RCDP Revista Catalana de Dret Públic

#71

www.rcdp.cat

L'AVALUACIÓ D'IMPACTE EN ELS DRETS FONAMENTALS DE L'ÚS D'INTEL·LIGÈNCIA ARTIFICIAL EN EL SECTOR PÚBLIC: MODELS I METODOLOGIES EN PERSPECTIVA COMPARADA*

Pere Simón Castellano**

Resum

L'article examina l'aparició i la regulació de l'avaluació d'impacte en els drets fonamentals (AIDF) en matèria d'intel·ligència artificial dins el marc normatiu europeu, especialment arran de l'aprovació del Reglament d'intel·ligència artificial (RIA). Primer, es contextualitza la necessitat d'aquest instrument *ex ante*, que amplia l'àmbit de les avaluacions de protecció de dades del Reglament general de protecció de dades (RGPD) per comprendre tots els drets consagrats a la Carta de la UE. Seguidament, s'analitza l'abast de l'obligació legal de l'article 27 del RIA –sistemes d'IA d'alt risc–, els subjectes obligats –organismes de dret públic i entitats privades que presten serveis públics– i el contingut mínim que han d'incloure les AIDF –descripció de processos, període d'ús, col·lectius afectats, riscos, garanties tècniques i mesures de governança. A continuació, es comparen la guia pionera de l'Autoritat Catalana de Protecció de Dades, l'eina neerlandesa AIIA 2.0 i les directrius italianes d'AgID posant de manifest enfocaments pràctics, matrius de risc i models participatius per verificar impactes i mitigar-los. Finalment, s'identifiquen reptes d'implementació al sector públic –capacitat tècnica, integració organitzativa, supervisió i participació ciutadana– i s'apunta la necessitat d'un marc metodològic estandarditzat i de recursos per a una governança proactiva i multidimensional de la IA. Aquest treball ofereix, doncs, una visió jurídica i pràctica de l'AIDF com a pilar d'una IA ètica i de confiança al servei de la ciutadania.

Paraules clau: intel·ligència artificial; drets fonamentals; avaluació d'impacte; dret de la Unió Europea; governança pública.

FUNDAMENTAL RIGHTS IMPACT ASSESSMENT FOR THE USE OF ARTIFICIAL INTELLIGENCE IN THE PUBLIC SECTOR: A COMPARATIVE APPROACH TO MODELS AND METHODOLOGIES

Abstract

This article examines the emergence and regulation of the fundamental rights impact assessment (FRIA) for artificial intelligence within the European regulatory framework, especially as a result of the adoption of the Artificial Intelligence Act (AI Act). First, it contextualises the need for this ex-ante instrument, which extends the scope of the data protection impact assessments of the General Data Protection Regulation (GDPR) in order to include all the rights enshrined in the EU Charter. It then analyses the scope of the legal obligation of Article 27 of the AI act (high-risk AI systems), the liable parties (bodies governed by public law and private bodies providing public services) and the minimum content that the FRIAs have to contain (description of processes, period of use, groups affected, risks, technical guarantees, and governance measures). Next, it compares the pioneering guide of the Catalan Data Protection Authority, the Dutch AIIA 2.0 tool, and the Italian AgID guidelines, indicating practical approaches, risk matrices, and participatory models to check for impacts and mitigate them. Finally, it identifies implementation challenges in the public sector (technical capacity, organisational integration, supervision, and citizen participation) and points out the need for a standardised methodological and resource framework for proactive and multidimensional governance of AI. The paper thus offers a legal and practical approach to the FRIA as a pillar of ethical AI and of trust in the service of citizens.

Keywords: artificial intelligence; fundamental rights; impact assessment; European Union law; public governance.

---

* Aquest treball s'emmarca dins del projecte "Derechos y garantías públicas frente a las decisiones automatizadas y el sesgo y discriminación algorítmicas (2023-2025) (PID2022-136439OB-I00)", finançat pel MCIN/AEI/10.13039/501100011033/ i FEDER.

** Pere Simón Castellano, professor titular de dret constitucional de la Universitat Internacional de La Rioja (UNIR). pere.simon@unir.net 0003-1722-6498

Recepció de l'article: 13.06.2025. Avaluacions: 18.07.2025 i 28.07.2025. Acceptació de la versió final: 18.09.2025.

Citació recomanada: Simón Castellano, Pere. (2025). L'avaluació d'impacte en els drets fonamentals de l'ús d'intel·ligència artificial en el sector públic: models i metodologies en perspectiva comparada. *Revista Catalana de Dret Públic*, *71*, 4-24. https://doi.org/10.58992/rcdp.i71.2025.4486

Sumari

1 Introducció

L'ús creixent de sistemes d'intel·ligència artificial (en endavant, IA) per part de les administracions públiques planteja nous reptes des de l'òptica de l'afectació i la tutela dels drets i les llibertats fonamentals. La gestió algorítmica i el seu ús potencial en el marc de la presa de decisions públiques –des de la distribució de recursos socials fins a l'avaluació de perfils de risc– pot comportar beneficis notables en eficiència i objectivitat. Tanmateix, també pot generar impactes negatius: discriminació indirecta, manca de transparència, opacitat decisional o vulneracions dels drets de l'esfera privada, entre d'altres (Cotino Hueso, 2025, p. 124-127). Davant d'aquesta realitat, el legislador europeu, després d'una tramitació parlamentària llarga i sinuosa, marcada per nombrosos girs i la pressió dels *lobbies* tecnològics a Brussel·les, ha aconseguit aprovar una normativa, el Reglament d'intel·ligència artificial (RIA),[1] que a través de les avaluacions d'impacte tracta de garantir que el progrés tecnològic sigui compatible amb el respecte dels drets fonamentals, com a resposta també a certs plantejaments doctrinals (Mantelero, 2022; Simón Castellano, 2023) que tindrien el seu reflex en la carta que diferents experts sobre la matèria, del Brussels Privacy Hub (2023), varen remetre a les institucions europees en el marc del procés d'elaboració del RIA.

La idea d'avaluar l'impacte potencial d'una nova tecnologia té precedents en els àmbits ambiental, social i de drets humans (Stahl et al., 2023). En l'àmbit de la privacitat i la protecció de dades, les avaluacions d'impacte ja s'han consolidat com una eina clau arran del Reglament general de protecció de dades de la Unió Europea (RGPD).[2] Ara bé, la complexitat i l'autonomia creixent dels sistemes d'IA han evidenciat les limitacions d'aquests instruments tradicionals enfocats sobretot en la protecció de dades, ja que les conseqüències de la IA s'estenen a un ventall molt més ampli de drets i interessos (Mantelero, 2022).

En aquest context, sorgeix amb força la figura de l'avaluació d'impacte en els drets fonamentals (en endavant, AIDF) en matèria d'IA, que també troba suport en la metodologia HUDERIA elaborada pel Comitè d'Intel·ligència Artificial del Consell d'Europa,[3] que defineix un marc basat en el risc i l'avaluació continuada de l'impacte dels sistemes d'IA sobre els drets humans, la democràcia i l'estat de dret (Consell d'Europa - CAI, 2024). Es tracta d'un instrument *ex ante*, destinat a identificar i mitigar els riscos que un sistema algorítmic pot suposar per als drets humans (discriminació, afectació a la privacitat, llibertat, seguretat, etc.) abans de la seva posada en marxa. Aquesta eina de governança algorítmica suposa un canvi de paradigma en la regulació: desplaça la responsabilitat de ponderar i prevenir impactes cap als desenvolupadors i, especialment, cap als responsables del desplegament, això és, en l'àmbit públic, les mateixes administracions (Simón Castellano, 2023). Ens trobem, per tant, davant d'un desafiament i alhora una oportunitat per reforçar les garanties de la IA en el sector públic tot integrant principis de transparència, traçabilitat i rendició de comptes en el cicle de vida dels sistemes algorítmics.

L'objectiu d'aquest article és analitzar, des d'una perspectiva jurídica rigorosa, l'abast de l'obligació continguda en el RIA i el règim jurídic de l'AIDF en l'àmbit del sector públic. S'examinarà el marc normatiu europeu resultant,

1 Reglament (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 es modifiquen diversos reglaments i directives. (DOUE, núm. 1689, 12.07.2024, p. 1-144)

2 Reglament (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). (DOUE, núm. 119, 04.05.2016, p. 1-88).

3 A banda del RIA, resulta indispensable fer referència al Conveni marc sobre IA del Consell d'Europa, un text extens que comprèn setze considerants i trenta-sis articles distribuïts en vuit capítols. Representa un esforç significatiu per harmonitzar i establir un marc mínim comú a Europa, amb una ambició global centrada en els drets humans, la democràcia i l'estat de dret. Destaca el seu enfocament basat en el risc, especialment a l'article 16, que exigeix l'avaluació i la mitigació contínua dels riscos i impactes adversos de qualsevol sistema d'IA. No obstant això, en el text del Conveni no hi figuren referències explícites a les avaluacions d'impacte en drets fonamentals o AIDF, tal com sí que les introdueix el RIA. Aquesta llacuna ha estat coberta pel Comitè d'Intel·ligència Artificial (CAI) del mateix Consell d'Europa mitjançant l'aprovació de la metodologia HUDERIA (Consell d'Europa - CAI, 2024).

així com les infructuoses propostes i esmenes del Parlament Europeu, i es concretarà l'obligació de realitzar una avaluació d'impacte algorítmic i les seves implicacions per a les administracions. Més enllà del que hauria pogut ser, i del que finalment és (el *què*, subjectes obligats i contingut de l'obligació), s'estudiarà i analitzarà el *com*. Així, es compararan diverses metodologies i pràctiques emergents en l'elaboració d'AIDF: des del model pioner desenvolupat a Catalunya per l'Autoritat Catalana de Protecció de Dades (en endavant, APDCAT) fins a experiències en altres jurisdiccions com els Països Baixos i Itàlia. Sobre aquesta base, s'identificaran els principals reptes d'implementació al sector públic, tenint en compte la necessitat de recursos especialitzats, de coordinació normativa i de cultura organitzativa orientada a l'ètica i els drets humans.

D'aquesta manera es pretén oferir una visió completa i actual de l'estat de la qüestió sobre l'avaluació d'impacte en drets fonamentals per a la IA pública i destacar-ne l'enfocament innovador i les incògnites que cal resoldre. Està el sector públic preparat per assumir aquest nou estàndard d'auditoria preventiva (*due diligence*) tecnològica? La resposta a aquesta pregunta condicionarà l'èxit de la implantació d'una IA ètica i digna de confiança al servei de la ciutadania.

