



IA urbana ètica a la pràctica

Mecanismes polítics per establir marcs de governança locals

Marta Galceran-Vercher i Alexandra Vidal D'oleo (eds.)

CIDOB CIUTATS GLOBALS

BARCELONA
CENTRE FOR
INTERNATIONAL
AFFAIRS



Ajuntament de
Barcelona

IA urbana ètica a la pràctica

Mecanismes polítics per establir marcs de governança locals

Marta Galceran-Vercher i Alexandra Vidal D'oleo (eds.)

CIDOB
BARCELONA
CENTRE FOR
INTERNATIONAL
AFFAIRS



© 2024 CIDOB

CIDOB edicions
Elisabets, 12
08001 Barcelona
Tel.: 933 026 495
www.cidob.org
cidob@cidob.org

Impressió: QP Print Global Services
ISBN: 978-84-18977-29-9
Dipòsit Legal: B 3198-2025

Barcelona, desembre 2024

Foto de la coberta:

Vehicles en una via envoltada d'edificis alts. 10th Ave.

<https://unsplash.com/photos/vehicles-on-pathway-surrounded-by-high-rise-buildings-3UnFTjhsWWA>

SUMARI

SOBRE ELS AUTORS	5
RESUMS PART I	9
INTRODUCCIÓ. IA URBANA ÈTICA A LA PRÀCTICA	11
Tanya Álvarez, Marta Galceran-Vercher i Alexandra Vidal D'oleo	
PART I. OPERACIONALITZACIÓ DE PRINCIPIS ÈTICS EN ENTORNS	17
Shazade Jameson	
Rendició de comptes i transparència en la IA urbana	19
Josuan Eguiluz Castañeira i Carlos Fernández Hernández	
Privacitat i governança de dades en la IA urbana	29
Leandry Junior Jieutsa	
Equitat i no-discriminació en la IA urbana	41
María Pérez-Ortiz	
Sostenibilitat en la IA urbana	51
PART II. ESTUDIS DE CAS DE MARCS DE GOVERNANÇA DE LA IA URBANA	63
Alexandra Vidal D'oleo	
Estudi de cas 1: Barcelona	66
Estudi de cas 2: Àmsterdam	68
Estudi de cas 3: Nova York	71
Estudi de cas 4: San José	73
Estudi de cas 5: Dubai	75
Estudi de cas 6: Singapur	76
CONCLUSIONS. MECANISMES POLÍTICS, REPTES I RECOMANACIONS EN LA IA URBANA	79
Marta Galceran-Vercher i Alexandra Vidal D'oleo	

SOBRE ELS AUTORS

Tanya Álvarez

És investigadora de l'Observatori del Mobile World Capital Foundation. Dirigeix investigacions sobre la inclusió digital i l'ús de sistemes de presa de decisions automatitzades en el sector públic. Defensa una perspectiva interdisciplinària de l'impacte de la tecnologia sobre la societat. També té un grau en Història de l'Art pel Swarthmore College i un màster en Gestió del Patrimoni Cultural per la Universitat de Barcelona.

Josuan Eguiluz Castañeira

Té un grau en Dret i TIC per la Universitat de Deusto i un màster doble en Accés a la Professió Jurídica i Propietat Intel·lectual, Noves Tecnologies i Protecció de Dades per Esade. Va començar la seva carrera professional com a advocat al Departament de Propietat Intel·lectual, Noves Tecnologies i Protecció de Dades de Cuatrecasas. Actualment, exerceix com a assessor jurídic a Adevinta. També és professor de Dret Digital a Esade, Deusto i The Legal School, entre altres. Cursa un doctorat industrial en IA, àrees d'alt risc i drets fonamentals a la Universitat Pompeu Fabra, el Barcelona Supercomputing Center i Infojobs/Adevinta.

Carlos Fernández Hernández

Té un grau en Dret i actualment fa el doctorat a la Universitat Carlos III de Madrid. Després d'exercir com a advocat durant una sèrie d'anys, ha dedicat la seva carrera professional a publicar continguts jurídics per a professionals del sector, sempre sota el segell La Ley. Des de 2013, està especialitzat en el camp del dret tecnològic, a l'inici al portal *Noticias Jurídicas* i des de 2016 a *Diario La Ley Ciberderecho*, en què ha publicat més de 1.500 articles, informes i entrevistes. Hernández també ha estat editor de les revistes *Derecho Digital e Innovación* i *La Ley Privacidad*. Ha impartit conferències especialitzades sobre aquests temes en diferents universitats i centres de postgrau i ha fet contribucions a diversos llibres col·laboratius.

Marta Galceran-Vercher

És investigadora sènior del Programa de Ciutats Globals del CIDOB (Barcelona Centre for International Affairs) i professora associada en relacions internacionals a la Universitat Pompeu Fabra i a CEI-International Affairs. És politòloga, té un doctorat en Relacions Internacionals per la Universitat Pompeu Fabra i un màster (amb menció especial) per la Universitat de Warwick. La seva recerca es centra en la diplomàcia de les ciutats, les aliances urbanes i la governança global, especialment en el camp de les transformacions digitals i l'humanisme tecnològic. És autora de diversos documents polítics, articles acadèmics i altres publicacions sobre aquests temes. També dirigeix la recerca del GOUAI, dins

la Coalició de Ciutats per als Drets Digitals. Abans de treballar al CIDOB, va ser coordinadora de programes del Congrés Mundial Smart City Expo i consultora sènior a Anteverti, on assessorava governs locals i organitzacions internacionals (l'ONU-Habitat, el BID, la Comissió Europea, etc.) sobre internacionalització i polítiques innovadores urbanes.

Shazade Jameson

És consultora sènior en governança digital per a organitzacions d'interès públic. Està especialitzada en administracions públiques urbanes. Com a investigadora en ciències socials, proposa noves perspectives en matèria de política digital i tendeix ponts per a una aplicació efectiva. Va codirigir l'informe «AI and Cities: Risks, Applications and Governance», editat per l'ONU-Habitat i el Mila Quebec AI Institute. També és estudiant de doctorat i analitza la política tecnològica de Singapur, en el marc del projecte Justícia Mundial de Dades, finançada pel Consell Europeu de Recerca de l'Institut de Dret, Tecnologia i Societat de Tilburg. Ha col·laborat amb el CIDOB en la investigació de polítiques per al GOUAI i treballa amb la iniciativa de la UIT United Smart Sustainable Cities, dins la temàtica de transformació digital per a les ciutats centrades en les persones. Des de 2024, treballa amb la Unitat d'Ètica de la IA de la UNESCO, facilitant el disseny i la implementació de la IA ètica en les administracions públiques.

Leandry Junior Jieutsa

És consultor en innovació urbana i investigador sobre la governança de la IA a les ciutats. Al llarg de la seva trajectòria professional ha exercit diferents papers en planificació i desenvolupament urbà, motivat per la seva passió per integrar tecnologies en l'urbanisme. Treballa com a consultor independent en innovació urbana i governança digital. Jieutsa també és el fundador de la xarxa Africa Innovation Network, un centre d'estudis dedicat als problemes urbans. Actualment amplia els seus coneixements com a estudiant de doctorat a la Càtedra UNESCO de Paisatge Urbà de la Universitat de Montreal, on s'especialitza en governança de IA a les ciutats. La seva recerca es centra en l'impacte de la governança de la IA sobre el benestar de les persones a les ciutats. Amb el seu treball, pretén ajudar els ajuntaments a construir ciutats intel·ligents més centrades en les persones mitjançant una governança de la IA responsable.

Maria Pérez-Ortiz

És professora adjunta del Centre d'IA del Departament d'Informàtica de la University College London (Regne Unit). També exerceix d'investigadora sènior a l'Oficina d'Assumptes Exteriors, de la Commonwealth i de Desenvolupament del govern britànic, on assessorava sobre les inversions en IA per al desenvolupament internacional. El 2022 va cofundar el primer programa de màster en IA per al Desenvolupament Sostenible, en la cruïlla de les tecnologies emergents d'IA, la sostenibilitat i l'ètica. Pérez-Ortiz també és adjunta de la Càtedra de la UNESCO sobre IA. El seu darrer informe polític «Challenging systematic prejudices: an investigation into bias against women and girls in large language models» mostra el grau en el qual els models del llenguatge presenten biaixos de gènere. La seva línia de treball actual és la IA responsable i com aquestes noves tecnologies poden ajudar els responsables polítics en escenaris complexos com el canvi climàtic.

Alexandra Vidal D'oleo

És investigadora i gestora de projectes del Programa de Ciutats Globals del CIDOB (Barcelona Centre for International Affairs). La seva recerca gira al voltant de la digitalització de les ciutats i els processos ètics i democràtics del desenvolupament de la intel·ligència artificial (IA) urbana. Actualment, forma part de l'equip de recerca de l'Observatori Global de la Intel·ligència Artificial Urbana (GOUAI, per la sigla en anglès) en el marc de la Coalició de Ciutats pels Drets Digitals. Té un postgrau en Planificació Territorial amb una menció en urbanisme sostenible per la Universitat Politècnica de Catalunya i un grau en Relacions Internacionals per Blanquerna - Universitat Ramon Llull, en el marc del qual va cursar part dels seus estudis a l'École de Gouvernance et d'Économie de Rabat. Abans de treballar al CIDOB, era investigadora i gestora de projectes a la Tanja Foundation, on desenvolupava projectes de cooperació entre el Marroc i Espanya.

Rendició de comptes i transparència en la IA urbana

Shazade Jameson

Moltes administracions públiques urbanes veuen el potencial que ofereix la implementació de la intel·ligència artificial (IA), però no se senten preparades per fer-ho de manera responsable. Tot i que els marcs i els enfocaments basats en una IA responsable estan cada cop més generalitzats, molts d'aquests se centren en el sector privat o en la indústria. Tot i que la IA responsable cada vegada rep més atenció, existeixen moltes menys directrius específiques per als governs locals. Aquest capítol pretén contribuir a resoldre aquesta mancança presentant definicions de rendició de comptes i transparència que incloguin tant perspectives tècniques restrictives com sociopolítiques més àmplies. Per facilitar la rendició de comptes i la transparència en el context de la implementació de la IA per part de les administracions públiques urbanes, hi ha dues preguntes aparentment senzilles però fonamentals que haurien de guiar la fase de disseny: «Cal usar la IA?» i «Com s'ha d'usar la IA?». Després d'haver reflexionat sobre el que impliquen aquestes qüestions per als professionals urbans, el capítol presenta un resum dels mecanismes polítics existents que poden adaptar-se en pro d'aquests objectius i assenyala algunes lliçons apreses de recerques anteriors.

Privacitat i governança de dades en la IA urbana

Josuan Eguiluz Castañeira i Carlos Fernández Hernández

El desenvolupament de la IA ètica a Europa, tal com la concep la Llei d'IA de la UE, ha d'incloure mecanismes robustos per a la privacitat i la gestió de les dades. En el context dels entorns públics urbans, el tractament de dades personals a través de sistemes d'IA presenta reptes específics que les autoritats públiques hauran d'abordar acuradament. Tenint en compte aquest nou marc legislatiu, aquest article té els objectius següents: (i) presentar els marcs jurídic i ètic que regulen el tractament de dades personals a través dels sistemes d'IA en entorns urbans, (ii) resumir els mecanismes clau per implementar el principi de privacitat i (iii) analitzar els reptes vinculats a aquestes pràctiques de tractament de dades i aportar una sèrie de recomanacions i bones pràctiques al respecte.

Equitat i no-discriminació en la IA urbana

Leandry Junior Jieutsa

La IA és una tecnologia emergent, disruptiva i ambivalent. Com a part del seu desplegament, les ciutats han de posar en marxa diversos mecanismes per garantir que aquesta tecnologia té el menor impacte negatiu possible sobre les persones i les comunitats. Aquest capítol formula recomanacions polítiques per integrar els aspectes de l'equitat i la no-discriminació en el desplegament de la IA per part de les ciutats. El primer apartat tracta dels conceptes d'equitat i no-discriminació en entorns urbans i presenta els factors que determinen la IA equitativa i no discriminatòria. El segon apartat analitza les oportunitats i els impactes de la IA a les ciutats i, per últim, el tercer apartat proposa recomanacions polítiques per a una IA més equitativa a les ciutats. Aquestes recomanacions tenen en compte els diferents papers que les ciutats poden tenir en el desplegament de la IA, com a desenvolupadores de solucions internes, implementadores i reguladores. Les ciutats han de ser àgils i, per això, s'han de basar en la participació, els enfocaments locals, la innovació sociotecnològica, la col·laboració, etc.

Sostenibilitat en la IA urbana

María Pérez-Ortiz

El capítol examina el potencial de la IA per fomentar el desenvolupament de ciutats sostenibles, abordant les dimensions social, mediambiental i econòmica de la sostenibilitat. Atès que la urbanització es va accelerant globalment, les ciutats s'enfronten a reptes cada cop majors en camps com la mobilitat, l'habitatge, la contaminació i la gestió dels recursos. La IA promet optimitzar la infraestructura urbana, reduint les emissions i millorant l'eficiència dels recursos. No obstant això, el seu desplegament també planteja problemes en termes d'equitat social, impacte mediambiental i perturbacions econòmiques. Es proposa la IA sostenible com a marc per concordar el desenvolupament i l'aplicació de la IA amb els objectius de sostenibilitat, garantint que opera dins els límits ecològics, que promou la inclusió i potencia el creixement econòmic circular i equitatiu. Un dels temes que rep una atenció especial és la reducció de la petjada de carboni de la IA a través de pràctiques eficients energèticament, incorporant l'equitat en els sistemes urbans impulsats per IA i garantint una governança transparent. El text aporta recomanacions polítiques per orientar el desplegament de la IA en entorns urbans, emfasitzant la col·laboració internacional, la governança ètica i les polítiques econòmiques per potenciar la resiliència i la inclusió a les ciutats del demà.

INTRODUCCIÓ. IA URBANA ÈTICA A LA PRÀCTICA

Tanya Álvarez

Investigadora, Observatori del Mobile World Capital Foundation

Marta Galceran-Vercher

Investigadora sènior, Programa de Ciutats Globals, CIDOB

Alexandra Vidal D'oleo

Investigadora i gestora de projectes, Programa de Ciutats Globals, CIDOB

La intel·ligència artificial (IA) és una de les tecnologies més revolucionàries del nostre temps i que promet transformar completament la societat. Aquesta transformació es produeix en diferents nivells i àmbits, i l'espai urbà no n'és una excepció. Des de que es concebés el paradigma de la «ciutat intel·ligent», empreses tecnològiques y actors municipals recorren cada cop més als avenços tecnològics per resoldre els reptes urbans més apressants als quals s'enfronten les nostres societats. En aquest procés, el desplegament de sistemes algorítmics per part dels governs locals es va generalitzant, fet que va configurant el procés de «fer ciutat» tal com l'entendem.

En aquest context, la «IA urbana» es pot definir com la relació entre els sistemes d'IA i el medi urbà. Aquests sistemes, juntament amb altres tecnologies, es van incrustant a tot tipus d'emplaçaments urbans, com ara habitatges, llocs de feina, espais públics i infraestructures. Així mateix, la digitalització d'aquestes experiències urbanes crea un entorn híbrid en el qual les tecnologies digitals tenen un paper a l'hora d'influir i augmentar l'experiència urbana (Aurigi i De Cindio, 2008). Els residents de les ciutats, tot just comencen a percebre com la IA, com a element immers als entorns urbans, té un efecte profund en la vivència urbana i en el propi procés de «fer ciutat».

Pel que fa als sistemes d'IA i l'automatització, les ciutats són un banc de proves ideal per desplegar aquestes tecnologies. El desenvolupament i la implementació de la IA requereixen un ventall de recursos que es poden trobar fàcilment en escenaris urbans: 1) un entorn físic sobre el qual actuar, 2) accés a activitats diverses, 3) dades abundants i d'alta qualitat i 4) infraestructures i instal·lacions (Cugurullo *et al.*, 2023). A la darrera dècada, ha sorgit una ingent quantitat de tecnologies basades en dades que aborden reptes urbans tals com el manteniment de la infraestructura, els serveis públics personalitzats, la salut, la millora del transport, l'urbanisme i l'ús eficient dels recursos, entre altres (Galceran-Vercher i Vidal, 2024).

Els residents de les ciutats, tot just comencen a percebre com la IA, com a element immers als entorns urbans, té un efecte profund en la vivència urbana i en el propi procés de «fer ciutat».

A mesura que les tecnologies algorítmiques es van generalitzant, urgeix que les administracions locals vetllin per un ús responsable i ètic d'aquests sistemes.

Cada cop hi ha més governs municipals que són conscients dels beneficis que la IA aporta a l'administració i a la prestació de serveis. Els sistemes d'IA s'adopten amb l'esperança que alleugin les càrregues rutinàries, automatitzant les tasques burocràtiques, cosa que permet als governs locals funcionar de manera més eficient. També pretenen ser més rendibles al prendre decisions més intel·ligents basades en dades per així deixar més temps als governs locals per respondre millor a les necessitats de la ciutadania.

No obstant això, a mesura que les tecnologies algorítmiques es van generalitzant, urgeix que les administracions locals vetllin per un ús responsable i ètic d'aquests sistemes. Cal destacar que la rellevància cada cop més gran que els governs locals van adquirint en l'escenari polític global, juntament amb el seu potencial impacte sobre la vida de milions de persones, requereix que la governança de la IA tingui en compte l'impacte sobre les persones, les comunitats i el medi ambient.

Amb l'aparició de la regulació sobre la IA, la governança de la IA a les ciutats ha esdevingut una preocupació especial per als defensors dels drets humans, les organitzacions de la societat civil i les minories urbanes, que han estat testimonis dels potencials perills del desplegament dels sistemes d'IA. Per exemple, tot i que la vigilància mitjançant la IA proporciona a les ciutats solucions en matèria de seguretat i protecció, gestió del trànsit o monitoratge de factors mediambientals, s'ha demostrat que poden ser solucions invasives i discriminatòries amb determinats sectors de la població. Aquest exemple en concret, entre altres, ha fet saltar les alarmes sobre el fet que la millora en eficiència gràcies a l'automatització pot implicar grans costos i, per tant, les administracions municipals han de ser conscients de les implicacions ètiques dels sistemes d'IA que pretenen implementar.

El repte de posar en pràctica principis ètics en la IA urbana

Tot i que les ciutats poder estar preocupades pels beneficis tècnics i operatius que promet la IA, els experts sostenen que, com a sistemes sociotecnològics, l'impacte de la IA va més enllà de la mateixa precisió tècnica dels sistemes. Per tant, els administradors i els responsables polítics locals que només se centren en la precisió tècnica o en l'equitat d'un sistema no aborden les implicacions més generals que aquests sistemes poden tenir a l'hora de garantir que es despleguen de manera responsable i ètica. Implementar la IA responsable va més enllà del desplegament de sistemes que produeixen uns resultats adequats o fiables. La IA ètica i responsable emfasitza la importància de l'ètica al llarg de tot el cicle de vida dels sistemes i assegura que les eines algorítmiques estan en consonància amb els valors democràtics i la protecció dels drets digitals de les persones.

És lloable que moltes ciutats ja hagin començat a implementar i desenvolupar mecanismes polítics d'IA responsable. La ciutat de Nova York, per exemple, ha implantat auditories obligatòries per a les eines de contractació de personal. A Finlàndia, tres ciutats s'han unit per fomentar la transparència de la IA a través de registres d'algoritmes, que posen la informació a disposició dels residents. Un altre exemple és la Policia de Toronto, que ha establert una política de compra per a les tecnologies d'IA. Així mateix, una investigació recent portada a terme per l'Observato-

ri Global de la IA Urbana (**GOUAI**, per les sigles en anglès) conclou que les ciutats d'arreu del món habitualment fomenten altres mecanismes polítics per incentivar els sistemes d'IA responsable, com ara el desenvolupament de directrius i principis concrets, prohibicions o moratòries de determinats sistemes algorítmics d'alt risc (per exemple, sistemes de reconeixement facial en temps real), registres públics d'algoritmes, auditories i avaluacions d'impacte, creació d'organismes supervisors independents i externs o clàusules d'adquisició pública que garanteixin el respecte pels drets humans. Aquestes experiències poden servir com a full de ruta a altres agents del sector públic per entendre quin tipus de mecanismes polítics funcionen a l'hora de desenvolupar sistemes d'IA responsable.

No obstant això, tot i que hi ha algunes ciutats que van donant els primers passos per desenvolupar pràctiques d'IA responsable, cada vegada és més necessari que els administradors de les ciutats i els responsables polítics municipals entenguin com es va desenvolupant la IA urbana i quines bones pràctiques s'han posat en marxa. L'esmentat GOUAI atén aquesta necessitat. Aquest projecte conjunt està liderat pel CIDOB, amb el suport de les ciutats de Barcelona, Amsterdam i Londres, i la **Coalició de Ciutats pels Drets Digitals** i ONU-Habitat. Amb aquest objectiu, l'Atlas del GOUAI recopila exemples globals d'IA urbana que estan en concordança amb els sis principis ètics següents: transparència i obertura, protecció de la privacitat, equitat i no-discriminació, seguretat i ciberseguretat, rendició de comptes i sostenibilitat. Una publicació recent basada en una anàlisi de l'Atlas (Galceran-Vercher i Vidal, 2024) va revelar que en la tendència cada cop major a l'ús de la IA urbana hi ha disparitats entre les ciutats que han adoptat eines algorítmiques i les que han establert polítiques o estratègies que garanteixen que la IA està en consonància amb els principis ètics.

En aquest context, aquesta monografia del CIDOB aprofundeix en els marcs de governança existents i en els mecanismes polítics específics per posar en pràctica principis ètics concrets i fomentar una IA urbana responsable sobre el terreny. L'objectiu és crear un document útil que pugui inspirar a l'acció i que serveixi com a full de ruta per a altres actors del servei públic.

Estructura de la publicació

La primera part d'aquesta publicació, que consta de quatre capítols, analitza com els principis ètics clau –és a dir, 1) la rendició de comptes i la transparència, 2) la privacitat i la governança de dades, 3) l'equitat i la no-discriminació, com també 4) la sostenibilitat– poden aplicar-se a la pràctica en entorns urbans a través de mecanismes polítics específics. Al respecte, **Shazade Jameson** argumenta que els governs locals no tenen orientacions clares per fomentar la IA urbana ètica en la seva jurisdicció, i presenta dues definicions pràctiques de rendició de comptes i transparència que inclouen tant el punt de vista tècnic com la perspectiva sociopolítica més general. L'autora afirma que, per tal de potenciar la rendició de comptes i la transparència en la implementació de la IA urbana, hi ha dues preguntes aparentment senzilles però alhora fonamentals que haurien de guiar la fase de disseny: «Cal usar la IA?» i «Com s'ha d'usar la IA?».

La IA ètica i responsable emfasitza la importància de l'ètica al llarg de tot el cicle de vida dels sistemes i assegura que les eines algorítmiques estan en consonància amb els valors democràtics i la protecció dels drets digitals de les persones.

Una investigació recent portada a terme per l'Observatori Global de la IA Urbana conclou que les ciutats d'arreu del món habitualment fomenten altres mecanismes polítics per incentivar els sistemes d'IA responsable.

En el segon capítol, **Leandry Junior Jieutsa** analitza els factors que contribueixen a la IA equitativa i no-discriminatòria. L'autor identifica dos factors principals que generen discriminació en els sistemes d'IA: els biaixos algorítmics i l'ús de tecnologies d'IA. Per això, aporta recomanacions polítiques amb l'objectiu de crear ciutats més equitatives impulsades amb IA i emfasitza la necessitat que els governs locals s'adaptin a les seves diverses funcions com a desenvolupadors, implantadors i reguladors. Sosté que les ciutats han de recórrer als processos participatius, els enfocaments locals, la innovació sociotecnològica i la *col·laboració* intersectorial per garantir un desplegament responsable i equitatiu de la IA.

A continuació, **Josuan Eguiluz Castañeira** i **Carlos Fernández Hernández** revisen els mecanismes per a la privacitat i la gestió robustes de les dades en el desplegament de la IA. Analitzen els marcs jurídics i ètics que regeixen el tractament de dades personals per part dels sistemes d'IA i se centren en la Llei d'IA de la UE. Els autors resumeixen els mecanismes clau per aplicar el principi de privacitat als entorns urbans, analitzen els reptes relacionats amb aquestes pràctiques de tractament de dades i ofereixen un seguit de recomanacions factibles. També destaquen que la governança de dades ha d'estar al centre de les estratègies de la IA urbana, prioritzant la qualitat, la rellevància i la protecció dels conjunts de dades usats en els sistemes d'IA, cosa que inclou dur a terme avaluacions d'impacte per protegir tant les dades personals com els drets fonamentals i així garantir que la privacitat i la seguretat dels ciutadans no es veuen compromeses.

Per acabar, el capítol de **María Pérez-Ortiz** analitza el potencial que té la IA per contribuir al desenvolupament de les ciutats sostenibles, per a la qual cosa aborda les dimensions social, mediambiental i econòmica de la sostenibilitat. L'autora sosté que, tot i que la IA és força prometedora, el seu desplegament també planteja problemes d'equitat, impacte mediambiental i perturbacions econòmiques. En aquest context, el marc de la IA sostenible constitueix una valuosa eina per concordar el desenvolupament i l'aplicació de la IA amb els objectius de sostenibilitat, la qual cosa garanteix que la IA operi dins els límits ecològics, potenciï la inclusió i fomenti el creixement econòmic equitatiu i circular.

La segona part de la publicació presenta sis estudis de cas com a exemples de marcs de governança de la IA local establerts per ciutats d'arreu del món que han adoptat mecanismes polítics concrets per aplicar la IA urbana ètica a la pràctica. En concret, en aquest apartat, **Alexandra Vidal D'oleo** avalua la governança d'IA de les ciutats de Barcelona, Amsterdam, Nova York, San José, Singapur i Dubai.

El volum es tanca amb un capítol final en el qual **Marta Galceran-Vercher** i **Alexandra Vidal D'oleo** presenten unes conclusions on categoritzen els mecanismes polítics derivats dels capítols, els estudis de cas i la literatura que s'han revisat. Aquesta anàlisi explora els mecanismes polítics més generalitzats i examina com concorden amb els diferents principis ètics. Les autores també identifiquen les tendències comunes i els reptes als quals s'enfronten les ciutats quan intenten aplicar aquests principis ètics a la pràctica i ofereixen un conjunt de recomanacions generals per abordar-los.

Referències bibliogràfiques

Aurigi, A. i De Cindio, F. «Augmented Urban Spaces: Articulating the Physical and Electronic City». Farnham: Ashgate Publishing, 2008

Cugurullo F. *et al.* «Artificial Intelligence and the City. Urbanistic perspectives on AI». Londres: Routledge, 2023

Galceran-Vercher, M. i Vidal, A. «Mapatge de la intel·ligència artificial urbana: primer informe de l'Atles de la Intel·ligència Artificial Urbana del GOUAI». *CIDOB Briefings*, n.º 56, 2024

PART I. OPERACIONALITZACIÓ DE PRINCIPIS ÈTICS EN ENTORNS URBANS

- RENDICIÓ DE COMPTES I TRANSPARÈNCIA
EN LA IA URBANA

Shazade Jameson

- PRIVACITAT I GOVERNANÇA DE DADES
EN LA IA URBANA

*Josuan Eguiluz Castañeira, Carlos Fernández
Hernández*

- EQUITAT I NO-DISCRIMINACIÓ EN LA IA URBANA

Leandry Junior Jieutsa

- SOSTENIBILITAT EN LA IA URBANA

María Pérez-Ortiz

Shazade Jameson

Consultora sènior en governança digital

1. Introducció

En aquest volum definim la intel·ligència artificial (IA) urbana ètica com la implementació d'enfocaments d'IA responsable al si de les administracions públiques urbanes. Per tant, qualsevol debat sobre la IA responsable ha de tenir en compte les necessitats concretes i les particularitats de les administracions públiques i els seu electorat.

