de recursos humans, que, en tot cas, lideren la selecció del personal. El sistema no pren decisions</p>	<p>resultat d'adequació incorrecte es podria corregir en el procés de gestió de la vacant, mitjançant l'actuació del tècnic de recursos humans; o bé a posteriori es qualifica com a alta, en el cas que es resolgués un cop adjudicada la vacant</p>	
--	--	---	---	--	---	---	--

Dret a la informació i a la consulta dels treballadors	El sistema s'inclou en el marc de la gestió de la selecció de personal de l'entitat. Per tant, un incompliment de la normativa laboral, en particular en relació amb les obligacions d'informació als treballadors o als seus representants, podria afectar aquest dret	[Baixa] Ja que tant en la creació com en l'ús del sistema hi han participat, entre d'altres, els equips d'assessoria jurídica en general, inclosa la laboral. Els procediments establerts garanteixen els deures d'informació tant dels empleats com dels seus representants	[Molt alta] L'impacte afecta potencialment totes les persones a qui s'aplica l'algorisme	[Mitjana]	[Baixa] El potencial perjudici d'aquesta afectació seria la manca d'informació preceptiva a la representació laboral	[Baix] La mitigació d'aquest perjudici potencial passaria per esmenar la informació deficient	[Baixa]
--	---	--	---	------------------	---	--	----------------

Taula 7. Impacte general del risc

		Gravetat			
		Baixa	Mitjana	Baixa	Mitjana
Probabilitat	Baixa				
	Mitjana				
	Alta				
	Molt alta				

Taula 2A. Gestió de riscos

Dret/llibertat afectat	Probabilitat	Gravetat	Impacte global	Mesures per prevenir/mitigar els impactes
Protecció de dades	[Mitjana]	[Baixa]	[Baix]	<ul style="list-style-type: none"> N/A. Mesures adoptades tant en el desenvolupament de l'algorisme com abans d'utilitzar-lo (com ara l'execució de l'AIPD, facilitar informació, etc.)
No discriminació i igualtat d'homes i dones	[Baixa]	[Mitjana]	[Baix]	<ul style="list-style-type: none"> Formació dels tècnics de recursos humans, per evitar la confiança excessiva en el resultat de l'algorisme
Dret a la informació i a la consulta dels treballadors	[Mitjana]	[Baixa]	[Baix]	<ul style="list-style-type: none"> N/A. Mesures adoptades tant en el desenvolupament de l'algorisme com abans d'utilitzar-lo. S'ha facilitat la informació preceptiva tant al personal de l'entitat com a la representació laboral, seguint els models establerts pel Ministeri de Treball

4. Comentaris

Tal com s'ha exposat, l'organització compta per defecte i des del disseny amb estructures, procediments i controls que se centren en diverses de les qüestions cobertes pel RIA. Això ha permès que l'EIDF sobre el cas d'ús i el seu impacte sobre els drets fonamentals donin un nivell de risc d'impacte baix. Aquestes estructures, procediments i controls permeten que el sistema es desenvolupi en un entorn de treball controlat i que porta el datòleg (*data scientist*, en anglès) a incloure en el desenvolupament i la implementació del sistema d'IA, per defecte, certes mesures que mitiguen els riscos que ha identificat l'entitat a priori.

Adicionalment, en cas que no s'hagin identificat riscos o mesures a priori, la participació dels diferents equips avaluadors i el seu acompanyament en el desenvolupament del sistema permet que s'identifiquin i mitiguin en el moment del desenvolupament/implementació.

Cas d'ús 3: Una eina d'imatges mèdiques impulsada per IA per a la detecció del càncer

1. El context

A Europa, s'està produint un avenç important en el desenvolupament d'eines d'intel·ligència artificial (IA) que utilitzen imatges de càncer. Ja al 2012, un equip investigador de la Universitat de Maastricht va proposar el concepte de "radiomics", per referir-se al mètode d'extracció d'un gran nombre de característiques de les imatges mèdiques emprant algoritmes de caracterització de dades. Aquest creixement es pot il·lustrar consultant el nombre de publicacions sobre "radiomics AI" al portal PubMed (42 resultats el 2019; 99 resultats el 2020; 165 resultats el 2021; 235 resultats el 2022; 309 resultats el 2023; i 338 resultats el 2024). Per tant, ens trobem davant un tipus de sistema d'IA que cada vegada serà més comú, no només en el món acadèmic sinó també en el món assistencial.

D'altra banda, hi ha tipus de càncer comuns en el món on els pacients reben un alt grau de sobretractament i d'efectes considerables, que es podrien evitar amb l'ús d'un sistema d'IA que analitzés les imatges mèdiques dels pacients. Això permetria als professionals sanitaris disposar d'una eina de suport per predir la resposta a la teràpia i, en conseqüència, ajustar el tractament perquè sigui al més eficient possible; és a dir, permetria assolir l'objectiu marcat amb el tractament mínim. A més, aquests sistemes d'IA també proporcionarien tant als professionals sanitaris com als pacients una predicció de l'evolució del pacient durant els propers anys.

2. El projecte

El projecte s'estructura en una primera fase, que consisteix a desenvolupar un sistema d'IA basat en imatges mèdiques, que s'entrena amb la informació de 5.000 pacients originaris de deu països d'Europa. Per tant, el conjunt de dades d'entrenament és multicèntric.

A més, es preveu una segona fase del projecte, que implicarà la validació del sistema d'IA en vuit centres sanitaris de tot el món, fora d'Europa. L'objectiu d'aquesta segona fase és provar el sistema d'IA en un centre sanitari ubicat a Àsia, un altre a Àfrica i un altre, a l'Amèrica del Sud.

3. L'AIDF

Planificació i abast

Secció A Descripció i anàlisi del sistema d'IA, inclosos els fluxos de dades relacionats	Quins són els principals objectius del sistema d'IA?	Millorar el tractament dels pacients amb càncer X, mitjançant la predicció del següent: a) Resposta del pacient al tractament. b) Efectes secundaris (toxicitat i sensibilitat). c) Previsions per als propers cinc anys.
	Quines són les principals característiques del sistema?	Reconeixement d'imatges basat en la IA. Ús d'imatges mèdiques per predir la resposta d'un pacient a un determinat tractament. També per ajudar els professionals sanitaris a determinar-ne l'aplicació en casos concrets, així com el nivell d'ús un cop s'apliqui.
	En quins països s'oferirà?	Distribució mundial.
	Quins tipus de dades es tracten (personals, no personals, categories especials)?	<ul style="list-style-type: none"> ▪ Dades demogràfiques (sexe, edat i país). ▪ Característiques del càncer (tipus i zona afectada). ▪ Estadi del càncer i subtipus molecular. ▪ Informació sobre el tractament previ. ▪ Règim de tractament (esquema i durada). ▪ Informe de patologia (post tractament).