2 Marc normatiu de les avaluacions d'impacte algorítmic

2.1 De l'oblit inicial del legislador europeu a l'obligació de l'article 27 del RIA

El RIA ha establert per primera vegada en l'ordenament comunitari l'obligació de realitzar una AIDF en relació amb determinats sistemes d'IA. Aquesta obligació es recull a l'article 27 del RIA i s'inscriu dins l'enfocament basat en el risc que articula tota la regulació. Des del primer considerant, el RIA inclou la idea de promoure una IA de confiança i centrada en l'ésser humà per garantir un elevat nivell de protecció de la salut, la seguretat i els drets fonamentals consagrats a la Carta[4] (vegeu també, sobre la idea d'un RIA centrat en l'ésser humà més enllà de la supervisió humana com a garantia específica dels sistemes d'IA d'alt risc, la contribució de Lazcoz Moratinos [2024, p. 697-699]). Consegüentment, moltes de les obligacions previstes –inclosa l'AIDF– es dissenyen amb la finalitat de materialitzar aquests objectius. La referència als drets fonamentals és constant tant als considerants com a l'articulat del RIA, cosa que evidencia la voluntat del legislador europeu de prevenir els riscos socials i jurídics associats a la IA, més enllà dels purament tècnics o de seguretat de productes i components.

Inicialment, en la primera proposta de RIA presentada per la Comissió Europea[5] no es feia cap referència explícita a l'AIDF, i mancaven al llarg de tot el text referències a l'impacte potencial en els drets fonamentals. L'única referència –i molt parcial– era la relativa a les avaluacions d'impacte en protecció de dades, en el marc de les obligacions dels usuaris de sistemes d'IA d'alt risc (a l'article 29.6 de la proposta inicial de la Comissió Europea). La situació va canviar arran de les esmenes proposades pel Parlament Europeu[6] que foren formalment aprovades en la sessió plenària el juny de 2023 i que incorporaren l'obligació de dur a terme una AIDF per a tots els sistemes d'IA d'alt risc, una visió molt ambiciosa que, no obstant això, va quedar finalment molt reduïda o descafeïnada, limitada pràcticament en exclusiva, com es veurà amb detall *infra*, al sector públic.

4 Carta dels Drets Fonamentals de la Unió Europea (2000/C 364/01). (DOUE, núm. 364, 18.12.2000).

5 Proposta de Reglament del Parlament Europeu i del Consell pel qual s'estableixen normes harmonitzades en matèria d'intel·ligència artificial (Llei d'intel·ligència artificial) i es modifiquen determinats actes legislatius de la Unió (COM(2021) 206 final), Brussel·les, 21 d'abril de 2021.

6 Resolució legislativa del Parlament Europeu, de 14 de juny de 2023, sobre la proposta de Reglament pel qual s'estableixen normes harmonitzades en matèria d'intel·ligència artificial (COM(2021) 206 final). Esmenes aprovades en primera lectura (P9_TA(2023)0238), Brussel·les, 14 de juny de 2023.

El resultat final, recollit a l'article 27 del RIA, institucionalitza les AIDF però en delimita l'abast: obliga únicament els organismes de dret públic i les entitats privades que presten serveis públics, així com els responsables que apliquin IA d'alt risc en solvència creditícia i assegurances de vida o salut. A més, restringeix l'obligació al primer ús i permet reutilitzar avaluacions prèvies "en casos similars",[7] fet que ha generat dubtes sobre el risc de trobar-nos davant d'una obligació de compliment normatiu (*compliance*) cosmètic davant la falta d'exigència o necessitat de garantir revisions periòdiques que reflecteixin l'evolució dels sistemes d'IA.[8]

2.2 Àmbit d'aplicació: subjectes obligats

La norma distingeix tres grups de responsables del desplegament[9] obligats a realitzar l'AIDF. En primer lloc, els organismes de dret públic que introdueixin i utilitzin sistemes d'IA d'alt risc[10] en qualsevol dels àmbits de l'annex III (educació, assistència sanitària, serveis socials, migració, justícia, etc.). En segon lloc, les entitats privades que presten serveis públics –centres sanitaris concertats, universitats privades amb subvenció pública, empreses de gestió social, etc.–, atès que assumeixen funcions d'interès general. En tercer lloc, els responsables del desplegament que utilitzen IA d'alt risc per avaluar la solvència de persones físiques o fixar preus i riscos en assegurances de vida i salut (annex III.5.*b* i *c*), independentment de la seva naturalesa pública o privada.

Aquesta delimitació fa recaure el pes de les AIDF sobretot sobre actors que interactuen directament amb drets bàsics de la ciutadania i, en especial, projecta els seus efectes vinculants sobre el sector públic. Tanmateix, l'exempció de la resta de sectors privats i la possibilitat de recórrer a les AIDF elaborades prèviament pel proveïdor plantegen reptes de coherència reguladora i de supervisió efectiva. En aquest sentit, el paper de l'Oficina Europea d'IA[11] –que haurà d'elaborar un model o plataforma per simplificar

7 L'article 27.2 explicita que l'obligació és aplicable "al primer ús del sistema d'IA d'alt risc" i que "en casos similars, el responsable del desplegament podrà basar-se en avaluacions d'impacte relatives als drets fonamentals realitzades prèviament o a avaluacions d'impacte existents realitzades pels proveïdors".

8 Un risc que dependrà de la interpretació que l'Oficina Europea d'IA realitzi de l'article 27.2 del RIA i, més concretament, de la part final, que sembla que és contradictòria amb la part inicial quan indica que "si, durant l'ús del sistema d'IA d'alt risc, el responsable del desplegament considera que alguns dels elements enumerats en l'apartat 1 han canviat o han deixat d'estar actualitzats, ha d'adoptar les mesures necessàries per actualitzar la informació". La terminologia emprada no sembla la més encertada, en la mesura que les AIDF efectivament necessiten estar adaptades al context d'ús de la IA i al context organitzacional, però això no exigeix una mera "actualització de la informació", sinó una revisió periòdica de la validesa del context, l'estructura, la naturalesa i les finalitats tant de la tecnologia com del seu ús concret, que pot, al seu torn i un cop actualitzats, esdevenir en resultats o nivells de risc inherents i residuals totalment diferents dels que es varen projectar en AIDF anteriors. La millora contínua és un principi fonamental de les avaluacions de risc i dels sistemes de gestió de riscos, tal com ha recordat la doctrina precisament en l'àmbit de les AIDF (Simón Castellano, 2023, p. 74-75), si bé sembla que aquesta és una qüestió que no ha quedat ben resolta en el redactat de l'article 27 del RIA i obre la porta, en el pitjor dels escenaris, al models de compliment normatiu cosmètic als quals fèiem referència al text del cos de l'article.

9 El terme *responsable del desplegament* (*deployer* en la terminologia original i *usuaris* en el marc de la proposta inicial de la Comissió Europea) fa referència a l'entitat que posa en funcionament el sistema en un context real, que en l'àmbit públic normalment serà l'òrgan o l'administració que utilitza l'algoritme per a la seva activitat. En aquest sentit, la norma especifica alguns supòsits concrets en què l'obligació s'activa: quan els responsables del desplegament siguin entitats de dret públic (administracions i organismes públics) o entitats privades que exerceixin funcions públiques, i també en certs sectors privats d'especial sensibilitat.

10 El RIA no ha optat per requerir una AIDF en tots els casos d'ús de la IA, sinó només en aquells que considera d'alt risc segons la classificació de riscos que estableix al capítol II. A grans trets, els sistemes d'IA d'alt risc són els enumerats a l'annex III del RIA, que inclou diverses categories vinculades majoritàriament a serveis essencials o decisions que afecten drets de les persones (per exemple, sistemes d'IA per a l'accés a l'educació o la formació professional, la selecció i la gestió de personal, l'avaluació de solvència creditícia, sistemes de puntuació social, sistemes d'identificació biomètrica en espais públics, determinats dispositius de seguretat en productes, etc.). A més de les categories de l'annex III, també es consideren d'alt risc aquells sistemes addicionals que es puguin designar seguint els criteris de l'article 7 del RIA en el futur.

11 Vegeu la Decisió de la Comissió (UE) 2024/1459, de 24 de gener de 2024, per la qual s'estableix l'Oficina Europea d'Intel·ligència Artificial. (DOUE, núm. 1459, 14.02.2024, p. 1-5).

el procés d'elaboració d'AIDF i centralitzar les notificacions dels seus resultats–[12] serà determinant per harmonitzar criteris i evitar disfuncions en l'aplicació pràctica de l'article 27 del RIA.

Mentre esperem aquests models i criteris per part de l'Oficina Europea d'IA o fins i tot els que pugui establir l'Agència Espanyola de Supervisió de la IA[13] (en endavant, AESIA), les administracions i els organismes públics responsables del desplegament de sistemes d'IA d'alt risc estan obligats a realitzar l'AIDF corresponent (Pedrós Ramos, 2025). Això significa que, per exemple, un ajuntament que vulgui implantar un algoritme per repartir recursos d'ajuda social o un servei de salut que introdueixi un sistema d'IA per triar l'ordre dels pacients en llistes d'espera haurien de realitzar prèviament l'AIDF, i no desplegar ni autoritzar l'eina basada en IA si els riscos no estan dins del llindar del que es considera acceptable o tolerable com a resultat de l'avaluació.

D'aquesta manera, el RIA combina un criteri subjectiu (naturalesa pública del subjecte que utilitza o desplega la tecnologia, o privada però exercint funció pública) i un d'objectiu (àmbits de crèdit i assegurances) per determinar qui està obligat a portar a terme una AIDF, i cobreix sectors en què l'impacte social de la IA es considera especialment crític.

És rellevant subratllar que aquesta obligació legal comporta una intensificació del principi de transparència i de responsabilitat en l'àmbit públic. L'article 27.3 del RIA estableix que el responsable del desplegament ha de notificar els resultats de l'AIDF a l'autoritat de vigilància del mercat mitjançant un model que ha d'emplenar, i res no impedeix que bona part del resum de resultats de l'AIDF i de l'eina d'IA en qüestió siguin compartits amb la ciutadania. De la mateixa manera, l'article 26.9 del RIA enllaça l'AIDF amb la transparència i estableix que els responsables del desplegament han d'utilitzar la informació recopilada en l'AIDF per complir també l'obligació de realitzar l'AIPD del RGPD, en cas que correspongui.