Les administracions públiques urbanes són un context particular, ja que representen l'interès públic i se circumscriuen en gran mesura a l'esfera local. Així, són un entorn especialment interessant i complex, atès que els toquen de prop les complexitats locals i alhora estan prou allunyades de les estratègies nacionals. A més, presenten una gran diversitat quant a mida i capacitats.

Consegüentment, tot i que aquestes administracions poden aprofitar moltes idees dels enfocaments d'IA ètica i IA responsable, la possibilitat de reutilitzar-los pot ser limitada perquè requereixen una perspectiva molt més àmplia del que permeten en gran mesura els recursos disponibles localment. Molts enfocaments d'IA responsable s'emmarquen en la governança empresarial i estan orientats al context industrial; per exemple, com poden les empreses aplicar la IA als seus productes i serveis de manera responsable? Però les administracions públiques urbanes tenen un model de negoci diferent, ja que se suposa que se centren per sobre de tot en l'interès públic.

Cada vegada hi ha més interès pel paper dels enfocaments d'IA responsable en el sector públic (vegeu, per exemple, OECD, 2024). No obstant això, l'assessorament sobre aquesta matèria adreçat concretament als governs locals és molt limitat, sobretot des d'un punt de vista global. Aquest capítol pretén contribuir a superar aquesta limitació presentant definicions dels principis de rendició de comptes i transparència en el context de la implementació de la IA per part de les administracions públiques urbanes i resumint els mecanismes polítics existents que poden adaptar-se per aconseguir-ho.

2. Principis de rendició de comptes i transparència

Cada vegada hi ha més interès pel paper dels enfocaments d'IA responsable en el sector públic. No obstant això, l'assessorament sobre aquesta matèria adreçat concretament als governs locals és molt limitat, sobretot des d'un punt de vista global.

2.1. Rendició de comptes

El concepte de la rendició de comptes es pot explicar tant amb definicions restrictives com amb definicions més generals. És important que els governs locals tinguin en compte ambdós tipus de definicions, atesa la seva naturalesa d'ens públics.

En essència, la rendició de comptes és una forma de relació. La teoria més acceptada sobre la rendició de comptes a l'administració pública (Bovens, 2007) és que la rendició de comptes és la *relació* entre un actor i un fòrum, el qual té l'autoritat per rebutjar-la. La rendició de comptes ha d'especificar *respecte de què* i *davant qui* es porta a terme. Com a relació, és un procés social que requereix compromís social i una visió compartida de la societat (Wieringa, 2020).

Respecte de què cal retre comptes sovint es determina a través de les normes procedimentals i substancials de l'administració pública i la capacitat per avaluar si aquestes normes s'han complert. El *davant qui* és summament important per als governs locals, i pot referir-se a diferents grups de públics: el patrocinador, l'interessat, els ciutadans afectats, etc. Atès que les administracions públiques han de considerar l'interès públic, el grup d'interessats i de responsables és molt més ampli (Jameson *et al.*, 2021). Alguns usos concrets de la IA poden afectar la rendició de comptes de naturalesa política, com l'escàndol que va motivar la dimissió en bloc del Govern dels Països Baixos arran d'unes ajudes a famílies per tenir cura dels fills (Dachwitz, 2022; Amaro, 2021).

Quan els governs locals dissenyen aplicacions de la IA adreçades als ciutadans, és important que col·laborin amb les comunitats afectades des de la fase de disseny del projecte (vegeu, per exemple, UN-Habitat i Mila Quebec AI Institute, 2022). Alguns marcs d'IA responsable tenen un abast limitat i poden estar mal preparats per satisfer les demandes del procés participatiu més ampli que es requereix en una administració pública. En concret, la codificació dels biaixos i les desigualtats en els algoritmes com a forma de governança planteja que cal preveure noves vies per fer arribar disputes i comentaris a l'hora de portar a terme la reestructuració organitzativa al voltant de la governança de la IA (Taylor, 2021).

2.2. Transparència

El principi de transparència en relació amb la IA presenta múltiples vessants. Igual que la rendició de comptes, aquest mecanisme fa anys que està establert als àmbits de l'administració pública, l'enginyeria de programari i la informàtica.

Des del punt de vista tècnic, la transparència té a veure amb la publicació de la informació relativa al sistema algorítmic al llarg del seu cicle de vida i permet dur a terme investigacions independents i revisar com s'usen els models i la seva qualitat, incloent-hi les finalitats del disseny, les fonts de les dades, els requisits quant a *hardware*, les condicions de funcionament, el rendiment esperat del sistema i —cosa que és impor-

tant en els sistemes algorítmics— la relació entre les variables del model i l'arquitectura, així com les característiques de les dades amb les quals aquest s'ha entrenat. La transparència implica documentar el procés de selecció dels conjunts de dades, les variables i els indicadors de qualitat per al desenvolupament del sistema.

La procedència (és a dir, l'origen de les dades) i la qualitat de les dades d'entrenament són molt importants a l'hora d'implementar la IA a les administracions públiques. És un factor limitant important per a la qualitat dels models algorítmics i la principal font de biaixos quan s'implementa la IA en les administracions públiques (UN-Habitat i Mila Quebec AI Institute, 2022; Longpre *et al.*, 2023). La transparència és un principi general en el camp de la IA explicable, que inclou els conceptes d'explicabilitat i interpretabilitat, els quals s'han popularitzat en poc temps com a mecanismes de transparència i rendició de comptes des dels punts de vista tant tècnic com sociopolític. L'objectiu general d'aquest camp és obrir la «caixa negra» dels algoritmes tancats que no revelen com funcionen internament (Adadia i Berrada, 2018).

Existeixen enfocaments diferents per a l'explicabilitat des del punt de vista tècnic que, a grans trets, es classifiquen en quatre categories (Wierenga, 2020). El primer consisteix a explicar el model, per exemple, donant indicacions clares sobre quin procediment segueixen els models algorítmics i fins a quin punt aquests es poden explicar a una persona no especialista utilitzant un llenguatge planer. El segon és explicar el resultat, la qual cosa significa aprofundir en les decisions concretes que prenen els algoritmes i si els mecanismes per prendre aquestes decisions es poden entendre i avaluar. El tercer es basa a inspeccionar la caixa negra, amb tècniques com ara la visualització del funcionament intern de l'algoritme, entre moltes altres. Per acabar, crear una caixa transparent és un principi de disseny que usa predictors explícits i visibles. En conjunt, el repte per a la transparència des del punt de vista tècnic és arribar a una solució de compromís entre la interpretabilitat i la precisió.

Deixant de banda les qüestions tècniques, la transparència també inclou una vessant sociopolítica, la qual es basa a donar a conèixer com s'usen els sistemes algorítmics, quines decisions de disseny s'han pres i qui les ha pres, i explicita les premisses relacionades amb la governança. D'aquesta manera, la transparència és una condició que afavoreix la rendició de comptes algorítmica, atès que esbossa vies per a la impugnació.

La transparència també inclou una vessant sociopolítica, la qual es basa a donar a conèixer com s'usen els sistemes algorítmics, quines decisions de disseny s'han pres i qui les ha pres, i explicita les premisses relacionades amb la governança.

2.3. Interrelació entre la transparència i la rendició de comptes

Els dos principis de transparència i rendició de comptes estan interrelacionats. Les solucions per a la rendició de comptes sovint es basen en un principi de transparència, que després s'ha d'emmarcar en un context institucional perquè es puguin crear relacions de rendició de comptes.

Per exemple, els registres algorítmics són eines per a la rendició de comptes. A la pràctica es basen a fer que la informació sobre els algoritmes i el seu ús sigui transparent i accessible de franc des d'un registre (Jameson i Leal, 2022; Cath i Jansen, 2021). D'aquesta manera, la transparència és un mitjà per a l'avaluació de la rendició de comptes en el disseny d'un sistema algorítmic.

Els dos principis de transparència i rendició de comptes estan interrelacionats. Les solucions per a la rendició de comptes sovint es basen en un principi de transparència, que després s'ha d'emmarcar en un context institucional perquè es puguin crear relacions de rendició de comptes.

No obstant això, tot i que la transparència pot ser una condició sine qua non per a la rendició de comptes, no és suficient per si sola. Per exemple, el simple fet que un sistema algorítmic estigui ben documentat i sigui transparent no explica per què es va decidir avaluar-lo com a “prou bo” per a l'objectiu en qüestió, qui ho va decidir ni qui va participar en el procés. Mentre que la transparència pot funcionar de manera passiva, la rendició de comptes és més activa: inclou no només com funciona un sistema, sinó també per què (Wierenga, 2020).

3. Implementació de la IA responsable a les administracions públiques urbanes

En considerar un ús responsable de la IA, hi ha dues preguntes fonamentals que les administracions públiques s'haurien de plantejar: «cal usar la IA?» i «com s'ha d'usar la IA?». Donar respostes clares a aquestes preguntes aparentment senzilles pot ser una de les maneres més efectives per assolir la transparència i la rendició de comptes, perquè fan visibles les premisses fonamentals. Aquest procés també requereix temps i recursos.

3.1. Cal usar la IA?

La IA no és neutra, sinó que més aviat incorpora i reforça les premisses establertes a les seves dades i el seu disseny. Si la IA no es dissenya amb cura pel que fa a un conjunt de valors que fomentin l'interès públic, les estructures de la IA i la seva governança incorporaran altres valors de manera inconscient, cosa que suposa riscos (vegeu, per exemple, UN-Habitat i Mila, 2022). Per tant, la pregunta de si cal usar la IA no s'ha de plantejar a la lleugera. Per a una vertadera rendició de comptes, l'opció de deixar de fer servir la IA ha d'estar sobre la taula i el «no» ha de ser una possibilitat. Altrament, la rendició de comptes es veuria reduïda a un principi, al qual s'eludiria com a virtut i no com a relació funcional (Wierenga, 2020).

A més a més, la qüestió de «si cal» no és només de caire normatiu, sinó també operatiu. Les administracions públiques tenen un objectiu particular i la IA podria ser la millor manera per assolir-lo. Però no sempre: altres solucions tecnològiques o basades en dades poden ser més adequades. En concret, la IA i les aplicacions d'aprenentatge automàtic requereixen grans quantitats de dades de qualitat, de manera que quan aquestes condicions no es compleixen podria haver-n'hi prou amb anàlisis de dades més simples.

Els projectes municipals basats en dades sovint han de fer front a infraestructures heretades, sensors obsolets i bases de dades inoperatives. Això significa que, perquè una aplicació d'aprenentatge automàtic sigui reeixida en un context urbà, cal perllongar les fases de descobriment del projecte, en ocasions fins al 30%-40% de la durada del projecte. Aquest període inclou la investigació sobre el problema en qüestió, l'estat actual de la infraestructura i el conjunt de dades i el tipus de solució que podria ser més apropiada. Els pressupostos i les expectatives de les parts interessades han de permetre tenir en consideració aquesta fase exploratòria ampliada.

Degut a l'entusiasme i l'atenció que desperta l'aplicació de la IA, hi ha un risc important de caure en el tecnosolucionisme: és a dir, la concepció antiga que les eines fan les feines. De vegades, però, un enfocament més social o conductual pot ser més apropiat per resoldre el problema en qüestió. Tot sovint les solucions diferents responen a perspectives diferents, la qual cosa significa que la perspectiva des de la qual plantejem el problema determina on hem de cercar la solució. És a dir, la nostra concepció sobre el problema ja defineix el tipus de solucions que podem crear. Aquesta circumstància no és exclusiva de la IA, sinó que s'aplica a totes les interaccions persona-tecnologia de manera més general. Vegem-ho amb un exemple simple: si el problema és que els ascensors són massa lents, en comptes d'intentar-ne optimitzar la velocitat gràcies a innovacions mecàniques de l'àmbit de l'enginyeria, n'hi ha prou a col·locar-hi un mirall per fer més amè el trajecte. Per tant, una fase exploratòria més llarga també permet a les parts interessades fer-se una pregunta fonamental: quin problema volem resoldre?

La fase exploratòria més llarga inclou també una col·laboració estreta amb les parts interessades locals. Crear una IA reeixida gairebé sempre és una tasca col·laborativa i implica treballar amb universitats, instituts d'estudis i empreses locals, sobretot tenint en compte el problema de manca de capacitat dels ajuntaments. A Barcelona, per exemple, l'algoritme d'aprenentatge automàtic creat per assistir a la presa de decisions en el procediment d'admissió als centres d'acollida de serveis socials va ser el resultat d'una col·laboració estreta entre entitats per elaborar un algoritme bilingüe i pertinent en l'àmbit local (Jameson i Leal, 2022).

3.2. Com s'ha d'usar la IA?

Tot i que la IA té moltes aplicacions a les ciutats, a les administracions públiques la tendència és usar la IA per a dues funcions concretes: l'automatització de processos ja existents i les prediccions basades en dades.

L'automatització implica automatitzar una part dels procediments burocràtics o dels serveis urbans ja existents. En aquest cas, partim d'una lògica o un procés preexistent i una part de la cadena d'accions es durà a terme de manera més ràpida o més eficient amb l'assistència de la IA. A l'hora de considerar com aplicar la IA, el punt de partida és el sistema vigent.

Les prediccions basades en dades són un enfocament diferent, perquè el punt de partida és un altre: un munt de dades. A partir d'aquestes dades, els analistes de dades obtindran informació i partint d'aquesta informació l'Administració dissenyarà nous procediments burocràtics per als serveis urbans. Així, els models predictius constitueixen una nova lògica basada en dades a l'Administració (Kitchin, 2016).

Tot i que aquestes dues categories poden usar el mateix tipus d'IA des del punt de vista tècnic (per exemple, ambdues poden utilitzar tècniques d'aprenentatge automàtic o de reconeixement d'imatges), la manera d'integrar la IA en els processos de la ciutat és diferent, pel fet que canvia el tipus d'efectes que pot tenir i, per consegüent, la nostra percepció sobre la rendició de comptes i la transparència.

En considerar un ús responsable de la IA, hi ha dues preguntes fonamentals: «cal usar la IA?» i «com s'ha d'usar la IA?». Donar respostes clares a aquestes preguntes aparentment senzilles pot ser una de les maneres més efectives per assolir la transparència i la rendició de comptes.

Degut a l'entusiasme i l'atenció que desperta l'aplicació de la IA, hi ha un risc important de caure en el tecnosolucionisme. De vegades, però, un enfocament més social o conductual pot ser més apropiat per resoldre el problema en qüestió.

Per exemple, quan la IA s'utilitza per automatitzar processos burocràtics ja existents, aquests poden millorar-se afegint-hi mecanismes de rendició de comptes. Per exemple, una revisió trimestral pot millorar-se si es combina amb una avaluació d'impacte. Introduir altres innovacions en el procés pot optimitzar pràctiques organitzatives existents a fi de tenir en compte ensenyaments extrets de la integració de la IA, com els comentaris del funcionariat que participa en el procés, com també dels suggeriments dels ciutadans.

D'altra banda, l'ús de prediccions basades en dades requereix un enfocament una mica més complex de la transparència i la rendició de comptes perquè són una nova forma de creació de coneixement, que les administracions públiques tradicionals no estan preparades per processar. En concret, els models predictius canvien la funció del coneixement local i el lloc on s'aplica (Kitchin, 2016). Una persona que fa 20 anys que camina per aquells carrers, posem per cas, pot tenir una perspectiva diferent del que mostren les dades. El coneixement computacional és diferent del coneixement basat en l'experiència (Van Ewijk i Baud, 2009), i la presa de decisions assistida amb algorismes pot fer decantar la balança cap a un tipus de coneixement o l'altre.

Els processos de rendició de comptes requeriran un diàleg entre diferents concepcions, com la diferència entre el coneixement computacional i el coneixement basat en l'experiència. Com trobem sentit al problema urbà actual que ens ocupa? Aquest fet de «trobar sentit» o «crear significat» implica decidir com s'avaluen les diferents opcions polítiques i resultats socials i es pot dir que és quelcom en què la IA depèn totalment dels humans (Tan, 2024). Rumiar i redisenyar els processos de rendició de comptes i els mecanismes polítics suposa l'oportunitat d'avaluar les diferents formes de crear significat en joc per garantir que es fa un ús ètic de la IA a l'Administració pública.

4. Mecanismes polítics

Un enfocament sociotecnològic de la IA reconeix que el que passa amb un sistema d'IA depèn de la interacció entre aspectes tècnics i aspectes socials, entre el sistema i la manera en què s'integra en el seu context. Això significa que, per entendre com funciona un sistema algorítmic, és important entendre com interactua el sistema algorítmic amb el seu entorn i quins mecanismes hi poden tenir un impacte més gran.

Un sistema algorítmic es pot descriure a partir del cicle de vida de la IA, que és una manera resumida de descriure el procés de disseny, desenvolupament i desplegament de la IA. Entendre-ho és útil perquè molts models de gestió de riscos es basen en variacions d'aquest cicle de vida de la IA.

A continuació presentem diferents opcions de mecanismes polítics disponibles en diferents fases del cicle de vida de la IA. També hi ha mecanismes generals de governança institucional que poden tenir lloc durant tot el cicle de vida de la IA i que en poden constituir el context.

Enquadrament i disseny:

- Les **avaluacions d'impacte** solen concretar-se en qüestionaris per analitzar les possibles conseqüències socials i ètiques abans del desplegament d'un sistema d'IA. Hi ha molts tipus d'avaluacions d'impacte, com ara les que es basen en l'ètica, en la privacitat, en l'equitat, etc. Vegeu, per exemple, [l'eina d'avaluació de l'impacte ètic de la UNESCO](#).
- Les **clàusules d'adquisició pública** són estipulacions que els governs inclouen en els contractes quan adquireixen béns i serveis, en aquest cas serveis d'IA o relacionats amb la IA. Tot i que puguin semblar una mera formalitat burocràtica, aquestes clàusules poden esdevenir una palanca estratègica per als objectius d'interès públic, per exemple, definint normes d'auditabilitat. És el cas de la [Coalició GovAI](#), encapçalada per la ciutat californiana de San José, que ha creat plantilles de polítiques perquè les administracions públiques les reutilitzin, incloent-hi una fitxa tècnica sobre IA i un acord per a proveïdors que els obliga a complir determinats requisits relacionats amb el funcionament, els biaixos algorítmics, la supervisió humana, entre altres. La xarxa Eurocities també elabora [plantilles de clàusules d'adquisició](#) de conformitat amb la Llei d'IA de la UE.

Un enfocament sociotecnològic de la IA reconeix que el que passa amb un sistema d'IA depèn de la interacció entre aspectes tècnics i aspectes socials, entre el sistema i la manera en què s'integra en el seu context.

Desenvolupament:

- Les **auditories externes d'algoritmes** són avaluacions independents sobre el funcionament dels sistemes algorítmics per garantir que aquests compleixen les normes jurídiques i ètiques. Vegeu, per exemple, la [llista de verificació per a les auditories del Comitè Europeu de Protecció de Dades](#).

Desplegament:

- Els **registres d'algoritmes i les normes de transparència** són llistes accessibles públicament que fan el seguiment de com les administracions públiques usen els algoritmes o la IA, perquè aquesta informació sigui accessible al públic i a les parts interessades. Aquests repositoris es basen en un model comú de metadades i informació sobre l'algoritme. Vegeu la [Norma de transparència d'algoritmes](#), incloent-hi la codificació usada per nou ciutats europees. Una iniciativa semblant és la [Norma de registre de la transparència algorítmica](#) del Regne Unit.

Context de polítiques i governança:

- Els **comitès multidisciplinaris de supervisió de la governança** apleguen experts procedents de camps de coneixement molt diversos, com ara el dret, l'ètica i les ciències socials, i representants de les comunitats afectades per aportar perspectives diverses al procés de supervisió. Perquè aquests comitès de supervisió siguin efectius, han de ser independents i tenir capacitat real de veto.
- Els **processos participatius**, especialment amb les comunitats afectades, impliquen de manera activa i significativa a les persones en totes les etapes del cicle de vida de la IA, des de les fases de contextualització

i disseny, i no només a posteriori. Mitjançant un procés més equitatiu, aquestes poden contribuir a codissenyar resultats més equitatius.

- El **disseny amb intervenció humana** (*human-in-the-loop design*) significa que els humans segueixen sent els principals responsables a l'hora de prendre decisions relatives a tots els aspectes del sistema per reduir errors i permetre'n el funcionament en mode manual. Tot i que els sistemes algorítmics mai no es poden deslligar totalment dels humans perquè tots els sistemes integren els valors de disseny humans (i molts són propietat d'empreses), aquest enfocament de disseny és útil per deixar clar que els humans han de seguir sent els responsables finals de la presa de decisions.

5. Lliçons apreses

Una investigació prèvia duta a terme pel CIDOB (Jameson i Leal, 2022) va analitzar estudis de cas i experiències d'administracions municipals que apliquen mecanismes de rendició de comptes i transparència en matèria d'IA urbana. Específicament, la recerca es va centrar en el registre d'algoritmes d'Amsterdam, el registre d'IA d'Hèlsinki i un cas d'aprenentatge automàtic explicable desenvolupat per als serveis socials de Barcelona. Aquest capítol destaca algunes de les recomanacions i les lliçons apreses per tal que les iniciatives de transparència i rendició de comptes tinguin èxit.

Disseny

- Les iniciatives de rendició de comptes i transparència van funcionar bé quan es van abordar des de la perspectiva de l'interès públic, lligades a aspectes socials generals, i no només com a qüestions tècniques.
- El fet d'identificar les prioritats per part dels ajuntaments és clau per a l'èxit local, ja que les iniciatives d'un lloc concret poden ser diferents de les d'un altre lloc. En diversos casos, aquestes variacions van ser fruit dels esdeveniments i les informacions de la zona.
- Hi ha expectatives diferents sobre el que pot aconseguir una iniciativa de rendició de comptes d'IA a l'administració pública. Per això, perquè un projecte tingués èxit, va caler dedicar-hi molts esforços i haver designat un «propietari o propietària» del projecte –la persona de referència principal–, que va dedicar gran part del seu temps a gestionar les expectatives de les parts interessades.

Procés

- Cal proposar definicions clares i comprensibles per a tot el funcionari no especialista en la matèria. Els termes clau per garantir l'harmonització són: algoritme, l'objecte de la transparència respecte, la data de publicació, el fòrum on es produeix la rendició de comptes, i el «propietari o propietària» del producte per a cada element del projecte.
- Cal identificar les pràctiques de la cultura empresarial que poden potenciar-se amb els processos de rendició de comptes. Per exem-

ple, les reunions per presentar els estats financers trimestrals eren el moment en què l'equip directiu ja estava reunit i es podien examinar altres innovacions tècniques.

- Cal començar a poc a poc i iterar. Els canvis en el funcionament de l'administració pública requereixen temps i tenen més èxit quan s'introdueixen de manera gradual i no pas tots alhora.

Capacitat

- Totes les iniciatives de rendició de comptes van requerir inversió en formació i reciclatge laboral del funcionariat per actualitzar-se i tenir temps per familiaritzar-se amb els nous enfocaments.
- Cal establir contactes amb xarxes d'intercanvi de coneixement, com la Coalició de Ciutats per als Drets Digitals, en què es comparteixen i intercanvien experiències relatives a l'adaptació dels mecanismes de rendició de comptes i transparència.

Referències bibliogràfiques

Adadi, A. i Berrada, M. «Peeking Inside the Black-Box: A Survey on Explainable Artificial Intelligence (XAI)», Dins *IEEE Access*, vol. 6 (2018), p. 52138–52160.

Amaro, S. «[Dutch government resigns after childcare benefits scandal](#)» *CNBC* (gener 2021) [Data de consulta: 18 de gener de 2021].

Bovens, M. «Analysing and Assessing Accountability: A Conceptual Framework». *European Law Journal* vol. 13, n.º 4 (2007), p. 447–68.

Cath, C. i Jansen, F. «Dutch Comfort: The Limits of AI Governance through municipal registers». *ArXiv*, (setembre 2021).

Dachwitz, I. «[Childcare benefits scandal: Dutch government to pay million Euro fine over racist data discrimination](#)». *netzpolitik.org* (gener 2022)

Van Ewijk, E. i Baud, I. «Partnerships between Dutch municipalities and municipalities in countries of migration to the Netherlands; knowledge exchange and mutuality.» *City-to-City Co-Operation*, vol. 33, n.º 2 (2009), p. 218–26

Jameson, S. i Leal, A. «Transparency and accountability in urban artificial intelligence: Lessons from city initiatives». *Global Observatory of Urban AI (GOUI)*, CIDOB (2022).

Jameson, S., Taylor, L. i Noorman, M. «Data Governance Clinics: a New Approach to Public-Interest Technology in Cities». *SSRN Scholarly Paper. Rochester, NY: Social Science Research Network* (setembre 2021)

Rob, K. «The ethics of smart cities and urban science». *Phil. Trans. R. Soc. A*, vol. 374, n.º 2083 (desembre 2016): 20160115

Longpre, S. *et al.* «The Data Provenance Initiative: A Large Scale Audit of Dataset Licensing & Attribution in AI». *ArXiv* (novembre 2023).

OECD, «[G7 Toolkit for Artificial Intelligence in the Public Sector](#)» (2024)

Tan, V. «[AI's meaning-making problem](#)». *Substack newsletter. The Uncertainty Mindset (Soon to Become Tbd)*, blog (maig 2024)

Taylor, L. «[Fairness and AI governance – responsibility and reality](#)». *Global Data Justice*, blog (abril 2021).

UN-Habitat i Mila Quebec AI Institute. «[AI & Cities: Risks, Applications and Governance](#)» (2022)

Wieringa, M. «What to account for when accounting for algorithms: a systematic literature review on algorithmic accountability». *FAT* '20: Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, (2020), p. 1-18. Nova York: Association for Computing Machinery.

Josuan Eguiluz Castañeira

Assessor legal, Adevinta

Carlos Fernández Hernández

Membre del consell assessor, Global LegalTech Hub

1. Introducció

Des que el 2018 la Unió Europea (UE) va començar a dissenyar el seu marc normatiu sobre la intel·ligència artificial (IA), va emfasitzar que aquesta tecnologia ha de ser «fiable». Es considera que una IA és fiable si respecta el marc normatiu aplicable i és ètica i robusta, tant des del punt de vista tècnic com social, ja que els sistemes d'IA, fins i tot si s'utilitzen amb bones intencions, poden provocar danys accidentals (Grup d'experts d'alt nivell sobre intel·ligència artificial, 2019).

En conseqüència, l'enfocament europeu en aquesta matèria encoratja el desenvolupament i l'adopció d'una IA ètica i fiable a tota l'economia de la UE, a partir del principi que aquesta tecnologia ha d'estar al servei de les persones i ser una força positiva per a la societat (Llibre blanc sobre la intel·ligència artificial, 2020: cap. 6).

Atès que disposar de dades és fonamental per entrenar els sistemes algorítmics, i que moltes d'aquestes dades són de caràcter personal, un component de la IA ètica és que ha d'incloure mecanismes de gestió de la privacitat i de les dades (Comissió Europea, revisió de 2021 del Pla coordinat sobre la IA). Aquesta exigència ha estat plenament inclosa en el Reglament europeu en matèria d'intel·ligència artificial (RIA), aprovat el juny de 2024, que fixa com a objectiu promoure l'adopció d'una IA fiable i centrada en la persona (art. 1), respectant alhora el marc normatiu vigent en matèria de protecció de dades, constituït –principalment, però no només– pel Reglament general de protecció de dades (RGPD) de 2016.