	Identificació dels possibles titulars de drets: qui són els individus o grups que es poden veure afectats pels sistemes d'IA? S'hi inclouen individus o grups vulnerables?	<p>Persones entre 18 i 85 anys.</p> <p>Totes les persones implicades estan afectades pel càncer. Per tant cal considerar-les vulnerables, ateses les seves condicions de salut i les relacions entre aquestes condicions i la finalitat del sistema d'IA.</p>
	Identificació dels responsables: quines persones/entitats estan implicades en el disseny, desenvolupament i desplegament dels sistemes d'IA? Quin és el seu paper?	<p>En el disseny hi participen hospitals i centres de recerca. Aquests darrers, només en el disseny del sistema d'IA; els primers, també en el tractament sanitari relacionat.</p>
Secció B Context de drets fonamentals	Quins drets fonamentals es veuen potencialment afectats per l'ús del sistema d'IA?	<input checked="" type="checkbox"/> Dret a la intimitat (inclosa la protecció de dades). <input checked="" type="checkbox"/> Llibertat davant la discriminació. <input checked="" type="checkbox"/> Dret a un nivell de vida adequat (inclòs el dret a la salut física i mental).
	Quins instruments jurídics internacionals/regionals de protecció dels drets humans/fonamentals s'han aplicat a nivell operatiu?	<p>Declaració universal dels drets humans, Carta de drets fonamentals de la Unió Europea, normatives de protecció de dades aplicables.</p>
	Quins són els tribunals o organismes més rellevants en matèria de drets fonamentals en el context d'ús?	<p>Autoritats de protecció de dades del país/regió on es desenvolupen i utilitzen els sistemes d'IA. Òrgans jurisdiccionals, Tribunal de Justícia de la Unió Europea i Tribunal Europeu de Drets Humans.</p>
	Quines són les decisions i disposicions més rellevants en matèria de drets humans/fonamentals?	<p>No aplicable</p>

Secció C Controls implementats	Quines polítiques i procediments s'han implementat per avaluar el possible impacte sobre els drets fonamentals, inclosa la participació de les parts interessades?	S'ha previst que es constitueixi un comitè d'ètica específic per al projecte.
	S'ha efectuat, desenvolupat i aplicat una avaluació d'impacte en relació amb qüestions específiques (per exemple, protecció de dades) o algunes característiques del sistema (per exemple, l'ús de la biometria)?	Cal fer una avaluació d'impacte sobre la protecció de dades (AIPD).
Secció D Intervenció de les parts interessades i diligència deguda	Quins són els principals grups o comunitats potencialment afectats pel sistema d'IA, inclòs el seu desenvolupament?	Pacients amb càncer X [anonimitzats].
	Quines parts interessades, a més dels individus o grups que poden veure's afectats pels sistemes d'IA, hi han de participar (per exemple, la societat civil i les organitzacions internacionals, experts, associacions industrials, periodistes)?	Associacions de pacients amb càncer.
	Hi ha altres titulars d'obligacions que hi hagin de participar, a part del proveïdor i el responsable del desplegament de la IA (per exemple, autoritats nacionals, organismes governamentals)?	Autoritat de protecció de dates, departament de salut local, comitè ètic d'investigació científica, autoritat de supervisió de la IA.
	Han participat en el procés d'avaluació els socis comercials, inclosos els proveïdors de serveis (per exemple, subcontractistes en sistemes d'IA i conjunts de dades)?	No

	<p>El proveïdor d'IA ha dut a terme una avaluació de la seva cadena de subministrament, per identificar si les activitats dels proveïdors/contractistes que participen en el desenvolupament de productes/serveis poden afectar els drets fonamentals?</p> <p>El proveïdor ha promogut estàndards o auditories de drets fonamentals, per garantir el respecte dels drets fonamentals entre els proveïdors?</p>	<p>N/A</p>
	<p>El proveïdor i el responsable del desplegament de la IA han comunicat públicament les possibles repercussions del sistema d'IA en els drets fonamentals?</p>	<p>No</p>
	<p>El proveïdor i el responsable del desplegament de la IA han proporcionat formació sobre estàndards de drets fonamentals al personal encarregat de la gestió i la contractació relacionades amb el sistema d'IA?</p>	<p>N/A</p>

Matrius del risc

Taula 1. Probabilitat

Baixa	El risc de perjudici és improbable o altament improbable
Mitjana	El risc es pot materialitzar
Alta	Hi ha una alta probabilitat que el risc es materialitzi
Molt alta	És molt probable que el risc es materialitzi

Taula 2. Exposició

Baixa	Pocs o molt pocs titulars de drets de la població identificada estan potencialment afectats
Mitjana	Part de la població identificada està potencialment afectada
Alta	La majoria de la població identificada està potencialment afectada
Molt alta	Quasi tota la població identificada està potencialment afectada

Taula 3. Probabilitat que es produeixi un perjudici (*likelihood*)

		Probabilitat			
		Baixa	Mitjana	Alta	Molt alta
Exposició	Baixa	B	B/M	B/A	B/MA
	Mitjana	M/B	M	M/A	M/MA
	Alta	A/B	A/M	A	A/MA
	Molt alta	MA/B	MA/M	MA/H	MA

Probabilitat			
Baixa	Mitjana	Alta	Molt alta

Taula 4. Gravetat del perjudici

Baixa	Els individus i grups afectats poden patir només perjudicis menors en l'exercici dels seus drets i llibertats
Mitjana	Els individus i grups afectats poden patir perjudicis importants
Alta	Els individus i grups afectats poden enfrontar-se a perjudicis greus
Molt alta	Els individus i grups afectats poden patir perjudicis greus, o fins i tot irreversibles

Taula 5. Esforç per superar el perjudici i revertir els efectes adversos

Baix	Els perjudicis soferts es poden superar sense problemes (com ara temps dedicat a modificar la informació, molèsties, irritacions, etc.)
Mitjà	Els perjudicis soferts es poden superar malgrat algunes dificultats (com ara costos addicionals, por, incomprensió, estrès, petits dolors físics, etc.)
Alt	Els perjudicis soferts es poden superar, tot i que amb dificultats serioses (com ara, pèrdues econòmiques, danys materials, empitjorament de la salut, etc.)
Molt alt	Els perjudicis soferts poden no superar-se (per exemple, dolors psicològics o físics de llarga durada, mort, etc.)