2.3 El contingut mínim de l'obligació

No existeix un model únic o universal d'AIDF, però qualsevol avaluació parteix de l'anàlisi del context organitzacional, l'abast i les finalitats de la tecnologia, la naturalesa de l'entitat que la implementa, el cicle de vida del sistema i els col·lectius potencialment afectats. A partir d'aquests elements, es detecten i es valoren els riscos pels drets i les llibertats, s'estableixen mesures de control i tractament, i s'assignen tasques i terminis d'implementació a responsables específics, d'acord amb un principi de millora contínua (Simón Castellano, 2023, p. 67-76, i Chaveli Donet, 2024, p. 519-528; vegeu també, sobre el contingut de l'avaluació, Menéndez Sebastián [2024, p. 204-206]).

El RIA defineix les AIDF a l'article 27 com una anàlisi estructurada i sistemàtica per identificar, avaluar i mitigar els riscos que els sistemes d'alt risc poden generar sobre els drets fonamentals. El contingut mínim inclou sis elements: (a) descripció detallada dels processos en què s'emprarà la IA; (b) període i freqüència d'ús prevists; (c) categories específiques de persones i col·lectius potencialment afectats; (d) identificació i valoració dels riscos de perjudici; (e) avaluació de la informació facilitada pel proveïdor i de les garanties tècniques, i (f) definició de mesures de control –supervisió humana, canals de reclamació i mecanismes de governança interna– que s'activaran si els riscos es materialitzen.

L'objectiu de les AIDF és que el responsable del desplegament identifiqui els riscos específics que el sistema d'IA d'alt risc pot generar per als drets de les persones o dels col·lectius que previsiblement en resultaran

12 L'article 27.5 del RIA estableix en sentit literal aquest mandat quan indica que l'Oficina Europea d'IA ha d'elaborar un model de "qüestionari, també mitjançant una eina automatitzada, a fi de facilitar que els responsables del desplegament compleixin amb les seves obligacions en virtut d'aquest article de manera simplificada".

13 Vegeu el Reial decret 729/2023, de 22 d'agost, pel qual s'aprova l'Estatut de l'Agència Espanyola de Supervisió de la Intel·ligència Artificial. (BOE, núm. 210, 02.09.2023, p. 122289-122316).

afectats i que defineixi, així mateix, les mesures que caldrà adoptar si aquests riscos es materialitzen. L'avaluació ha de delimitar els processos del responsable en què es farà servir el sistema d'IA d'acord amb la finalitat prevista i ha d'incloure una descripció del període i de la freqüència d'ús, així com de les categories concretes de persones físiques i de col·lectius que probablement es puguin veure afectats en aquell context específic. Igualment, ha de determinar els riscos de perjudici que presumiblement poden incidir sobre els seus drets fonamentals.

En vista dels riscos detectats, el responsable del desplegament ha d'establir les mesures que s'aplicaran si aquests riscos esdevenen realitat: mecanismes de governança adequats al context d'ús, sistemes de supervisió humana de conformitat amb les instruccions d'ús (vegeu, pel que fa al cas, Cerrillo Martínez [2024, p. 126-135]) i procediments de tramitació de reclamacions i recursos, entre altres instruments cabdals per mitigar l'impacte sobre els drets fonamentals. Un cop completada l'AIDF, el responsable del desplegament ha de notificar-ne el resultat a l'autoritat de vigilància del mercat corresponent.

El considerant 96 del RIA reforça aquesta arquitectura mínima: l'AIDF ha de determinar els riscos concrets per a cada dret, establir sistemes de supervisió humana adaptats al context d'ús i preveure procediments de tramitació de reclamacions (vegeu Gatt et al. [2024], sobre el model d'implementació de les AIDF conforme al RIA).

Cal destacar que l'exigència d'una AIDF no substitueix altres avaluacions d'impacte ja existents. Així, quan un sistema d'IA comporti tractament de dades personals, caldrà complir tant l'AIDF prevista pel RIA com, si escau, l'avaluació d'impacte relativa a la protecció de dades de l'article 35 del RGPD (també sobre aquest punt, vegeu Simón Castellano [2024, p. 563-564]; Chaveli Donet [2024, p. 529-531], i Pedrós Ramos [2025]). La necessitat de coordinar ambdues avaluacions és un primer exemple de la governança integral que requereix la IA: implica abordar simultàniament els requeriments del RGPD i del RIA, sense duplicitats però sense buits de control.

L'obligació o l'exigència normativa de l'AIDF arriba abans que existeixi un consens clar sobre la metodologia concreta que s'ha de seguir per dur-la a terme. El legislador enumera els elements que cal considerar (context, impactes i mesures), però deixa als operadors i eventualment als reguladors el desenvolupament d'eines pràctiques per realitzar l'anàlisi. Alguns autors, com Mantelero (2024), es plantegen vies de resposta a aquesta mancança i proposen blocs fonamentals d'un model d'AIDF en el seu estudi amb l'objectiu de facilitar la feina futura de les autoritats i les entitats obligades, de tal manera que col·loquen aquesta eina al cor de l'enfocament europeu de la IA de confiança. Segons Mantelero (2024, p. 30-31), l'AIDF ha de ser concebuda com una eina versàtil i aplicable més enllà dels casos estrictes de l'article 27, i ha de servir com a full de ruta per a qualsevol iniciativa reguladora que busqui alinear la IA amb els drets humans.

Més enllà del compliment formal, la metodologia ha d'acreditar tres vectors clau: (i) una identificació traçable dels impactes sobre els drets, amb criteris de probabilitat i gravetat, si bé s'hi poden afegir altres variables com ara l'escalabilitat o la dependència de tercers; (ii) la justificació de la necessitat i la proporcionalitat en el context administratiu –incloent-hi alternatives menys intrusives–, i (iii) mecanismes efectius de supervisió humana i de transparència activa al llarg de tot el cicle de vida del sistema (Lazcoz i De Hert, 2023).

La literatura recent ofereix pautes concretes amb propostes i models que graduen la gravetat dels impactes i n'orienten la mitigació (Malgieri i Santos, 2025), enfocaments empírics que integren evidència i governança (Mantelero i Esposito, 2021), propostes de formalització de l'AIDF mitjançant la lògica derrotable[14] (Garcia-Godinez, 2025) i guies pràctiques per operativitzar l'AIDF en organitzacions públiques (Janssen et al., 2022).

14 La lògica derrotable és no monòtona, de tal manera que les conclusions provisionals poden retirar-se davant d'excepcions o de regles prioritàries noves. Es tracta d'una proposta útil per a la traçabilitat de la motivació i la consistència decisòria, però la proporcionalitat en conflictes de drets fonamentals –o entre drets i interessos constitucionals– exigeix una ponderació contextual que no es pot reduir a esquemes purament lògics o mecànics. Referent a això, vegeu Garcia-Godinez (2025).

Aquest marc es pot i s'hauria d'alinear amb l'estandardització tècnica; en particular, amb la norma tècnica ISO/IEC 42005:2025, que proposa un procés d'avaluació d'impacte de sistemes d'IA integrat amb el sistema intern de gestió de riscos, i amb l'ISO/IEC TS 12791:2024,[15] que ofereix tècniques transversals de tractament del biaix en tasques de classificació i regressió, útils per avaluar i mitigar discriminacions en serveis públics. Des d'una perspectiva de dret administratiu, aquestes exigències es connecten naturalment amb el dret a la bona administració i amb els deures de transparència activa. L'AIDF ha d'incorporar microavaluacions de necessitat i proporcionalitat degudament motivades, informació comprensible per a les persones afectades i registres publicables sobre finalitat, dades, governança i controls del sistema (Ponce Solé, 2024). Això exigeix que l'AIDF no es redueixi a un mer qüestionari o llista de control (*checklist*): cal preveure espais deliberatius reals, interdisciplinaris, i punts de control vinculats a canvis substantius del cas d'ús (Mir, 2025).

En suma, i com es veurà a continuació, som en una fase en què la teoria jurídica i l'enginyeria social s'estan posant al servei de l'execució pràctica d'aquestes avaluacions. La construcció d'una cultura de compliment proactiu –en què les administracions integrin l'avaluació de riscos en la seva presa de decisions tecnològiques– és un procés incipient. El *JuLIA Handbook* (Colombi Ciacchi et al., 2025), dedicat a la IA i l'Administració pública, aborda precisament els límits jurídics de la governança algorítmica i reflexiona sobre com les administracions poden incorporar principis de legalitat, transparència i control en l'ús d'algoritmes. Les obligacions de l'AIDF del RIA representen un instrument innovador per avançar en aquesta direcció, però el seu èxit dependrà en bona mesura de com s'acabin concretant aquestes metodologies i de com es duguin a la pràctica, tal com analitzarem tot seguit a través de diversos casos i models.

3 Metodologies d'avaluació d'impacte: casos i pràctiques comparades

L'obligació legal d'efectuar l'AIDF planteja immediatament com a derivada la pregunta de *com* dur-la a terme de manera eficaç. Quina metodologia s'ha de seguir? Quines preguntes cal respondre, quins criteris s'han d'aplicar per mesurar el risc i determinar les mesures de mitigació adequades? Atès el caràcter "nou" de l'eina[16] i de l'obligació, fins fa poc no es disposava de models consolidats. Tanmateix, en els últims anys han emergit diverses iniciatives pioneres que han desenvolupat metodologies pràctiques per avaluar l'impacte de sistemes algorítmics i, més concretament, avaluar-ne l'impacte en els drets fonamentals de les persones.

A continuació, analitzarem tres experiències particularment rellevants i complementàries que han estat escollides i seleccionades expressament per actualitzar el debat i introduir models aprovats recentment que, en qualsevol cas, han estat dissenyats post-RIA. Un exercici similar però previ a l'aprovació del RIA es pot trobar en la monografia de Simón Castellano (2023, p. 103-114), on s'analitzen els models de l'Ava Lovelace Institute, l'eina AIA del Govern dels Estats Units, l'Algorithmic Impact Assessment Tool del Canadà, la versió 1.0 del model FRAIA dels Països Baixos i el model PIO de l'Observatori d'Ètica en Intel·ligència Artificial de Catalunya, entre d'altres. Amb l'objectiu de compartir models recents i actualitzats a la realitat del RIA, proposem l'estudi de tres mètodes, com són el model català d'AIDF impulsat per l'Autoritat Catalana de Protecció de Dades (en endavant, APDCAT) (2025), el model neerlandès d'AI Impact Assessment (AIIA) versió 2.0 (2024) i les directrius italianes de l'AgID (2023-2025), que integren l'avaluació de drets fonamentals en el cicle de vida de la IA pública. Cadascun d'aquests casos il·lustra un enfocament amb característiques pròpies, però tots contribueixen a perfilar com podria ser una AIDF robusta i efectiva, perquè cal recordar

15 ISO/IEC 42005:2025 Information technology – Artificial intelligence (AI) – AI System impact assessment.

ISO/IEC TS 12791:2024 Information technology – Artificial intelligence – Treatment of unwanted bias in classification and regression machine learning tasks.