Com han assenyalat alguns autors (Almonacid Lamelas, 2024), el RIA representa un repte important per als governs locals, ja que han d'adaptar els seus processos, les seves polítiques i les seves estratègies per complir les noves exigències. Tanmateix, el RIA també suposa una oportunitat per millorar el funcionament dels ajuntaments, així com la qualitat i la confiança dels serveis basats en IA oferts als ciutadans (*ibíd.*). Això explica la proliferació de sistemes d'«intel·ligència artificial urbana», un concepte que fa referència a «la recopilació, la interpretació i l'anàlisi

Els sistemes d'IA han de garantir la protecció de les dades al llarg de tot el seu cicle de vida. Això inclou tant la informació inicialment facilitada per l'usuari com la que es genera sobre ell en el context de la seva interacció amb el sistema.

de dades urbanes per tal de donar suport a la presa de decisions relacionades amb les polítiques, així com el desenvolupament de solucions que s'utilitzen, o podrien utilitzar-se, en un context urbà» (Galcerán-Vercher, 2023).

Amb tot, el tractament de dades personals en l'àmbit públic de l'urbanisme pot plantejar problemes específics, des de la legitimitat del tractament per a una finalitat per a la qual originalment no es va consentir, fins a la necessitat de fer avaluacions de l'impacte que té sobre els drets fonamentals de les persones. Aquests factors, inequívocament, els han de tenir en compte els organismes públics.

D'acord amb el nou marc legislatiu, l'objecte d'aquest article és (i) presentar el marc jurídic i ètic que regula el tractament de dades personals en l'àmbit urbanístic per mitjà de sistemes d'IA, especialment a Europa (RIA); (ii) identificar els principals mecanismes per implementar el principi de privacitat, i (iii) analitzar els reptes que planteja aquest tipus de tractament de dades i oferir un conjunt de recomanacions i bones pràctiques per minimitzar-los o eliminar-los.

2. IA ètica i privacitat

Una IA fiable ha de ser ètica i per ser-ho ha de respectar la privacitat de les persones, entre altres requisits. El RIA estableix com a objectiu específic promoure l'adopció d'una IA fiable i centrada en l'ésser humà. Per això, les normes comunes que estableix per als sistemes d'IA d'alt risc han de ser coherents amb la Carta dels Drets Fonamentals de la Unió Europea (2000) i han de tenir en compte tant la Declaració Europea sobre els Drets i els Principis Digitals per a la Dècada Digital (2022) com les Directrius ètiques per a una IA fiable del Grup d'experts d'alt nivell sobre intel·ligència artificial (2019). Segons aquestes directrius, en un context de canvi tecnològic ràpid, «la fiabilitat és un prerrequisit perquè les persones i les societats desenvolupin, despleguin i utilitzin sistemes d'IA. Si aquests sistemes –i les persones que es troben al darrere– no demostren ser mereixedors de confiança, hi pot haver conseqüències no desitjades que n'obstaculitzin l'adopció, cosa que impediria l'assoliment dels enormes beneficis econòmics i socials que poden comportar els sistemes d'IA». (Introducció, punt 13)

La fiabilitat de la IA es basa en tres components, que s'han de tenir en compte al llarg de tot el cicle de vida del sistema d'IA:

1. Ha de ser lícita, de manera que es garanteixi el respecte de tot el marc jurídic i normatiu aplicable.
2. Ha de ser ètica, és a dir, ha d'assegurar el compliment dels principis i els valors ètics.
3. Ha de ser robusta, tant des del punt de vista tècnic com social, ja que els sistemes d'IA, fins i tot si s'utilitzen amb bones intencions, poden provocar danys accidentals.

Per tant, cal establir l'ètica com a pilar fonamental per garantir i expandir una IA fiable. Això implica garantir que es compleixin unes normes ètiques bàsiques, així com les mesures que estableix el RIA per a la protecció dels drets fonamentals.

En aquest sentit, la protecció de dades és un dret fonamental que es veu especialment afectat pels sistemes d'IA i que té una relació estreta amb el principi de prevenció del dany. Aquest principi comença per una gestió adequada d'aquestes dades, que abasti la qualitat i la integritat de les que s'utilitzen, la seva pertinència en contrast amb l'àmbit en què es desplegaran els sistemes d'IA, els seus protocols d'accés i la capacitat per processar dades sense vulnerar la privacitat.

Entre aquestes mesures s'inclou el fet que els sistemes d'IA disposin d'un mecanisme de gestió de la privacitat i de les dades que inclogui tant el respecte de la privacitat com la qualitat i la integritat de les dades i l'accés a aquestes dades.

A més, els sistemes d'IA han de garantir la protecció de les dades al llarg de tot el seu cicle de vida. Això inclou tant la informació inicialment facilitada per l'usuari com la que es genera sobre ell en el context de la seva interacció amb el sistema (per exemple, els productes que genera el sistema d'IA per a determinats usuaris o la resposta d'aquests usuaris a certes recomanacions). Els registres digitals del comportament humà poden fer possible que els sistemes d'IA no només infereixin les preferències de les persones, sinó també l'orientació sexual, l'edat, el sexe o les opinions polítiques i religioses. Perquè els individus confiïn en el procés de recopilació de dades, cal garantir que la informació recollida sobre ells no s'utilitzarà per discriminar-los de manera injusta o il·legal.

Del compliment d'aquests requisits se n'han d'encarregar els operadors, en particular, els desenvolupadors dels sistemes d'IA i els responsables del desplegament (que cal que s'assegurin que els sistemes que utilitzen i els productes i els serveis que ofereixen compleixen els requisits establerts). D'altra banda, les persones que es vegin afectades pel funcionament d'un sistema d'IA tenen dret a estar informades d'aquesta afectació i, si escau, presentar una reclamació per infracció del RIA (arts. 85 i 86).

2.1. La privacitat en el RIA

L'article 2.7 del RIA recull el principi general que el propi RIA respecta íntegrament el marc regulador de la UE en matèria de protecció de dades que estableix l'RGPD.

En primer lloc, les normes harmonitzades que s'estableixen en el RIA s'han d'aplicar a tots els sectors i s'han d'entendre sens perjudici del dret vigent de la UE. És important destacar, doncs, que el RIA no pretén afectar l'aplicació del dret de la UE que regula el tractament de dades personals, incloses les funcions i les competències de les autoritats de supervisió independents que vigilen el compliment d'aquests instruments. De la mateixa manera, tampoc no afecta les obligacions prèvies dels proveïdors i els encarregats del desplegament de sistemes d'IA com a responsables del tractament de dades personals. En particular, el RIA no ha d'afectar les pràctiques actualment prohibides pel dret de la UE, incloent-hi els drets en matèria de protecció de dades.

En paral·lel, el fet que un sistema d'IA sigui classificat com d'alt risc no s'ha d'interpretar com un indicador que el seu ús sigui lícit d'acord amb altres actes del dret de la UE o del dret nacional, per exemple, en matèria

El article 2.7 del RIA recull el principi general que el propi RIA respecta íntegrament el marc regulador de la UE en matèria de protecció de dades que estableix l'RGPD.

Per a les ciutats, garantir que els seus sistemes d'IA compleixen regulacions com l'RGPD o el RIA al llarg de tot el cicle de vida de la IA és fonamental per salvaguardar els drets dels ciutadans i mantenir la confiança pública

de protecció de dades personals. Tots els usos d'aquest tipus de sistemes d'IA s'han de continuar fent exclusivament d'acord amb els requisits oportuns derivats de la Carta, el dret derivat de la UE i el dret nacional.

A més, el RIA no constitueix un fonament jurídic per al tractament de dades personals, incloses les categories especials de les dades esmentades, llevat que es disposi específicament una altra cosa. Per això, arran de l'entrada en vigor del RIA les persones interessades continuen gaudint de tots els drets i les garanties que els confereix el dret de la UE, inclosos els que estan relacionats amb les decisions individuals totalment automatitzades, com ara l'elaboració de perfils. Així doncs, les normes harmonitzades que estableix el RIA han de permetre l'exercici dels drets i d'altres vies de recurs dels interessats garantits pel dret de la UE en matèria de protecció de dades personals i altres drets fonamentals.

Finalment, per tal de facilitar el compliment del dret de la UE en matèria de protecció de dades, en determinades condicions, el RIA proporciona la base jurídica perquè, en un entorn controlat de proves, els proveïdors (també els potencials) utilitzin dades personals recollides per a altres finalitats per desenvolupar determinats sistemes d'IA en favor de l'interès públic.

3. Mecanismes polítics per implementar el principi de privacitat en l'àmbit urbanístic

La protecció de la privacitat i de les dades a l'hora d'implementar la IA urbana requereix l'adopció de mecanismes polítics específics, els quals permeten a les ciutats complir les normatives vigents i assegurar que la IA es desplega de manera ètica i responsable, respectant els drets dels ciutadans. Tot seguit, s'identifiquen i expliquen els principals mecanismes polítics per implementar aquest principi ètic.

a) Garantia de la conformitat legal

El compliment de la regulació és un requisit ètic essencial en la protecció de la privacitat i les dades a l'hora d'implementar sistemes d'IA en entorns urbans per part de les autoritats públiques. Per a les ciutats, garantir que els seus sistemes d'IA compleixen regulacions com l'RGPD o el RIA al llarg de tot el cicle de vida de la IA és fonamental per salvaguardar els drets dels ciutadans i mantenir la confiança pública. Això inclou l'adhesió a requeriments clau com la qualitat i la integritat de les dades utilitzades, la seva pertinència en contrast amb l'àmbit en què es desplegaran els sistemes d'IA, els seus protocols d'accés i la capacitat per processar dades sense vulnerar la privacitat (Grup d'experts d'alt nivell sobre intel·ligència artificial, 2019).

Precisament, aquests requeriments es materialitzen en obligacions concretes al mateix RIA, específicament dissenyades per a casos d'alt risc, com ara els sistemes d'IA d'identificació biomètrica remota —per exemple, el programa ABIS (Pascual, 2024)— o els que s'utilitzen per avaluar si les persones físiques compleixen els requisits per beneficiar-se de serveis i prestacions essencials d'assistència pública —per exemple, el cas Syri (Digital Future Society, 2022).

b) Sistemes de gestió de risc i governança de dades

El RIA inclou obligacions específiques (arts. 9 i 10) estretament vinculades al principi de privacitat i protecció de dades. D'una banda, l'article 9 se centra en la creació d'un sistema de gestió de riscos que sigui capaç d'identificar, documentar i mitigar els riscos associats a l'ús d'IA a les ciutats. Aquests sistemes de gestió de riscos han d'establir processos iteratius continus, planificats i executats al llarg de tot el cicle de vida de les tecnologies d'IA, que, per descomptat, requeriran revisions i actualitzacions sistemàtiques periòdiques. De fet, no es tracta només d'avaluar els possibles riscos abans de la introducció al mercat o posada en servei d'aquests sistemes d'IA, sinó també establir i/o supervisar el funcionament d'un sistema de vigilància postcomercialització per gestionar riscos emergents –arts. 17.1 h), 26.5 i 72 RIA–.

D'altra banda, la governança de dades regulada a l'article 10 exigeix que els conjunts de dades d'entrenament, validació i prova utilitzats en sistemes d'IA d'alt risc se sotmetin a pràctiques de governança i gestió de dades adequades per a la finalitat prevista. Les pràctiques que han d'implementar les ciutats per assegurar una governança de dades efectiva i legal s'han de centrar en qüestions com ara els processos de recollida i origen de les dades, la finalitat del tractament, l'avaluació de la disponibilitat, la quantitat i l'adequació dels conjunts de dades necessàries, l'examen de possibles biaixos que puguin afectar la salut, la seguretat o els drets fonamentals de les persones, etc.

c) Avaluacions d'impacte

L'article 35 de l'RGPD imposa als responsables del tractament (per exemple, ajuntaments) l'obligació de fer una avaluació d'impacte relativa a la protecció de dades (AIPD). Aquesta avaluació s'ha de fer quan sigui probable que un tipus de tractament, per la seva naturalesa, abast, context o finalitat (en particular si utilitza noves tecnologies), comporti un alt risc per als drets i les llibertats de les persones físiques (AEPD, 2018; Grup de Treball de l'Art. 29, 2017; Friedewald *et al.*, 2022). Aquest enfocament preventiu és fonamental als entorns urbans per anticipar possibles vulnerabilitats en la protecció de dades i prendre les mesures necessàries per corregir-les a temps.

Així mateix, per als sistemes d'IA d'alt risc, l'article 27 del RIA introdueix l'obligació de fer una avaluació d'impacte relativa als drets fonamentals (FRAI) (Govern dels Països Baixos, 2022; Institut Danès de Drets Humans, 2020), que complementi l'AIPD. Aquesta avaluació té com a objectiu determinar els riscos específics per als drets de les persones que probablement se'n vegin afectades i definir les mesures que s'han d'adoptar en cas que es materialitzin aquests riscos (considerant 96 RIA). Cal destacar que les avaluacions d'impacte (Manzoni *et al.*, 2022) s'han de centrar no només en el retorn de la inversió, sinó també en la sostenibilitat i l'impacte ètic de la tecnologia, abordant aspectes financers, humans i mediambientals (OECD, 2024).

d) Auditories

A continuació, cal poder demostrar davant les autoritats, les parts interessades i els ciutadans que es compleix la legislació i tots els seus requisits específics d'implementació. En aquest sentit, cal dur a terme

auditories internes i externes i obtenir certificacions que verifiquin que els sistemes operen dins els marcs legals establerts. Per això, les ciutats europees, per exemple, han de fer avaluacions de conformitat (art. 43 RIA) per tal de garantir i demostrar que han complert els requisits associats a sistemes d'alt risc, de conformitat amb les normes harmonitzades publicades en el *Diari Oficial de la Unió Europea* (art. 41 RIA). També han de seguir les especificacions comunes establertes per la Comissió Europea, assegurant així una implementació estandarditzada i segura dels sistemes d'IA (per exemple, certificacions ISO).

Les auditories d'IA es consideren un mecanisme de governança fonamental per assegurar que la implementació i l'operació dels sistemes d'IA compleixen les normatives legals i els estàndards ètics i tècnics establerts (Fernández i Eguíluz, 2024). Amb caràcter general, aquestes auditories les han de dur a terme entitats independents i competents. El procés d'auditoria inclou metodologies que incorporen avaluacions d'impacte ètic (UNESCO, 2024; CEN-CENELEC, 2017), que assegurin que els sistemes d'IA es comporten de manera responsable i que els seus efectes en la societat i els individus són degudament monitorats i mitigats. No obstant això, és recomanable plantejar les auditories d'IA des d'un punt de vista multidisciplinari –legal, tècnic i ètic– (Mökander, 2023). En aquest sentit, sorgeixen propostes com la d'Algo Score per classificar i avaluar de manera accessible el nivell de conformitat dels sistemes algorítmics en matèries com el compliment ètic, la governança de la IA, l'equitat del model i la seva vigilància posterior, seguint un enfocament similar al de les etiquetes d'eficiència energètica (Galdon, 2024).

e) Repositoris d'algoritmes i registres de sistemes d'IA

En darrer lloc, convé recordar la importància dels repositoris d'algoritmes públics i els registres de sistemes d'IA (art. 49 RIA), que promouen la transparència en la presa de decisions automatitzades en el sector públic i tenen un paper crucial en la protecció de la privacitat i les dades personals. En fer accessibles els detalls sobre com es dissenyen, implementen i operen aquests sistemes, els repositoris i els registres permeten als ciutadans i a les organitzacions entendre com s'utilitzen les dades personals en aquests processos i amb quines finalitats. Aquests repositoris inclouen també informació sobre les fonts de les dades utilitzades i els mecanismes de supervisió, cosa que és essencial per avaluar l'impacte dels sistemes en la privacitat dels individus i garantir que les mesures de protecció de dades siguin efectives (Gutiérrez i Muñoz-Cadena, 2024).

4. Reptes i recomanacions

La gestió inadequada de les dades és una de les principals limitacions per a la implementació d'IA en el sector públic. Igualment, també ho és la manca d'accés a volums suficients de dades d'alta qualitat. Aquest problema es veu agreujat per l'intercanvi insatisfactori de dades entre organitzacions per la manca d'estàndards unificats i una governança de dades poc desenvolupada. A més, la desconfiança en els sistemes d'IA agreuja aquests reptes. La dispersió de les lleis i el coneixement insuficient sobre els impactes de la IA també generen barreres significatives (Manzoni *et al.*, 2023). Així mateix, l'augment de ciberatacs ha fet que

la [Directiva NIS 2](#) (2022) reforci la seguretat i la responsabilitat legal per als administradors. L'any 2023, l'Administració pública va ser un dels sectors més afectats, amb el 19 % dels incidents reportats, entre els quals destaca l'augment del nombre d'atacs, com ara el programari de segrest (o *ransomware*) i l'atac DDoS (ENISA, 2023).

El complex panorama regulador també suposa un repte significatiu. La interacció entre les normatives urbanístiques europees, estatals i locals crea un entramat de regles que complica la implementació efectiva d'IA a les ciutats. La legislació urbanística i les regulacions específiques de cada municipi s'han d'adaptar a les normatives europees com el Reglament sobre l'Europa Interoperable (Interoperable Europe Act) (2022), que cerca millorar la interoperabilitat dels serveis públics digitals (Tangi *et al.*, 2023).

Una altra limitació important és la manca d'experiència i de coneixement tècnic que hi ha al si de les administracions locals, cosa que dificulta la implementació de la IA de manera adequada. L'escassetat de professionals especialitzats en aquesta matèria arreu del món i la creixent competència pel talent representen una barrera significativa per a les ciutats que intenten desenvolupar i desplegar aquests sistemes de manera efectiva (OECD, 2024).

D'altra banda, la recollida massiva de dades personals, necessària per entrenar aquests sistemes, pot vulnerar el dret dels ciutadans a controlar les seves dades, ja que poden ser dades sensibles o gestionades de manera inapropiada. A més, les aplicacions d'IA, com les que es fan servir en el control policial, poden intensificar la vigilància massiva i comprometre encara més la privacitat dels individus (Véliz, 2020; Agarwal, 2018; Dwivedi *et al.*, 2019).

Per superar aquestes barreres, és essencial promoure mecanismes d'innovació, com els *sandboxes* reguladors (Madiega, 2022), que permeten a les ciutats experimentar amb IA en un entorn controlat mentre es garanteix el compliment regulador (Tangi *et al.*, 2023). Així mateix, la coordinació entre les autoritats estatals (en el cas espanyol, l'Agència Espanyola de Supervisió de la Intel·ligència Artificial - AESIA) i europees (l'Oficina Europea d'Intel·ligència Artificial) és crucial per garantir que els sistemes d'IA compleixin les normatives vigents i s'implementin de manera segura i responsable.

La interoperabilitat i la col·laboració són igualment clau. Iniciatives com el sistema SALER - Sistema d'Alertes Ràpides, utilitzat a la Comunitat Valenciana per prevenir la corrupció a l'Administració, demostren que la IA es pot utilitzar de manera efectiva per millorar els processos de governança (Digital Future Society, 2023). Igualment, és essencial que el finançament públic estigui condicionat a l'assoliment de certs resultats per part de les diferents administracions (per exemple, generar conjunts de dades públiques) (Comissió Europea, 2022). En aquest sentit, la Comissió Europea ha publicat, mitjançant el [Reglament d'execució \(UE\) 2023/138](#), una llista de conjunts de dades específiques d'alt valor que han d'estar disponibles per a la reutilització gratuïta, cosa que destaca el potencial de les dades públiques en benefici de la societat, el medi ambient i l'economia (Comissió Europea, 2022). A més, l'accés a dades multilingües per entrenar models locals d'IA que reflecteixin les carac-

És essencial promoure mecanismes d'innovació, com els *sandboxes* reguladors, que permeten a les ciutats experimentar amb IA en un entorn controlat mentre es garanteix el compliment regulador.

terístiques específiques de cada regió (OECD, 2024) i la recopilació de casos d'ús d'IA en el sector públic europeu (Comissió Europea, 2021) milloraran l'efectivitat i l'equitat dels sistemes d'IA i alhora seran una valuosa font d'informació sobre com es van implementant aquestes tecnologies en diversos contextos.

5. Conclusions

El marc general de protecció de dades a la UE ja està establert en uns principis coneguts i sòlidament interpretats pels organismes administratius i jurisdiccionals de la UE. Tot i això, la IA planteja problemes específics, de caràcter tecnològic i jurídic, el coneixement i el tractament dels quals encara són incipients.

Per això, encara calen nombrosos estudis, experiències i precisions per dotar-los d'un marc jurídic que garanteixi la finalitat proclamada que la IA ha d'estar centrada en l'ésser humà, ser una eina per a les persones i tenir com a objectiu últim millorar el seu benestar.

La implementació de mecanismes polítics específics és essencial per garantir que les ciutats utilitzin sistemes d'IA de manera ètica i respectuosa amb els drets de la ciutadania. El compliment de regulacions com l'RGPD i el RIA resulta indispensable per salvaguardar la privacitat i les dades personals en entorns urbans. De la mateixa manera, és fonamental que les ciutats estableixin sistemes de gestió de riscos que abordin de manera iterativa les contingències associades a tot el cicle de vida de la IA, incloent-hi revisions periòdiques i auditories externes que assegurin el compliment normatiu.

D'altra banda, la governança de dades ha de ser el centre de les estratègies urbanes d'IA. Les ciutats han d'implementar pràctiques de governança i gestió de dades sòlides, centrades en la qualitat, la pertinència i la protecció dels conjunts de dades emprats als sistemes d'IA. Això inclou dur a terme avaluacions d'impacte tant per a la protecció de dades personals com per als drets fonamentals per assegurar que les tecnologies implementades no vulnerin la privacitat ni la seguretat de la ciutadania.

En definitiva, aconseguir una IA centrada en l'ésser humà exigeix un esforç conjunt entre els responsables de desenvolupar polítiques públiques, les institucions acadèmiques i els sectors privats, que han de col·laborar per assegurar que els sistemes d'IA que implementin les ciutats concordin amb els valors i els principis ètics fonamentals.

Com hem assenyalat, el futur de les ciutats intel·ligents es caracteritzarà per la combinació de múltiples tecnologies orientades a satisfer l'intricat mosaic de necessitats humanes. Aquesta convergència requereix una optimització precisa de les tecnologies aplicades que assegurin que la digitalització dels espais urbans està en consonància amb pràctiques sostenibles i equitatives, així com l'atenció a les dimensions ètiques que aquestes innovacions comporten. Per això, és imperatiu que la integració de la IA al si de les ciutats intel·ligents es regeixi per principis que defensen la privacitat, la seguretat i la inclusió. En aquest sentit, i com apunten Zhenjun *et al.* (2023), "garantir que els beneficis del desenvolupament de

les ciutats intel·ligents es comparteixin equitativament és essencial per evitar fractures socials i fomentar un entorn en què la tecnologia serveixi de pont cap a una vida urbana més il·lustrada i harmoniosa”.

Referències bibliogràfiques

Agencia Española de Protección de Datos (AEPD). «[Gestión del riesgo y evaluación de impacto en tratamientos de datos personales](#)». Madrid, 2018 [Data de consulta: 02.09.2024]

Agencia Española de Protección de Datos. «[Guía sobre protección de datos y administración local](#)». Actualitzada el 2023 [Data de consulta: 02.09.2024]

AI Ethics Impact Group. «[From Principles to Practice: An interdisciplinary framework to operationalise AI ethics](#)». 1 d'abril de 2020 [Data de consulta: 02.09.2024]

Alan Turing Institute. «[Urban analytics](#)» [Data de consulta: 02.09.2024]

Alan Turing Institute. «[Why the public sector needs to know about AI ethics \(and how we're helping\)](#)». 2 de novembre de 2023 [Data de consulta: 02.09.2024]

Almonacid Lamelas, V. «[Reglamento \(europeo\) de Inteligencia Artificial: impactos y obligaciones que genera en los Ayuntamientos](#)». El Consultor de los Ayuntamientos, LA LEY, 15 de juliol de 2024 [Data de consulta: 02.09.2024]

Burbano, L. (1). «[Privacy protection in smart cities: How are they taking care of citizens' most precious information?](#)» Tomorrow.city, 23 de gener de 2024 [Data de consulta: 02.09.2024]

Burbano, L. (2). «[AI urbanism: risks and benefits of a seemingly untoppable movement](#)». Tomorrow.city, 22 de febrer de 2024 [Data de consulta: 02.09.2024]

Canda, J. «[AI in Urban Planning and Smart City Development](#)». Medium, 7 d'abril de 2024 [Data de consulta: 02.09.2024]

CEN-CENELEC. «[Ethics assessment for research and innovation - Part 2: Ethical impact assessment framework](#)». CWA 17145-2, juny de 2017 [Data de consulta: 02.09.2024]

Centro Latinoamericano de Administración para el Desarrollo (CLAD). «[Carta Iberoamericana de Inteligencia Artificial en la Administración Pública](#)». 20 de novembre de 2023 [Data de consulta: 02.09.2024]

Comissió Europea (1). «[Selected AI cases in the public sector \(JRC129301\)](#)». Joint Research Centre (JRC), 2021 [Data de consulta: 02.09.2024]

Comissió Europea (2). «[Revisión de 2021 del plan coordinado sobre la inteligencia artificial](#)», Anexos de la Comunicación de la Comisión

al Parlamento Europeo, al Consejo Europeo, al Consejo, al Comité Económico y Social Europeo y al Comité de las Regiones. Fomentar un planteamiento europeo en materia de inteligencia artificial, COM(2021) 205 final, 2021 [Data de consulta: 02.09.2024]

Comissió Europea (3). «[Opportunities and challenges of artificial intelligence technologies for the cultural and creative sectors](#)». Oficina de Publicaciones de la UE, Luxemburg, 2022 [Data de consulta: 02.09.2024]

Comissió Europea (4). «[Second Report on the application of the General Data Protection Regulation](#)», COM(2024) 357 final, Brussel·les, 25 de juliol de 2024 [Data de consulta: 02.09.2024]

Digital Future Society (1). «[Algorithmic discrimination in Spain: limits and potential of the legal framework](#)», agost de 2022 [Data de consulta: 02.09.2024]

Digital Future Society (2). «[El acceso digital en las ciudades, entendido como algo más que un derecho fundamental](#)», juny de 2023 [Data de consulta: 02.09.2024]

Digital Future Society (3). «[El uso de algoritmos en el sector público en España: cuatro estudios de caso sobre ADMS](#)», febrer de 2023 [Data de consulta: 02.09.2024]

European Data Protection Board (EDPB). «[Statement 3/2024 on data protection authorities' role in the Artificial Intelligence Act framework](#)», 16 de juliol de 2024 [Data de consulta: 02.09.2024]

European Union Agency for Cybersecurity (ENISA). «[Threat Landscape 2023](#)», octubre de 2023 [Data de consulta: 02.09.2024]

Fernández, C. i Eguiluz, J. A. «[Diez puntos críticos del Reglamento europeo de Inteligencia Artificial](#)». Diario LA LEY, núm. 85, Sección Ciberderecho, 28 de juny de 2024 [Data de consulta: 02.09.2024].

Friedewald, M. *et al.* «[Data Protection Impact Assessments in Practice: Experiences from Case Studies](#)». Computer Security, ESORICS 2021 International Workshops, febrer de 2022, p. 424-443 [Data de consulta: 02.09.2024]

Galceran-Vercher, M. «[Trustworthy Cities: Ethical Urban Artificial Intelligence](#)». The GovLab, Course «AI Ethics, Global perspectives», desembre de 2023 [Data de consulta: 02.09.2024]

Galceran-Vercher, M. i Vidal, A. «[Mapping urban artificial intelligence: first report of GOUAI's Atlas of Urban AI](#)». Global Observatory of Urban Artificial Intelligence (GOUAI), 2024 [Data de consulta: 02.09.2024]

Galdon, G. «[AI Auditing. Proposal for Algo-scores](#)». EDPB, 27 de juny de 2024 [Data de consulta: 02.09.2024]

Ghisleni, C. «[Artificial Intelligence and Urban Planning: Technology as a Tool for City Design](#)». ArchDaily, 8 de febrer de 2024 [Data de consulta: 02.09.2024]

Govern dels Països Baixos. «[Impact Assessment Fundamental Rights and Algorithms](#)», 31 de març de 2022 [Data de consulta: 02.09.2024].