Taula 6. Gravetat (severity)

		Gravetat			
		Baixa	Mitjana	Alta	Molt alta
Esforç	Baix	B	B/M	B/A	B/MA
	Mitjà	M/B	M	M/A	M/MA
	Alt	A/B	A/M	A	A/MA
	Molt alt	MA/B	MA/M	MA/A	MA

Gravetat			
Baixa	Mitjana	Alta	Molt alta

Taula 1A. Recollida de dades i anàlisi dels riscos

Drets/llibertats potencialment afectats	Descripció de l'impacte	Probabilitat			Gravetat		
		Probabilitat de resultats adversos	Exposició	Probabilitat	Gravetat	Esforç	Gravetat
Protecció de dades/privacitat	<p>El desenvolupament del sistema d'IA es basa en l'ús de categories especials de dades personals i altra informació personal dels pacients.</p> <p>Qualsevol operació de tractament que no compleixi la normativa aplicable en matèria de protecció de dades personals podria afectar aquest dret</p>	<p>[Baixa]</p> <p>El projecte està subjecte a avaluacions específiques d'impacte ètic i de protecció de dades</p>	<p>[Molt alta]</p> <p>L'impacte afecta potencialment totes les persones a les quals s'aplica l'algoritme</p>	[Mitjana]	<p>[Mitjana]</p> <p>El tractament il·legal de dades sanitàries relacionades amb el càncer i l'ús il·lícit d'aquesta informació poden ser invasius i afectar la intimitat de las persones</p>	<p>[Mitjà]</p> <p>La recollida i el tractament il·lícit de dades es poden detectar i aturar, suprimint la informació que s'hagi recollit il·legalment</p>	[Mitjana]
No discriminació	<p>L'algoritme es va entrenar amb dades de centres sanitaris europeus. Per això, es poden produir discriminacions quan s'utilitzi en els tres centres sanitaris que no pertanyen a la UE</p>	<p>[Alta]</p> <p>L'ètnia pot provocar algunes diferències en les imatges mèdiques, que poden afectar la precisió del diagnòstic</p>	<p>[Molt alta]</p> <p>Totes les persones del grup pertinent (grup ètnic) a les quals s'aplica l'algoritme</p>	[Molt alta]	<p>[Molt alta]</p> <p>Impacte negatiu en la igualtat d'accés a la assistència sanitària i en la qualitat del tractament oncològic rebut</p>	<p>[Alt]</p> <p>Caldria adaptar o fins i tot tornar a entrenar l'algoritme amb dades que evitin la discriminació</p>	[Molt alta]

Dret a la salut física i mental	El funcionament incorrecte de l'algoritme pot donar lloc a un tractament sanitari ineficaç i perjudicial per al pacient, amb el consegüent perjudici per al dret a la salut	[Mitjana] Quan s'utilitza en pacients europeus [Alta] Quan s'utilitza en pacients no europeus.	[Molt alta] L'impacte afecta potencialment totes les persones a les quals s'aplica l'algoritme	[Alta] Quan s'utilitza en pacients europeus [Molt alta] Quan s'utilitza en pacients no europeus	[Molt alta] El funcionament incorrecte de l'algoritme pot donar lloc a un tractament sanitari ineficaç i perjudicial per al pacient	[Mitjà] Patologies en les quals un seguiment posterior pot corregir l'error del sistema [Alt] Patologies en les quals el control posterior no pot corregir l'error del sistema	[Alta] Quan el control posterior del càncer pot corregir l'error del sistema [Molt alta] Quan el control posterior del càncer no pot corregir l'error del sistema
---------------------------------	---	--	---	--	--	---	--

Taula 7. Impacte general del risc

		Gravetat			
		Baixa	Mitjana	Alta	Molt alta
Probabilitat	Baixa				
	Mitjana				
	Alta				
	Molt alta				

Taula 2A. Gestió dels riscos (I)

Dret afectat/Llibertat afectada	Probabilitat	Gravetat	Impacte global	Mesures per prevenir/mitigar els impactes
Protecció de dades/privacitat	[Mitjana]	[Mitjana]	[Mitjà]	<ul style="list-style-type: none"> Publicar informació sobre els procediments utilitzats per obtenir i processar les dades originals emprades per entrenar el sistema d'IA
No discriminació	[Molt alta]	[Molt alta]	[Molt alt]	<ul style="list-style-type: none"> Ampliar el conjunt de dades d'entrenament, per evitar una baixa representació dels grups rellevants
Dret a la salut física i mental	<p>[Alta]</p> <p>Quan s'utilitza en pacients europeus</p> <p>[Molt alta]</p> <p>Quan s'utilitza en pacients no europeus</p>	<p>[Alta] Quan el control posterior del càncer pot corregir l'error del sistema</p> <p>[Molt alta] Quan el control posterior del càncer no pot corregir l'error del sistema.</p>	<p>[Alt]</p> <p>Només quan s'utilitza en pacients europeus i quan la patologia permet que el tractament posterior corregeixi l'error del sistema</p> <p>[Molt alt]</p> <p>Per als altres tres escenaris</p>	<ul style="list-style-type: none"> Informar els professionals sanitaris de les limitacions de l'eina. Per exemple, indicant el tipus d'errors Diferenciar entre patologies amb o sense evolució ràpida Informar els professionals sanitaris que cal tenir en compte la taxa d'error de l'equip de diagnòstic per imatge que s'utilitzi

Taula 3A. Gestió dels riscos (II)

Dret afectat/Llibertat afectada	Probabilitat (residual)	Gravetat (residual)	Impacte residual
Protecció de dades/privacitat	[Mitjana]	[Baixa] ja que l'esforç s'ha reduït a baix	[Mitjà]
No discriminació	[Mitjana] ja que la probabilitat s'ha reduït a baixa	[Mitjana] ja que l'esforç s'ha reduït a baix	[Mitjà]
Dret a la salut física i mental	<p>[Mitjana]</p> <p>Quan s'utilitza en pacients europeus, ja que la probabilitat s'ha reduït a baixa</p> <p>[Alta]</p> <p>Quan s'utilitza en pacients no europeus, ja que la probabilitat s'ha reduït a mitjana</p>	<p>[Mitjana]</p> <p>Si es tracta d'una patologia en què un seguiment posterior pot corregir l'error del sistema, ja que l'esforç s'ha reduït a baix</p> <p>[Alta]</p> <p>Si es tracta d'una patologia en què el seguiment posterior no pot corregir l'error del sistema, ja que l'esforç s'ha reduït a mitjà</p>	<p>[Mitjà]</p> <p>Només quan s'utilitzi en pacients europeus i quan la patologia permeti que el seguiment posterior corregeixi l'error del sistema</p> <p>[Alt]</p> <p>Per als tres escenaris restants</p>

4. Comentaris

Les principals dificultats trobades en l'ús de la metodologia en el cas d'ús plantejat es descriuen a continuació.

Aquesta metodologia requereix, com a mínim, identificar els drets fonamentals i llibertats sobre els quals impactarà el sistema d'intel·ligència artificial. Per tant, cal tenir un coneixement expert sobre el contingut essencial de cadascun dels drets fonamentals i llibertats objecte d'examen. Es pot partir de la base que els delegats de protecció de dades tenen aquests coneixements, en relació amb el dret fonamental a la protecció de dades i amb el dret fonamental a la intimitat personal i familiar. Tanmateix, un delegat de protecció de dades no necessàriament ha de tenir aquesta expertesa en la resta de drets fonamentals i llibertats. En conseqüència, la primera dificultat és la necessitat de disposar d'un coneixement certament profund sobre cadascun dels drets fonamentals i llibertats, per poder identificar en primer lloc en quins drets/llibertats impactarà. Un cop identificats, sempre es pot sol·licitar a persones expertes assessorament sobre el seu contingut essencial.