16 Utilitzem aquesta terminologia tenint present que el mateix legislador europeu empeny cap a un model de qüestionaris i eines automatitzades, d'acord amb el redactat de l'article 27.5 del RIA.

que no hi ha un model universal o únic, sinó diferents metodologies que són perfectament vàlides per aconseguir els objectius i les finalitats que persegueix l'article 27 del RIA, respectant, això sí, els elements o el contingut mínim de l'obligació que ha estat objecte d'estudi *ut supra*.

3.1 El model proposat per l'Autoritat Catalana de Protecció de Dades

L'APDCAT (2025) ha estat una de les primeres autoritats de control, encara que les seves funcions es projecten en exclusiva en l'àmbit de la protecció de dades personals i no pròpiament en l'àmbit de la IA,[17] que ha dissenyat i proposat un model aplicat d'AIDF específic per a sistemes d'IA d'alt risc, amb casos d'ús reals. Ho ha fet a través d'una guia o "model", que incorpora una metodologia aplicada de l'AIDF en el disseny i el desenvolupament de la IA, que fou presentada formalment en un acte institucional[18] coincidint amb el Dia Internacional de la Protecció de Dades de 2025. Es tracta de la primera metodologia a Europa que aplica l'AIDF a casos concrets amb l'objectiu de formular les noves obligacions establertes pel RIA. De fet, el model s'ha dissenyat expressament per ajudar el sector públic a complir amb l'article 27 del RIA, incloent-hi les fases de detecció i de tractament (mitigació) de riscos, de prevenció de biaixos i discriminacions, etc.

En absència de directrius europees encara en aquest àmbit i, en especial, a l'espera de l'activitat proactiva i dels criteris de l'Oficina Europea d'IA i de l'AESIA, l'APDCAT ha pres la iniciativa i ha contribuït a omplir el buit amb un model de referència desenvolupat localment però amb aspiració internacional. Tant és així que en el mateix text s'indica aquesta voluntat o aspiració quan s'afirma que "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" (APDCAT, 2025, p. 21).

El projecte s'ha materialitzat en una guia metodològica no precisament extensa (vint-i-dues pàgines sense comptar les experiències o casos d'ús). El model s'ha testat en quatre casos pilot (educació, gestió de treballadors, assistència sanitària i serveis socials), que han permès verificar-ne l'aplicabilitat.

Pel que fa a l'estructura metodològica, el model català proposa un procediment dividit en tres fases successives, alineades precisament amb els factors que marca l'article 27 del RIA. Segons exposa la guia, la primera fase és la de planificació, definició de l'abast i identificació de riscos (APDCAT, 2025, p. 10-11). En aquesta fase inicial es descriuen detalladament el sistema d'IA (finalitat, funcionament, dades emprades) i el seu context d'ús, i s'identifiquen els riscos intrínsecs (vinculats al sistema mateix, com ara l'opacitat de l'algoritme o els potencials biaixos en dades) i els riscos extrínsecs (derivats de la interacció amb l'entorn on s'implementa, tals com l'impacte en un col·lectiu vulnerable). També s'hi analitzen les categories de persones potencialment afectades i les diferents dimensions de l'impacte.

La segona fase és la d'anàlisi de riscos (APDCAT, 2025, p. 12-14), en la qual es realitza ja una ponderació més profunda: es va més enllà d'una identificació general de possibles impactes i es determina el nivell d'impacte sobre cada dret o llibertat afectat emprant criteris d'estimació (probabilitat d'ocurrència, severitat del dany, extensió de les persones afectades, etc.). Es tracta, en definitiva, de quantificar o qualificar el risc en termes de grau utilitzant una matriu de quatre nivells (baix, mitjà, alt o molt alt) per a cada tipus de dret en joc (tant pel que fa a la probabilitat com a la gravetat i a l'exposició). Aquest model de matriu de

17 Mentre no s'aprovi l'Avantprojecte de llei per al bon ús i la governança de la intel·ligència artificial (versió sotmesa a informació pública, 12 de març de 2025), que en el seu article 6.3 preveu que l'Agència Espanyola de Protecció de Dades i les autoritats autonòmiques de protecció de dades seran autoritats de vigilància del mercat dels sistemes d'IA només per a aquells sistemes que realitzin pràctiques d'IA prohibides recollides en els apartats *d*, *g* i *h* de l'article 5.1 del RIA i dels sistemes d'IA d'alt risc relatius a l'ús de la biometria en la gestió de fronteres, relatius a l'àmbit de migració, asil i gestió de control fronterer i vinculats a l'àmbit de la garantia del compliment del dret.

18 Vegeu l'enregistrament de l'acte institucional de presentació del model/guia en el marc de la jornada "Intel·ligència artificial i drets fonamentals: una mirada més enllà de la privacitat" (Parlament de Catalunya, 28 de gener de 2025).

4x4 és una opció més ambiciosa que el 3x3 que ha proposat el legislador amb l'estàndard de l'Esquema Nacional de Seguretat.[19] El model de matriu 4x4 segueix la línia marcada prèviament per l'Agència Espanyola de Protecció de Dades (AEPD) amb el seu model d'avaluació d'impacte en protecció de dades (Agencia Española de Protección de Datos, 2021, p. 75-77), si bé és cert que en altres models de gestió de riscos, més vinculats al compliment normatiu penal, també s'han utilitzat en moltes ocasions les matrius de 5x5, més completes i que ofereixen un nivell de detall més gran en la identificació i la delimitació dels riscos inherents i residuals (Simón Castellano, 2020).

En l'àmbit de les AIDF, la matriu més completa ha estat la proposada a escala doctrinal per Bertaina et al. (2025, p. 8-15), que formula un marc innovador que combina un qüestionari contextual amb una matriu quantitativa per mesurar l'impacte en cada un dels drets fonamentals. Un cop avaluats els cinquanta drets de la Carta de la UE, s'assigna a cadascun un indicador RISA (*risk of impact on a specific article*) que combina diversos factors: exposició segons percentatge de població afectada; probabilitat d'ocurrència (probabilitat que el sistema cometi un error que derivi en un perjudici); intensitat (magnitud del dany físic o simbòlic causat per la decisió algorítmica); gravetat (impacte en la qualitat de vida o en els drets de la persona); durada dels efectes (persistència temporal del perjudici), i esforç de recuperació (accions requerides per revertir o mitigar el dany). A partir d'aquestes magnituds s'obté per a cada dret un nivell de risc o valor numèric. Atès que la població rarament és homogènia, cal, a la pràctica, dividir-la en subgrups amb nivells de risc similars, calcular un *class risk score* per a cadascun i agregar-los en un indicador RISA final que reflecteixi adequadament el risc global. Aquesta metodologia (Bertaina et al., 2025), híbrida entre anàlisi qualitativa –experts legals, ètics i de domini– i mètrica quantitativa –científics de dades–, assegura que tant la naturalesa no absoluta dels drets fonamentals com el seu valor social i jurídic quedin degudament ponderats.

La tercera fase correspon a la mitigació i la gestió dels riscos identificats prèviament (APDCAT, 2025, p. 14-15). Aquí es revisen mesures preventives o correctores que ja hagin estat introduïdes pel proveïdor o pel responsable del desplegament i, si escau, se n'estableixen de noves per a cadascun dels riscos detectats i es recalcula el nivell de risc residual després d'aplicar-les. Només comparant el nivell de risc abans i després de les mesures es pot demostrar que aquestes són efectives i adequades, tal com subratlla la guia. Aquesta comprovació és essencial per dotar l'AIDF de credibilitat: no n'hi ha prou amb llistar actuacions, cal evidenciar que redueixen el risc a un nivell acceptable (sobre la noció de risc acceptable, vegeu APDCAT [2025, p. 12-13] i nota núm. 8).

El model català també aporta reflexions importants sobre aspectes innovadors del procés d'AIDF. En particular, a partir dels casos pràctics s'han identificat qüestions crucials, com, per exemple, determinar quines són les variables rellevants que cal tenir en compte en valorar l'impacte (més enllà de l'obvietat d'incloure la no-discriminació, la privacitat, etc., es debat com mesurar-les i ponderar-les); establir quina metodologia es pot seguir per avaluar el risc i fins i tot crear índexs de risc quantitatius que facilitin comparacions; fixar quin és el paper dels qüestionaris estàndard en aquest exercici i les seves limitacions (la guia constata que les típiques llistes de control o qüestionaris de verificació poden ser útils per no oblidar cap àrea, però insuficients per ells mateixos per captar tota la complexitat), i, finalment, delimitar quin paper han de tenir els experts independents dins l'avaluació. La metodologia també preveu la revisió d'un expert extern independent a fi d'aportar objectivitat.[20]

El model d'AIDF de l'APDCAT representa un referent pioner que demostra la possibilitat de racionalitzar els requisits del RIA i traduir-los eficaçment en un procés d'anàlisi i gestió de riscos respectuós amb els

19 Vegeu l'estàndard de seguretat i el complet llistat de mesures de seguretat previst a l'annex II del Reial decret 311/2022, de 3 de maig, pel qual es regula l'Esquema Nacional de Seguretat. (BOE, núm. 106, 04.05.2022, p. 61715-61804).

20 Sobre això, vegeu també l'interessant exemple que s'il·lustra en l'article de Muis et al. (2025, p. 4).

drets fonamentals. Caldrà veure, doncs, com evoluciona i s'actualitza aquesta metodologia a mesura que s'analitzin nous casos. El projecte català preveu continuar incorporant-ne més en el futur, i la idea és que especialment el sector públic acabi abraçant i emprant el model proposat per l'APDCAT.

3.2 L'eina *AI Impact Assessment* dels Països Baixos (AIIA 2.0)

Els Països Baixos han estat un altre dels països capdavanters en el desenvolupament d'eines per avaluar l'impacte de la IA, si bé ho han fet des d'una perspectiva lleugerament diferent i més àmplia. Ja des de 2019, el Govern neerlandès va començar a promoure una llista de control per mesurar l'impacte algorítmic en projectes públics, integrat en iniciatives de govern obert i transparència. Recentment, el desembre de 2024, s'ha publicat la versió 2.0 de la guia *AI Impact Assessment* (en endavant, AIIA), una eina exhaustiva per assegurar projectes d'IA responsables en el sector públic (Govern dels Països Baixos, 2024). L'AIIA 2.0 s'ha elaborat amb la col·laboració de diversos ministeris (notablement el d'Infraestructura i Gestió de l'Aigua, que ha liderat el projecte) i actualitza l'eina anterior incorporant-hi novetats alineades amb el RIA.