Grup d'experts d'alt nivell sobre intel·ligència artificial. «[Directrices éticas para una IA fiable](#)», abril de 2019 [Data de consulta: 02.09.2024]

Grup de treball de l'Art. 29. «[Guidelines on Data Protection Impact Assessments \(DPIA\) and determining whether processing is “likely to result in a high risk” for the purposes of Regulation 2016/679](#)», Brussel·les, 2017 [Data de consulta: 02.09.2024]

Gutiérrez, J. D. i Muñoz-Cadena, S. «[Algorithmic transparency in the public sector: A state-of-the-art report of algorithmic transparency instruments](#)». Global Partnership on Artificial Intelligence, maig de 2024 [Data de consulta: 02.09.2024]

Hadec, J., Di Leo, M. i Kotsev, A. «[AI generated synthetic data in policy applications](#)». *Science for Policy Brief*, European Commission, Joint Research Center, 2024 [Data de consulta: 02.09.2024]

Imdat As, P., Basu, P. i Talwar, P. «[Artificial Intelligence in Urban Planning and Design. Technologies, Implementation, and Impacts](#)». Amsterdam: Elsevier, 2022.

Institut Danès de Drets Humans. «[Guidance on Human Rights Impact Assessment of Digital Activities](#)». Copenhaguen, 2020 [Data de consulta: 02.09.2024]

Madiega, T. i Van De Pol, A. L. «[Artificial intelligence act and regulatory sandboxes](#)». European Parliamentary Research Service (EPRS), PE 733.544, juny de 2022 [Data de consulta: 02.09.2024]

Manzoni, M. *et al.* «[AI Watch. Road to the adoption of Artificial Intelligence by the public sector](#)». Oficina de Publicaciones de la UE. Luxemburg, 2022. JRC129100 [Data de consulta: 02.09.2024]

Mökander, J. «[Auditing of AI: Legal, Ethical and Technical Approaches](#)». *Digital Society*, vol. 2, art. núm. 49, 2023 [Data de consulta: 02.09.2024]

Organización de Naciones Unidas (ONU). «[Recommendation on the Ethics of Artificial Intelligence](#)», 23 de novembre de 2021 [Data de consulta: 02.09.2024]

Organización para la Cooperación y el Desarrollo Económico (OCDE) (1). «[Advancing Accountability in AI: governing and managing risks throughout the lifecycle for trustworthy AI](#)», *OECD Artificial Intelligence Papers*, núm. 349, 23 de febrer de 2023 [Data de consulta: 02.09.2024]

Organización para la Cooperación y el Desarrollo Económico (OCDE) (2). «[Governing with artificial intelligence: Are governments ready?](#)», *OECD Artificial Intelligence Papers*, núm. 20, juny de 2024 [Data de consulta: 02.09.2024]

Pascual, M. G. «La Policía española ya usa en sus investigaciones un sistema automático de reconocimiento facial». El País, 28 de maig de 2024 [Data de consulta: 02.09.2024]

Pellegrin, J., Colnot, L. i Delponte, L. «Artificial Intelligence and Urban Development». Research for REGI Committee, European Parliament, Policy Department for Structural and Cohesion Policies, Brussel·les, 2021 [Data de consulta: 02.09.2024]

Tangi, L. *et al.* «Artificial Intelligence for Interoperability in the European Public Sector: an exploratory study», Oficina de Publicaciones de la UE, Luxemburg, 2023 [Data de consulta: 02.09.2024]

Tangi, L. *et al.* «AI Watch. European Landscape on the Use of Artificial Intelligence by the Public Sector», Oficina de Publicaciones de la UE, Luxemburg, 2022 [Data de consulta: 02.09.2024]

Timan, T., Van Veenstra, A. F. i Bodea, G. «Artificial Intelligence and public services». Policy Department for Economic, Scientific and Quality of Life Policies, PE 662.936, juliol de 2021 [Data de consulta: 02.09.2024]

Véliz, C. «Privacidad es poder: Datos, vigilancia y libertad en la era digital». Madrid: Debate, 2021.

Verhulst, S. G. «Are we entering a Data Winter? On the urgent need to preserve data access for the public interest» [Data de consulta: 02.09.2024]

Yan, Z. *et al.* «Intelligent urbanism with artificial intelligence in shaping tomorrow's smart cities: current developments, trends, and future directions». Journal of Cloud Computing, 18 de desembre de 2023 [Data de consulta: 02.09.2024]

Leandry Junior Jieutsa

*Investigador en governança de la IA a les ciutats,
Càtedra UNESCO de Paisatge Urbà*

1. Introducció

L'equitat i la no-discriminació són valors fonamentals de la IA urbana a les ciutats intel·ligents centrades en les persones. El debat cada vegada major entre investigadors i legisladors palesa la importància creixent d'abordar els biaixos i la discriminació en els sistemes d'IA. L'equitat deriva del judici moral, és a dir, el procés pel qual els individus determinen què és moralment correcte o reprobable (Weinkauff, 2023). Tot i que la IA presenta molts avantatges per a les ciutats, el seu desplegament posa a prova l'objectiu d'aconseguir una ciutat justa, en generar o reforçar la discriminació i les desigualtats. Així doncs, integrar els principis d'equitat i no-discriminació en els cicles de vida de la IA urbana és fonamental per garantir el benestar de les persones i les comunitats a les ciutats intel·ligents. No obstant això, la posada en pràctica d'aquest principi segueix sent complexa i poc concreta. Per aconseguir-ho, les ciutats han d'articular diferents papers en la governança de la IA, com a desenvolupadores de solucions internes, responsables del desplegament de sistemes externs o reguladores. Per això, cal adoptar diferents mecanismes, incloent-hi innovacions sociotecnològiques, creació de normes locals d'equitat en la IA i estàndards d'adquisició pública. A més, cal introduir la legislació urbana per protegir els més vulnerables i garantir que els ciutadans poden exercir els seus drets digitals. Tanmateix, aquestes mesures requereixen recursos, que les ciutats poden mobilitzar fomentant la cooperació i el treball en xarxa.

2. Un sistema d'IA equitatiu ha de ser imparcial i utilitzar-se de manera responsable

L'equitat i la no-discriminació són conceptes complexos i fonamentals en la societat contemporània (Barocas et al., 2023a). L'equitat, que el diccionari Cambridge defineix com la qualitat de tractar les persones per igual de manera justa o raonable, respecta les persones com a individus i com a membres de la societat. Tres elements primaris, articulats en dimensions distributiva i sociorelacional, constitueixen aquest concepte que es refereix als individus o als grups (Barocas et al., 2023b): la igualtat

Les ciutats han d'articular diferents papers en la governança de la IA, com a desenvolupadores de solucions internes, com a responsables del desplegament de sistemes externs o com a reguladores.

d'oportunitats, el dret a la justificació i la igualtat en les relacions (Giovanola i Tiribelli, 2022). Una societat equitativa ha de considerar cada individu o grup d'individus tenint en compte les seves circumstàncies i característiques particulars per garantir un tracte i uns resultats equitatius (Giovanola i Tiribelli, 2022; Lyu et al., 2023). Així doncs, l'equitat abraça el concepte de no-discriminació, que implica que ningú no ha de ser exclòs. Els grups o les persones vulnerables són més susceptibles a patir discriminació.

El sorgiment de tecnologies disruptives com la IA suposa un repte per als aspectes relacionats amb l'equitat. En la discriminació relacionada amb la IA hi intervenen dos factors principals: els biaixos algorítmics i la utilització de sistemes basats en IA (Ferrara, 2023; O'Neil, 2016; Wachter et al., 2021).

El primer factor, el biaix algorítmic, distorsiona les dades d'entrenament originals o l'algoritme d'IA, cosa que genera resultats esbiaixats i potencialment perjudicials (Holdsworth, 2023). Aquests biaixos redueixen la precisió i el potencial de la IA en diferent grau segons l'aplicació. Hi ha dos tipus principals de biaixos en la IA: biaixos d'automatització i biaixos per correlació (Barocas et al., 2023a; González-Sendino et al., 2023). Els biaixos d'automatització consisteixen en la propagació a gran escala a través dels processos del sistema d'IA de biaixos socioculturals que estan profundament enquistats en les dades d'entrenament històriques utilitzades per alimentar el sistema d'IA. S'hi inclouen els biaixos humans, els biaixos de les dades, els biaixos d'aprenentatge i els biaixos de desplegament. El biaix per correlació, per la seva banda, té lloc quan la correlació involuntària de variables protegides (per exemple, el sexe o la raça) permeten inferir biaixos, tot i els esforços per excloure'ls de les dades d'entrenament.

El segon factor és la utilització de sistemes basats en IA. De fet, quan s'utilitzen per elaborar perfils o per al control social, els sistemes vulneren els drets digitals (Calzada, 2021; Cugurullo et al., 2022). Per exemple, en recollir i utilitzar informació personal, les tecnologies de reconeixement facial vulneren la intimitat i les dades personals dels ciutadans (UN-Habitat, 2023). Els drets digitals es consideren drets humans existents que cal protegir en el context de les tecnologies digitals, atès que l'espai físic i el digital cada cop estan més entrellaçats (UN-Habitat, 2020).

L'equitat algorítmica es basa en variables interrelacionades (Weinkauf, 2023). Un sistema de decisions automatitzades es considera que és equitatiu quan no depèn de dades sensibles com el sexe o la religió, no desfavoreix les minories i es fa servir de manera responsable.

3. El dilema de la IA: l'equilibri entre les oportunitats i els impactes dels sistemes d'IA a les ciutats

L'urbanisme històricament ha contribuït a crear i reforçar diferents formes de desigualtat i discriminació urbanes (Fainstein, 2009; Hall, 2014). Les poblacions més afectades en són principalment les minories i les persones més vulnerables, que poden variar en funció del context.

En conseqüència, han sorgit nombrosos conceptes, com «dret a la ciutat» o «ciutat justa» de Henry Lefebvre (Fainstein, 2009; Fincher i Iveson, 2012; Harvey i Potter, 2009; Lefebvre, 1968), que aspiren a fer que les ciutats siguin més equitatives, especialment proporcionant oportunitats i serveis urbans que millorin la qualitat de vida.

El sorgiment de la IA suposa un repte per a la ciutat justa, ja que proporciona oportunitats per aconseguir ciutats més inclusives i alhora crea i reforça diferents formes de desigualtats i discriminació. De fet, els sistemes d'IA tenen la capacitat per filtrar i processar grans volums de dades connectades amb xarxes extenses i a l'entorn urbà. Per això, faciliten que es prenguin decisions complexes de manera autònoma o semiautònoma (Marvin et al., 2022; Sherman, 2023; Yigitcanlar et al., 2021). Les metodologies d'IA explicable poden ajudar els ajuntaments a entendre el càlcul de l'equitat i a millorar-la (Lyu et al., 2023). La implementació d'infraestructures basades en IA ajuda a créixer la participació entre els ciutadans i la municipalitat i optimitza la prestació de serveis, especialment per a les poblacions més vulnerables.

Per exemple, les eines d'aprenentatge profund milloren la gestió de les dades espacials per optimitzar la prestació de serveis en barris desafavorits de [Durban \(Sud-àfrica\)](#). La [IA generativa](#) facilita els processos de planificació participatius generant escenaris urbans en temps real, cosa que permet un urbanisme més inclusiu que incorpora perspectives diverses. A més, els xats de conversa municipals, com els que s'han implementat a [Hèlsinki \(Finlàndia\)](#) o a [Saint-Lin-Laurentides \(Canadà\)](#), automatitzen la interacció dels ciutadans amb els ajuntaments, cosa que millora la gestió dels serveis públics, especialment per als individus que no estan familiaritzats amb els tràmits administratius complexos o per a les persones que tenen dificultats per accedir als serveis en persona.

Tanmateix, com hem assenyalat anteriorment, els sistemes d'IA i l'accent en la competitivitat econòmica de les ciutats suposen un repte per a la ciutat justa en produir resultats discriminatoris i poc equitatius. A més, a diferència de les formes tradicionals de discriminació, la discriminació automatitzada generada pels algoritmes resulta més abstracta o opaca i poc intuïtiva, subtil, intangible, difícil de detectar i a gran escala (Kleinberg et al., 2018; O'Neil, 2016; Sanchez et al., 2024; Wachter et al., 2021).

Per exemple, els algoritmes fiscals que singularitzaven «noms que sonaven estrangers» i persones amb «doble nacionalitat» van provocar que [milers de famílies racialitzades fossin acusades falsament de frau](#) als Països Baixos. Globalment, els sistemes de vigilància predictiva, com [Clearview AI](#), suposen problemes de privacitat i alhora reforcen els biaixos (Dauvergne, 2022; O'Neil, 2016). El 2021, la revista *Forbes* va informar de [biaixos algorítmics en les sol·licituds d'hipoteques](#), de manera que al 80 % dels sol·licitants negres se'ls va denegar. De la mateixa manera, *The Markup* (2021) va descobrir que els sol·licitants de color tenien un [40-80 % més de probabilitats que els deneguessin els préstecs](#), la qual cosa emfasitza l'impacte discriminatori de la IA.

A més, la concentració de la riquesa a les grans ciutats, degut a la IA urbana, comporta gentrificació urbana (Sanchez et al., 2024). Especialment, la població amb rendes baixes té cada vegada més difícil, si no impossible,

A diferència de les formes tradicionals de discriminació, la discriminació automatitzada generada pels algoritmes resulta més abstracta o opaca i poc intuïtiva, subtil, intangible, difícil de detectar i a gran escala.

Degut a la seva opacitat i magnitud, és difícil per a les persones afectades fer valer els seus drets [...] perquè els sistemes de presa de decisions algorítmics alteren els recursos i els procediments jurídics tradicionals que detecten, investiguen, prevenen i corregeixen la discriminació.

accedir a un habitatge. Així doncs, el desplegament de polítiques d'IA urbana contribueix a reforçar les asimetries entre territoris i les desigualtats urbanes.

Per tant, els sistemes d'IA tenen efectes importants sobre les ciutats i les societats. Aquesta ambivalència planteja la necessitat d'una governança efectiva. A més, degut a la seva opacitat i a la magnitud del seu impacte, és difícil per a les persones afectades defensar-se o fer valer els seus drets. Això posa entredit el dret a la no-discriminació del qual gaudeixen els ciutadans perquè els sistemes de presa de decisions algorítmics alteren els recursos i els procediments jurídics tradicionals que detecten, investiguen, prevenen i corregeixen la discriminació (Wachter et al., 2021).

4. Recomanacions polítiques

Segons les recomanacions de la UNESCO, els agents implicats en la IA han d'adoptar un enfocament inclusiu per tal que els beneficis de les tecnologies de la IA arribin a tothom, tenint en compte les necessitats concretes dels diferents grups (UNESCO, 2023). Des del punt de vista de les ciutats, la implementació de l'equitat i la no-discriminació en els sistemes d'IA urbana requereix l'articulació dels diversos papers que els ajuntaments assumeixen com a desenvolupadors de solucions internes (tot i que relativament infreqüents degut a limitacions financeres i tècniques), responsables del desplegament i reguladors. A més, millorar l'equitat dels sistemes d'IA implica tenir en compte tot el seu cicle de vida, abordant els diversos aspectes al llarg dels processos de disseny, desenvolupament i implementació. Així mateix, oferir solucions eficaces per solucionar les desigualtats en els resultats del sistema d'IA passa per identificar-ne les causes subjacents.

4.1. Recomanacions generals

- **Definir una estratègia:** les ciutats han d'implementar estratègies d'IA que girin al voltant dels principis d'equitat i no-discriminació. Aquests documents estratègics permeten a les ciutats establir uns fonaments robustos i dur a terme una avaluació precisa dels seus objectius relacionats amb la IA. Aquest enfocament és fonamental per planificar la integració de la IA amb l'objectiu de maximitzar els beneficis i alhora mitigar els riscos potencials. Aquestes estratègies s'han de desenvolupar a través d'un procés participatiu juntament amb plans d'acció que defineixin mesures concretes per garantir la integració equitativa de la IA que no deixi enrere ningú.
- **Establir nivells de risc segons les aplicacions:** les ciutats han d'identificar les aplicacions d'IA de risc alt dins el marc de les seves competències, tenint en compte les disparitats i les desigualtats existents al territori. Un cop identificades aquestes aplicacions de risc alt cal implementar mecanismes de protecció. Les aplicacions relacionades amb serveis socials essencials s'han de considerar de risc alt i se n'ha de prohibir el funcionament de manera totalment autònoma. Per exemple, la [ciutat californiana de San José](#) ha posat en marxa un registre d'IA estructurat al voltant d'una avaluació rigorosa dels sistemes d'IA. Aquest procés inclou l'anàlisi de riscos i una posterior avaluació més exhaustiva de l'impacte, en funció del nivell de risc, tot documentat en

un «Full d'impacte» i una «Fulla informativa d'IA».

4.2. Recomanacions concretes per a les ciutats com a desenvolupadores de solucions internes

- **Fomentar la innovació socioecològica inclusiva.** Cal comptar amb diverses parts interessades de perfil no tècnic al llarg del cicle de vida de la IA. Segons ONU-Habitat, aquest cicle de vida de la IA consta de cinc fases: contextualització, disseny, implementació, desplegament i manteniment. Si les decisions en aquestes etapes les prenen majoritàriament tècnics o grups homogenis, hi ha un risc important que els seus biaixos s'integrin en el sistema d'IA. Aquest risc és especialment pronunciat si l'eina s'aplica posteriorment de manera generalitzada a sectors més amplis de població. Els governs locals han de posar més èmfasi en la interdisciplinarietat i la multidisciplinarietat que integren els grups socials en el cicle de vida de la IA urbana.
- **Aplicar tècniques d'equitat,** com ara el pretractament de dades (que implica identificar i abordar biaixos en les dades abans de l'entrenament del model), la selecció del model (que utilitza mètodes de selecció de models que prioritzen l'equitat) i les decisions de postprocessament (que consisteix en ajustar el resultat dels models de IA per mitigar els biaixos i garantir l'equitat) (Ferrara, 2023).
- **Fomentar la diversitat a l'hora de construir bases de dades en tres aspectes: equips, dades i models.** Crear equips diversos i interdisciplinaris i proporcionar formació contínua en equitat i ètica és fonamental per reduir els biaixos. Pel que fa a les dades, millorar la recollida d'atributs sensibles (per exemple, sexe, raça, ètnia) i documentar les decisions relacionades amb les dades fomenta la transparència i ajuda a fer front a les desigualtats del món real. Per als models, proporcionar accés obert a la comunitat per al testatge, garantir la documentació transparent i utilitzar IA explicable pot ajudar a identificar i reduir els biaixos, garantint uns resultats equitatius (González-Sendino et al., 2023).
- **Integrar la correlació compensatòria en els sistemes d'IA.** Tal com indiquen Giovanola i Tiribelli (2022), garantir la igualtat d'oportunitats en els sistemes d'IA no es pot limitar a eliminar els biaixos discriminatoris en les dades d'entrenament. Cal dissenyar els sistemes d'IA urbana per considerar les desigualtats existents en el seu context i incorporar-hi mecanismes per compensar-les. Per exemple, en una ciutat amb diferències entre comunitats o grups socials, les IA han de tenir en compte aquestes disparitats i aplicar-hi mesures compensatòries, cosa que es pot materialitzar en contingut personalitzat, per exemple.
- **Integrar tècniques de mitigació en el cicle de vida de la IA.** Neutralitzar els efectes discriminatoris de les dades durant la fase de preentrenament a través de mètodes com el remostreig (és a dir, modificar la mida del conjunt de dades que afecta la distribució sense transformar les dades), la representació justa (que s'aconsegueix eliminant informació que permeti associar un individu amb un grup protegit) i la reponderació (utilitzada per transformar les dades modificant el seu pes dins el conjunt de dades). Durant la

fase d'entrenament, cal utilitzar la regularització i l'entrenament adversari, que són els mètodes més comuns per a aquest objectiu. Altres enfocaments emergents són l'aprenentatge descentralitzat, la regressió lineal justa, el DeepFair, els models multimodals i la clusterització de fairlets. Durant la fase de postentrenament, cal implementar probabilitats igualades, probabilitats igualades calibrades i rebutjar la classificació d'opcions.

4.3. Recomanacions concretes per a les ciutats com a responsables de desplegament i reguladores

- **Determinar normes locals per a la IA equitativa.** La discriminació i les desigualtats es poden manifestar de maneres diferents segons el context, afectant individus o grups socials de maneres diferents i a diferents escales. Per tant, les ciutats han d'aplicar normes d'equitat per a la IA urbana que tinguin en compte aquestes particularitats locals. Aquestes normes han d'incloure principis generals i alhora qüestions d'abast local amb l'objectiu de garantir que la IA urbana no reforça la discriminació ja existent ni crea nous biaixos que afectin de manera desproporcionada els més vulnerables. Cal consultar les comunitats locals a l'hora de desenvolupar aquestes normes i tenir en compte tot el cicle de vida de la IA.
- **Establir normes d'adquisició pública per a una IA equitativa.** Les ciutats han de garantir que les entitats que els proporcionen serveis compleixin els principis de la IA equitativa, cosa que requereix establir mecanismes d'adquisició pública que obliguin els proveïdors de serveis a complir les normes d'IA equitativa de la ciutat. Els proveïdors han de complir els requisits de conformitat relatius als seus algorismes si els ha d'usar la ciutat. Per exemple, la [Coalició GOV AI](#), encapçalada per la ciutat de San José, ha adoptat i introduït l'esmentada «Fulla informativa d'IA» per a sistemes de tercers, una plantilla unificada perquè els proveïdors facilitin informació detallada sobre els productes d'IA, que inclou, entre altres aspectes, la finalitat del sistema, les dades d'entrenament, els detalls del model, els indicadors per mesurar el rendiment, la gestió de biaixos, la robustesa i la interacció persona-ordinador.
- **Implementar una legislació urbana que garanteixin el dret a la justificació.** Aquest dret permet a les persones afectades per un sistema d'IA entendre el raonament subjacent a una decisió algorítmica, cosa que possibilita als ciutadans entendre i controlar com els tracten aquests sistemes. Quan aquest dret no es respecta de manera adequada, s'ha de tenir la capacitat per qüestionar i modificar els paràmetres subjacents de la decisió. Per tant, les ciutats han de considerar al llarg del procés si despleguen o retiren un sistema d'IA, especialment si no es pot atendre la petició d'explicacions per part d'algú. Aquesta mesura permet a les persones que pateixen discriminació fer valer els seus drets digitals.
- **Crear òrgans assessors per investigar, prevenir i mitigar els potencials usos maliciosos de la IA.** Els governs municipals han de crear òrgans assessors multidisciplinars que incloguin organitzacions comunitàries, el món acadèmic, empreses i altres parts interessades de la comunitat. Aquests organismes tindran un paper de control per limitar la discriminació relacionada amb la IA. Avaluaran els models d'IA de la

ciutat basant-se en indicadors que mesurin l'equitat. Aquesta avaluació (1) identificarà els potencials biaixos que poden afectar l'equitat, (2) seleccionarà indicadors per mesurar l'equitat dels sistemes, i (3) mitigarà l'impacte generat per aquests biaixos. A més, actuaran com a organismes assessors per orientar les ciutats a l'hora d'emprendre accions i dur a terme polítiques relacionades amb la IA equitativa.

5. Limitacions

Assolir l'equitat en la IA és un procés complex, ja que les intervencions que tenen aquesta meta en contextos d'IA urbana poden entrar en conflicte amb els objectius mateixos dels algoritmes. Per aquest motiu, les ciutats han d'adoptar solucions de compromís per arribar a un equilibri entre els guanys i els beneficis i alhora prioritzar el benestar de les persones i de les comunitats. Tanmateix, aquest principi pot semblar abstracte i pot donar lloc a interpretacions divergents, cosa que complica l'aplicació de mesures d'èxit i impacte (Sadek et al., 2024). Per això, les ciutats han d'adoptar un enfocament local per posar en pràctica l'equitat i la no-discriminació en la IA urbana. Aquest enfocament integral considera la configuració socioeconòmica i cultural de la ciutat al llarg de tot el cicle de vida de la IA.

Des del punt de vista tècnic, la IA urbana equitativa requereix recursos humans diversos i infraestructures adaptades (Du et al., 2023; Marvin et al., 2022; Yigitcanlar et al., 2020, 2023), per a la qual cosa es necessiten considerable inversions econòmiques (Bettoni et al., 2021). També es generen més despeses per la capacitat i la formació del personal i les comunitats (Sadek et al., 2024; Varanasi, 2023). Les ciutats també han de preveure els costos jurídics i de conformitat, com ara auditories i ajustos del sistema per complir les normes reglamentaries, inversions que poden suposar despeses importants, especialment per a les ciutats petites i mitjanes.

Per superar aquestes limitacions, les ciutats poden recórrer al treball en xarxa a través de xarxes que proporcionen oportunitats per a l'intercanvi de coneixement, la innovació política i les respostes coordinades a problemes globals. En són exemples:

Coalició de Ciutats per als Drets Humans Digitals: plataforma per fomentar un desenvolupament inclusiu i democràtic de les noves tecnologies a les ciutats.

City AI Connect: comunitat d'aprenentatge global i plataforma digital perquè les ciutats provin i promoguin l'ús de IA generativa per millorar els serveis públics.

GovAI: coalició formada per més de 1.000 membres i més de 350 entitats locals, estatals i federals unides amb l'objectiu de promoure una IA responsable i amb un propòsit clar en el sector públic.

AI4Cities: projecte que va permetre a Hèlsinki, Amsterdam, Copenhagen, el Gran París, Stavanger i Tallinn desafiar el mercat a trobar solucions d'IA per reduir les emissions de CO₂ en els camps de l'energia i la mobilitat.

Referències bibliogràfiques

Barocas, S., Hardt, M., i Narayanan, A. «Classification». Dins *Fairness and Machine Learning : Limitations and Opportunities*. MIT Press, 2023a

Barocas, S., Hardt, M., i Narayanan, A. «Relative notions of fairness». Dins *Fairness and Machine Learning: Limitations and Opportunities*. MIT Press, 2023b

Bettoni, A. et al. «An AI adoption model for SMEs: A conceptual framework». *IFAC-PapersOnLine*, 2021, vol. 54, n.º 1, p. 702-708.

Calzada, I. «The right to have digital rights in smart cities». *Sustainability (Switzerland)*, 2021, vol. 13, n.º 20.

Cugurullo, F. et al. «Urban AI in China: Social control or hyper-capitalist development in the post-smart city? ». *Frontiers in Sustainable Cities*, 2022

Dauvergne, P. «Facial recognition technology for policing and surveillance in the Global South: A call for bans». Dins *Third World Quarterly*. Routledge, 2022, vol. 43, n.º 9, p. 2325-2335.

Du, J. et al. «Artificial intelligence enabled participatory planning: A review». *International Journal of Urban Sciences*, 2023

Fainstein, S. «Planning and the Just City». Dins P. Marcuse (Ed.), *Searching for the just city: Debates in urban theory and practice*. Routledge, 2009

Ferrara, E. «Fairness and bias in artificial intelligence: A brief survey of sources, impacts, and mitigation strategies». *Sci*, 2023, vol 6, n.º 1, p. 3.

Fincher, R., i Iveson, K. «Justice and injustice in the city». *Geographical research*, 2012, vol 50, n.º.3, p. 231-241.

Giovanola, B., i Tiribelli, S. «Weapons of moral construction? On the value of fairness in algorithmic decision-making». *Ethics and Information Technology*, 2022, vol. 24, n.º 1, p. 3.