La dificultat anterior es torna més complexa, quan es preveu utilitzar el sistema d'intel·ligència artificial fora de la Unió Europea. Si bé es pot arribar a definir un marc comú sobre el contingut dels drets fonamentals i llibertats dins de la Unió Europea, tenint en compte la Carta de drets fonamentals de la UE i els pronunciaments del TJUE, sense arribar al detall de les diferències establertes pels òrgans jurisdiccionals nacionals, difícilment es pot parlar d'un marc comú quan s'examina el contingut dels drets fonamentals i llibertats en tot el món. La aplicació d'aquesta metodologia a la UE i fora de la UE, amb el mateix nivell de detall, requeriria una anàlisi de dret comparat que difícilment podria assumir la gran majoria d'organitzacions, atesos els recursos humans, temporals i econòmics que caldrien. Per tant, seria útil disposar d'una guia amb el contingut mínim de cada dret fonamental i llibertat a nivell mundial, o bé per regions o tradicions jurídiques, que permetés utilitzar aquesta metodologia sense necessitat de

recursos desproporcionats en relació amb la seva finalitat; es a dir, disposar d'una anàlisi *ex ante* que faciliti el disseny del sistema d'intel·ligència artificial.

L'última dificultat trobada ha estat identificar les persones que han d'intervenir en l'avaluació d'impacte, tant pels seus coneixements com pel seu paper en el desenvolupament i execució d'aquest tipus d'intel·ligència artificial. Per exemple, identificar quines persones expertes en el desenvolupament del sistema d'intel·ligència artificial amb finalitats sanitàries hi havien de participar, sense que la seva participació impliqués revelar informació del projecte a la "competència".

Cas d'ús 4: ATENEA, IA al servei de les persones grans

1. El context

Les Administracions públiques, en general, es troben immerses en un procés de transformació digital, amb la idea de reformar els serveis públics aprofitant els avantatges que aporta l'evolució exponencial de les tecnologies. Però aquestes polítiques de transformació digital s'han de formular i implementar de manera que generin un impacte positiu en termes d'inclusió social, combinant l'impuls de la digitalització amb les polítiques socials per a minimitzar al màxim la bretxa digital que, inevitablement, es genera en aquests processos de canvi tan disruptius. Es tracta, doncs, de posar la tecnologia al servei de les persones, amb l'objectiu de millorar les relacions amb la ciutadania, l'atenció social, fer front a les desigualtats que origina la digitalització creixent, garantir la igualtat d'oportunitats i, en general, millorar les condicions de vida de la ciutadania.

És en aquest context, i sota aquestes premisses i compromisos socials, que s'ha impulsat el projecte ATENEA (en fase de prova pilot) que té com a finalitat contribuir a la transformació digital dels territoris, posant en el centre a la ciutadania més vulnerable, amb els objectius específics de reduir la bretxa existent i la solitud no desitjada, augmentar la seguretat de les persones, especialment dins de la llar i, fomentar la inclusió, el benestar, la salut i, en definitiva, la qualitat de vida de les persones. El projecte va adreçat específicament a persones més grans de 65 anys, que viuen soles i que pateixen amb més incidència les vulnerabilitats originades per la bretxa digital.

2. El projecte

ATENEA és un projecte basat en el desenvolupament d'un sistema d'IA generativa (xarxes neuronals), assistent de veu, biometria per veu i automatització de processos robotitzada, combinades amb tecnologies madures com *data analytics*, informàtica al núvol i telefonia intel·ligent. Mitjançant el reconeixement biomètric de la veu permet donar resposta a les peticions de les persones majors usuàries amb relació a diferents casos d'ús: trucada i videotrucada a un familiar, trucada a emergències (112), trucada a l'ajuntament de referència i/o reserva automàtica de cita prèvia amb els serveis socials, reserva de cita prèvia amb el centre d'atenció primària de referència, recordatoris d'agenda, indicacions de rutes de transport i, en un futur, compra en línia, gestions bancàries i gestions de subministres. La solució ATENEA no requereix habilitats digitals ni manipulació per part de l'usuari, la identificació és biomètrica amb la finalitat de garantir l'ús individual exclusiu i la seguretat.

Amb tot, ATENEA és una intel·ligència artificial en forma de tauleta, sense botons i sense pantalles tàctils, que funciona només amb la interacció tan senzilla com la veu. A una persona gran, probablement en situació de dependència, li dona respostes àgils a les seves necessitats bàsiques. Aquesta intel·ligència artificial permet fer amb una conversa les gestions diàries, com ara conèixer l'extracte bancari, demanar una cita prèvia al metge, fer una videoconferència amb un familiar i trucar a emergències o serveis socials.

ATENEA ha estat dissenyada amb la cocreació de persones grans, cuidadors professionals i l'entorn de la persona usuària, establint un vincle de confiança i recolzament en cas de necessitat.

La iniciativa està liderada per una aliança estratègica entre entitats tecnològiques i socials que s'encarreguen, respectivament, del disseny de la solució, el contacte amb els usuaris i el desplegament d'integradors sociodigitals sobre el terreny, els serveis de núvol, la dotació de dispositius

(tauletes), la tecnologia de la parla, l'automatització robòtica de processos, la provisió de comunicacions, la seguretat, l'avaluació de resultats d'impacte i la garantia dels drets de les persones usuàries. Es tracta d'una col·laboració público-privada. La fase actual del projecte està subvencionada pel Departament de Drets Socials del Govern de la Generalitat de Catalunya en el marc del Pla de recuperació, transformació i resiliència, finançat per la Unió Europea, Next Generation EU. Aquesta experiència té el suport de diferents ajuntaments i Administracions públiques que posen el seu territori com a prova pilot per al testatge de la solució i s'encarreguen de trobar usuaris potencials.

3. L'AIDF

Planificació i abast

Secció A Descripció i anàlisi del sistema d'IA, inclosos els fluxos de dades relacionats	Quins són els principals objectius del sistema d'IA?	Donar resposta automatitzada a les peticions d'atenció realitzades mitjançant la veu de les persones usuàries: <ul style="list-style-type: none"> ▪ Trucada i videotrucada a un familiar. ▪ Trucada a emergències (112). ▪ Trucada a l'ajuntament de referència i/o reserva automàtica de cita prèvia amb els serveis socials. ▪ Reserva de cita prèvia amb el centre d'atenció primària de referència ▪ Recordatoris d'agenda ▪ Informació de rutes de transport ▪ Recordatoris de medicació ▪ Agenda de la ciutat ▪ En desenvolupament: compra en supermercat en línia, televisió i ràdio a la carta
	Quines són les principals característiques del sistema?	<p>Un assistent de veu basat en intel·ligència artificial que interactua amb els usuaris mitjançant comandaments de veus, proporcionant informació, executant tasques i oferint serveis en temps real. El seu disseny combina tecnologies de reconeixement de veu, processament del llenguatge natural (NLP) i síntesis de veu per a entendre, processar i respondre a les sol·licituds dels usuaris de manera natural.</p> <p>Les peticions per veu de les persones usuàries es processen mitjançant un sistema d'automatització robòtica de processos, que és una tecnologia que permet automatitzar tasques repetitives, basades en regles i que generalment</p>