El tret distintiu del model neerlandès és que aborda l'impacte de la IA des d'una òptica integral, no centrada només en drets fonamentals, sinó incloent-hi també aspectes ètics, socials i de sostenibilitat (Govern dels Països Baixos, 2024, part A, p. 9-13), i també comprèn elements per a una avaluació des del disseny, el desenvolupament i el primer ús de la tecnologia (Govern dels Països Baixos, 2024, part B, p. 16-20). El contingut de l'AIIA s'organitza en diverses àrees, que cobreixen drets fonamentals, sostenibilitat, altres efectes socials, decisió sobre la conveniència o no d'usar la IA en qüestió i mesures de control durant el cicle de vida del projecte.

És interessant ressaltar que en la versió 2.0 l'AIIA ha reforçat precisament la part de drets fonamentals integrant el que anomenen *Fundamental Rights and Algorithms Impact Assessment* (FRAIA),[21] que seria l'equivalent a l'AIDF que preveu l'article 27 del RIA. Això significa que ara l'eina neerlandesa no només pregunta com aplicar la IA de manera responsable, sinó també si és realment desitjable emprar IA per a una finalitat concreta, tenint-ne en compte els impactes en drets i valors democràtics (Govern dels Països Baixos, 2024, p. 12). Aquesta orientació reflecteix un nivell de precaució elevat i reconeix que hi pot haver casos en què la millor decisió sigui no utilitzar IA en absolut si els riscos per als drets humans són massa greus o impossibles de mitigar.

Pel que fa al tractament específic dels drets fonamentals dins l'AIIA, la guia dedica un apartat breu (capítol 2.1) a aquesta qüestió. S'hi estableix que qualsevol projecte ha de vetllar perquè els drets de totes les parts interessades siguin respectats. Es parteix del marc de drets consagrats a la Constitució neerlandesa i al Conveni Europeu de Drets Humans, i s'hi adjunta fins i tot un annex amb un llistat de clústers de drets fonamentals rellevants per ajudar els avaluadors a no descuidar-se cap dimensió important. L'eina proporciona preguntes guia o model, com ara "quin serà el potencial impacte sobre els drets fonamentals dels ciutadans si usem aquest sistema d'IA?" o "quina base jurídica fonamenta l'ús del sistema d'IA i les decisions que es preveuen adoptar a partir del seu funcionament?".

Pel que fa a la gestió i el tractament dels riscos, l'AIIA (a la part B) inclou tot un seguit de mesures de salvaguarda o de control per protegir drets, que caldrà aplicar quan correspongui segons els nivells de risc residual detectats. Si, malgrat les mesures, es detecta que hi ha un risc de violació greu de drets fonamentals o fins i tot que el projecte pot infringir lleis vigents, l'eina és taxativa i assenyala que no s'hauria de permetre desplegar aquell sistema d'IA.

21 Vegeu *Impact assessment. Fundamental rights and algorithms* (en anglès), o *Impact assessment. Mensenrechten en algoritmes* (en neerlandès).

Un altre aspecte notable del model holandès és la insistència en el caràcter participatiu i deliberatiu de l'avaluació. L'eina anima a involucrar diferents agents en el procés: des de diàlegs ètics amb grups d'interès fins a consultes amb experts en drets. L'objectiu és que l'AIDF no sigui un exercici mecànic d'una llista de control, sinó una conversa estructurada sobre els aspectes ètics i jurídics del projecte d'IA. D'aquesta manera, es promou una cultura d'autoreflexió dins l'organització que desenvolupa o posa en marxa la IA, així com la transparència i la rendició de comptes davant la societat. Cal destacar que el Govern neerlandès ha fet pública aquesta guia AIIA 2.0 en anglès per facilitar que altres països o entitats se'n puguin beneficiar, la qual cosa demostra també una voluntat de lideratge i cooperació internacional en la matèria.

L'experiència neerlandesa aporta un model d'AIDF en què la dimensió dels drets fonamentals apareix com un element nuclear (a partir de la versió 2.0), però combinat amb d'altres com ara la seguretat, la sostenibilitat i la viabilitat tècnica. Aquest model reforça algunes idees clau com ara la necessitat de determinar la proporcionalitat de l'ús del sistema d'IA, la conveniència d'escalar els riscos i preveure accions segons la seva gravetat i la importància d'incorporar l'avaluació a un diàleg multidisciplinari dins de l'organització. L'orientació pràctica i l'èmfasi a fer preguntes als responsables dels projectes públics fan de l'AIIA un complement valuós a la visió més estrictament jurídica de l'AIDF.

3.3 La incorporació de l'AIDF en les directrius italianes (AgID, 2023-2025)

Itàlia ha afrontat la qüestió de l'impacte de la IA en drets fonamentals a través de la generació de línies directrius nacionals per a la utilització de la IA a l'Administració pública. L'Agència per a la Itàlia Digital (AgID) –organisme públic encarregat de la transformació digital– va iniciar un procés de consulta pública d'unes *Linee guida per l'adozione dell'Intelligenza Artificiale nella Pubblica Amministrazione* (AgID, 2025),[22] en el marc del *Pla triennal per a la informàtica a l'Administració Pública 2024-2026*,[23] amb l'objectiu d'orientar tant els proveïdors com els usuaris públics de tecnologies d'IA. Les directrius, encara pendents d'aprovació formal final, de 119 pàgines i acompanyades de nombrosos annexos tècnics, ofereixen una visió integral del cicle de vida de la IA en el sector públic, des de la planificació fins al desplegament, el monitoratge continu i l'eventual retirada del sistema. L'enfocament italià és, doncs, bastant complet i operatiu, i incorpora tant les exigències del RIA com aspectes organitzatius i de capacitats que han de desenvolupar les administracions públiques.

Un dels punts centrals del text és, precisament, l'AIDF prèvia a l'adopció o l'ús de qualsevol sistema d'IA. Un valor afegit de la proposta italiana és que clarifica la diferent funció de diverses avaluacions d'impacte que poden entrar en joc –la relativa a les dades personals, la vinculada a l'impacte sobre els drets fonamentals– i, fins i tot, introdueix el concepte més genèric d'avaluació d'impacte de la intel·ligència artificial (*artificial intelligence impact assessment*) com una avaluació més àmplia (seguretat, explicacions, fiabilitat, danys potencials, col·lectius vulnerables, sostenibilitat). El document reforça la necessitat d'AIDF per a aquells sistemes d'IA que poden incidir en les llibertats civils, l'accés a serveis públics o l'equitat en processos de presa de decisions.

Pel que fa al contingut de l'avaluació, aquesta ha d'analitzar en profunditat els possibles impactes del sistema considerant aspectes com ara la durada i la freqüència d'ús de l'algoritme, els subjectes involucrats (per exemple, grups de ciutadans afectats) i els riscos tant per a la privacitat com per a altres drets fonamentals. És remarcable que es posi l'accent també en factors operatius (durada, freqüència), perquè això connecta l'avaluació amb la proporcionalitat: no és el mateix un algoritme utilitzat esporàdicament que un que decideix de manera contínua en processos massius.

22 En fase de consulta pública des del 18 de febrer fins al 20 de març de 2025, d'acord amb la Decisió núm. 17, de 17 de febrer de 2025. Inici del tràmit de consulta i informació de les Directrius per a l'adopció de la intel·ligència artificial a l'Administració pública.

23 Agència per a la Itàlia Digital. (2023). *Piano triennale per l'informatica nella Pubblica Amministraziones. Edizione 2024-2026*.

Cal esmentar que Itàlia ha creat fins i tot noves figures i eines organitzatives per afrontar aquests reptes. En les directrius apareix, per exemple, la figura de l'*AI architect* (arquitecte d'IA) dins les administracions com l'expert encarregat de dissenyar i avaluar l'arquitectura dels sistemes d'IA per assegurar-ne una escalabilitat, una seguretat i un rendiment adequats. També es proposa un *maturity model* (model de maduresa) per avaluar el grau de preparació de cada organisme públic en l'adopció de la IA i identificar-hi àrees de millora.

Entre els annexos pràctics destaquen un annex B dedicat a l'avaluació de riscos i un annex C específic sobre l'avaluació d'impacte, que inclou plantilles i exemples per guiar les administracions públiques en la realització d'aquestes avaluacions. L'esforç de sistematització reflecteix la voluntat que l'AIDF no sigui un exercici improvisat, sinó integrat en una estratègia definida amb objectius clars, recursos assignats i un sistema de monitoratge continu.

El cas italià mostra una aproximació de governança pública de la IA en què l'AIDF és una peça central però inserida en un engranatge més ampli de gestió de riscos i compliment normatiu. S'anticipen així moltes de les qüestions que totes les administracions europees hauran d'afrontar a curt termini: com conjugar les diverses avaluacions (dades personals, drets humans, ètica, impacte sociotècnic, etc.), com dotar-se d'estructures i perfils competents per fer-les (AgID, 2025, p. 45-47), i com assegurar que no es tracta d'un tràmit formal, sinó que realment es demostri haver analitzat els riscos i haver garantit l'ús justificat i proporcional de la IA (AgID, 2025, p. 85-91). La integració holística de les avaluacions d'impacte de diferent naturalesa podria servir de model per a altres estats que necessitin aclarir la relació entre diferents marcs normatius. I, sobretot, posa en relleu que la gestió de la IA requereix tant coneixement tècnic per entendre el funcionament i les limitacions dels algoritmes com coneixement jurídic per comprendre els impactes en drets i les obligacions legals, així com una bona dosi de cultura ètica dins de les organitzacions. Sense aquests ingredients, les millors directrius i metodologies de les autoritats de control poden quedar en paper mullat.

3.4 Anàlisi comparada i valoració crítica

En termes d'exhaustivitat i control material, el model català (APDCAT, 2025) destaca per l'enfocament sistemàtic de riscos sobre drets, amb matrius que utilitzen les tradicionals variables de probabilitat per gravetat i que, al seu torn, estan orientades per a l'elaboració d'un pla d'acció amb mesures de mitigació, fet que facilita una traçabilitat justificativa alineada amb el dret a la bona administració i amb la gestió recurrent dels riscos.