González-Sendino, R. et al. «A review of Bias and Fairness in Artificial Intelligence». *International Journal of Interactive Multimedia and Artificial Intelligence* (en premsa), 2023

Hall, P. «Cities of tomorrow: An intellectual history of urban planning and design since 1880 (4ª edició)». *Wiley-Blackwell*, 2014

Harvey, D., i Potter, C. «The right to the Just City». Dins P. Marcuse (Ed.), *Searching for the just city: Debates in urban theory and practice*. Routledge, 2009

Holdsworth, J. «What is AI bias?», 2023

Kleinberg, J. et al. «Discrimination in the Age of Algorithms». *Journal of Legal Analysis*, 2018, vol.10, 2005, p. 113-174.

- Lefebvre, H. «[The Right to the City](#)», 1968
- Lyu, Y. *et al.* «IF-City: Intelligible fair city planning to measure, explain and mitigate inequality». *IEEE Transactions on Visualization and Computer Graphics*, 2023
- Martinez, E., i Kirchner, L. «[The secret bias hidden in Mortgage-Approval algorithms](#)», 2021
- Marvin, S. *et al.* «Urban AI in China: Social control or hyper-capitalist development in the post-smart city? ». *Frontiers in Sustainable Cities*, 2022, vol. 4, p. 1030318.
- O'Neil, C. «Weapons of math destruction: How big data increases inequality and threatens democracy (First edition) ». *Crown*, 2016
- Rezende, I. N. «Facial recognition in police hands: Assessing the 'Clearview case' from a European perspective». *New Journal of European Criminal Law*, 2020, vol. 11, n.º 3, p. 375-389.
- Sadek, M. *et al.* «Challenges of responsible AI in practice: Scoping review and recommended actions». *AI & SOCIETY*, 2024
- Sanchez, T. W., Brenman, M., i Ye, X. «The Ethical Concerns of Artificial Intelligence in Urban Planning». *Journal of the American Planning Association*, 2024, vol. 0, n.º 0.
- Sherman, S. «The Polyopticon: A diagram for urban artificial intelligences». *AI and Society*, 2023, vol. 38, n.º 3, p. 1209-1222.
- UNESCO. «Readiness assessment methodology: A tool of the recommendation on the ethics of artificial intelligence». *UNESCO*, 2023
- UN-Habitat. «[Mainstreaming human rights in the digital transformation of cities-A guide for local governments](#)». *United Nations Human Settlements Programme*, 2020.
- UN-Habitat. «[Human rights in the digital era](#)». *United Nations Human Settlements Programme*, 2023, p. 1-56.
- Varanasi, R. A. «“It is currently hodgepodge”: Examining AI/ML Practitioners' Challenges during Co-production of Responsible AI Values», 2023
- Wachter, S., Mittelstadt, B., i Russell, C. «Why fairness cannot be automated: Bridging the gap between EU non-discrimination law and AI». *Computer Law & Security Review*, 2021, vol. 41, p. 105567.
- Weinkauff, D. «[Privacy Tech-Know Blog: When Worlds Collide – The Possibilities and Limits of Algorithmic Fairness \(Part 1\)](#)». *Office of the Privacy Commissioner of Canada*, 2023.
- Yigitcanlar, T. *et al.* «Artificial intelligence in local governments: Perceptions of city managers on prospects, constraints and choices». *AI and Society*, 2023, vol. 38, n.º 3, p. 1135-1150.

Yigitcanlar, T. *et al.* «Responsible urban innovation with local government artificial intelligence (Ai): A conceptual framework and research agenda». *Journal of Open Innovation: Technology, Market, and Complexity*, 2021, vol. 7, n.º 1, p. 1-16.

Yigitcanlar, T. *et al.* «Contributions and Risks of Artificial Intelligence (AI) in Building Smarter Cities: Insights from a Systematic Review of the Literature», 2020

María Pérez-Ortiz

*Profesora titular, Centro de Inteligencia Artificial
de University College London (Reino Unido)*

1. Introducció

Segons previsions de les Nacions Unides, l'any 2050 prop del 70% de la població mundial viurà a zones urbanes (enfront del 56% actual) (ONU-Habitat, 2022). Com que el món se segueix urbanitzant a un ritme sense precedents, els problemes als quals s'enfronten les ciutats –mobilitat, prestació de serveis i habitatge, contaminació i salut urbana i ús de recursos, entre altres– esdevenen cada cop més apressants. Degut a aquest ràpid creixement urbà, sorgeix la necessitat de desenvolupar solucions innovadores per garantir que les ciutats siguin habitables, favorables al desenvolupament humà, eficients i respectuoses amb el medi ambient.

Un dels avenços més controvertits a l'hora de fer front a aquests reptes és la integració de la tecnologia i, més concretament, la intel·ligència artificial (IA) als entorns urbans. A més, a mesura que es va generalitzant l'ús de la IA, van sorgint dubtes sobre la seva sostenibilitat, tant mediambiental com social. En aquest capítol explorem el concepte d'IA sostenible i ens centrem en el paper que aquesta ha de tenir en el desplegament d'aquestes tecnologies en entorns urbans. Analitzem les qüestions mediambientals, socials i econòmiques relacionades amb el desplegament de la IA a les ciutats i emfasitzem els avantatges, els reptes i les orientacions per al futur de la IA a la recerca de futurs urbans més sostenibles i equitatius.

2. La recerca de la sostenibilitat

La sostenibilitat, sobretot la relativa al desenvolupament, tal com la defineixen les Nacions Unides (Keeble, 1988), té a veure amb la capacitat per fer que el desenvolupament satisfaci les necessitats humanes presents i futures (per exemple, la salut, el benestar, l'educació de qualitat, la feina digna o la igualtat social) i que ho faci dins els límits socioecològics actuals. Sovint es divideix en els tres pilars següents:

La IA sostenible inclou l'ús de la IA per a la sostenibilitat social, mediambiental i econòmica, però també la sostenibilitat de la mateixa IA, en abordar una sèrie de problemes, com ara el seu consum d'energia, l'ús de recursos i l'equitat social.

- **Sostenibilitat mediambiental:** viure tenint en compte els nostres recursos, protegir els nostres ecosistemes i permetre que es desenvolupin. Als entorns urbans, un repte fonamental és reduir les emissions de gasos d'efecte hivernacle i millorar la qualitat de l'aire urbà. Per exemple, les ciutats poden declarar zones de baixes emissions, ampliar el transport públic i fomentar els edificis energèticament sostenibles per reduir la contaminació urbana i pal·liar l'efecte d'illa de calor.
- **Sostenibilitat social:** assolir de manera permanent un bon benestar social. A les ciutats això sol implicar potenciar la inclusió, l'accessibilitat, l'equitat i una vida més sostenible als processos d'urbanització. Un repte actual és crear habitatge assequible i assegurar un accés equitatiu a les zones verdes, el transport i els serveis essencials. Per exemple, els governs de les ciutats poden dissenyar barris d'ús mixt pels quals sigui agradable caminar, que prioritzin la integració social i enjardinar zones amb plantes autòctones per crear entorns més ecològics.
- **Sostenibilitat econòmica:** fer un ús eficient i responsable dels recursos. La dimensió econòmica de la sostenibilitat se centra en les economies justes, verdes i circulars en detriment del creixement econòmic constant. En el context de la sostenibilitat urbana, el quid de la qüestió és el desenvolupament d'economies resilents i que generin baixes emissions de carboni. N'és un exemple la promoció d'infraestructura sostenible que potencii l'ocupació verda i els principis de l'economia circular, com ara la creació de sistemes de mobilitat urbana sostenible (per exemple, la bicicleta compartida o les infraestructures per al vehicle elèctric) i el foment de les empreses locals ecològiques.

Tot i que sovint concebem la sostenibilitat com si es referís només al pilar mediambiental, els tres exemples anteriors de sostenibilitat mediambiental, social i econòmica palesen que els tres pilars estan interconnectats i que cal considerar-los en conjunt. Tal com ho va expressar el filòsof mediambiental John Muir (2023[1911]), «Quan hom intenta aïllar una sola cosa, descobrim que aquesta està connectada amb tot el que hi ha a l'Univers». Igualment, les ciutats són sistemes complexos i, com a tal, cal considerar-les en conjunt i adoptar plantejaments que resolguin objectius múltiples i tinguin en compte els pilars social, mediambiental i econòmic.

3. La confluència de la tecnologia i la sostenibilitat

Podrien els avenços tecnològics influir en la nostra capacitat per garantir un futur sostenible? (GACGC, 2019). És fonamental que entenguem l'impacte que té la tecnologia en la sostenibilitat, sobretot a mesura que tecnologies disruptives com la IA acceleren el canvi global impulsat per l'ésser humà a un ritme sense precedents i sense una agenda unificada clara.

En aquests darrers anys, ha sorgit un nou conjunt de principis desitjables per a la sostenibilitat en relació amb els sistemes d'IA (Vinueza, 2020; Van Wynsberghe, 2021). En concret, la IA sostenible és un marc en ràpida evolució que pretén donar forma al desenvolupament, el desplegament i l'ús de tecnologies d'IA de manera mediambientalment, socialment i èticament responsable cercant l'equilibri entre els riscos i les oportunitats de la IA.

En la seva essència, la IA sostenible també té en compte les tres dimensions principals de la sostenibilitat: el pilar social, el mediambiental i l'econòmic. No obstant això, la IA sostenible és més que la suma de les seves parts: inclou l'ús de la IA per a la sostenibilitat social, mediambiental i econòmica, però també —i això és important— la sostenibilitat de la mateixa IA (Van Wynsberghe, 2021), en abordar una sèrie de problemes, com ara el seu consum d'energia, l'ús de recursos i l'equitat social, entre altres, per garantir que les solucions d'IA contribueixin de manera positiva a la societat i al medi ambient al llarg del temps.

Un dels reptes més importants a l'hora de desenvolupar IA sostenible és fer front a la petjada ecològica de la mateixa IA. Els sistemes d'IA, en especial els models a gran escala com els que s'utilitzen en l'aprenentatge profund, necessiten una enorme capacitat de càlcul, cosa que provoca un gran consum d'energia i altes emissions de carboni. Per exemple, segons una estimació de 2019, entrenar un sol gran model de llenguatge pot emetre tant carboni com cinc cotxes en tot el seu cicle de vida (Strubell et al., 2019) i és probable que aquesta estimació hagi augmentat considerablement des d'aleshores. Això té en compte només l'entrenament, però no la petjada derivada del seu ús (per exemple, s'estima que crear una imatge amb IA generativa consumeix tanta energia com carregar un telèfon mòbil (Heikkilä, 2023). A més de les emissions derivades del funcionament de la IA, és important tenir en compte l'energia incorporada, és a dir, la petjada de carboni relacionada amb la producció, el transport i l'eliminació del *hardware* utilitzat per a la IA, com ara els servidors i els processadors gràfics. Així, la IA sostenible pretén minimitzar tant l'impacte mediambiental del funcionament com el derivat de l'energia incorporada fomentant uns algorismes més eficients energèticament com ara la quantització o la destil·lació de models.

Deixant de banda els seus impactes mediambientals, la sostenibilitat social de la IA també és un problema apressant. Si no es dissenyen amb cura, els sistemes d'IA poden reforçar les desigualtats socials preexistents, degut als biaixos dels algorismes o a un accés desigual a les tecnologies d'IA. Per tant, es requereix la creació de sistemes d'IA que fomentin la inclusió social i l'equitat, de manera que es garanteixi que els grups marginats no resultin perjudicats per les decisions impulsades per la IA en àmbits com la contractació de treballadors, l'habitatge o la justícia penal. Per això, cal incloure qüestions ètiques i relacionades amb els drets humans en el disseny i la implementació de la IA, amb marcs de transparència, rendició de comptes i governança robustos que protegeixin contra danys no intencionats. Per a una anàlisi més aprofundida de l'equitat en els sistemes d'IA, vegeu el capítol sobre equitat d'aquesta monografia.

Pel que fa a la sostenibilitat econòmica, la IA té el potencial tant d'impulsar com de trastocar el desenvolupament sostenible. La IA pot optimitzar les infraestructures urbanes, les xarxes elèctriques i els sistemes de transport, fomentant un ús més eficient dels recursos i reduint els residus a les ciutats. També pot facilitar la transició a una economia circular millorant processos com la gestió de la cadena de subministrament, la reducció de residus i l'optimització del cicle de vida dels productes. No obstant això, si no es gestiona amb cura, la IA pot agreujar les desigualtats econòmiques, en automatitzar llocs de feina sense crear noves oportunitats

La IA sostenible no només és un subconjunt de tecnologies dissenyades específicament per a la sostenibilitat, sinó que també implica reconfigurar tot el camp de la IA per garantir que contribueix sistemàticament al benestar social, mediambiental i econòmic a llarg termini.

laborals o en augmentar la concentració de la riquesa. Per tant, la IA sostenible aposta per models econòmics impulsats per IA que donin prioritat als beneficis socials a llarg termini en detriment dels guanys a curt termini, cosa que garanteix que els guanys econòmics derivats de la IA es distribueixen de manera equitativa.

Algunes de les preguntes que caldria plantejar-se en el disseny de sistemes d'IA en el context de la sostenibilitat són les següents (Vinuesa, 2020; Van Wynsberghe, 2021):

- Com es compensen els impactes directes i indirectes de la tecnologia de la IA sobre la societat, el medi ambient i l'economia? Com podem dissenyar sistemes d'IA que siguin més sostenibles des del començament? Quin marc per a l'avaluació dels riscos ens pot ajudar a preveure'n les conseqüències involuntàries abans que sorgeixin?
- Com podem tenir en compte el sistema sociotecnològic més ampli que envolta la IA, que inclou les repercussions socials sobre les persones que usen aquestes tecnologies i hi estan afectades? Quines accions podem emprendre per desenvolupar una IA que estigui en consonància amb la conservació dels recursos naturals per a les generacions actuals i futures, fomenti models econòmics sostenibles i respecti els valors socials fonamentals de les diferents comunitats?
- Com podem potenciar el canvi al llarg de tot el cicle de vida de la IA –des de la generació d'idees, l'entrenament i el refinament fins a l'avaluació, la implementació i la governança– per assolir major sostenibilitat ecològica i equitat social? Quines mesures són necessàries per garantir que els sistemes d'IA operen dins els límits ecològics del nostre planeta, pel que fa, per exemple, al consum d'energia, l'ús d'aigua dolça i la dependència en minerals escassos?

En definitiva, la IA sostenible fomenta un enfocament en què la IA actua com a palanca per assolir objectius de sostenibilitat, augmentant els esforços per mitigar el canvi climàtic, reduir la desigualtat i potenciar economies inclusives i resilients, alhora que garanteix que el desenvolupament i l'ús de les mateixes tecnologies d'IA compleixin els principis de la sostenibilitat. Cal reconèixer que les tecnologies, incloent-hi la IA, poden suposar riscos extrínsecs i intrínsecs per a la sostenibilitat. Pel que fa als riscos intrínsecs, tot i que la tecnologia no s'apliqui de manera directa a problemes de sostenibilitat, si és defectuosa, poc robusta o poc equitativa, pot perjudicar de manera no intencionada els objectius de sostenibilitat agreujant la desigualtat, els danys ecològics o la inestabilitat econòmica per vies indirectes. Per tant, la IA sostenible no només és un subconjunt de tecnologies dissenyades específicament per a la sostenibilitat, sinó que també implica reconfigurar tot el camp de la IA per garantir que contribueix sistemàticament al benestar social, mediambiental i econòmic a llarg termini.

4. IA sostenible en entorns urbans

Tot i que hi comença a haver més marcs per a la sostenibilitat de la IA (Vinuesa, 2020; Van Wynsberghe, 2021; Wu, 2022; Wilson, 2022; Nishant, 2020), actualment molts pocs treballs se centren en els futurs

urbans (Yigitcanlar, 2020; Bibri, 2021; Pastor-Escuredo, 2022). Amb aquest objectiu aquí volem presentar una perspectiva que exemplifica els fonaments que serien necessaris per garantir que els sistemes d'IA desplegats a les ciutats no només són avançats tecnològicament, sinó també responsables, equitatius i beneficiosos tant per al medi ambient com per a les poblacions urbanes.

Impacte mediambiental en els sistemes d'IA urbans. Un dels problemes rellevants de la IA és el seu impacte mediambiental, que inclou, entre altres, el seu considerable consum d'energia i d'aigua dolça (Luccioni, 2024). En entorns urbans, on la IA s'utilitza cada cop més en aplicacions com sistemes intel·ligents de regulació del trànsit, xarxes elèctriques i gestió dels edificis, les demandes acumulades d'aquests sistemes poden ser molt importants. Les tecnologies d'IA a les ciutats poden tenir un gran impacte mediambiental degut a la seva dependència dels centres de dades i dels recursos computacionals. Un enfocament sostenible de la IA urbana implicaria optimitzar aquests sistemes perquè siguin eficients energèticament, amb tècniques com l'aprenentatge automàtic en dispositius petits (*tiny machine learning*), les pràctiques de computació ecològica de *software*, la destil·lació del coneixement, i la poda o quantització de models. Això també pot implicar aprendre a partir de conjunts de dades més petits, però d'alta qualitat (és a dir, fer més amb menys), usant energia renovable, fomentant models sostenibles de producció i consum i minimitzant la petjada de carboni a l'hora de desplegar la IA a tota una ciutat. Per exemple, les xarxes intel·ligents que funcionen amb IA poden ajustar de manera dinàmica el consum d'energia basant-se en dades obtingudes en temps real, cosa que redueix els residus i fomenta la integració de fonts d'energia renovable com la solar i l'eòlica. L'aprenentatge federat obre una altra via prometedora per a la sostenibilitat de la IA urbana. En comptes de dependre de centres de dades centralitzats, amb aquesta tècnica es poden entrenar models més petits directament en dispositius descentralitzats, com els sensors de l'Internet de les Coses (IdC) incrustats en la infraestructura urbana. Així es redueix la necessitat de transmetre grans quantitats de dades a servidors centralitzats, fet que redueix el consum d'energia en el processament i l'emmagatzemament de dades. Aprofitant els recursos informàtics locals ja existents, l'aprenentatge federat també redueix la demanda agregada de *hardware* i mitiga la petjada ecològica de les operacions d'IA massives. A més, millora la privacitat i la seguretat de les dades, ja que manté la informació sensible en dispositius locals, cosa que redueix la necessitat de compartir dades i alhora fomenta pràctiques d'IA sostenibles.

Sostenibilitat social: igualtat i equitat. A mesura que les ciutats adopten cada cop més la IA per impulsar serveis com la seguretat, l'assistència sanitària i la distribució de recursos, és fonamental que aquests sistemes contribueixin a la sostenibilitat social fomentant la igualtat, l'equitat i la inclusió. En l'urbanisme, per exemple, la IA sostenible es pot aprofitar per identificar i abordar les desigualtats, com ara garantint als barris desfavorits un accés equitatiu al transport, a l'assistència sanitària i a l'educació. No obstant això, el component social de la IA sostenible no només implica la finalitat del seu ús, el disseny

Les tecnologies d'IA a les ciutats poden tenir un gran impacte mediambiental degut a la seva dependència dels centres de dades i dels recursos computacionals. Un enfocament sostenible de la IA urbana implicaria optimitzar aquests sistemes perquè siguin eficients energèticament.

La veritable sostenibilitat social d'aquests sistemes d'IA també ha de tenir en compte l'ètica del seu procés de desenvolupament, garantint pràctiques justes al llarg de tot el cicle de vida de la IA, potenciant ciutats més justes i inclusives i abordant alhora les desigualtats globals en la producció de la IA.

d'algoritmes que minimitzin els biaixos i la garantia que les comunitats marginades no es vegin perjudicades per les decisions basades en la IA, sinó també abordar les implicacions ètiques del desenvolupament i el desplegament dels sistemes d'IA. Molts d'aquests són entrenats i mantinguts per treballadors de comunitats del Sud Global que estan mal pagats i fan jornades maratonianes (Rowe, 2023), els quals sovint estan contractats per terceres empreses. Aquesta mà d'obra, imprescindible per entrenar molts dels sistemes d'IA, posa en relleu les profundes desigualtats existents en la cadena de subministrament de la IA global, atès que aquests operaris sovint treballen en condicions precàries i alhora paguen el peatge de dur a terme tasques repetitives i infravalorades. La veritable sostenibilitat social d'aquests sistemes també ha de tenir en compte l'ètica del seu procés de desenvolupament, garantint pràctiques justes al llarg de tot el cicle de vida de la IA, cosa que contribueix a fomentar la sostenibilitat social potenciant ciutats més justes i inclusives i abordant alhora les desigualtats globals en la producció de la IA.

Governança ètica i rendició de comptes. Els sistemes d'IA urbans s'han de regir per marcs ètics robustos que donin prioritat a la transparència i la rendició de comptes. Les parts interessades i els governs municipals han de garantir que els sistemes d'IA siguin explicables i els processos de presa de decisions clars; això generaria confiança i garantiria que es puguin identificar i abordar de seguida tots els errors o conseqüències no desitjades. Per exemple, els sistemes d'IA usats per a la vigilància o l'ordre públic a les ciutats s'han de dissenyar amb estructures clares de rendició de comptes, que protegeixin la privacitat i els drets civils dels ciutadans. Per a una anàlisi més aprofundida sobre l'aplicació pràctica d'aquests principis als sistemes d'IA, vegeu el capítol sobre transparència i rendició de comptes d'aquest volum.

Sostenibilitat econòmica a les ciutats. Perquè la IA sigui sostenible en entorns urbans, també ha de ser econòmicament viable a llarg termini, cosa que implica desenvolupar sistemes d'IA que s'integrin sense problemes amb la infraestructura urbana ja existent, s'adaptin a la demanda futura i es construeixin per a un ús a llarg termini. Les ciutats poden potenciar models d'economia circular fomentant la reutilització i el reciclatge de les tecnologies, les dades i el *hardware* de la IA, amb la qual cosa es redueixen els residus i els costos.

Concordança amb els objectius de sostenibilitat urbana. Els sistemes d'IA desplegats a les ciutats no només han de potenciar els objectius de sostenibilitat urbana –com ara reduir la contaminació o millorar la salut pública i la qualitat de vida–, sinó també garantir que el mateix ús de les tecnologies d'IA contribueixi a la sostenibilitat. Una manera d'aconseguir-ho és reutilitzant l'energia i els recursos que fa servir la infraestructura d'IA. Per exemple, reutilitzar l'excedent de calor procedent dels centres de dades –un important subproducte derivat de les demandes computacionals de la IA– pot reduir el consum total d'energia. A Estocolm, el projecte de Parcs de Dades d'Estocolm¹ ha demostrat que la calor residual dels centres de dades es pot desviar per utilitzar-la com a calefacció en edificis residencials i comercials, cosa que palesa que la infraestructura d'IA es pot integrar en un model d'economia circular en concordança amb els objectius climàtics i alhora reduir

1. Vegeu <https://stockholmdataparks.com/>

les necessitats d'energia de la societat. A part de l'eficiència de recursos (que es relaciona amb el concepte de la sostenibilitat de la pròpia IA), la IA pot servir per optimitzar els sistemes urbans quant a la sostenibilitat. Aprofitant la IA per millorar la gestió dels recursos, reduir el consum d'energia i fomentar les iniciatives de resiliència climàtica, les ciutats poden fer front a problemes apressants com el canvi climàtic i la urbanització. Per exemple, les zones urbanes són especialment vulnerables als efectes del canvi climàtic, com ara temperatures més altes o fenòmens meteorològics extrems. La IA pot proporcionar analítica predictiva avançada i sistemes d'alerta primerenca per als riscos climàtics. Aquests sistemes poden avisar les autoritats de potencials perills mediambientals i permetre una resposta ràpida, fet que ajuda a mitigar els efectes de la contaminació urbana sobre la salut i el medi ambient.

La IA per a la planificació i la urbanització de les ciutats. La IA transforma la planificació i la urbanització de les ciutats (Jha, 2021) i fomenta una major sostenibilitat social, mediambiental i econòmica. Analitzant grans conjunts de dades sobre creixement demogràfic, ús del terreny, patrons de mobilitat i factors mediambientals, entre molts altres, la IA pot ajudar urbanistes i legisladors a dissenyar ciutats més sostenibles i eficients usant els anomenats bessons digitals. Per exemple, els models d'IA poden predir com els canvis en la infraestructura, com ara la construcció de noves carreteres o l'ampliació de la xarxa de transport públic, afectaran els patrons del trànsit, els nivells de contaminació i el consum d'energia. Això ajuda els planificadors a anticipar-se al futur i a dur a terme una prospectiva responsable, la qual cosa permet prendre decisions amb major coneixement de causa que promoguin la sostenibilitat a llarg termini. La IA també es pot usar per optimitzar l'ús del terreny i les polítiques de zonificació, de manera que es garanteixi que s'arriba a un equilibri entre la urbanització i la preservació de les zones verdes i els recursos naturals. Això és especialment important a les ciutats que experimenten un creixement ràpid, on la demanda d'habitatge i infraestructura sovint comporta la dispersió urbana i la pèrdua de valuosos ecosistemes. No obstant això, més enllà de la seva aplicació als sistemes urbans, és fonamental que el desenvolupament i el desplegament de la mateixa IA estiguin en consonància amb les pràctiques sostenibles. Tot i que els bessons digitals són eines molt potents per simular diferents escenaris de l'urbanisme, la seva sostenibilitat depèn de l'eficàcia dels models de la IA subjacent i de la infraestructura que hi dona suport. Tanmateix, tal com apunten molts estudis (Andersson, 2021), cal assenyalar que els bessons digitals construïts amb IA poden constituir un enfocament més eficient dels recursos que les seves simulacions físiques equivalents, atès que es poden executar en unitats centrals de processament (CPU) d'ordinadors portàtils en uns pocs segons, en contraposició de la necessitat de superordinadors durant dies.

Perquè la IA sigui sostenible en entorns urbans, també ha de ser econòmicament viable a llarg termini, cosa que implica desenvolupar sistemes d'IA que s'integrin sense problemes amb la infraestructura urbana ja existent, s'adaptin a la demanda futura i es construeixin per a un ús a llarg termini.