		són realitzades per humans, utilitzant robots de programari o bots. Aquests bots imiten i executen accions humanes en entorns digitals, com fer clics, moure arxius, completar formularis, copiar i enganxar informació, o processar dades de forma massiva. En suma, ATENEA permet fer les gestions diàries amb una conversa amb la IA.
	En quins països s'oferirà?	Espanya
	Quins tipus de dades es tracten (personals, no personals, categories especials)?	<ul style="list-style-type: none"> ▪ Dades demogràfiques (edat, ciutat, situació familiar, país) ▪ Informació personal, sense incloure les dades relacionades que les persones usuàries tracten amb els serveis acceditis a través de la solució. ▪ Configuració del dispositiu ▪ Idioma dels diàlegs (CAT/SPA) ▪ Informació d'accés a la cita prèvia del metge de família ▪ Telèfon de cita prèvia serveis socials municipals/Oficina d'Atenció Ciutadana (OAC) ▪ Telèfon d'emergències/teleassistència ▪ Telèfon i grau de parentiu de famílies ▪ Telèfon dels integradors sociotecnològics assignats. ▪ Informació relacionada recordatoris i alertes personalitzades.
	Identificació dels possibles titulars de drets: qui són els individus o grups que es poden veure afectats pels sistemes d'IA? S'hi inclouen individus o grups vulnerables?	<p>Les persones usuàries, més grans de 65 anys, amb un nivell cognitiu suficient per relacionar-se amb el dispositiu.</p> <p>Atès que totes les persones implicades estan afectades per la bretxa digital i la solitud a la llar, s'han de considerar vulnerables, ateses les seves condicions socio-demogràfiques i a les relacions entre aquestes condicions i la finalitat del sistema d'IA.</p>

	<p>Identificació dels responsables: quines persones/entitats estan implicades en el disseny, desenvolupament i desplegament dels sistemes d'IA? Quin és el seu paper?</p>	<p>La iniciativa està liderada per una aliança estratègica entre entitats tecnològiques i socials que s'encarreguen del disseny de la solució, el contacte amb els usuaris i el desplegament d'integradors sociodigitals sobre el terreny, serveis de núvol, dotació de dispositius (tauletes), tecnologia de la parla, l'automatització robòtica de processos, provisió de comunicacions, seguretat, avaluació de resultats i impacte i garantia dels drets de les persones usuàries. Es tracta d'una col·laboració públic privada. La fase actual del projecte està subvencionada pel Departament de Drets Socials del Govern de Catalunya en el marc del Pla de recuperació, transformació i resiliència, finançat per la Unió Europea, Next Generation EU. Aquesta experiència té el suport de diferents ajuntaments i administracions públiques que posen el seu territori com a prova pilot per al testatge de la solució i s'encarreguen de trobar els usuaris potencials.</p>
<p>Secció B Context de drets fonamentals</p>	<p>Quins drets fonamentals es veuen potencialment afectats per l'ús del sistema d'IA?</p>	<ul style="list-style-type: none"> <input checked="" type="checkbox"/> Dret a la intimitat (inclosa la protecció de dades). <input checked="" type="checkbox"/> Dret a la privacitat <input checked="" type="checkbox"/> Llibertat davant la discriminació. <input checked="" type="checkbox"/> Dret a la salut <input checked="" type="checkbox"/> Dret a l'assistència social <input checked="" type="checkbox"/> Dret a l'accés als serveis d'interès econòmic general <input checked="" type="checkbox"/> Dret a l'accés als serveis d'interès general

	<p>Quins instruments jurídics internacionals/regionals de protecció dels drets humans/fonamentals s'han aplicat a nivell operatiu?</p>	<p>El projecte es va iniciar abans de la publicació del Reglament d' intel·ligència artificial que també serà aplicable a aquesta solució. En el seu moment, la normativa aplicable va ser el Reglament (UE) 2016/679, del Parlament Europeu i del Consell, de 27 de abril de 2016, relatiu a la protecció de les persones físiques pel que fa al tractament de dades personals i a la lliure circulació d'aquestes dades (Reglament general de protecció de dades (RGPD) i la Llei orgànica 3/2018, de 5 de desembre, de protecció de dades personals i garantia dels drets digitals (LOPDGDD) i normativa complementària.</p>
	<p>Quins són els tribunals o organismes més rellevants en matèria de drets fonamentals en el context d'ús?</p>	<p>A Espanya, la protecció dels drets fonamentals es garanteix a través d'un marc normatiu nacional i institucions judicials i no judicials. A més, està vinculada als sistemes de protecció internacionals i regionals, especialment al sistema europeu de drets fonamentals.</p> <p>A continuació, es presenten els tribunals i organismes més rellevants en matèria de drets fonamentals a Espanya:</p> <ul style="list-style-type: none"> ▪ Tribunal Constitucional ▪ Tribunal Suprem ▪ Audiència Nacional ▪ Jutjats i Tribunals ordinaris ▪ Defensor del poble i institucions anàlogues en les regions ▪ Agències autonòmiques de defensa dels drets ▪ Fiscalia General de l'Estat ▪ Autoritats de control en matèria de protecció de dades del país/regió

	<p>Quines són les decisions i disposicions més rellevants en matèria de drets humans/fonamentals?</p>	<p>Tribunal Constitucional:</p> <ul style="list-style-type: none">▪ Sentència 135/2024: aborda la vulneració del dret a la tutela judicial efectiva, en aplicar preceptes legals que havien estat declarats inconstitucionals prèviament.▪ Sentència 113/2021: el Tribunal va reconèixer la vulneració del dret a la tutela judicial efectiva en relació amb la protecció a la família i els menors, i va emfatitzar la necessitat d'una motivació reforçada en casos que afecten drets fonamentals substantius.▪ Sentència 58/2018: tracta sobre la protecció de dades personals i la llibertat d'expressió, i estableix criteris sobre l'equilibri entre ambdós drets. <p>Disposicions legislatives:</p> <ul style="list-style-type: none">▪ Llei orgànica 3/2007, de 22 de març, per a la igualtat efectiva de dones i homes: estableix mesures per eliminar la discriminació per raó de gènere.▪ Llei 15/2022, de 12 de juliol, integral per a la igualtat de tracte i la no-discriminació: reforça el marc legal contra qualsevol forma de discriminació.▪ Llei orgànica 1/2004, de 28 de desembre, de mesures de protecció integral contra la violència de gènere: aborda la prevenció, protecció i sanció de la violència contra les dones.
--	--	--