Per la seva banda, el model neerlandès d'AIIA integra, com a principal aportació, una cobertura molt operativa del cicle de vida, de tal manera que diferencia una part A vinculada a la necessitat, els valors públics i la proporcionalitat de l'ús de la IA, d'una part B que es refereix a qüestions més tècniques: de disseny, robustesa tècnica, governança i monitoratge dels riscos. Aquesta metodologia ordena l'avaluació en fases amb punts de control decisori (*stage-gates*): en cada fase, l'òrgan responsable ha de motivar, a partir del test de necessitat i proporcionalitat i de la ponderació dels impactes que s'espera que la IA projecti sobre els drets de les persones, si escau autoritzar la continuació, imposar salvaguardes addicionals o interrompre l'ús d'IA d'alt risc. Aquestes decisions queden formalitzades i arxivades a l'efecte de rendició de comptes, supervisió interna i control jurisdiccional, de manera que connecten l'avaluació amb els principis de bona administració i transparència activa.

Pel que fa al model italià (AgID, 2025), l'avaluació d'impacte forma part d'un procés estructurat al llarg del cicle de vida tecnològic, que coordina tres instruments: el perfil tecnicoorganitzatiu, l'impacte esperat sobre els drets fonamentals i l'avaluació específica de l'impacte en matèria de protecció de dades personals. La recomanació operativa és integrar les diferents avaluacions –i, més concretament, l'AIDF i l'AIPD– en un únic expedient i en una única metodologia, que en qualsevol cas exigeix la participació del delegat de protecció de dades, dels serveis jurídics i també de perfils tècnics. Amb aquesta unitat metodològica de les diferents

avaluacions es pretén evitar duplicitats, alinear mesures de mitigació i ancorar l'avaluació a decisions clau de compra, desplegament i monitoratge de la IA. La principal utilitat d'aquest plantejament rau a simplificar el compliment *ex ante* en matèria de contractació pública, però al mateix temps exigeix inversió organitzativa i capacitat interna estable per sostenir un equip multidisciplinari de governança i seguiment de la IA.

En síntesi comparada, el model català és el que aporta més granularitat jurídica i traçabilitat en la ponderació de drets i eficàcia de les mesures de mitigació de riscos; el model neerlandès sobresurt en governança per fases amb punts de decisió motivats i en la connexió entre necessitat i proporcionalitat i control tècnic; el model italià és el més consistent en institucionalització (procés integrat de diferents avaluacions i encaix amb els processos públics de compra i gestió). Per a administracions amb recursos limitats, un enfocament híbrid que combini la matriu de riscos en drets de l'APDCAT, els *stage-gates* motivats de l'AIIA i la integració procedimental de l'AgID oferiria la millor relació impacte-cost i facilitaria la rendició de comptes.

4 Reptes d'implementació al sector públic

Les experiències analitzades mostren camins viables per realitzar AIDF, però alhora evidencien que implementar aquesta exigència en el dia a dia del sector públic comporta reptes importants. En aquesta secció identifiquem i discutim alguns dels desafiaments principals que afronten les administracions públiques a l'hora de donar compliment efectiu a l'article 27 del RIA i, en general, d'integrar l'ètica i els drets en la governança dels sistemes d'IA.

4.1 Capacitat tècnica i metodològica

Un primer repte és purament tècnic i metodològic: les administracions disposen de les eines i els coneixements necessaris per dur a terme una AIDF de qualitat? A diferència de les grans corporacions tecnològiques, moltes entitats públiques, especialment a escala local o regional, no disposen d'equips multidisciplinaris especialitzats en IA, ètica i drets. La realització d'una AIDF exigeix dominar aspectes que van des de l'arquitectura de l'algoritme fins a la interpretació jurídica de conceptes com la discriminació indirecta, la desinformació i l'impacte en la dignitat humana. Aquesta transversalitat requereix equips de treball amb perfils especialitzats molt diversos: informàtics, juristes, estadístics, sociòlegs i, si escau, experts sectorials (educació, salut, etc.).

Reunir i coordinar perfils especialitzats tan diversos no és senzill en entorns administratius sovint organitzats en sitges d'informació o en unitats sense connexió. És per això que iniciatives com la italiana (AgID, 2025, p. 45-46), que preveuen la figura de l'*AI architect* o la creació de comitès d'ètica i experts, seran probablement necessàries a escales més àmplies. En la implementació concreta, una administració pot optar per capacitar personal intern o bé recórrer a consultors externs especialitzats en AIDF. Ambdues opcions tenen avantatges i riscos: la formació interna reforça la competència de l'organització a llarg termini, però a curt termini pot no cobrir totes les necessitats; la contractació externa aporta expertesa immediata, però pot ser costosa i generar dependència de tercers. Serà clau, doncs, que els estats destinin recursos i formació perquè les administracions desenvolupin aquesta capacitat interna en la mesura que sigui possible, encara que inicialment l'assessorament i l'acompanyament per part d'externs siguin més intensos o significatius (Muis et al., 2025, p. 4-5).

En paral·lel, caldrà refinar les metodologies d'AIDF disponibles. Com hem vist, hi ha un debat obert entre doctrina, administracions i autoritats de control i de vigilància del mercat sobre la millor manera de quantificar riscos que es projecten sobre els drets fonamentals i sobre l'ús potencial de llistes de control predefinides i parcialment genèriques per oposició a una anàlisi experta qualitativa, entre altres qüestions pendents. Un risc real seria que, a la pràctica i per complir formalment, es recorregués a models o *templates*

molt simplificats que no capturen la complexitat real, per exemple, mitjançant simples llistes de control sense anàlisi profunda. Si l'AIDF es banalitzés en una llista de control burocràtica, se'n desvirtuaria la finalitat. En canvi, tampoc es pot pretendre que cada administració desenvolupi des de zero la seva pròpia metodologia complexa. Tampoc és acceptable un model en el qual les administracions públiques o les entitats del sector públic es limitin exclusivament a replicar o remetre a una AIDF que han realitzat prèviament els proveïdors, que, lògicament, no han tingut en compte necessàriament el context i les finalitats d'ús específiques de la IA per a aquella entitat o administració concreta.

Davant aquests riscos que amenacen de buidar de contingut real l'obligació de l'article 27 del RIA, sorgeix una solució o oportunitat per avançar cap a l'estandardització de certes guies i plantilles bàsiques, entre les quals destaca l'eina que pugui proveir l'Oficina Europea d'IA o les corresponents autoritats nacionals, com l'AESIA, combinada amb la possibilitat d'adaptar-les al context i afegir-hi una capa d'anàlisi experta. La creació d'espais per compartir millors pràctiques entre administracions, com ara entorns controlats de proves (*sandbox*) i repositoris de casos d'ús, pot ser de gran ajuda: per exemple, que una entitat pugui inspirar-se en com una altra va avaluar un algoritme similar en un context anàleg. Així es guanya en eficiència i coherència. El repte metodològic implica evitar tant l'*over-engineering*, això és, fer l'AIDF tan complexa que resulti impracticable, com l'*under-engineering,* és a dir, reduir-la a un tràmit superficial. Trobar l'equilibri entre una cosa i l'altra requerirà prova i error, orientació especialitzada i flexibilitat local.

4.2 Integració en els processos administratius i presa de decisions

Un segon repte és la integració organitzativa de l'AIDF dins dels processos públics. És a dir, no n'hi ha prou amb saber com fer l'avaluació; cal determinar quan i en quin marc es farà, i com els seus resultats influiran en la presa de decisions. Per exemple, si un organisme públic vol implantar un sistema d'IA, hauria de realitzar l'AIDF en fases primerenques i, fins i tot, idealment, durant la fase de disseny o, si escau, abans de la contractació de la tecnologia, per valorar-ne la proporcionalitat i la necessitat, entre d'altres. Això suposa planificar l'AIDF dins el projecte: definir un calendari, responsabilitats i lliurables.

De les experiències i les propostes en perspectiva comparada n'hem extret una segona idea present en les directrius italianes (AgID, 2025), que insisteixen en el concepte de cicle de vida de la IA i en la necessitat de monitoratge continu. En efecte, l'AIDF no hauria de ser un esdeveniment aïllat, sinó part d'un procés cíclic d'auditoria algorítmica permanent i, per tant, diferenciat però alhora connectat amb el sistema de gestió de riscos que preveu l'article 9 del RIA com a obligació per als sistemes d'alt risc.[24] Un cop desplegat el sistema, és probable que calgui revisar periòdicament l'AIDF, en especial quan es produeixen canvis substancials en l'algoritme o en el context d'ús. Aquesta actualització contínua o revisió periòdica pot suposar un canvi cultural per a unes administracions acostumades a projectes que s'acaben amb la posada en producció o en la presa de decisió favorable a la utilització d'una tecnologia. La gestió de la IA no pot acabar amb la seva implementació; cal un seguiment constant i preveure'n fins i tot la retirada quan el sistema esdevingui obsolet. Això planteja reptes de recursos, ja que caldrà mantenir equips de supervisió actius, i també voluntat política, així com capacitat per aturar un projecte si es demostrés perjudicial, amb al·lucinacions o danys que afecten el contingut dels drets fonamentals.

Com a derivada d'aquest últim punt, caldria assegurar que l'AIDF tingui impacte real en les decisions. Què passa si l'avaluació conclou que el sistema comporta riscos inacceptables? Tindran les administracions la determinació d'abandonar o de redissenyar el projecte? Aquí entra en joc la governança interna: pot ser necessari establir comitès ètics o de drets que validin els resultats de les AIDF i facin recomanacions vinculants als directius. Algunes organitzacions i especialment el sector privat han començat a establir

24 Sobre la diferenciació entre les obligacions dels articles 9 i 27 del RIA, que es projecten sobre subjectes diferents, proveïdors i responsables del desplegament, respectivament, vegeu Simón Castellano (2024, p. 563-564).

protocols interns d'ús responsable de la IA amb comitès o figures, com ara el director d'IA (*chief AI officer*), per supervisar els seus projectes d'IA i, mitjançant un sistema de registre i autoritzacions, controlar la tecnologia i com s'utilitza en el si de la companyia. El repte és evitar que l'AIDF sigui simplement un document que s'arxiva en un calaix. Ha de ser un instrument que informi decisions estratègiques. Per això, s'ha d'enllaçar amb altres procediments, com ara la fase de contractació pública, i evitar adjudicar sense una AIDF prèvia per part del proveïdor. Una bona pràctica seria que els resultats de l'AIDF es fessin públics, almenys l'informe executiu de resum o les conclusions de l'avaluació, de manera que l'opinió pública i els òrgans de control extern (si escau, parlaments, síndics de greuges, etc.) poguessin conèixer-los i exigir comptes. De fet, aquesta transparència és en si mateixa una garantia: si se sap que l'avaluació transcendirà, hi haurà més incentiu per fer-la rigorosa i no maquillar els riscos.