5. Recomanacions polítiques i observacions finals

La innovació i la tecnologia tindran un paper cada cop més central en la planificació de futurs urbans sostenibles (ONU-Habitat, 2022). Tal com analitzem amb la llista de recomanacions polítiques que presentem a continuació, el disseny i el desplegament de la tecnologia ha de dissenyar-se per adaptar-se a la gran diversitat de contextos urbans:

a. Recomanacions sobre sostenibilitat mediambiental

- La urgència per descarbonitzar les economies urbanes ha d'impulsar la **convergència de les tecnologies verdes i les tecnologies intel·ligents**. Les polítiques s'han de centrar en l'eficiència energètica, la conservació del medi ambient i la resiliència, cosa que inclou l'establiment de normes per a la IA verda que donin prioritat als algoritmes i el *hardware* eficients energèticament, així com la creació d'economies circulars al voltant dels centres de dades, per exemple, reciclant la calor residual. Les polítiques de gestió del cycle de vida han de fomentar el subministrament responsable, la reutilització i el reciclatge del *hardware* per a la IA per minimitzar els residus electrònics.
- Les avaluacions d'impacte han de ponderar acuradament si el desplegament de la IA per a projectes de sostenibilitat justifica els costos mediambientals de la tecnologia, com destaquen treballs previs (Dixon, 2022). És essencial disposar de nous marcs per mesurar i comparar tots els costos de la IA al llarg de tot el cycle de vida i garantir una avaluació exhaustiva de la sostenibilitat.
- Comencen a sorgir sistemes de classificació amb estrelles per a l'eficiència energètica de la IA (Luccioni, 2024), els quals s'han d'afegir als dispositius urbans de l'Internet de les Coses per proporcionar informació valuosa als usuaris perquè entenguin millor l'impacte mediambiental de les eines que utilitzen i així les puguin adoptar de manera més responsable.
- La col·laboració i les associacions publicoprivades poden impulsar el desenvolupament de tecnologies d'IA sostenibles a zones urbanes. Crear **consorcis de ciutats per a la IA sostenible** pot fomentar l'associació entre governs, empreses tecnològiques i institucions de recerca per abordar reptes urbans com ara la gestió de l'energia, el transport i la reducció de residus. Les ciutats han d'incentivar el desenvolupament de la IA sostenible establint desgravacions fiscals o subvencions per a empreses que desenvolupin solucions d'IA respectuoses amb el medi ambient.
- La resiliència urbana i les infraestructures intel·ligents han de tenir el suport de **polítiques que fomentin la IA per a la resiliència climàtica i de la biodiversitat**, cosa que inclou sistemes d'alerta primerenca potenciats per IA per a catàstrofes naturals, punts d'inflexió quant a la pèrdua de biodiversitat i fenòmens climàtics extrems.

b. Recomanacions sobre sostenibilitat social

- Atès que totes les dimensions de la IA sostenible estan interrelacionades, el desplegament de la IA ètica i responsable també és una dimensió fonamental als entorns urbans. Els sistemes d'IA s'han d'auditar pel que fa a l'equitat per evitar la discriminació i la desigualtat social. La creació de **junes d'ètica de la IA local** ha de garantir que els projectes urbans d'IA estan en consonància amb les normes de privacitat, equitat i rendició de comptes.
- La seguretat i la privacitat de les dades també són àmbits prioritaris. Cal aprovar **lleis contundents sobre la privacitat de les dades**

urbanes per protegir les dades personals que recopilen els sensors, les càmeres i les aplicacions mòbils, i això n'inclou l'anonimització i l'ús amb consentiment explícit. A més, són necessaris marcs transparents i segurs per a l'intercanvi de dades entre els governs, les empreses privades i els desenvolupadors d'IA per assegurar un ús responsable de les dades dels ciutadans que no en comprometi la privacitat.

- Per potenciar la comprensió i el suport de la ciutadania, les polítiques han de promoure **la participació pública i l'alfabetització digital**. Les ciutats han de fomentar **models de governança participativa** que impliquin els ciutadans en els processos de presa de decisions en matèria d'IA i alhora han de llançar campanyes d'alfabetització digital per formar la ciutadania en les tecnologies d'IA, l'impacte d'aquestes tecnologies i la protecció dels seus drets.
- Garantir un **accés igualitari als serveis públics potenciats per IA**, especialment per a les comunitats marginades i desafavorides, és crucial per assolir la inclusió.
- Fomentar la **ciència oberta** és prioritari, a fi de permetre les auditories públiques d'aquests sistemes i alhora garantir mesures robustes de ciberseguretat per protegir les dades sensibles i els serveis públics. La transparència també és clau, per la qual cosa calen regulacions que exigeixin que els sistemes d'IA usats en els serveis públics siguin explicables perquè tant les parts interessades com la ciutadania pugin entendre com es prenen les decisions.

c. Recomanacions generals

- Per tal d'abordar l'impacte econòmic de la IA, les polítiques han de **fomentar la transició laboral i el desenvolupament de la mà d'obra**. Cal destinar finançament públic a programes de reciclatge laboral que ajudin els treballadors en la transició a noves feines, sobretot en sectors emergents en què l'automatització pot generar pèrdues de llocs de treball. Fomentar el creixement del nombre de llocs de treball verds basats en IA, com ara els que estan relacionats amb la gestió de l'energia renovable i la infraestructura urbana sostenible, pot impulsar més el creixement econòmic sostenible.
- Són essencials els marcs de supervisió i rendició de comptes per garantir que els sistemes d'IA concorden amb els objectius de sostenibilitat a llarg termini. Les **avaluacions obligatòries d'impacte de la IA**, semblants a les avaluacions d'impacte mediambiental, han d'avaluar els efectes socials, econòmics i mediambientals del desplegament de la IA a les ciutats. La supervisió i l'auditoria contínues dels sistemes d'IA urbans poden ajudar a garantir que segueixen sent adaptables als nous reptes i a les consideracions ètiques.
- Cal **incloure qüestions sobre normes mediambientals, socials i de governança** que tinguin en compte la sostenibilitat de les dades, els algorismes i els recursos computacionals usats per les empreses, així com el suport donat a les fonts d'energia renovables i les economies computacionals circulars.

- La normativa reguladora de les infraestructures urbanes intel·ligents ha d'assegurar que les **tecnologies d'IA són adaptables, interoperables i redimensionables per a les necessitats urbanes futures**, especialment en camps com la gestió del trànsit, la reducció de residus i l'eficiència energètica.
- Finalment, des de la perspectiva global, cal incentivar la **col·laboració i l'estandardització internacionals**. Les ciutats han de cooperar per desenvolupar normes de sostenibilitat globals per a la IA i compartir bones pràctiques (Strubell, 2019) que garanteixin la concordança amb els objectius internacionals. Les plataformes per compartir coneixement entre ciutats poden contribuir a accelerar l'adopció de pràctiques d'IA sostenible arreu del món.

Referències bibliogràfiques

Andersson, T. R., et al. «Seasonal Arctic sea ice forecasting with probabilistic deep learning». *Nature Communications*, vol. 12, n.º 1 (2021), p. 5124.

Bibri, S. E. «Data-driven smart sustainable cities of the future: Urban computing and intelligence for strategic, short-term, and joined-up planning». *Computational Urban Science*, vol. 1, n.º 1 (2021), p. 8.

Dixon, B., Pérez-Ortiz, M. i Bieker, J. «Comparing the carbon costs and benefits of low-resource solar nowcasting». *NeurIPS Workshop on Tackling Climate Change with Machine Learning* (2022).

German Advisory Council on Global Change, *Towards Our Common Digital Future*, Flagship Report, 2019.

Heikkilä, M. «Making an image with generative AI uses as much energy as charging your phone», *MIT Technology Review* (2023).

Jha, A. K. et al. «A review of AI for urban planning: Towards building sustainable smart cities». *2021 6th International Conference on Inventive Computation Technologies (ICICT)*. IEEE, 2021.

Keeble, B. R. «The Brundtland Report: "Our Common Future"». *Medicine and War*, vol. 4, n.º 1 (1988), p. 17-25.

Luccioni, S., Trevelin, B. i Mitchell, M. «The Environmental Impacts of AI—Primer», 2024.

Muir, J. «My first summer in the Sierra», Dins: *British Politics and the Environment in the Long Nineteenth Century*. Routledge, 1911 (reeditat en 2023), p. 291-296.

Nishant, R., Kennedy, M. i Corbett, J. «Artificial intelligence for sustainability: Challenges, opportunities, and a research agenda». *International Journal of Information Management*, vol. 53 (2020), p. 102104.

ONU- Habitat, «Envisaging the Future of Cities», *World Cities Report 2022*.

Pastor-Escuredo, D., Treleaven, P. i Vinuesa. R. «An Ethical framework for artificial intelligence and sustainable cities». *Ai*, vol. 3, n.º 4 (2022), p. 961-974.

Rowe, N. «Millions of Workers Are Training AI Models for Pennies», *WIRED* (2023).

Strubell, E., Ganesh, A. i McCallum, A. «Energy and policy considerations for modern deep learning research». *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 34, n.º 9 (2020).

Van Wynsberghe, A. «Sustainable AI: AI for sustainability and the sustainability of AI». *AI and Ethics* 1.3 (2021), p. 213-218.

Vinuesa, R. *et al.* «The role of artificial intelligence in achieving the Sustainable Development Goals». *Nature Communications*, vol. 11, n.º 1 (2020), p. 1-10.

Wilson, C., i Van Der Velden, M. «Sustainable AI: An integrated model to guide public sector decision-making». *Technology in Society*, vol. 68 (2022), p. 101926.

Wu, C-J. *et al.* «Sustainable AI: Environmental implications, challenges and opportunities». *Proceedings of Machine Learning and Systems*, vol. 4 (2022), p. 795-813.

Yigitcanlar, T., i Cugurullo, F. «The sustainability of artificial intelligence: An urbanistic viewpoint from the lens of smart and sustainable cities». *Sustainability*, vol. 12, n.º 20 (2020), p. 8548.

PART II. ESTUDIS DE CAS DE MARCS DE GOVERNANÇA DE LA IA URBANA

- ESTUDI DE CAS 1: BARCELONA
- ESTUDI DE CAS 2: AMSTERDAM
- ESTUDI DE CAS 3: NOVA YORK
- ESTUDI DE CAS 4: SAN JOSÉ
- ESTUDI DE CAS 5: DUBAI
- ESTUDI DE CAS 6: SINGAPUR

PART II. ESTUDIS DE CAS DE MARCS DE GOVERNANÇA DE LA IA URBANA

Alexandra Vidal D'oleo

Investigadora i gestora de projectes, Programa de Ciutats Globals, CIDOB

El primer informe de l'Atles d'intel·ligència artificial (IA) urbana del GOUAI (Galceran-Vercher i Vidal, 2024) revela que, tot i que moltes ciutats experimenten actualment amb la IA, només un petit percentatge ha implementat polítiques o estratègies generals per regular-ne l'ús i garantir-ne la concordança amb principis ètics fonamentals. Des de fa poc, l'atenció ha recaigut a abordar els reptes urbans immediats en un pragmatisme orientat a solucions. En conseqüència, hi ha una bretxa significativa entre l'adopció de la IA i l'establiment de marcs de governança efectius. No obstant això, motivats pel discurs públic i per un corrent dominant que impulsa la regulació global dels algoritmes i la IA, alguns governs locals han pres la iniciativa de crear els seus propis marcs de governança. S'espera que aquesta tendència vagi a més en els propers anys.

En les pàgines següents es presenta un seguit d'estudis de cas de ciutats d'arreu del món que s'han esforçat per establir marcs de governança local de IA, adoptant diferents mecanismes polítics per governar la IA de manera integral. Tots els mecanismes polítics s'emmarquen en una de les categories següents: (1) principis, estratègies i directrius; (2) normativa i legislació local; (3) mecanismes de transparència i explicabilitat; (4) avaluació de l'impacte algorítmic; (5) auditories i inspeccions reglamentàries; (6) supervisió humana, rendició de comptes, procediments d'audiència i apel·lació; (7) condicions d'adquisició pública; (8) òrgans assessors i supervisors externs/independents; (9) aliances, comunitats de pràctiques i grups d'aprenentatge; (10) programes de capacitat; (11) foment de la innovació, el coneixement i l'experimentació local; (12) participació de la comunitat; (13) governança de dades, i (14) altres polítiques i mesures. Cal assenyalar que, tot i que algunes de les polítiques que es presenten figuren en desenvolupament, potser ja han estat implementades sense previ avís.

Les ciutats triades presenten diferències quant a ubicació geogràfica, mida i renda per capita. En concret, es presenten els estudis de cas següents:

1. Barcelona (Espanya)
2. Amsterdam (Països Baixos)
3. Nova York (Estats Units)
4. San José (Estats Units)
5. Dubai (Emirats Àrabs Units)
6. Singapur (República de Singapur)

Estudi de cas 1: Barcelona

Població: 1.655.956 habitants (2023)	Renda per capita: 31.531 € (2022)	Regió: Europa
---	--	----------------------

MARC DE GOVERNANÇA DE LA IA		
[1] Principis, estratègies i directrius	Estratègia d'IA	<p>Estratègia municipal d'algoritmes i dades per a l'impuls ètic de la intel·ligència artificial: defineix un conjunt de principis guies i 20 accions per al desplegament ètic de la IA, com ara:</p> <ul style="list-style-type: none"> • L'ús de la IA per als sistemes de recomanacions automatitzades, en comptes de sistemes de presa de decisions. • La transparència i l'auditabilitat: les bases de dades i els models algorítmics han de ser accessibles, entenibles i auditables pel públic general. • L'establiment del règim de responsabilitat pels danys o les pèrdues que es puguin causar.
	Principis locals per a la IA	Inclusos en l'estratègia municipal d'IA: (1) acció i supervisió humana; (2) robustesa tècnica i seguretat; (3) privacitat i governança de les dades; (4) transparència; (5) diversitat, inclusió i equitat; (6) compromís social i ambiental, i (7) responsabilitat, rendició de comptes i control democràtic.
	Protocols interns per a l'ús de la IA	Definició de metodologies de treball i protocols per a la implementació de sistemes algorítmics: defineix els mecanismes de cada etapa de la licitació i implementació dels sistemes d'IA per part de l'Ajuntament i estableix els òrgans de governança i supervisió que vetllaran perquè l'impacte de la IA estigui en consonància amb els principis ètics.
[3] Mecanismes de transparència i explicabilitat	Registre públic d'algoritmes (*)	Creació d'un registre municipal d'algoritmes actuals i futurs que afecten els serveis i els tràmits municipals. El registre serà públic i també servirà per classificar els algoritmes segons el risc que suposen, amb explicacions clares per als ciutadans i per a altres parts interessades. Per a cada algoritme registrat el registre inclourà un punt de contacte públic al qual la ciutadania es podrà adreçar.
	Estàndard de transparència algorítmica	Desenvolupament d'un model algorítmic comú que garanteixi l'ús adequat de les dades. Projecte desenvolupat juntament amb 8 ciutats europees més en el marc d'Eurocities.
	Web municipal amb tota la informació pertinent sobre la IA	Barcelona Ciutat Digital: portal municipal amb tots els recursos disponibles a la ciutat per potenciar la digitalització, incloent-hi tota la informació pertinent referent a projectes i iniciatives d'IA.
[4] Avaluació de l'impacte algorítmic	Avaluació i gestió del risc	Anàlisi del risc (inclosa en el protocol): l'oficina tècnica d'IA avalua els algoritmes que utilitza el consistori i emet un informe amb l'avaluació del risc. Segons el nivell de risc determinat per l'oficina, s'estableixen les accions següents. Tots els algoritmes considerats en la categoria de «risc inacceptable» són rebutjats, mentre que els algoritmes d'«alt risc» han de passar per una avaluació obligatòria d'impacte algorítmic per part de la Comissió Transversal municipal (vegeu-ne més detalls a continuació).
	Avaluació d'impacte en matèria de drets humans	Avaluacions obligatòries de l'impacte dels algoritmes per als sistemes d'alt risc (incloses en l'estratègia i en el protocol).
[5] Auditories i inspecció reglamentària	Auditories	Auditories obligatòries per als sistemes d'alt risc (incloses en l'estratègia i en el protocol). Les conclusions es faran públiques a través del registre d'algoritmes.
[7] Condicions d'adquisició pública	Clàusules d'adquisició pública*	Inclusió de clàusules relatives als drets digitals en la licitació de solucions basades en la IA.

[8] Òrgans assessors i supervisors externs/independents	Consell assessor	El Consell Assessor en Intel·ligència Artificial, Ètica i Drets Digitals està integrat per 15 membres, que són persones expertes i independents procedents de diferents disciplines. Té com a objectius assessorar el consistori en l'ús de la IA, dur a terme estudis d'impacte algorítmic dels sistemes algorítmics d'alt risc i avaluar el desenvolupament de l'estratègia municipal d'IA, entre altres.
[9] Aliances, comunitats de pràctiques i grups d'aprenentatge	Membre d'una comunitat de pràctiques	Coalició de les Ciutats pels Drets Digitals (CC4DR, per la sigla en anglès) (membre fundador): xarxa de ciutats a escala global amb més de 60 membres que treballen en el camp del disseny de polítiques basades en drets digitals.
[10] Programes de capacitat	Capacitació municipal	Formació del personal municipal (inclosa en l'estratègia).
	Òrgan municipal	Comissió transversal per a l'impuls de la IA ètica : constituïda per 25 membres, té l'objectiu d'orientar i alinear les polítiques municipals que desenvolupen eines que utilitzen IA i promoure la col·laboració interdepartamental. Ha desenvolupat l'estratègia d'IA.
[11] Promoció de la innovació, el coneixement i l'experimentació local	Observatoris de IA local	Observatori Global d'IA Urbana (GOUAI, per la sigla en anglès) : una iniciativa de la ciutat de Barcelona, impulsada pel CIDOB amb la col·laboració d'Amsterdam i Londres, en el marc de la CCD4R, amb el suport de l'ONU-Habitat. Fa recerca sobre IA urbana des d'un punt de vista ètic.
[12] Participació de la comunitat	Participació popular	Canals de comunicació amb la ciutadania (inclosos en l'estratègia). Promoció d'espais per a la reflexió i el debat sobre l'impacte de la IA en els serveis públics (inclosos en l'estratègia).
[13] Governança de dades	Transparència de les dades	Open Data BCN : el servei de dades obertes de la ciutat. Té l'objectiu de fer públic el màxim nombre de recursos, fent que les dades generades o custodiades per organismes públics siguin accessibles, gratuïts i fàcils d'usar per tothom.
	Drets sobre les dades	Projecte Decode : iniciativa col·laborativa de la UE per reforçar els drets sobre les dades dels ciutadans i donar-los el control de les seves dades, així com permetre'n l'intercanvi en favor del bé comú.

* Mecanisme polític planificat, però encara no implementat completament el desembre de 2024.

Estudi de cas 2: Amsterdam

Població: 921.402 habitants (2022)	Renda per capita: 54.700 € (2022)	Regió: Europa
---	--	----------------------

MARC DE GOVERNANÇA DE LA IA		
[1] Principis, estratègies i directrius	Agenda d' IA	<p>Agenda d'intel·ligència d'Amsterdam (2020-2024): destacava els objectius de la ciutat en matèria d'algoritmes, especialment d'IA. Els objectius es centraven a adoptar accions per a la millora de la qualitat de vida dels residents a Amsterdam, reduir els efectes perjudicials de la digitalització i potenciar-ne els resultats positius.</p> <p>Visió d'Amsterdam sobre la IA (2024): en aquesta nova política, la ciutat destaca com caldria integrar la IA en la vida urbana i com hauria d'influir en la ciutat. La visió es va desenvolupar parlant amb treballadors municipals, experts i residents de la ciutat. Per posar-la en pràctica, l'administració municipal esta creant una nova Agenda d'IA, que s'espera que estigui publicada el 2025. Proporcionarà orientacions per a l'ús responsable de la IA dins el municipi, garantint que la tecnologia s'aplica de manera ètica, inclusiva i sostenible.</p>
	Principis locals per a la IA	<p>L'aprovació dels principis Tada (del Manifest Tada), que guien l'Agenda digital d'Amsterdam i també s'apliquen al camp de la IA: (1) inclusiva, (2) controlada, (3) adaptada a les persones; (4) legítima i supervisada, (5) oberta i transparent i (6) de tothom i per a tothom. Els principis que regeixen la IA local (tal com s'indiquen en la «visió d'Amsterdam sobre la IA» (2024)): (1) centrada en les persones, (2) fiable i (3) amb garantia de futur.</p>
	Directrius, guies i manuals	<p>Guia d'algoritmes: guia que estableix l'enfocament integral de la ciutat i les eines polítiques per fer un ús responsable dels algoritmes. L'enfocament al llarg del cicle de vida dels algoritmes consta de les 7 eines següents per gestionar i avaluar els riscos i estudiar els algoritmes al llarg de tot el seu cicle de vida: (1) registre d'algoritmes, (2) condicions de contractació, (3) procediment d'objecció, (4) definició de governança i model del cicle de vida, (5) auditoria, (6) model d'anàlisi de biaixos i (7) model d'anàlisi d'impacte sobre els drets humans.</p> <p>Manual d'equitat: normes per a l'anàlisi de biaixos, pla detallat per avaluar un model de biaixos i mitigar-ne els efectes.</p>
	Protocols interns per a l'ús de la IA	<p>Creació de la governança i model de cicle de vida: especifica les tasques i les responsabilitats, així com les mesures que cal prendre per evitar riscos a l'hora d'aplicar algoritmes, la informació que cal documentar i la persona que és responsable en cas que un algoritme no compleixi l'objectiu previst. L'enfocament del cicle de vida dels algoritmes descriu el procés d'un algoritme des del principi fins al final (inclòs en el Manual).</p>
[3] Mecanismes de transparència i explicabilitat	Registre públic d'algoritmes	Registre d'algoritmes: resum dels algoritmes i sistemes d'IA que utilitza l'ajuntament.
	Estàndard de transparència algorítmica	Desenvolupament d'un model algorítmic comú que garanteixi l'ús adequat de les dades. Projecte desenvolupat juntament amb 8 ciutats europees més en el marc d'Eurocities.
	Web municipal amb tota la informació pertinent sobre la IA	Portal municipal accessible amb tota la informació i els recursos rellevants sobre IA, així com els principals projectes i documents estratègics.

	Avaluació d'impacte en matèria de drets humans	Model d'avaluació d'impacte en matèria de drets humans: basat en les lliçons pràctiques apreses i en les eines ja existents (inclòs en el Manual).
[4] Avaluació de l'impacte algorítmic	Anàlisi de biaixos	Pla detallat pas a pas per avaluar un model d'anàlisi de biaixos, que inclou els components següents, entre altres: (a) definir els grups («sensibles») que cal analitzar, (b) esbossar hipòtesis sobre els trets que podrien generar biaixos indirectes, (c) seleccionar indicadors que escaven amb el projecte, (d) analitzar biaixos directes i indirectes, (e) analitzar biaixos sobre variables no mesurables, (f) sospesar i revisar els biaixos amb els responsables de la gestió, (g) reduir els biaixos, si escau, i (h) redactar les conclusions. El Manual per a l'equitat esmentat aporta orientacions per dur a terme l'anàlisi de biaixos.
[5] Auditories i inspecció reglamentària	Auditories	Auditories anuals: encarregades pel cap de gestió informàtica i portades a terme pel servei d'auditoria d'ACAM. S'ha creat un marc normatiu per a aquestes auditories (inclòs en el Manual).
[6] Supervisió humana, rendició de comptes, procediments d'audiència i apel·lació	Procediments de resposta i objeccions per a la ciutadania	Orientacions i procediments per a les objeccions per als encarregats de gestionar-les (inclòs en el Manual).
[7] Condicions d'adquisició pública	Clàusules d'adquisició pública	Clàusules estàndard per a l'adquisició de sistemes algorítmics fiables: el primer estàndard que estipula les condicions contractuals i els requisits d'informació que han de complir els subministradors de sistemes d'IA que es comprin.
[8] Òrgans assessors i supervisors externs/independents	Òrgan assessor i supervisor extern	Comitè de Dades Personals d'Amsterdam: assessora l'administració municipal sobre algorismes, ètica de les dades, drets humans digitals i exposició de dades personals. Inclou l'avaluació ètica de l'ús dels algorismes. El comitè defensa la transparència organitzant reunions públiques i emetent dictàmens.
[9] Aliances, comunitats de pràctiques i grups d'aprenentatge	Membre d'una comunitat de pràctiques	Coalició de Ciutats per als Drets Digitals (CC4DR) (fundador)
	Col·laboracions multiactor	Amsterdam AI: col·laboració entre el municipi, centres mèdics, centres de recerca, centres de coneixement d'Amsterdam i la Junta Econòmica d'Amsterdam. Aquesta col·laboració es basa en la IA responsable des d'un enfocament centrat en les persones. Coalició NL IA: la ciutat forma part del grup de treball nacional sobre IA en el sector públic, juntament amb la coalició Amsterdam AI. Xarxa Smart Health Amsterdam: la xarxa regional per a les dades i la innovació impulsada per la IA en el sector de la salut i les ciències de la vida a Amsterdam (també forma part de la coalició Amsterdam AI).
[10] Programes de capacitació	Capacitació municipal	Formació del personal municipal: tots els funcionaris han de fer el curs d'IA nacional. El personal dels serveis jurídics, d'adquisició i d'auditoria ha de dur a terme periòdicament cursos d'actualització. Creació d'un equip municipal d'IA.
[11] Promoció de la innovació, el coneixement i l'experimentació local	Laboratoris, hubs i centres d'innovació d'IA	Civic IA Lab: establert pel municipi, la universitat pública i un ministeri del país, té l'objectiu d'investigar com la IA pot contrarestar la desigualtat social o evitar que la reforci. DataLab: laboratori de dades que crea solucions tecnològiques responsables i en obert.

[13] Governança de dades	Transparència de les dades	Data Amsterdam (versió beta): el servei de dades obertes de la ciutat. Té l'objectiu de fer públic el màxim nombre de recursos, fent que les dades generades o custodiades per organismes públics siguin accessibles, gratuïts i fàcils d'usar per a tothom.
	Drets sobre les dades	Projecte Decode
	Mecanismes d'intercanvi de dades	Intercanvi de Dades Amsterdam (AMdEX, per la sigla en anglès) : per intercanviar dades entre organitzacions amb la creació d'un notari digital.

Estudi de cas 3: Nova York

Població: 8.258.000 habitants (2023)	Renda per capita: 44.537 € (2022)	Regió: Amèrica del Nord
---	--	--------------------------------

MARC DE GOVERNANÇA DE LA IA		
[1] Principis, estratègies i directrius	Estratègia d'IA	Estratègia d'IA (2021-2023): la iniciativa pionera per fomentar un ecosistema local d'IA intersectorial sòlida. El document va establir la informació de referència sobre la IA per contribuir a garantir que els responsables polítics treballen partint d'una concepció comuna de la tecnologia i dels problemes que presenta. Inclou un conjunt de principis ètics, governança i marc polític.
	Pla d'acció per a la IA	Pla d'acció per a la IA (2023-2025): inclou 37 accions pertanyents a 7 iniciatives per crear la governança en l'ús de la IA per part de la ciutat. Encarrega un informe anual sobre el progrés del pla.
	Principis locals per a la IA	Principis d'IA (inclosos en el Pla d'acció per a la IA): (1) validesa i fiabilitat, (2) responsabilitat social, (3) privacitat de la informació, (4) ciberseguretat i (5) confiança i transparència. Principis inclosos en l'antiga Estratègia d'IA: (1) rendició de comptes, (2) equitat, (3) privacitat i seguretat, (4) implicació i participació de la comunitat.
	Directrius, guies i manuals	Directrius sobre principis i definicions de la IA: especifica els principis de la IA de la ciutat, aporta definicions concretes de termes relacionats amb la IA i detalla les lleis, les polítiques, els requisits i els processos que s'hi apliquen. Guia d'ús preliminar: intel·ligència artificial generativa: inclou condicions i definicions, rols i responsabilitats, assessorament per a la IA generativa que fa servir la municipalitat i detalla les lleis, les polítiques, els requisits i els processos que s'hi apliquen.
	Protocols interns per a l'ús de la IA	Els protocols interns especificats al Pla d'acció, incloent-hi un informe anual elaborat pels organismes municipals per ampliar les informacions públiques notificades sobre IA.
[2] Normativa i legislació local	Regulació sobre l'aplicació d'IA controvertida	Llei de protecció de dades biomètriques per a empreses Auditoria obligatòria per detectar biaixos en tecnologia usada per a la contractació de treballadors.
[3] Mecanismes de transparència i explicabilitat	Web municipal amb tota la informació pertinent sobre la IA	Portal municipal que conté tota la informació pertinent i recursos sobre IA, així com documents estratègics i projectes importants.
	Directorio municipal d'eines d'IA adquirides (per a ús intern)	Creació d'un directorio intern d'eines d'IA adquirides i indicacions per a usar-les de manera adequada, compartit per tots els organismes per facilitar-ne la visibilitat i l'accés. Inclòs en el pla d'acció.
[4] Avaluació de l'impacte algorítmic	Avaluació i gestió del risc (*)	Procés d'avaluació del risc en matèria d'IA i revisió de projectes (inclòs en el pla d'acció).
[7] Condicions d'adquisició pública	Clàusules d'adquisició pública (*)	Estàndards d'adquisició pública específiques per a la IA.
[8] Òrgans assessors i supervisors externs/independents	Òrgan assessor i supervisor extern	Xarxa assessora sobre IA: aplega experts independents procedents del sector privat, l'acadèmia, les organitzacions sindicals i cíviques per donar suport a les iniciatives municipals en matèria d'IA amb caràcter consultiu.