Secció C Controls implementats	Quines polítiques i procediments s'han implementat per avaluar el possible impacte sobre els drets fonamentals, inclosa la participació de les parts interessades?	<p>ATENEA és un projecte de col·laboració voluntària de les persones usuàries per millorar el seu benestar i qualitat de vida. Les persones usuàries van participar des de l'inici del desenvolupament del pla funcional del projecte.</p> <p>De fet, per poder participar al projecte, les persones usuàries van haver de donar el seu consentiment informat, i exercir el dret a la seva autonomia. Per exercir-lo la persona va ser informada verbalment i per escrit. Posteriorment, aquest consentiment informat inicial va ser validat per les mateixes persones usuàries i se'ls va tornar a informar dels riscos i beneficis de l'ús d'aquest sistema.</p>
	S'ha efectuat, desenvolupat i aplicat una avaluació d'impacte en relació amb qüestions específiques (per exemple, protecció de dades) o algunes característiques del sistema (per exemple, l'ús de la biometria)?	<p>S'ha realitzat l'avaluació d'impacte en protecció de dades personals (AIPD) i compliment de la normativa vigent en matèria de seguretat i protecció de dades personals. Així es recull als convenis de col·laboració amb els territoris pilot.</p>
Secció D Intervenció de les parts interessades i diligència deguda	Quins són els principals grups o comunitats potencialment afectats pel sistema d'IA, inclòs el seu desenvolupament?	<p>Persones més grans de 65 anys que viuen soles.</p>
	Quines parts interessades, a més dels individus o grups que poden veure's afectats pels sistemes d'IA, hi han de participar (per exemple, la societat civil i les organitzacions internacionals, experts, associacions industrials, periodistes)?	<p>Administracions públiques, empreses privades i entitats del tercer sector.</p>

	<p>Hi ha altres titulars d'obligacions que hi hagin de participar, a part del proveïdor i el responsable del desplegament de la IA (per exemple, autoritats nacionals, organismes governamentals)?</p>	<p>Autoritat de protecció de dades, Administracions públiques locals i autonòmiques, comitè ètic d'investigació científica d'una Universitat pública, autoritat de supervisió d'IA.</p>
	<p>Han participat en el procés d'avaluació els socis comercials, inclosos els proveïdors de serveis (per exemple, subcontractistes en sistemes d'IA i conjunts de dades)?</p>	<p>En el procés d'avaluació hi han participat els <i>partners</i> del projecte constituïts com una agrupació d'empreses sense entitat jurídica.</p> <p>Des del primer moment, totes les parts que han participat en el projecte (es tracta d'una col·laboració público-privada), han tingut molt present la necessitat de vetllar pel compliment normatiu, així com per la garantia dels drets fonamentals de la ciutadania que poguessin veure's afectats per la posada en marxa d'aquest projecte. Per aquest motiu, per al projecte ATENEA es va comptar amb la participació d'un comitè extern (universitat pública de Catalunya) que va valorar aquest projecte i li va proporcionar l'acreditació necessària per poder dur-lo a terme.</p> <p>D'altra banda, en els convenis signats amb les administracions públiques participants, es preveuen com a obligació la realització d'avaluacions d'impacte, el seguiment continu, la formació, la transparència i la rendició de comptes d'aquest projecte.</p>
	<p>El proveïdor d'IA ha dut a terme una avaluació de la seva cadena de subministrament, per identificar si les activitats dels proveïdors/contractistes que participen en el desenvolupament de productes/serveis poden afectar els drets fonamentals?</p> <p>El proveïdor ha promogut estàndards o auditories de drets fonamentals, per garantir el</p>	<p>No (no és un requisit legal)</p>

	respecte dels drets fonamentals entre els proveïdors?	
	El proveïdor i el responsable del desplegament de la IA han comunicat públicament les possibles repercussions del sistema d'IA en els drets fonamentals?	No (no és un requisit legal)
	El proveïdor i el responsable del desplegament de la IA han proporcionat formació sobre estàndards de drets fonamentals al personal encarregat de la gestió i la contractació relacionades amb el sistema d'IA?	Només en matèria de protecció de dades

Matrius del risc

Taula 1. Probabilitat

Baixa	El risc de perjudici és improbable o altament improbable
Mitjana	El risc es pot materialitzar
Alta	Hi ha una alta probabilitat que el risc es materialitzi
Molt alta	És molt probable que el risc es materialitzi

Taula 2. Exposició

Baixa	Pocs o molt pocs titulars de drets de la població identificada estan potencialment afectats
Mitjana	Part de la població identificada està potencialment afectada
Alta	La majoria de la població identificada està potencialment afectada
Molt alta	Quasi tota la població identificada està potencialment afectada

Taula 3. Probabilitat (*likelihood*)

		Probabilitat			
		Baixa	Mitjana	Alta	Molt alta
Exposició	Baixa	B	B/M	B/A	B/MA
	Mitjana	M/B	M	M/A	M/MA
	Alta	A/B	A/M	A	A/MA
	Molt alta	MA/B	MA/M	MA/H	MA

Probabilitat			
Baixa	Mitjana	Alta	Molt alta

Taula 4. Gravetat del perjudici

Baixa	Els individus i grups afectats poden patir només perjudicis menors en l'exercici dels seus drets i llibertats
Mitjana	Els individus i grups afectats poden patir perjudicis importants
Alta	Els individus i grups afectats poden enfrontar-se a perjudicis greus
Molt alta	Els individus i grups afectats poden patir perjudicis greus, o fins i tot irreversibles

Taula 5. Esforç per superar el perjudici i revertir els efectes adversos

Baix	Els perjudicis soferts es poden superar sense problemes (com ara temps dedicat a modificar la informació, molèsties, irritacions, etc.)
Mitjà	Els perjudicis soferts es poden superar malgrat algunes dificultats (com ara costos addicionals, por, incomprensió, estrès, petits dolors físics, etc.)
Alt	Els perjudicis soferts es poden superar, tot i que amb dificultats serioses (com ara, pèrdues econòmiques, danys materials, empitjorament de la salut, etc.)
Molt alt	Els perjudicis soferts poden no superar-se (per exemple, dolors psicològics o físics de llarga durada, mort, etc.)

Taula 6. Gravetat (severity)

		Gravetat			
		Baixa	Mitjana	Alta	Molt alta
Esforç	Baix	B	B/M	B/A	B/MA
	Mitjà	M/B	M	M/A	M/MA
	Alt	A/B	A/M	A	A/MA
	Molt alt	MA/B	MA/M	MA/A	MA

Gravetat			
Baix	Mitjà	Alt	Molt alt

Taula 1A. Recollida de dades i anàlisi dels riscos

Drets/llibertats potencialment afectats	Descripció de l'impacte	Probabilitat			Gravetat		
		Probabilitat de resultats adversos	Exposició	Probabilitat	Gravetat	Esforç	Gravetat
Dret a la protecció de dades	El desenvolupament del sistema d'IA es basa en l'ús del reconeixement de la veu. En aquest sentit, la IA s'utilitza per elaborar/enviar les sol·licituds de l'usuari	[Mitjana] El projecte està subjecte a avaluacions específiques d'impacte ètic i de protecció de dades.	[Molt alta] L'impacte afecta potencialment tots els individus als	[Alta]	[Alta] Els riscos associats a l'ús de la biometria són limitats a causa de la naturalesa opcional de la	[Mitjà] La recollida i el tractament il·lícit de dades es poden detectar i aturar, suprimint la informació que s'hagi	[Mitjana]