4.3 Supervisió, rendició de comptes i participació ciutadana

Un altre repte important se situa en l'esfera de la supervisió externa i la legitimitat democràtica. Les administracions s'autoavaluaran, però qui vigilarà el vigilant? En altres àmbits, com el mediambiental, les avaluacions d'impacte solen ser revisades per un organisme independent o sotmeses a informació pública. Seria recomanable que les AIDF comptessin amb algun grau de verificació independent. L'article 27.3 del RIA estableix que el responsable del desplegament notificarà el resultats de l'AIDF a l'autoritat de vigilància del mercat mitjançant un model que ha d'emplenar, però ara cal dotar de contingut aquest mandat i res no impedeix que aquest model exigeixi informació suficient per tal de verificar no només els resultats, sinó les metodologies utilitzades i evidències concretes de compliment normatiu o controls i mesures aplicades per mitigar els nivells de risc.

L'Oficina Europea d'IA, l'AESIA i la resta d'autoritats de vigilància del mercat han de ser les encarregades de custodiar i verificar els resultats de les AIDF i, si cal, poden arribar a requerir modificacions a un projecte si consideren que els resultats de l'avaluació evidencien riscos no assumibles. A més, cal recordar el poder sancionador davant l'incompliment de les previsions del RIA, que s'estén també als responsables del desplegament que no facin l'AIDF quan sigui preceptiva, que és el cas de les administracions i els organismes públics i que utilitzen o pretenen utilitzar sistemes d'IA catalogats d'alt risc. No obstant això, interessa assenyalar que el règim sancionador en l'àmbit del sector públic no suposarà en cap cas una sanció econòmica, sinó un avís que cal entendre, en aquest cas, com una supervisió per part d'un ens públic orientada a la millora contínua, de tal manera que l'òrgan supervisor dialogui amb l'Administració, li assenyali febleses en l'avaluació i li suggereixi millores o controls addicionals.

La participació ciutadana és un altre element que cal considerar. En la mesura que els sistemes d'IA públics poden afectar drets col·lectius, seria coherent amb l'estat de dret democràtic obrir espais de participació en l'avaluació. Això es pot fer de diverses maneres: posant en consulta pública els esborranys d'AIDF, tal com es fa amb els projectes normatius; organitzant tallers o grups de discussió amb usuaris o col·lectius destinataris potencials, o permetent que organitzacions de la societat civil aportin informes alternatius (ONG de drets humans, col·legis professionals, etc.). Algunes metodologies, com la neerlandesa, ja inclouen la interacció amb les parts interessades i diàlegs ètics (Govern dels Països Baixos, 2024, p. 10 i 32). En contextos locals, podria ser fins i tot un exercici dins de les iniciatives de govern obert; per exemple, els ajuntaments podrien publicar la seva AIDF i recollir observacions dels veïns abans de continuar amb el projecte d'IA. Això augmentaria la transparència i la confiança en la IA pública, a més de permetre detectar riscos que als tècnics se'ls hagin pogut escapar gràcies a la participació de comunitats afectades. Per descomptat, cal equilibrar-ho amb l'eficiència, per tal com no tots els projectes menors d'IA justificaran un gran debat públic. Però en aquells en els quals es pressuposi un alt impacte, o en àrees i sectors tradicionalment polèmics –com la policia predictiva o la IA en la prestació de serveis socials sensibles– sí que semblaria adequat buscar una major legitimitat a través de la participació.

Un darrer aspecte que s'ha de tenir en compte és la responsabilitat derivada de danys potencials. Si, malgrat totes les cauteles contingudes en l'AIDF, finalment es produeix un dany a drets fonamentals (discriminació, desinformació, indefensió, vulneració de drets, etc.), quines conseqüències hi haurà? L'existència de l'AIDF pot jugar a favor o en contra de l'Administració en termes de responsabilitat. D'una banda, haver fet l'avaluació i haver implementat mesures podria servir com a prova de compliment normatiu i, per tant, com a eximent o atenuant davant de reclamacions. De l'altra, si l'AIDF era deficient, podria considerar-se negligència i fins i tot agreujar la responsabilitat, atès que l'Administració hauria reconegut un risc i no l'hauria abordat adequadament. En qualsevol cas, per als ciutadans afectats, el fet que existeixi una AIDF hauria de facilitar-los informació per exercir els seus drets, i no és estrany incloure, a més, en els resultats de l'AIDF i en l'apartat de tractament dels riscos, mesures i controls que estiguin precisament orientats a garantir respostes i la responsabilitat per danys en l'ús de la IA d'alt risc. D'aquí la importància que aquestes avaluacions reuneixin un estàndard mínim de transparència, estiguin disponibles i siguin, almenys parcialment, accessibles i comprensibles per als interessats.

5 Reflexions finals

L'aparició de l'AIDF en el panorama jurídic europeu de regulació de la IA marca un punt d'inflexió en la manera d'entendre la responsabilitat de qui dissenya i aplica sistemes algorítmics. Igual que en el seu moment el RGPD va revolucionar la cultura organitzativa en matèria de protecció de dades amb els principis de privacitat per defecte i en el disseny, o la responsabilitat proactiva i les avaluacions d'impacte en la privacitat, avui ens trobem iniciant un camí semblant però més ampli: integrar els drets fonamentals en el disseny, en el desenvolupament i en tot el cicle de vida útil de la IA. Ja no es tracta només d'evitar infraccions flagrants un cop la tecnologia està en funcionament, sinó de prevenir proactivament qualsevol efecte advers previsible adoptant les garanties necessàries. És un enfocament alineat amb la idea d'una IA ètica i centrada en l'ésser humà que inspira tant la normativa europea com nombroses declaracions internacionals (UNESCO, 2021; Consell d'Europa - CAI, 2024).

En aquest article hem analitzat el marc jurídic emergent –encapçalat pel RIA– i les primeres experiències de desplegament de l'AIDF en l'àmbit públic. Del recorregut fet, en podem extreure diverses reflexions finals que, al seu torn, podrien ser considerades com el punt de partida per a debats futurs sobre la matèria.

Primera. Sobre la necessitat de criteris i models metodològics per complir amb una obligació legal complexa. L'obligatorietat de l'AIDF per a tot el sector públic davant el potencial ús d'IA catalogada d'alt risc suposa una novetat legislativa significativa que reflecteix la voluntat de la UE de liderar el debat en matèria de drets humans i IA. No obstant això, la norma per si sola no garanteix l'eficàcia i caldrà un desenvolupament més detallat mitjançant guies, estàndards i cooperació interadministrativa. La doctrina nacional i internacional citada en aquest treball (Stahl et al., 2023; Chaveli Donat, 2024; Mantelero, 2024; Simón Castellano, 2024; Bertaina et al., 2025; Cotino Hueso, 2025) ja proporciona idees valuoses sobre com omplir aquest buit metodològic, però és essencial que les autoritats nacionals i europees en prenguin el relleu. La guia de l'APDCAT (2025) és un instrument pioner que aprofita parcialment aquest buit actual i tracta de consolidar-se com un estàndard vàlid més enllà de les nostres fronteres, a l'espera de les directrius i les guies de l'AESIA i de l'Oficina Europea d'IA.

Segona. Cap a un enfocament multidimensional de l'impacte. Les experiències analitzades (Catalunya, Països Baixos, Itàlia) mostren que l'avaluació de l'impacte algorítmic no pot limitar-se a un sol àmbit (privacitat, desinformació, indefensió, etc.). La IA genera efectes que són alhora jurídics, ètics, socials i tècnics. Per tant, el marc d'avaluació òptim serà aquell que combini diferents perspectives. L'AIDF, de naturalesa jurídica, inspirada parcialment en els tradicionals judicis de proporcionalitat (necessitat, valoració de les alternatives, idoneïtat, etc.), també haurà d'incloure consideracions ètiques (valors com la justícia,

l'autonomia, la beneficència, la sostenibilitat, etc.) i qüestions derivades d'una anàlisi tècnica (fiabilitat, ciberseguretat, rendiment, obsolescència, etc.). Només així es tindrà una imatge completa dels riscos que l'ús de determinades tecnologies projecten sobre els drets fonamentals. L'AIDF ha d'actuar com a pont entre el dret i la tecnologia traduint els requisits legals en controls tècnics concrets, i viceversa, i ha de fer que els resultats tècnics tinguin també una lectura jurídica.

Tercera. La importància d'allunyar-se de models de compliment normatiu cosmètic mitjançant l'establiment d'un sistema de governança algorítmica intern. Implantar l'AIDF exigirà reformes en la governança interna de les administracions, la qual cosa requereix processos, persones i controls. És per això que, al marge de l'avaluació pròpiament dita, s'haurà de definir qui s'encarrega de fer-la i de revisar-la, i crear nous rols o unitats o assignar-la als existents, com ara els delegats de protecció de dades o els comitès d'ètica i privacitat, i ampliar-ne les funcions. També caldrà delimitar els efectes dels resultats i del resum executiu de l'AIDF, que eleva a direcció la decisió sobre continuar o no un projecte basat en IA, i com es documenta i supervisa tot el procés a través del monitoratge recurrent o continu. Tanmateix, s'hauran d'integrar les garanties de la IA –transparència, traçabilitat, supervisió humana, etc.– en les estructures de gestió pública. Com s'assenyala en el *JuLIA Handbook* (2025), cal delimitar els límits jurídics de la governança algorítmica i establir marcs perquè l'ús d'IA en el sector públic no desbordi l'estat de dret. L'AIDF és una peça clau d'aquests marcs, però ha d'anar acompanyada d'altres mecanismes, com ara els registres d'algoritmes autoritzats, auditories independents i vies de reclamació per als ciutadans, per tal de conformar un ecosistema complet de compliment normatiu algorítmic.