[9] Aliances, comunitats de pràctiques i grups d'aprenentatge	Membre d'una comunitat de pràctiques	Coalició de Ciutats per als Drets Digitals (CC4DR) (fundador).
[10] Programes de capacitació	Òrgan municipal	NYC Automated Decision Systems Task Force : creat per emetre recomanacions sobre com la ciutat ha de gestionar l'ús dels algorismes. Va ser el primer d'aquest tipus al país i va culminar amb la publicació d'un informe accessible. Comitè Directiu sobre la IA a la Ciutat : aplega les parts interessades de tot el govern municipal perquè facin aportacions i supervisin les activitats d'IA.
[12] Participació de la comunitat	Participació popular	Audiències públiques .
[13] Governança de dades	Transparència de les dades	NYC Open Data : el servei de dades obertes de la ciutat. Té l'objectiu de fer públic el màxim nombre de recursos, fent que les dades generades o custodiades per organismes públics siguin accessibles, gratuïts i fàcils d'usar per tothom.

* Mecanisme polític planificat, però encara no implementat completament el desembre de 2024.

Estudi de cas 4: San José

Població: 969.655 habitants (2023)	Renda per capita: 139.761 € (2023)	Regió: Amèrica del Nord
---	---	--------------------------------

MARC DE GOVERNANÇA DE LA IA		
[1] Principis, estratègies i directrius	Manual d'IA	Manual d'IA: proporciona assessorament integral sobre com complir la política d'IA de la ciutat. Inclou: (1) política d'IA, (2) revisió de la IA (obligatòria per a totes les adquisicions públiques i iniciatives de dades), (3) governança de la IA (el marc per gestionar i supervisar el cicle de vida de la IA) i (4) orientacions per a la IA generativa.
	Directrius, guies i manuals	Directrius per a la IA generativa: va ser el primer pas d'un procés col·laboratiu per desenvolupar la política general de la IA de la ciutat. Els usuaris registrats podien formar part d'un grup de treball del Departament de TIC per compartir experiències i cocrear les polítiques d'IA de la ciutat.
	Principis locals per a la IA	Principis d'IA: (1) efectivitat (fiabilitat), (2) transparència, (3) equitat, (4) rendició de comptes, (5) disseny centrat en les persones, (6) privacitat (7), seguretat i protecció (8) apoderament dels treballadors.
[3] Mecanismes de transparència i explicabilitat	Registre públic d'algoritmes	Inventari d'IA (pas 5 del procés de revisió): resum dels sistemes i algoritmes d'IA que utilitza la ciutat.
	Web municipal amb tota la informació pertinent sobre la IA	Portal municipal que conté tota la informació pertinent i recursos sobre IA, així com documents estratègics i projectes importants.
[4] Avaluació de l'impacte algorítmic	Avaluació i gestió del risc	Model d'anàlisi del llinar de risc de la IA: elaborat pel Departament de Privacitat Digital (pas 2 del procés de revisió).
	Avaluació d'impacte en matèria de drets humans	Avaluació d'impacte algorítmic: portada a terme pel consistori municipal quan va adquirir sistemes d'IA considerats de risc mitjà-alt segons l'anàlisi del risc (pas 3 del procés de revisió).
[6] Supervisió humana, rendició de comptes, procediments d'audiència i apel·lació	Procediments de resposta i objeccions per a la ciutadania	Formulari de comentaris públics: a través del formulari els ciutadans poden fer arribar comentaris sobre els projectes que impliquin un nou ús d'informació pública. Podeu trobar la informació sobre els nous projectes en línia.
	Elaboració d'informes i supervisió interna	Informe anual d'ús: el departament propietari del sistema d'IA ha de lliurar un informe anual d'ús que detalli la informació següent: 1. Resum del projecte 2. Indicadors per avaluar el resultat requerit 3. Plans de futur per a la iniciativa tecnològica Els ciutadans poden fer comentaris en línia sobre l'ús de les dades i les actualitzacions anuals (pas 6 del procés de revisió).
[7] Condicions d'adquisició pública	Protocols interns per a l'adquisició de serveis d'IA	Marc de revisió de la IA: per avaluar els avantatges i els riscos de l'adquisició municipal. Procés de revisió: 1. Sol·licitud d'adquisició. 2. Anàlisi del risc. 3. Avaluació de l'impacte algorítmic (per a sistemes de risc mitjà-alt): inclou l'avaluació de l'impacte dels algoritmes municipals i la fulla informativa sobre IA per a proveïdors. 4. Revisió final. 5. Preparació prèvia al llançament: protocol d'ús de les dades, formació dels usuaris i comptabilització de l'inventari d'IA. 6. Supervisió contínua.
	Clàusules d'adquisició pública	Fulla informativa per a proveïdors d'IA: inclou una fulla informativa i un qüestionari d'avaluació de l'impacte algorítmic per al proveïdor (pas 3 del procés de revisió).

[8] Òrgans assessors i supervisors externs/ independents	Òrgan assessor i supervisor multiactor	Grup Assessor en IA : dirigit pel consistori municipal, les parts interessades externes assessoren els departaments municipals i el director d'informació sobre les polítiques i les activitats relacionades amb la governança de la IA. Està constituït per experts en IA procedents del sector industrial, l'acadèmia, els drets civils i membres de la societat. El Grup Assessor es reuneix trimestralment i el poder de decisió segueix estant en el municipi.
[9] Aliances, comunitats de pràctiques i grups d'aprenentatge	Aliances multiactor	Coalició GovAI : la coalició liderada per San José aplega entitats públiques, la societat civil, institucions acadèmiques i empreses per fomentar la IA responsable en el sector públic. Està constituïda per més de 1.500 membres i més de 500 entitats federals, estatals i locals.
	Membre d'una comunitat de pràctiques	Coalició de Ciutats per als Drets Digitals (CC4DR) .
[10] Programes de capacició	Grup de treball municipal	Grup de Treball en IA (AIWG, per la sigla en anglès) : empleats de diversos departaments municipals aborden projectes i problemes relacionats amb la IA de la ciutat. Constituït per responsables d'IA dels departaments i potencialment per representants d'altres departaments.
[12] Participació de la comunitat	Participació popular	Si es considera que un sistema d'IA adquirit és d'interès públic, l'ajuntament porta a terme activitats de divulgació en línia i presencials (adreçades a comunitats amb un accés a internet limitat). Els comentaris dels ciutadans posteriorment s'incorporen al Protocol d'ús de les dades.
[13] Governança de dades	Transparència de les dades	Portal de dades obertes : el portal de dades obertes de la ciutat. Té l'objectiu de fer públic el màxim nombre de recursos, fent que les dades generades o custodiades per organismes públics siguin accessibles, gratuïts i fàcils d'usar per tothom.
	Protocol d'ús de les dades	Protocol d'ús de les dades : protocol per a sistemes amb risc mitjà-alt que ha de regir la recollida, l'accés, el tractament i l'intercanvi de dades i garantir el compliment de la Política de privacitat digital de la ciutat (pas 5 del procés de revisió).

Estudi de cas 5: Dubai

Població: 3.051.000 habitants (2023)	Renda per capita: 47.995 € (2023)	Regió: Orient Mitjà i Nord d'Àfrica
---	--	--

MARC DE GOVERNANÇA DE LA IA		
[1] Principis, estratègies i directrius	Plan i full de ruta d'IA	Full de ruta d'IA (2024) forma part del <i>Dubai Universal Blueprint for Artificial Intelligence</i> dels Emirats Àrabs i fomenta els objectius de l'Agenda econòmica de Dubai D33.
	Principis locals per a la IA	Principis i directrius sobre l'ètica de la IA s'expliquen a continuació: <ul style="list-style-type: none"> • Principis d' IA ètica: (1) ètica (equitativa, responsable, transparent i explicable); (2) seguretat (segura i protegida), (3) humanitat i (4) inclusivitat. • Directrius d'IA ètica, per fer que els sistemes d'IA siguin: (1) equitatius, (2) responsables, (3) transparents i (4) explicables.
[3] Mecanismes de transparència i explicabilitat	Web municipal amb tota la informació pertinent sobre la IA	Portal municipal que conté tota la informació pertinent i recursos sobre IA, així com documents estratègics i projectes importants.
[5] Auditories i inspecció reglamentària	Eina per a l'autoavaluació	Eina d'autoavaluació ètica d'IA: creada perquè les organitzacions que desenvolupen o que operen els sistemes d'IA puguin avaluar l'ètica, fent servir les Directrius sobre l'ètica de la IA de Dubai.
[11] Promoció de la innovació, el coneixement i l'experimentació local	Laboratoris, hubs i centres d'innovació	Laboratori d'IA: creat en associació amb IBM, funciona amb una xarxa cada vegada més gran de socis del govern i del sector privat. Full de ruta per als responsables d'IA de Dubai.
[13] Governança de dades	Transparència de les dades	Dubai Pulse: el portal de dades obertes de la ciutat. Té l'objectiu de fer públic el màxim nombre de recursos, fent que les dades generades o custodiades per organismes públics siguin accessibles, gratuïts i fàcils d'usar per tothom.
	Mecanismes d'intercanvi de dades	Eines per a l'intercanvi de dades: proporciona assessorament i recursos a particulars i organitzacions públiques i privades per preparar i dissenyar una iniciativa d'intercanvi de dades.
	Privacitat de les dades	Marc de dades sintètiques: dissenyat per ajudar les organitzacions a adoptar la tecnologia de la IA evitant qualsevol tipus de violació de la privacitat.

Estudi de cas 6: Singapur

Població: 5.918.000 habitants (2023)	Renda per capita: 79.996 € (2023)	Regió: Àsia Oriental i Pacífic
---	--	---------------------------------------

MARC DE GOVERNANÇA DE LA IA		
[1] Principis, estratègies i directrius	Estratègia d'IA	Estratègia nacional d'IA 2.0 (inclou un compendi sobre la IA).
	Principis locals per a la IA	Principis guia del marc de governança de la IA: (1) explicabilitat, transparència i equitat; (2) solucions centrades en les persones. Principis de governança de AI Verify: (1) transparència, (2) explicabilitat, (3) repetibilitat/reproductibilitat, (4) seguretat, (5) protecció, (6) robustesa, (7) equitat, (8) governança de les dades, (9) rendició de comptes, (10) agència i supervisió humana i (11) creixement inclusiu i benestar social i mediambiental.
	Directrius, guies i manuals	Manual d' IA del sector públic: recurs de l'estratègia d'IA per al govern. El manual explica la IA, presenta les aplicacions més comunes al sector públic, indica els passos per iniciar un projecte d'IA i desenvolupar les capacitats municipals en aquest àmbit. Model del marc de governança de la IA (2a edició): proporciona assessorament detallat i fàcil d'implementar sobre com traduir els principis ètics en recomanacions pràctiques que les organitzacions poden adoptar per desplegar la IA de manera responsable. Model de marc de governança de la IA per a la IA generativa Directrius relatives a l'ús de dades personals en sistemes de recomanació i de presa de decisions basats en IA: a l'hora d'elaborar-les s'han tingut en compte les consultes públiques. Altres directrius sectorials: Directrius per a la IA en l'àmbit de la salut: aporten recomanacions per fomentar el desenvolupament i la implementació de manera segura de dispositius mèdics basats en IA i altres usos de la IA en el sector sanitari. Guia per redissenyar l'ocupació en l'era de la IA
[3] Mecanismes de transparència i explicabilitat	Web municipal amb tota la informació pertinent sobre la IA	Portal municipal que conté tota la informació pertinent i recursos sobre IA, així com documents estratègics i projectes importants.
[8] Òrgans assessors i supervisors externs/independents	Òrgan assessor i supervisor extern	Consell assessor per a l'ús ètic de la IA i les dades: format per 11 membres multidisciplinaris, ajuda el govern a l'hora de desenvolupar les normes ètiques i els marcs de governança, i publica directrius, guies pràctiques i codis de pràctica perquè siguin adoptats de manera voluntària per part de la indústria.
[9] Aliances, comunitats de pràctiques i grups d'aprenentatge	Coalició multiactor	AI Singapore: aplega institucions de recerca i el món empresarial per portar a terme recerca sobre IA fiable i governança ètica, crear eines de codi obert i fomentar el talent per als esforços nacionals en matèria d'IA. Consorci Veritas: format per socis industrials i l'Autoritat Monetària del Govern de Singapur, té l'objectiu de possibilitar que les institucions financeres avaluin les seves solucions basades en IA i dades pel que fa als principis d'equitat, ètica, rendició de comptes i transparència.
	Membre d'una comunitat de pràctiques	Fundació Verify AI: comunitat global de codi obert que aplega propietaris d'IA, proveïdors de solucions, usuaris i responsables polítics per crear una IA fiable.
[10] Programes de capacitació	Capacitació municipal	Formació del personal municipal: personalitzada per als diferents tipus d'usuaris municipals, Singapur compta amb un directori de cursos per adquirir diverses competències (inclosa en el Manual d'IA del sector públic).

[11] Promoció de la innovació, el coneixement i l'experimenta local	Laboratoris, <i>hubs</i> i centres d'innovació	Centre per a la IA i la Governança de Dades (CAIDG, per la sigla en anglès) : centre de recerca interdisciplinària amb associacions de múltiples parts interessades, des d'organismes governamentals fins a organitzacions intergovernamentals, empreses, universitats, grups de reflexió, ONG i responsables de seguretat.
	Entorns aïllats de proves reguladors (<i>regulatory sandboxes</i>)	Entorns aïllats de proves (<i>sandboxes</i>): Sanbox per a la IA generativa ; Sandbox tecnològics de protecció de la privacitat .
[13] Governança de dades	Transparència de les dades	Portal de dades obertes : el portal de dades obertes de la ciutat. Té l'objectiu de fer públic el màxim nombre de recursos, fent que les dades generades o custodiades per organismes públics siguin accessibles, gratuïts i fàcils d'usar per tothom.
[14] Altres mesures i polítiques	Marc i eines de prova	Verify AI : marc de proves de la governança de la IA i conjunt d'eines de <i>software</i> per a empreses que valida el rendiment dels sistemes d'IA tenint en compte un conjunt de principis ètics reconeguts internacionalment mitjançant proves estàndards. Guia d'implementació i autoavaluació per a organitzacions (ISAGO, per la sigla en anglès) : ajuda les organitzacions a avaluar si les seves pràctiques de governança de la IA concorden amb el model marc. Kit d'eines de codi obert Veritas : possibilita l'ús responsable de la IA en el sector financer.
	Certificacions verdes	Full de ruta per la certificació verda de centres de dades : presenta una via sostenible per al creixement continuat de centres de dades a Singapur. Certificat verd de centres de dades : per als operadors que han desplegat de manera reeixida bones pràctiques ecològiques a centres de dades, que demostren una sostenibilitat i uns resultats mediambientals superiors.

CONCLUSIONS. MECANISMES POLÍTICS, REPTES I RECOMANACIONS EN LA IA URBANA

Marta Galceran-Vercher

Investigadora sènior, Programa de Ciutats Globals, CIDOB

Alexandra Vidal D'oleo

Investigadora i gestora de projectes, Programa de Ciutats Globals, CIDOB

Els governs locals d'arreu del món van adoptant cada cop més els sistemes algorítmics per millorar la prestació dels serveis públics. No obstant això, l'evidència creixent assenyalava que aquests sistemes poden provocar danys no intencionats i ser implementats de manera poc transparent. En conseqüència, l'adopció de sistemes algorítmics sovint ha anat acompanyada del desenvolupament de principis rectors per a l'ús responsable de les tecnologies d'intel·ligència artificial (IA), principalment a escala nacional, supranacional o global. En són exemples notables els Principis d'IA de l'OCDE (2019), els Principis d'IA del G20 (2019), el grup de redacció del Conveni d'IA del Consell d'Europa (2022-2024), la declaració ministerial de l'Associació Global sobre Intel·ligència Artificial (2022), la declaració dels ministres del G7 (2023), la Declaració de Bletchley (2023), la Declaració Ministerial de Seül (2024), la Llei d'IA de la UE (2024) o l'Informe de l'ONU «Governing AI for Humanity» (2024). Tanmateix, aquests marcs normalment només proporcionen orientacions generals sobre el que constitueix un ús responsable de la IA i ofereixen poca orientació pràctica sobre com s'han d'aplicar aquests principis en contextos reals.

En resposta a aquests reptes, molts governs recorren a marcs normatius i eines polítiques per posar en pràctica aquests principis. Iniciatives com aquestes van sorgint amb molta rapidesa, tot i que varien considerablement quant a abast i enfocament. A més, gran part de l'anàlisi existent de les eines de polítiques del sector públic se sol centrar en perspectives d'abast nacional (per exemple, OECD, 2024), que sovint passen per alt el context i els reptes singulars als quals s'enfronten els governs locals.

Aquesta monografia CIDOB ha intentat superar aquesta limitació identificant els principals mecanismes polítics i marcs de governança que els governs locals aprofiten per garantir que adopten els sistemes algorítmics en consonància amb principis ètics bàsics, com ara la transparència i la rendició de comptes, l'equitat i la no-discriminació, la privacitat i la governança de dades, i la sostenibilitat. Aquesta anàlisi es complementa amb una sèrie d'estudis de casos que il·lustren com les principals ciutats implementen aquests mecanismes polítics a la pràctica, cosa que dona lloc a marcs locals integrals d'IA.

Aquesta monografia CIDOB identifica els principals mecanismes polítics i marcs de governança que els governs locals aprofiten per garantir que adopten sistemes algorítmics en consonància amb principis ètics fundamental.

Els criteris utilitzats per establir aquesta categorització es van centrar en la funció i els objectius principals dels mecanismes polítics.

En aquest capítol final, presentem una categorització dels mecanismes polítics identificats en aquesta publicació, juntament amb informació sobre quins mecanismes utilitzen més els governs locals i com concorden amb els principis ètics esmentats. També analitzem els reptes als quals s'enfronten els governs locals i oferim recomanacions per avançar cap a un ús més responsable dels sistemes d'IA en entorns urbans.

1. Categorització dels mecanismes polítics

Gràcies a una revisió exhaustiva dels mecanismes polítics presentats en els capítols anteriors d'aquesta monografia CIDOB, complementada amb una revisió bibliogràfica de les publicacions rellevants sobre mecanismes polítics per a governs locals i/o administracions públiques (com ara els informes de l'Ada Lovelace Institute, l'AI Now Institute i l'Open Government Partnership, 2021; Ben Dhaou *et al.*, 2024; Jordan *et al.*, 2024; United 4 Smart Sustainable Cities, 2024), hem identificat catorze categories diferents de polítiques que actualment implementen els governs locals arreu del món (vegeu la taula 1).¹

Els criteris utilitzats per establir aquesta categorització es van centrar en la funció i els objectius principals dels mecanismes polítics. Això inclou proporcionar orientació normativa per al desenvolupament i l'ús de sistemes d'IA, avaluar els riscos potencials dels algoritmes, garantir l'accés públic a la informació sobre els sistemes algorítmics i exigir rendició de comptes.

Altres possibles criteris s'haurien pogut centrar en si els mecanismes s'orienten cap a processos administratius interns (per exemple, directrius per al personal municipal o creació de comissions municipals d'IA) o accions externes, com la publicació de registres algorítmics públics o la imposició de prohibicions sobre aplicacions d'IA controvertides. A més, la nostra categorització no distingeix entre els mecanismes específics d'IA (per exemple, una estratègia d'IA) i els mecanismes indirectes que contribueixen a la governança ètica de la IA (per exemple, la importància de les dades en els sistemes d'IA).

1. Cal tenir en compte que la literatura sobre aquest tema és limitada i utilitza una sèrie de noms i termes per als diferents mecanismes polítics, però no hi ha un vocabulari comú per als seus components bàsics.

Taula 1. Categories de mecanismes polítics per a una governança responsable de la IA a les ciutats

CATEGORIA	DESCRIPCIÓ I FINALITAT	MECANISMES POLÍTICS
[1] Principis, estratègies i directrius	Documents polítics que ofereixen orientacions normatives no vinculants sobre els valors i els principis ètics a les administracions locals, i que presenten indicacions generals per desenvolupar i utilitzar la IA i alhora gestionar-ne els riscos associats. Tot i que el format varia, aquests documents solen identificar objectius polítics d'alt nivell i la seva rellevància per a l'ús dels sistemes algorítmics per part dels organismes públics. En alguns casos, també proporcionen orientacions pràctiques per implementar aquests principis a l'hora de dissenyar i desplegar aquests sistemes. En darrera instància, aquestes directrius estableixen estàndards normatius que permeten als organismes i a la ciutadania avaluar l'ús ètic dels sistemes algorítmics.	<ul style="list-style-type: none"> Plans d'acció, agendes, fulls de ruta, estatuts, manuals, estratègies d'IA ètica, etc. Definició de principis locals ètics per a la IA: declaració i/o aprovació. Directrius, guies i manuals sobre el desplegament de la IA ètica. Protocols interns per a l'ús de la IA. Marcos estratègics que no són específicament sobre IA que tenen un impacte en la governança de la IA (per exemple, marcs de governança de les dades o digitalització)
[2] Normativa i legislació local	Eines destinades a establir estàndards, lleis i regulacions que garanteixin el compliment i abordin els impactes socials.	<ul style="list-style-type: none"> Regulació i lleis locals (per exemple, normatives per garantir el dret a la justificació) Mecanismes de compliment legal: per garantir el compliment de marcs normatius regionals, nacionals o supranacionals Estàndards reguladors (per exemple, estàndards d'IA verda, estàndards de transparència) Adhesió als estàndards normatius internacionals Regulació de l'aplicació d'IA controvertida: prohibicions, moratòries, etc.
[3] Mecanismes de transparència i explicabilitat	Mecanismes per establir l'accés públic a la informació sobre sistemes i processos algorítmics. Tenen l'objectiu de proporcionar informació sobre sistemes algorítmics al públic en general (per exemple, les persones afectades, els mitjans de comunicació o la societat civil) perquè pugui conèixer l'ús d'aquests sistemes i exigir explicacions i justificacions relacionades amb aquest ús. Aquests mecanismes poden funcionar de manera independent o formar part de marcs més amplis per a la rendició de comptes algorítmica. Cal distingir els mecanismes de transparència pública dels drets d'audiència i d'explicació, que atorguen a les persones el dret a una explicació sobre les decisions algorítmiques concretes que s'han pres sobre elles.	<ul style="list-style-type: none"> Registres públics d'algorítmics Web municipal amb tota la informació pertinent sobre la IA Estàndards de transparència algorítmica Directori municipal d'eines d'IA adquirides per a ús intern Requisits per a la transparència del codi font Explicacions de lògiques algorítmiques
[4] Avaluació de l'impacte algorítmic	Instrument polític utilitzat pels organismes públics per avaluar els possibles riscos i danys dels sistemes algorítmics. Aquestes avaluacions tenen com a objectiu entendre, categoritzar i abordar els possibles efectes negatius dels algorítmics abans de ser desplegats o durant el procés de desplegament. Les avaluacions d'impacte algorítmic (AIA) es basen en marcs establerts d'altres camps, com ara les avaluacions d'impacte mediambiental i en matèria de drets humans i les avaluacions d'impacte de protecció de dades (AIPD).	<ul style="list-style-type: none"> Procediments d'avaluació i gestió del risc (inclouent-hi l'anàlisi de biaixos) Avaluació d'impacte en matèria de drets humans Mecanismes d'avaluació d'impacte mediambiental
[5] Auditories i inspecció reglamentària	Les auditories inclouen una sèrie de pràctiques adreçades a examinar com funciona un sistema algorítmic concret. Tenen l'objectiu d'entendre el funcionament del sistema i avaluar-ne el rendiment en funció d'uns estàndards normatius predefinits. Tot i que les auditories presenten algunes semblances amb les AIA, el context històric és diferent i normalment es duen a terme durant la implementació del sistema o després de la implementació. En canvi, les AIA normalment es duen a terme abans del desplegament o durant el desplegament. Les auditories les poden dur a terme entitats internes, externes o de tercers, en funció de l'abast i la naturalesa de l'avaluació. En una auditoria de tercers, una organització externa avalua el sistema basant-se únicament en els resultats del sistema. Una auditoria de segona part la du a terme un avaluador extern que té accés tant al fons (<i>backend</i>) del sistema com als seus resultats. Les auditories internes les duen a terme membres interns de l'organització.	<ul style="list-style-type: none"> Auditories de sistemes algorítmics Avaluació del procés Eines per a l'autoavaluació

[6] Supervisió humana, rendició de comptes, procediments d'audiència i apel·lació	Mecanismes per supervisar els sistemes d'IA i fer que siguin responsables. Concretament, es tracta de mecanismes que requereixen que les decisions preses amb l'ajuda de sistemes algorítmics segueixin procediments específics dissenyats per salvaguardar l'equitat i proporcionar vies perquè les persones cerquin reparacions en casos de resultats esbiaixats o erronis. Aquestes garanties procedimentals ofereixen a les persones o grups afectats la possibilitat de qüestionar o impugnar les decisions que els afecten.	<ul style="list-style-type: none"> • Elaboració d'informes i supervisió interna • Requisits de la intervenció humana (<i>human-in-the-loop</i>) • Procediments de resposta i objeccions per a la ciutadania • Deures de notificació de la resolució i audiència dels afectats • Obligacions d'aportar decisions motivades i explicacions d'una decisió • Mecanismes per garantir el dret de les parts afectades a presentar proves, apel·lar i impugnar decisions automatitzades
[7] Condicions d'adquisició pública	Les normes que regeixen l'adquisició de sistemes algorítmics per part dels governs i els organismes públics són fonamentals per garantir-ne l'ús responsable. Molts sistemes algorítmics utilitzats pels governs són subcontractats a proveïdors privats, ja sigui mitjançant la compra de productes o els contractes de servei. En conseqüència, els proveïdors tenen un paper important a l'hora de dissenyar i desplegar aquests sistemes. Les condicions dels contractes d'adquisició són fonamentals per configurar el desenvolupament i la implementació d'aquests sistemes. Es poden aplicar condicions específiques d'adquisició, com ara requisits de transparència i no-discriminació, per garantir que els sistemes adquirits compleixen els estàndards ètics i s'utilitzen de manera responsable.	<ul style="list-style-type: none"> • Clàusules d'adquisició pública • Directrius, marcs i protocols interns per a l'adquisició d'IA
[8] Òrgans assessors i supervisors externs/independents	Òrgans supervisors independents, que tenen la finalitat de supervisar i guiar l'ús de sistemes algorítmics per part dels organismes públics. Aquests mecanismes de supervisió independent tenen com a objectiu garantir la rendició de comptes mitjançant el seguiment de les actuacions dels organismes públics i la formulació de recomanacions, sancions o decisions sobre l'ús que fan dels sistemes algorítmics.	<ul style="list-style-type: none"> • Òrgans assessors: consells, comitès, juntes, xarxes, grups, etc.
[9] Aliances, comunitats de pràctiques i grups d'aprenentatge	Mecanismes adreçats a fomentar la cooperació i les associacions a escala local, nacional i internacional.	<ul style="list-style-type: none"> • Comunitats de pràctiques d'aprenentatge locals/nacionals/internacionals: xarxes de ciutats, grups de treball, etc. • Col·laboracions locals/nacionals/internacionals multiactors en l'àmbit de la IA: xarxes, plataformes, coalicions, etc. • Associació i col·laboracions publicoprivades
[10] Programes de capacitat	Mecanismes per millorar el coneixement i adquirir habilitats sobre la IA ètica. Aquestes iniciatives poden adreçar-se al personal municipal implicat, directament o indirectament, en el disseny, el desplegament o l'ús de sistemes algorítmics, així com al públic en general, per promoure una ciutadania informada i fomentar una comprensió més àmplia de l'ètica de la IA.	<ul style="list-style-type: none"> • Formació del personal municipal (enfocament sociotecnològic) • Creació d'un equip municipal d'IA • Òrgan municipal d'IA: òrgan o comissió transversal que coordina/supervisa l'ús municipal de la IA • Enfocament multidisciplinari: creació d'equips diversos
[11] Foment de la innovació, el coneixement i l'experimentació local	Mecanismes que ofereixen un espai per a l'experimentació, la innovació i les proves en entorns del món real.	<ul style="list-style-type: none"> • Promoció i col·laboració amb laboratoris, <i>hubs</i> i centres locals d'innovació d'IA • Observatori d'IA local • Entorns aïllats de proves reguladors locals d'IA (<i>regulatory sandboxes</i>) • Iniciatives per promoure els ecosistemes locals d'IA i donar-hi suport
[12] Participació de la comunitat	Eines per implicar els ciutadans, les comunitats i les parts interessades en els processos de presa de decisions en IA, fomentar converses i debats i garantir que les polítiques d'IA reflecteixin les preocupacions i les aportacions de la ciutadania.	<ul style="list-style-type: none"> • Participació popular: processos participatius, models de govern participatiu, sessions d'audiència pública, promoció d'espais de reflexió i debat, canals de comunicació amb la ciutadania, etc. • Formació dels ciutadans (alfabetització digital) • Comitès locals d'ètica d'IA
[13] Governança de dades		<ul style="list-style-type: none"> • Mesures de transparència de les dades com ara portals de dades obertes • Mecanismes d'intercanvi de dades • Drets sobre les dades • Protocols d'ús de les dades • Privacitat de les dades: lleis de privacitats de dades, marcs de dades sintètiques, etc. • Sistemes de governança de dades
[14] Altres mesures i polítiques		<ul style="list-style-type: none"> • Marcs i eines de prova • Incentius fiscals com ara desgravacions fiscals, subvencions, etc. • Programes de requalificació de treballadors • Marcs d'avaluació (per exemple, marcs d'avaluació de la IA mitjançant estrelles, certificacions verdes, etc.)