	<p>(de veu a text) a altres tecnologies basades en IA que executaran la tasca sol·licitada.</p> <p>ATENEA escolta totes les converses en espera del comandament wake up word “Hola ATENEA”. Aquest procés d’escolta es manté únicament al dispositiu (tauleta), no es processa ni emmagatzema al núvol.</p> <p>El comandament “Hola ATENEA” activa el servei i la seva comunicació al núvol, on es processa i emmagatzema la conversa per entrenar l’algoritme.</p> <p>Totes les converses de l’usuari relacionades amb un servei específic accedit a partir d’ATENEA, per</p>	<p>Les persones usuàries han de tenir un nivell cognitiu adequat per prestar el seu consentiment informant la utilització d’aquest sistema.</p> <p>El projecte ha estat avaluat per un comitè d’ètica reconegut d’una universitat pública de Catalunya i el sistema compleix amb les mesures de seguretat establertes en les normatives tècniques vigents a Espanya (ENS, normes UNE...) segons l’informe elaborat per l’Agència de Ciberseguretat de Catalunya</p>	<p>quals s’aplica l’algoritme</p>		<p>identificació biomètrica; La identificació biomètrica pot ser una solució proporcionada per a persones grans amb mobilitat reduïda. Hi ha el risc que les interaccions de veu es puguin utilitzar per a la creació de perfils.</p> <p>El principal risc és el tractament il·lícit de dades, incloses categories especials de dades relacionades amb la situació vulnerable dels usuaris, que pot ser invasiu i afectar la privacitat de les persones.</p>	<p>recollit il·legalment</p>	
--	--	---	-----------------------------------	--	--	------------------------------	--

	<p>exemple, trucades d'emergència o altres trucades, es mantenen privades.</p> <p>La biometria per veu només s'utilitza per a identificar la persona, i s'activa sobre la base de la decisió de la persona usuària.</p> <p>La identificació es pot basar en biometria de veu o en la tradicional combinació d'usuari i contrasenya.</p> <p>Qualsevol operació de tractament que no compleixi la normativa aplicable en matèria de protecció de dades personals podria afectar a aquest dret.</p>						
No discriminació	El mecanisme d'IA funciona amb tecnologia de la parla, per la qual cosa és possible que es produeixin discriminacions quan l'assistent de veu no	[Alta] Possibles problemes en la veu i la parla de les persones usuàries poden generar que quedin excloses	[Mitjana] Hi haurà un nombre limitat de persones amb problemes de veu o d'idioma	[Mitjana]	[Molt alta] Impacte negatiu en la igualtat d'accés a la assistència que ofereix la solució i	[Alt] Caldria adaptar l'assistent de veu per a integrar les casuístiques de	[Molt alta]

	<p>entengui el parlant per algun problema de comunicació (per exemple, impediment o deteriorament de la parla).</p> <p>Es poden produir discriminacions quan les persones usuàries no parlin correctament l'idioma del sistema (CAT/ES).</p>				<p>en la qualitat del servei</p>	<p>problemes en la parla o l'idioma.</p> <p>Les persones usuàries han de disposar de cert nivell cognitiu. Es podria excloure també per a l'ús les persones que tinguin algun tipus de problema en la parla o en l'idioma.</p>	
Dret a la salut	<p>Un funcionament inadequat del sistema pot provocar que no es garanteixi adequadament aquest dret</p>	<p>[Baix]</p> <p>El projecte està subjecte a avaluacions contínues i a un seguiment específic per a detectar possibles errors en el funcionament. Només en alguns casos el funcionament inadequat pot afectar</p>	<p>[Molt alta]</p> <p>L'impacte afecta potencialment totes les persones a les quals s'aplica l'algoritme</p>	<p>[Molt Alta]</p> <p>L'impacte afecta potencialment totes les persones a les quals s'aplica l'algoritme</p>	<p>[Alta]</p> <p>No poder accedir als serveis de salut quan es necessita pot causar un greu perjudici per a usuaris que potencialment estan més exposades a situacions de risc</p>	<p>[Mitjà]</p> <p>Aquest sistema no és l'únic al qual poden recórrer les persones en cas de necessitat, ja que hi ha altres canals d'accés a sistemes sanitaris (així com a altres serveis), tenint en compte que</p>	<p>[Mitjana]</p>

		significativament al dret a la salut (per exemple, trucada d'emergència).				els usuaris potencials han de tenir un nivell cognitiu determinat	
Dret a l'assistència social	Un funcionament inadequat del sistema pot provocar que no es garanteixi adequadament aquest dret	<p>[Baix]</p> <p>El projecte està subjecte a avaluacions contínues i a un seguiment específic per detectar possibles errors en el funcionament.</p> <p>Només en alguns casos d'ús el funcionament inadequat pot afectar aquest dret.</p>	<p>[Molt alt]</p> <p>L'impacte afecta potencialment totes les persones a les quals s'aplica l'algoritme</p>	[Baix]	<p>[Alt]</p> <p>No poder accedir als serveis d'assistència social quan es necessita pot causar un greu perjudici per a aquestes persones que potencialment estan més exposats a situacions de risc.</p>	<p>[Baix]</p> <p>Aquest sistema no és l'únic al qual poden recórrer les persones, ja que hi ha altres canals d'accés a serveis d'assistència social, tenint en compte que els usuaris potencials han de tenir un grau cognitiu determinat. A més, aquestes persones compten amb una persona de contacte dels serveis socials que fa un seguiment de la seva situació, i</p>	[Mitjana]

						no es tracta de serveis d'emergència.	
Dret d'accés als serveis d'interès econòmic general	Aquest dret es podria veure vulnerat si totes les persones no poden tenir accés a aquest sistema per falta de recursos de les administracions públiques	[Mitjana] El projecte es troba en fase pilot amb finançament públic provinents dels fons Next Generation.	[Mitjana] L'impacte afecta potencialment totes les persones que no poden beneficiar-se encara d'aquest projecte, per falta de recursos o d'informació.	[Mitjana]	[Mitjana] No poder accedir a aquest sistema bé per falta d'informació o de mitjans, pot vulnerar la igualtat d'oportunitats i la cohesió social del territori.	[Mitjà] Es tractaria de posar prou recursos perquè totes les persones susceptibles de disposar d'aquest sistema hi puguin accedir	[Mitjana]

Taula 7. Impacte general del risc

		Gravetat			
		Baixa	Mitjana	Alta	
Probabilitat	Baixa				
	Mitjana				
	Alta				
	Molt alta				

Taula 2A. Gestió de riscos (I)

Dret afectat/llibertat afectat	Probabilitat	Gravetat	Impacte global	Mesures per a prevenir/mitigar els impactes
Dret a la protecció de dades	[Alta]	[Mitjana]	[Alta]	<ul style="list-style-type: none"> Publicar la informació sobre els procediments utilitzats per a obtenir i processar les dades originals utilitzades per a entrenar el sistema d'IA (transparència algorítmica i registre públic d'aquests algoritmes) La identificació biomètrica mitjançant la veu és opcional i es basa en el consentiment previ. Aquesta identificació no s'utilitzarà per a altra finalitat, inclosa la identificació compartida amb tercers.