Quarta. Sobre la capacitat adaptativa i la millora contínua del sistema de compliment. La realitat de la IA evoluciona ràpidament; per tant, l'AIDF no pot ser estàtica. Les metodologies i les eines s'hauran d'anar revisant i actualitzant tenint en compte l'experiència i l'evidència empírica. Els casos pilot catalans, per exemple, han fet aflorar limitacions dels qüestionaris estàndard i han reforçat la idea i la utilitat de fer participar experts externs (APDCAT, 2025, p. 6 i 19). Aquestes lliçons haurien de retroalimentar les guies futures. Igualment, la regulació haurà d'estar disposada a adaptar-se en la mesura que, si es detecta que certs sectors no coberts inicialment pel RIA presenten riscos, potser s'hauran d'incloure en l'àmbit de l'AIDF; si sorgeixen tecnologies noves amb un impacte potencial sectorial rellevant des de l'òptica dels drets fonamentals, aquestes requeriran enfocaments d'impacte específics. La capacitat adaptativa serà clau per no quedar obsolets davant la innovació.

Cinquena. Consolidació d'una cultura ètica a l'Administració. Més enllà de les qüestions formals, l'èxit de l'AIDF es mesurarà en la manera com canvia la cultura institucional. Si les autoritats i els organismes públics assumeixen que cada projecte tecnològic comporta un exercici de reflexió ètica i jurídica i ho integren de forma natural en la seva manera de treballar, haurem fet un salt qualitatiu. Passar de la cultura de "primer fer i després ja veurem" a la cultura de "fer-ho amb garanties des del principi" (garanties *by design*) és un procés llarg, però l'existència d'un mandat legal específic (article 27 del RIA) ajuda a impulsar-lo. En aquest sentit, l'AIDF pot actuar com a catalitzador d'un canvi cultural, especialment si els lideratges polítics i administratius l'abracen no només com a obligació, sinó com a oportunitat per millorar les polítiques públiques.

L'AIDF es presenta com una eina indispensable per garantir que la transformació digital de l'Administració es faci de manera legítima, transparent i respectuosa amb els valors democràtics. És un camp en construcció, en què s'entrellacen innovació jurídica i experimentació pràctica. N'hem vist els primers passos prometedors, però també som conscients de les incògnites que s'hi obren. Com qualsevol novetat, comporta riscos i resistències, però alhora obre la porta a noves garanties per al ciutadà. Aconseguiran les AIDF que els sistemes d'IA públics mereixin veritablement la confiança de la ciutadania? La resposta dependrà de com de rigorosos i honestos siguem en la seva aplicació. Si fem de la IA una aliada –i no una amenaça– d'aquests

principis, haurem aprofitat l'oportunitat. Les AIDF són un instrument poderós per avançar en aquesta direcció; utilitzar-les bé és ara la nostra responsabilitat col·lectiva.

6 Referències

Agència per a la Itàlia Digital. (2025). *Linee guida per a l'adozione dell'Intelligenza Artificiale nella Pubblica Amministrazione*. [esborrany per a consulta pública]

Agencia Española de Protección de Datos. (2021). *Gestión del riesgo y evaluación de impacto en tratamientos de datos personales*.

Bertaina, Samuele, Biganzoli, Ilaria, Desiante, Rachele, Fontanella, Dario, Inverardi, Nicole, Penco, Ilaria Giuseppina, i Cosentini, Andrea Claudio. (2025). Fundamental rights and artificial intelligence impact assessment: A new quantitative methodology in the upcoming era of AI Act. *Computer Law & Security Review: The International Journal of Technology Law and Practice*, *56*. https://doi.org/10.1016/j.clsr.2024.106101

Brussels Privacy Hub. (2023, 12 de setembre). *Brussels Privacy Hub and other academic institutions ask to approve a Fundamental Rights Impact Assessment in the EU Artificial Intelligence Act*.

Cerrillo Martínez, Agustí. (2024). Automatización, discrecionalidad y supervisión humana a la vista del Reglamento de Inteligencia Artificial. Dins Antonio José Sánchez Sáez (dir.), *Claves del Reglamento Europeo de Inteligencia Artificial. Análisis y comentarios* (p. 105-145). Aranzadi.

Chaveli Donet, Eduard. (2024). La evaluación de impacto de derechos fundamentales por quienes despliegan sistemas de inteligencia artificial en el Reglamento. Dins Lorenzo Cotino Hueso i Pere Simón Castellano (dir.), *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea* (p. 495-533). Aranzadi.

Colombi Ciacchi, Aurelia, Flórez Rojas, María Lorena, Lane, Lottie, i Nowak, Tobias (ed.). (2025). *AI and Public Administration: The (legal) limits of algorithmic governance* [JuLIA Handbook]. Comissió Europea. Projecte JuLIA (101046631) "Justice, Fundamental Rights and Artificial Intelligence".

Comitè d'Intel·ligència Artificial. (2024). *Methodology for the risk and impact assessment of artificial intelligence systems from the point of view of human rights, democracy and the rule of law (HUDERIA methodology)* (Document CAI-2024-16rev2). Consell d'Europa.

Cotino Hueso, Lorenzo. (2025). Cómo abordar jurídicamente el impacto de la inteligencia artificial en los derechos fundamentales. Dins María Emilia Casas Baamonde (dir.) i Daniel Pérez del Prado (coord.), *Derecho y tecnologías* (p. 123-176). Fundación Ramón Areces.

Garcia-Godinez, Miguel. (2025). A default framework for fundamental rights impact assessment. *AI and Ethics*, *5*, 5149-5163. https://doi.org/10.1007/s43681-025-00761-1

Gatt, Lucilla, Gaggiano, Ilaria Amelia, Gaeta, Maria Cristina, Troisi, Emiliano, Savella, Roberta, Pratesi, Francesca, i Trasarti, Roberto. (2024). FRIA implementation model according to the AI Act. *European Journal of Privacy Law and Technologies*, *2*, 193-204. https://doi.org/10.57230/EJPLT242LGIACMCGETRSRTFP

Govern dels Països Baixos. (2024). *AI Impact Assessment. The tool for a responsible AI project* [versió 2.0]. Ministeri d'Infraestructura i Gestió de l'Aigua.

Janssen, Heleen, Seng Ah Lee, Michelle, i Singh, Jatinder. (2022). Practical fundamental rights impact assessments. *International Journal of Law and Information Technology*, *30*(2), 200-232. https://doi.org/10.1093/ijlit/eaac018

Lazcoz Moratinos, Guillermo, i De Hert, Paul. (2023). Humans in the GDPR and AIA governance of automated and algorithmic systems. Essential pre-requisites against abdicating responsibilities. *Computer Law and Security Review*, *50*. https://doi.org/10.1016/j.clsr.2023.105833

Lazcoz Moratinos, Guillermo. (2024). La vigilancia o supervisión humana en el artículo 14 del Reglamento de inteligencia artificial: ¿un mero requisito obligatorio para los sistemas de alto riesgo? Dins Lorenzo Cotino Hueso i Pere Simón Castellano (dir.), *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea* (p. 681-700). Aranzadi.

Malgieri, Gianclaudio, i Santos, Cristiana. (2025). Assessing the (severity of) impacts on fundamental rights. *Computer Law & Security Review*, *56*. https://doi.org/10.1016/j.clsr.2025.106113

Mantelero, Alessandro, i Esposito, Maria. (2021). An evidence-based methodology for human rights impact assessment (HRIA) in the development of AI data-intensive systems. *Computer Law & Security Review*, *41*. https://doi.org/10.1016/j.clsr.2021.105561

Mantelero, Alessandro. (2022). *Beyond data: Human rights, ethical and social impact assessment in AI*. T.M.C. Asser Press.

Mantelero, Alessandro. (2024). The fundamental rights impact assessment (FRIA) in the AI Act: Roots, legal obligations and key elements for a model template. *Computer Law & Security Review*, *54*. https://doi.org/10.1016/j.clsr.2024.106020

Mantelero, Alessandro (coord.), Marí, Joana (coord.), Guzmán, Cristina, Garcia, Esther, Ortiz, Ruben, Moro, M. Ascensión. (2025). *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*. DPD en xarxa. Autoritat Catalana de Protecció de Dades. Generalitat de Catalunya.

Menéndez Sebastián, Eva María. (2024). La prevención de los riesgos del uso de IA para la ciudadanía: el informe de evaluación de impacto relativo a los derechos fundamentales. Dins Antonio José Sánchez Sáez (dir.), *Claves del Reglamento Europeo de Inteligencia Artificial. Análisis y comentarios* (p. 175-213). Aranzadi.

Mir, Oriol. (2025). The AI Act from the perspective of administrative law: Much ado about nothing? *European Journal of Risk Regulation*, *16*(1), 63-75. https://doi.org/10.1017/err.2024.54

Muis, Iris, Straatman, Julia, i Kamphorst, Bart. (2025). Responsible AI innovation in the public sector: Lessons from and recommendations for facilitating fundamental rights and algorithms impact assessments. *Journal of Responsible Technology*, *22*. https://doi.org/10.1016/j.jrt.2025.100118

Pedrós Ramos, Núria. (2025, 7 d'abril). *L'avaluació d'impacte sobre els drets fonamentals en el disseny i el desenvolupament de la intel·ligència artificial*. Escola d'Administració Pública de Catalunya. (Espai temàtic "Gestió de la informació: transparència i protecció de dades")

Ponce Solé, Juli. (2024). *El Reglamento de Inteligencia Artificial de la Unión Europea de 2024, el derecho a una buena administración digital y su control judicial en España*. Marcial Pons.

Simón Castellano, Pere. (2020). Responsabilidad penal de las personas jurídicas, mapa de riesgos y cumplimiento en la empresa. Dins Pere Simón Castellano (coord.) i Alfredo Abadías Selma (coord.), *Mapa de riesgos penales y prevención del delito en la empresa* (p. 31-76). Wolters Kluwer - Bosch.

Simón Castellano, Pere. (2023). *La evaluación de impacto algorítmico en los derechos fundamentales*. Aranzadi.

Simón Castellano, Pere. (2024). Los sistemas de gestión de riesgos como obligación específica para los sistemas de inteligencia artificial de alto riesgo en el artículo 9 del Reglamento. Dins Lorenzo Cotino Hueso i Pere Simón Castellano (dir.), *Tratado sobre el Reglamento de Inteligencia Artificial de la Unión Europea* (p. 535-564). Aranzadi.

Stahl, Bernd Carsten, Antoniou, Josephina, Bhalla, Nitika, Brooks, Laurence, Jansen, Philip, Lindqvist, Blerta, Kirichenko, Alexey, Marchal, Samuel, Rodrigues, Rowena, Santiago, Nicole, Warso, Zuzanna, i Wright, David. (2023). A systematic review of artificial intelligence impact assessments. *Artificial Intelligence Review*, *56*. https://doi.org/10.1007/s10462-023-10420-8

UNESCO. (2021). Recommendation on the ethics of artificial intelligence. Conferència General de la UNESCO, 41a sessió.