Font: Autores

2. Concordança dels mecanismes polítics amb principis ètics

L'anàlisi dels mecanismes polítics identificats al llarg d'aquesta publicació² palesa que es poden establir diverses tipologies tenint en compte el grau de concordança amb principis ètics concrets (vegeu la Taula 2), que s'examinen a continuació.

- a) **Mecanismes polítics que serveixen per defensar determinats principis ètics.** Per exemple, una avaluació d'impacte mediambiental compleix principalment propòsits de sostenibilitat, mentre que una mesura d'intervenció humana (*human-in-the-loop*) pot fomentar tant el principi de rendició de comptes com el d'equitat en garantir que algú supervisa el funcionament correcte d'un sistema d'IA i, en darre-ra instància, deriva la decisió final a un humà.
- b) **Mecanismes polítics que defensen tots els principis ètics:** degut al caràcter transversal que presenten o gràcies a la personalització, alguns mecanismes polítics poden fomentar tots els principis ètics en conjunt. Per exemple, mecanismes polítics com les estratègies d'IA es poden personalitzar per incloure tots els principis ètics. De la mateixa manera, un comitè de supervisió es pot encarregar de supervisar la protecció de la privacitat, la rendició de comptes o tot el paquet de principis ètics.

Malgrat el seu potencial, no tots són aplicats amb la mateixa freqüència per les ciutats que s'esforcen per establir marcs locals d'IA ètica. Alguns exemples destacats d'aquests mecanismes són:

- **Principis, estratègies i directrius:** un dels mecanismes més utilitzats per les ciutats d'arreu del món. Les ciutats implementen sistemàticament aquests mecanismes polítics des d'un punt de vista ètic per proporcionar-los una base i orientació. Aquests mecanismes són especialment utilitzats pels municipis per demostrar la seva voluntat de comprometre's amb un desplegament responsable de l'IA.
- **Clàusules d'adquisició públiques:** com que els municipis sovint no tenen els recursos necessaris per crear els seus propis sistemes d'IA interns, un altre mecanisme polític que s'aplica habitualment són les clàusules d'adquisició pública, que acaben esdevenint fonamentals i són una solució pràctica i popular. Permeten als municipis aprofitar el seu poder adquisitiu quan adquireixen sistemes d'IA i alhora promouen un desenvolupament ètic de la IA per part dels proveïdors del sector privat.
- **Mecanismes externs:** cada cop hi ha més ciutats que recorren a òrgans assessors i supervisors constituïts per experts externs i independents que assessoren el municipi sobre problemes ètics i supervisen l'ús que fan dels sistemes algorítmics. De la mateixa manera, molts municipis formen part d'**aliances, comunitats de pràctiques i grups d'aprenentatge** per abordar conjuntament els reptes i identificar maneres d'utilitzar la IA de forma segura.

L'anàlisi dels mecanismes polítics identificats al llarg d'aquesta publicació palesa que es poden establir diverses tipologies tenint en compte el grau de concordança amb principis ètics concrets.

2. Extrets únicament dels capítols (Part I) i dels casos pràctics (Part II) d'aquesta monografia.

- **Governança de dades:** tot i que la governança de les dades pot no percebre's d'entrada com un mecanisme polític directe per a la IA, assenta les bases per gestionar de manera correcta les dades dels ciutadans i salvaguardar-les, i és fonamental perquè els sistemes no siguin discriminatoris, la qual cosa la converteix en una peça clau per a un desplegament ètic de la IA. La governança de les dades és un pilar fonamental que permet la transparència de les dades, la protecció dels drets en matèria de dades, la privacitat de les dades i l'ús sostenible de les dades per als sistemes d'IA. En són exemples els protocols per a l'anonimització de dades personals o l'ús de dades sintètiques per entrenar els sistemes d'IA, per tal de resoldre problemes de privacitat i equitat.
- **Auditories:** tot i que els experts i els treballadors públics les consideren un dels mecanismes polítics fonamentals per garantir l'ètica, les auditories segueixen sent poc utilitzades. S'utilitzen poc en part degut a les limitacions imposades pels proveïdors privats d'IA i a la manca de capacitació tècnica interna.

c) Mecanismes polítics no associats a cap principi ètic específic.

Tot i que no estan vinculats a cap principi ètic específic, aquests mecanismes polítics es consideren clau per a un desplegament ètic de la IA. Estableixen els processos estructurats que cal seguir, en coordinen el desplegament o proporcionen l'especialització i els coneixements necessaris per prendre decisions de manera fonamentada. Alguns exemples destacats d'aquests mecanismes són:

- **Protocols interns per a l'ús de la IA:** la majoria de les ciutats desenvolupen protocols interns per guiar el seu ús de la IA, que proporcionen estructures pas a pas per facilitar-ne la implementació per part del municipi. Algunes ciutats, encara que una minoria, els complementen, a més, amb **protocols integrals per a l'adquisició pública d'IA**. Aquests protocols poden incloure avaluacions obligatòries d'impacte, anàlisi de biaixos i altres mecanismes polítics per garantir el respecte de principis ètics concrets.
- **Centres, hubs i laboratoris innovadors d'IA:** un nombre important de ciutats col·laboren amb centres, *hubs* o laboratoris innovadors d'IA, els fomenten o n'han establert per crear coneixement pràctic i desenvolupar solucions d'IA. La recerca que duen a terme aquestes institucions té una orientació ètica.
- **Programes de capacitació:** un dels mecanismes menys implementats és la creació d'**equips municipals destinats a la IA** amb l'especialització necessària per auditar o desenvolupar sistemes d'IA interns. Això es deu principalment a les limitacions tècniques i econòmiques d'una part dels municipis. En canvi, molts han creat òrgans municipals d'IA encarregats de coordinar l'ús de la IA entre els departaments. Aquests òrgans tenen un paper fonamental a l'hora de facilitar el seguiment transversal del desplegament de la IA a la ciutat, garantint un enfocament més organitzat i responsable de la governança de la IA.

Taula 2. Concordança dels mecanismes polítics amb principis ètics

MECANISMES POLÍTICS		PRINCIPIS ÈTICS			
		Rendició de comptes i transparència	Privacitat i governança de dades	Equitat i no-discriminació	Sostenibilitat
[1]	Estratègies d'IA	X	X	X	X
	Principis locals per a la IA	X	X	X	X
	Directrius, guies i manuals	X	X	X	X
	Protocol intern per a l'ús de la IA				
[2]	Regulacions i lleis locals	X	X	X	X
	Mecanismes de compliment legal	X	X	X	X
	Estàndards reguladors	X	X	X	X
	Estàndards normatius internacionals	X	X	X	X
	Regulació sobre l'aplicació d'IA controvertida		X	X	
[3]	Registre públic d'algoritmes	X	X	X	
	Web municipal amb tota la informació pertinent sobre la IA	X			
	Directorí municipal d'eines d'IA adquirides per a ús intern	X			
[4]	Avaluació i gestió del risc	X	X	X	X
	Avaluacions d'impacte en matèria de drets humans	X	X	X	X
	Avaluació d'impacte ambiental				X
[5]	Auditories	X	X	X	X
	Eines per a l'autoavaluació	X	X	X	X
[6]	Elaboració d'informes i supervisió interna				
	Intervenció humana (human-in-the-loop)	X		X	
	Procediments de resposta i objeccions per a la ciutadania	X		X	
[7]	Clàusules d'adquisició pública	X	X	X	X
	Protocols interns per a l'adquisició de serveis d'IA				
[8]	Òrgans assessors i supervisors	X	X	X	X
[9]	Aliances, comunitats de pràctiques i grups d'aprenentatge	X	X	X	X
[10]	Formació del personal municipal	X	X	X	X
	Equip municipal d'IA				
	Òrgan municipal d'IA				
	Enfocament multidisciplinari			X	
[11]	Laboratoris, hubs i centres d'innovació d'IA				
	Observatori d'IA local				
	Entorns aïllats de proves reguladors (regulatory sandboxes)		X		
[12]	Participació popular	X		X	X
	Formació del públic (alfabetització digital)		X	X	X
	Comitès locals d'ètica d'IA	X	X	X	X
[13]	Governança de dades	X	X	X	X
[14]	Marcos i eines de prova	X	X	X	X
	Incentius fiscals (desgravacions fiscals, subvencions...)				X
	Programes de requalificació de treballadors				X
	Marcos d'avaluació	X	X	X	X

Font: les autores

Llegenda de la taula: groc (mecanismes polítics en consonància amb un principi ètic concret o diversos alhora); blau (mecanismes polítics transversal o personalitzables que serveixen per a tots els principis ètics); verd (mecanismes polítics no associats a cap principi ètic específic, però rellevants per a un desplegament responsable dels sistemes algorítmics en general). Gris clar (mecanismes polítics que no s'esmenten en els capítols); gris fosc (mecanismes polítics esmentats explícitament en els capítols; vegeu l'annex I).

Independentment del nivell d'especificitat, abast o eficàcia dels mecanismes polítics, les ciutats no els apliquen de manera uniforme, i s'observen diferències quant a la freqüència a causa de limitacions de recursos, limitacions tècniques o prioritats locals. La taula 3 compara les ciutats incloses en els estudis de cas i destaca la freqüència amb què s'apliquen determinats mecanismes polítics, quins són els més utilitzats i a quins es recorre poc sovint. Tot i que s'ha considerat rellevant comparar les ciutats analitzades a la part II d'aquesta publicació, cal reconèixer que les conclusions extretes dels estudis de cas seleccionats són limitades pel nombre reduït de les ciutats estudiades.

MECANISME POLÍTIC		CIUTATS					
		Barcelona	Amsterdam	Nova York	San José	Dubai	Singapur
[1]	Estratègies d'IA, agendes, plans d'acció, manuals, fulls de ruta, etc.	x	x	x	x	x	x
[1]	Principis locals per a la IA	x	x	x	x	x	x
[3]	Web municipal amb tota la informació pertinent sobre la IA	x	x	x	x	x	x
[13]	Portal de dades obertes (transparència de les dades)	x	x	x	x	x	x
[1]	Directrius, guies i manuals		x	x	x	x	x
[8]	Òrgans assessors i supervisors	x	x	x	x		x
[9]	Membre d'una comunitat de pràctiques	x	x	x	x		x
[1]	Protocols interns per a l'ús de la IA	x	x	x	x		
[4]	Avaluació i gestió del risc	x	x	x	x		
[7]	Clàusules d'adquisició pública	x	x	x	x		
[12]	Participació popular	x	x	x	x		
[3]	Registre públic d'algoritmes	x	x		x		
[4]	Avaluació d'impacte en matèria de drets humans	x	x		x		
[9]	Col·laboracions multiactors en l'àmbit de la IA		x		x		x
[10]	Formació del personal municipal (capacitació municipal)	x	x				x
[10]	Òrgan municipal d'IA	x		x	x		
[11]	Laboratoris, hubs i centres d'innovació d'IA		x			x	x
[13]	Altres polítiques de governança de dades (drets en matèria de dades, mecanismes d'intercanvi de dades, protocols de dades...)	x			x	x	
[2]	Estàndards de transparència algorítmica	x	x				
[6]	Elaboració d'informes i supervisió interna			x	x		
[6]	Procediments de resposta i objeccions per a la ciutadania		x		x		
[5]	Auditories	x	x				
[2]	Regulació sobre l'aplicació d'IA controvertida			x			
[3]	Director municipal d'eines d'IA adquirides per a ús intern			x			
[10]	Creació d'un equip municipal d'IA (capacitació municipal)		x				
[11]	Observatori d'IA local	x					
[11]	Entorns aïllats de proves reguladors (regulatory sandboxes)						x
[14]	Marc i eines de prova						x
[14]	Marc d'avaluació						x

Font: les autores

Nota: la llista de mecanismes polítics s'ha ordenat, en primer lloc, per freqüència descendent i, en segon lloc, per categorització.

3. Reptes i recomanacions

- a. **Hi ha poques referències sobre com aplicar de manera efectiva els principis ètics d'IA a escala local.** La majoria de les directrius, els estudis i els programes de capacitació existents no tenen en compte els reptes específics als quals s'enfronten les administracions urbanes, que sovint estan desconnectades de les estratègies i les polítiques nacionals. Aquesta discrepància s'agreuja encara més per la naturalesa heterogènia dels governs locals, que presenten importants diferències quant a mida, recursos i capacitació.

Recomanacions:

- Localitzar (és a dir, adaptar al context local) els principis ètics regionals, nacionals o globals i els mecanismes polítics de governança de la IA establint definicions locals d'èxit i identificant les prioritats locals.
- b. **Les administracions locals pateixen una important manca de talent i coneixements tècnics,** un problema que s'agreuja encara més per l'escassetat global de professionals d'IA, que dificulta atraure localment persones qualificades. En conseqüència, els governs municipals solen tenir un coneixement limitat de les possibles repercussions i implicacions dels sistemes algorítmics.

Recomanacions:

- Prioritzar els programes de capacitació com a part de les estratègies municipals i els marcs polítics per controlar i regular els sistemes algorítmics. Això hauria d'incloure l'assignació de recursos específics per a programes municipals de formació, la inversió en campanyes de conscienciació pública i la promoció d'iniciatives per adquirir coneixements i habilitats fonamentals al voltant de la IA ètica a l'Administració pública.
 - Per superar els reptes d'atraure talent local, els governs locals han d'invertir en estratègies que facilitin l'intercanvi i l'adaptació del coneixement dels agents locals. A més, forjar aliances i connexions sòlides amb xarxes d'intercanvi de coneixements pot ajudar a resoldre les deficiències quant a especialització.
 - Adoptar un enfocament holístic de la capacitació fomentant els debats públics i les iniciatives de sensibilització dins de les comunitats locals. Cal centrar aquests esforços a formar els ciutadans sobre les oportunitats i els riscos associats a l'ús de sistemes algorítmics.
- c. **Garantir la transparència i la rendició de comptes dels sistemes algorítmics** utilitzats pels governs locals planteja diversos reptes, com ara: la gestió de l'opinió pública i les possibles reaccions negatives respecte als sistemes d'IA externs, l'adaptació de la cultura organitzativa i de les pràctiques laborals per als sistemes d'IA interns, i el foment de la responsabilitat compartida a tota l'Administració pública (és a dir, la IA no hauria de ser considerada una responsabilitat exclusiva del departament de TIC).

Independentment del nivell d'especificitat, abast o eficàcia dels mecanismes polítics, les ciutats no els apliquen de manera uniforme, i s'observen diferències quant a la freqüència a causa de limitacions de recursos, limitacions tècniques o prioritats locals.

Recomanacions:

- Incorporar la transparència com a objectiu bàsic més enllà de les polítiques específiques d'IA. Això inclou fomentar una cultura de la transparència i rendició de comptes al llarg de tot el cicle de vida de la IA.
 - Fomentar l'ús de codi obert, que pot augmentar la confiança i permetre un major control i col·laboració.
 - Deixar clares les responsabilitats designant una persona de contacte específica o «líder del projecte d'IA» per a totes les iniciatives d'IA, de manera que es garanteixi la rendició de comptes i es faciliti la comunicació.
 - Crear múltiples canals per fer arribar comentaris i integreu avaluacions en les etapes clau del projecte per garantir la millora contínua i la capacitat de resposta.
 - Destinar part del pressupost a fases explicatives exhaustives, de manera que es garanteixi que les parts interessades, tant internes com externes, entenguin completament els sistemes d'IA i les seves implicacions.
- d .** Assegurar la **protecció de la privacitat** presenta reptes específics, com ara un panorama regulador complex en el qual els governs locals sovint tenen problemes per desenvolupar-se. Els problemes relacionats amb les dades estan estretament lligats als mecanismes polítics dissenyats per salvaguardar la protecció de dades. En particular, als entorns urbans hi ha disponibles poques dades d'alta qualitat, cosa que es pot atribuir a diversos factors: pràctiques inadequades de gestió de les dades, qüestions ètiques i riscos relacionats amb la recollida de dades a gran escala i un intercanvi de dades inadequat entre administracions per l'absència d'estàndards unificats i de marcs ben desenvolupats de governança de les dades.

Recomanacions:

- Generar conjunts de dades públiques d'alt valor millorant les pràctiques de recollida i gestió de dades per millorar la qualitat i la utilitat de les dades.
 - Promoure la interoperabilitat i la col·laboració intersectorial i entre organismes per facilitar un intercanvi fluid de dades.
 - Crear marcs segurs i transparents per compartir dades que garanteixin la protecció de la privacitat i alhora permetin la innovació.
 - Fomentar la innovació i l'experimentació en entorns controlats, com ara entorns aïllats de proves reguladors, per provar noves solucions basades en dades, de manera segura i responsable.
- e .** Els sistemes d'algoritmes poden reforçar les desigualtats urbanes existents i alhora crear noves formes de discriminació, d'aquí la importància de parar atenció a la **noció d'equitat i no-discrimi-**

nació quan l'Administració local desplega sistemes d'IA. Un repte concret en aquest sentit és que la discriminació automatitzada provocada per la IA és més abstracta, opaca, difícil de detectar (caixes negres) i a gran escala. Per tant, altera els recursos i els procediments jurídics tradicionals que solen utilitzar els governs locals per detectar-la, prevenir-la i corregir-la.

Recomanacions:

- Considerar les múltiples funcions de les administracions públiques –com a desenvolupadores, responsables del desplegament i reguladores– en dissenyar iniciatives per millorar l'equitat i la no-discriminació dels sistemes algorítmics.
 - Localitzar els marcs polítics existents per abordar la discriminació no intencionada en els sistemes algorítmics i garantir que s'adapten als reptes específics dels entorns urbans.
 - Incorporar un enfocament holístic a la governança de la IA dins de les administracions locals, tenint en compte els impactes socioeconòmics al llarg de tot el cicle de vida de la IA, des del disseny fins al desplegament.
 - Garantir la diversitat entre els equips implicats en el disseny i el desplegament de sistemes algorítmics per reduir el risc de resultats esbiaixats i promoure solucions inclusives.
- f. El repte més important associat al principi de **sostenibilitat mediambiental de la IA** té a veure amb el fet que la IA per a la sostenibilitat sovint entra en conflicte amb la sostenibilitat de la pròpia IA. A més, hi ha pocs marcs per a la sostenibilitat de la IA amb un enfocament urbà.

Recomanacions:

- Avaluar l'impacte del cicle de vida complet dels sistemes d'IA per determinar si els seus beneficis superen els seus costos mediambientals. Minimitzar la petjada de carboni dels desplegaments de la IA a tota la ciutat prioritant els sistemes eficients energèticament, adoptant pràctiques de computació ecològica, utilitzant aprenentatge automàtic en dispositius petits (*tiny machine learning*) i alimentant els centres de dades amb energia renovable.
- Fomentar l'economia circular al voltant dels centres de dades reduint els residus electrònics. Promoveu l'aprovisionament responsable, la reutilització i el reciclatge del *hardware* de la IA. Fomentar la reutilització i el reciclatge de tecnologies, dades i infraestructures d'IA.
- Reutilitzar l'energia i els recursos utilitzats per la infraestructura d'IA i desplegar sistemes d'IA que s'integrin perfectament amb la infraestructura urbana existent, optimitzant tant el consum energètic com l'eficiència del sistema.

Finalment, és important reconèixer les limitacions de la recerca presentada en aquesta monografia CIDOB. L'estudi es va veure limitat tant pel període de temps relativament curt en què es va dur a terme com pels reptes inherents a la recollida d'informació en el context del GOUAI

(vegeu Galceran-Vercher i Vidal, 2024). En conseqüència, es reconeix que hi poden faltar alguns exemples clau de mecanismes polítics i marcs de governança algorítmics. A més, la majoria de les evidències que es presenten s'han extret de polítiques promogudes pel Nord Global, principalment a través d'intervencions liderades pels governs locals dels EUA i Europa. Aquest enquadrament geogràfic és una altra limitació que la futura recerca en el context del GOUAI intentarà abordar. Es reconeix que una anàlisi més sistemàtica de les polítiques i pràctiques de governança del Sud Global podria proporcionar nous coneixements i revelar diferents enfocaments polítics, prioritats i reptes de la implementació.

Referències bibliogràfiques

Ada Lovelace Institute, AI Now Institute and Open Government Partnership. «Algorithmic Accountability for the Public Sector», 2021

Ben Dhaou, S., et al. «Global Assessment of Responsible Artificial Intelligence in Cities: Research and recommendations to leverage AI for people-centred smart cities». Nairobi: Programa de las Naciones Unidas para los Asentamientos Humanos (ONU-Habitat), 2024

Jordan, C.; Glickman, J. i Panettieri, A. «AI in cities: Report and Toolkit». Washington: Liga Nacional de Ciudades i Google, 2024

OECD «OECD AI Policy Observatory: Catalogue of Tools & Metrics for Trustworthy AI», 2024

United 4 Smart Sustainable Cities. «Guiding principles for artificial intelligence in cities». Ginebra: Unión Internacional de Telecomunicaciones, 2024

ANNEX 1. Llista de mecanismes polítics esmentats en els capítols dels principis ètics (Part I)

Principis ètics	Mecanismes polítics
Rendició de comptes i transparència	<ul style="list-style-type: none"> • Avaluacions d'impacte [4] • Clàusules d'adquisició pública [7] • Auditories externes algorítmiques [5] • Registres d'algoritmes [3] • Normes de transparència [3] • Comitès multidisciplinaris de supervisió de la governança [8] • Processos participatius (en totes les etapes del cicle de vida de la IA) [12] • Disseny amb intervenció humana (<i>human-in-the-loop</i>) [6] • Formació i reciclatge laboral del funcionariat [10] • Contactes amb xarxes d'intercanvi de coneixement [9] • Col·laboració amb les parts interessades locals [9]
Privacitat i governança de dades	<ul style="list-style-type: none"> • Conformitat legal [2] • Sistemes de gestió de risc [4] • Sistemes de governança de dades [13] • Avaluacions d'impacte [4] • Auditories [5] • Repositoris d'algoritmes i registres de sistemes d'IA [3] • Sandboxes reguladors [11] • Estratègies urbanes d'IA [1] • Col·laboració multiactor [9]
Equitat i no-discriminació	<ul style="list-style-type: none"> • Estratègies d'IA [1] • Anàlisi de riscos i mecanismes de protecció [4] • Avaluacions d'impacte [4] • Normes locals per a la IA equitativa [2] • Normes d'adquisició pública per a una IA equitativa [7] • Legislació urbana per garantir el dret a la justificació [2] • Òrgans assessors multidisciplinaris [8] • Equips diversos i interdisciplinaris • Auditories [5] • Tècniques de mitigació en el cicle de vida de la IA [14] • Xarxes d'intercanvi de coneixement [9] • Formació del personal [10] • Formació de les comunitats ciutadania [12] • Principis [1] • Anàlisi de biaixos [4] • Protecció dels drets digitals [14]
Sostenibilitat	<ul style="list-style-type: none"> • Normes mediambientals, socials i de governança (per exemple, normes d'IA verda) [2] • Avaluacions d'impacte [4] • Supervisió i auditoria [5] • Consorci de ciutats per a la IA sostenible [9] • Incentius fiscals: desgravacions fiscals o subvencions [14] • Junes d'ètica de la IA local [12] • Lleis sobre la privacitat de les dades urbanes [13] • Participació pública [12] • Models de governança participativa [12] • Campanyes d'alfabetització digital [12] • Programes de reciclatge laboral [14] • Sistemes de classificació amb estrelles per a l'eficiència energètica de la IA [14] • Col·laboració internacional [9] • Estandardització internacional [2]

Font: les autores

Nota: els mecanismes polítics que es descriuen aquí mantenen la redacció original dels capítols dels principis ètics (Part I) i s'han classificat segons la categorització de les autores que es proporciona en el capítol final de la Monografia. Pot ser que els mecanismes no apareguin a l'apartat de «mecanismes polítics» dels capítols, però es troben al llarg del capítol en qüestió.

La intel·ligència artificial (IA) promet revolucionar els espais urbans aportant solucions als reptes metropolitans més importants als quals s'enfronta la humanitat avui dia. Les ciutats són un camp de proves ideal per a la IA, atès que els governs locals adopten tecnologies basades en dades per automatitzar tasques rutinàries, augmentar l'eficiència i prendre decisions rendibles. Tanmateix, el desplegament accelerat dels sistemes algorítmics planteja dilemes ètics, especialment pel que fa als drets dels ciutadans i als impactes socials i ambientals de la IA. En aquest sentit, és fonamental una governança d'IA responsable per evitar-ne les conseqüències no desitjades i alhora aprofitar els beneficis d'aquesta tecnologia. Les ciutats que ja han implementat polítiques d'IA ètica poden oferir exemples molt valuosos, ressaltant la necessitat que els responsables polítics adoptin enfocaments que prioritzin el desplegament d'una IA inclusiva, transparent i ètica.

En aquest context, el CIDOB presenta aquesta monografia amb l'objectiu d'examinar mecanismes polítics específics i marcs de governança existents que fomenten una IA urbana responsable. El volum analitza, d'una banda, com els principis ètics fonamentals –com ara la rendició de comptes i la transparència, la privacitat i la governança de dades, la justícia i la no-discriminació, o la sostenibilitat– poden implementar-se de manera efectiva en entorns urbans a través de polítiques concretes. De l'altra, proporciona estudis de cas de ciutats d'arreu del món que han establert marcs integrals de governança d'IA. En definitiva, vol ser un document pràctic que motivi a l'acció i serveixi de guia per a altres agents del sector públic local en el llarg camí que cal recórrer per garantir un desplegament ètic de la IA urbana.