				<ul style="list-style-type: none"> • Excloure qualsevol activitat d'elaboració de perfils basada en l'ús de serveis accessibles a través d'ATENEA. • Implementar i mantenir d'un sistema de gestió de riscos i de la seguretat durant tot el cicle de vida del sistema. • Garantir la qualitat de les dades d'entrenament, validació i prova mitjançant practiques adequades de governança i gestió de dades. • <i>Accountability</i>: elaborar i mantenir actualitzada la documentació tècnica pel sistema. • Permetre una supervisió humana efectiva durant el seu ús, amb personal capacitat. • Garantir l'exercici dels drets de les persones usuàries.
No discriminació	[Mitjana]	[Molt Alta]	[Alta]	<ul style="list-style-type: none"> • Entrenar l'assistent de veu per evitar algunes de les situacions de discriminació descrites. • Restringir les persones usuàries potencials per evitar aquestes situacions. • Implantar i mantenir un sistema de gestió de riscos i de la seguretat durant tot el cicle de vida del sistema. • Garantir la qualitat de les dades d'entrenament, validació i prova mitjançant practiques adequades de governança i gestió de dades.

				<ul style="list-style-type: none"> ▪ <i>Accountability</i>: elaborar i mantenir actualitzada la documentació tècnica pel sistema. ▪ Permetre una supervisió humana efectiva durant el seu ús, amb personal capacitat i proves periòdiques d'eficàcia del sistema en els casos crítics. ▪ Garantir l'exercici dels drets de les persones usuàries.
Cura de la salut	[Baixa]	[Mitjana]	[Mitjana]	<ul style="list-style-type: none"> ▪ Garantir l'assistència tècnica adequada i millorar la qualitat tècnica del sistema. ▪ Informar de tots els canals que tenen a la seva disposició aquestes persones per accedir a cura de la salut.
Assistència social	[Baixa]	[Mitjana]	[Mitjana]	<ul style="list-style-type: none"> ▪ Garantir l'assistència tècnica adequada i millorar la qualitat tècnica del sistema. ▪ Informar de tots els canals que tenen a la seva disposició aquestes persones per accedir a assistència social.
Accés als serveis d'interès econòmic general	[Mitjana]	[Mitjana]	[Mitjana]	<ul style="list-style-type: none"> ▪ Garantir els recursos i mitjans suficients perquè totes les persones usuàries potencials tinguin accés a aquest servei .

Taula 3A. Gestió de riscos (II)

Dret afectat/llibertat afectada	Probabilitat (residual)	Gravetat (residual)	Impacte residual
Protecció de dades/Privacitat	[Baixa] La probabilitat que es produeixi un impacte sobre aquest dret (probabilitat de resultats adversos) ha disminuït a baixa, ja que les mesures adoptades han limitat la possibilitat de conseqüències negatives.	[Mitjana] La gravetat de l'impacte ha disminuït a mitjana, ja que les mesures adoptades han restringit l'àmbit de les conseqüències negatives que podrien produir-se. El nivell de gravetat continua essent mitja, però d'un nivell més baix que l'inicial.	[Mitjana]
No discriminació	[Baixa] La probabilitat s'ha reduït a baixa perquè la probabilitat d'efectes adversos s'ha reduït a mitjana, a causa de la disminució dels casos possibles de funcionament inadequat, i l'exposició s'ha reduït a baixa, a causa de la reducció del nombre de persones afectades (s'han identificat algunes limitacions d'ús relacionades amb el coneixement d'idiomes).	[Mitjana] El nivell de gravetat s'ha reduït de molt alt a mitjà gràcies a la implementació de mesures complementàries/alternatives per accedir als serveis (reduint la gravetat de l'impacte de molt alt a alt) i l'esforç menor requerit per reaccionar degut a les mesures ja implementades (reduint l'esforç d'alt a mitjà).	[Mitjana]
Cura de la salut	[Baixa]	El nivell de gravetat ha disminuït, a causa de les mesures complementàries/alternatives per accedir als serveis (reducció de la gravetat de l'impacte de molt alt a alt) i al reduït esforç necessari per reaccionar degut a les mesures ja implementades (reducció d'esforç de mitjà a baix). La gravetat segueix sent mitjana, però d'un nivell més baix que l'inicial.	[Mitjana]

Assistència social	[Baixa]	<p>[Mitjana]</p> <p>La gravetat ha disminuït, a causa de les mesures complementàries/alternatives per a accedir als serveis (reducció de la gravetat de l'impacte de molt alt a alt) i al reduir l'esforç necessari per reaccionar gràcies a les mesures ja implementades (reducció d'esforç de mitjà a baix). La gravetat segueix sent mitjana, però d'un nivell més baix que l'inicial.</p>	[Mitjana]
Accés als serveis d'interès econòmic general	[Mitjana]	<p>[Mitjana]</p> <p>Es manté el nivell de gravetat ja que garantir aquest dret depèn dels recursos que els governs hi destinin.</p>	[Mitjana]

4. Comentaris

L'aparició de noves tecnologies no només està causant una revolució en la manera en què treballem i prestem serveis públics, sinó també en la necessitat de repensar, des de l'etapa de disseny, l'impacte d'aquestes aplicacions en els drets fonamentals dels ciutadans. De fet, aquest canvi de paradigma ja es va fer evident amb l'entrada en vigor del Reglament general de protecció de dades (RGPD) i ara està guanyant impuls.

Una metodologia per a dur a terme una avaluació d'impacte sobre els drets fonamentals (AIDF) és una eina essencial per identificar els riscos i les mesures organitzatives i tècniques a aplicar, ha de tenir un enfocament transversal que ha d'integrar en els equips de treball multidisciplinaris responsables de la implementació de la IA a les administracions públiques. Amb aquesta finalitat, com ja s'ha observat amb l'aplicació de l'RGPD, és absolutament necessari que, des del primer moment, els delegats de protecció de dades s'impliquin en les estratègies i projectes de transformació digital a implementar a cada organització. D'aquesta manera els riscos es poden analitzar des del principi i per defecte, i es poden prendre les mesures tècniques i organitzatives necessàries.

D'altra banda, també es recomanable revisar les descripcions de llocs de treball per incloure aquest tipus de competències i responsabilitats, invertir en formació i desenvolupar perfils professionals adequats. A més, en aquest moment de transició cap a "l'administració intel·ligent" (basada en l'ús de la IA), s'hauria de desenvolupar algun tipus de guia pràctica com a instrucció interna, de manera que tothom a les organitzacions en sigui conscient, incloses les directrius sobre les millors pràctiques en l'ús de la IA en l'activitat administrativa.

