

ALGORITMES I INTEL·LIGÈNCIA ARTIFICIAL AL SECTOR PÚBLIC:

Recomanacions des de la Justícia global



AUTOR

Pablo Jiménez Arandia. Periodista especialitzat en l'impacte social i polític de la tecnologia. Col·labora amb mitjans com El Confidencial, AlgorithmWatch i El Salto Diario. També ha realitzat diverses investigacions sobre l'ús de la IA i la digitalització al sector públic i privat, juntament amb el col·lectiu Lighthouse Reports, l'autora nord-americana Virginia Eubanks i la Fundació Mobile World Capital Barcelona. És autor del llibre *Transparència algorítmica al sector públic*, editat per la Generalitat de Catalunya, i dels pòdcasts *Algoritmos y gobiernos* i *La plataforma y yo*. També ha treballat per a la International Telecommunications Union (ITU), l'agència de les Nacions Unides especialitzada en les TIC.

EDICIÓ

Lafede.cat, Gener de 2024

Lafede.cat – organitzacions per a la Justícia global. Federació de més de 120 organitzacions catalanes que treballa activament per aconseguir la Justícia global i l'eradicació de les desigualtats a tot arreu, a d'altres llocs del món i a casa nostra, mitjançant la cooperació al desenvolupament, la defensa i promoció dels drets humans i el foment de la pau.

APORTACIONS I REVISIÓ

Sara Suárez-Gonzalo. Doctora en Comunicació per la Universitat Pompeu Fabra i investigadora Juan de la Cierva a l'Internet Interdisciplinary Institute de la Universitat Oberta de Catalunya. Estudia les implicacions socials i polítiques de les tecnologies basades en dades, amb un interès especial en els seus efectes sobre la democràcia, la justícia social i els drets fonamentals.

Javier Sánchez Monedero. Doctor en Ciències de la Computació i Intel·ligència Artificial i Investigador Distingit “Beatriz Galindo” a la Universitat de Còrdova. Prèviament va ser investigador associat al Data Justice Lab de la Universitat de Cardiff. El seu treball actual se situa en la bretxa de coneixement entre les ciències socials, la biomedicina i la tecnologia amb propostes d'auditoria i disseny a la intersecció dels sistemes d'informació intel·ligents i la justícia social.

Paula Boet Serrano. Assessora de polítiques urbanes a l'Ajuntament de Barcelona. Ha estat tècnica de drets digitals al Comissionat d'Innovació Digital - BIT Habitat, de l'Ajuntament de Barcelona, i a la Coalició de Ciutats pels Drets Digitals. És graduada en Filosofia, Política i Economia, i màster en Ciutadania i Drets Humans.

Coordinació i revisió final: Judith Membrives i Llorens

MAQUETACIÓ I IL·LUSTRACIÓ

Fàbrica Gràfica Coop V. www.fabrikagrafika.com



Aquesta obra està subjecta a la llicència de Reconeixement -NoComercial-SenseObraDerivada 4.0 de Creative Commons. Si voleu veure una còpia d'aquesta llicència, accediu a : <https://creativecommons.org/licenses/by-nc-nd/4.0/deed.ca>

1. [INTRODUCCIÓ] Nous relats per a una tecnologia pel bé comú **4**

2. Paraules per entendre aquesta guia **6**

3. Automatitzar què i per a què?
Per on començar? **9**

4. De les dades a l'aplicació: de què es compon un SAD? **12**

5. Principis aplicats a la pràctica: com ho fem? **17**

Recursos **25**

1—[INTRODUCCIÓ]

Nous relats per a una tecnologia pel bé comú



Les tecnologies i l'ús que se'n fa **no escapen de les dinàmiques polítiques i ideològiques que les envolten**. El 2019 el llavors relator de les **Nacions Unides** per a l'extrema pobresa i els drets humans, **Philip Alston**, va alertar sobre com els governs de bona part del món **estaven avançant cap a "una distopia de benestar digital"**. Un escenari en què els algorismes i altres sistemes basats en dades es fan servir per "automatitzar, predir, identificar, vigilar, detectar, singularitzar i castigar" a la ciutadania.

Les polítiques de digitalització, alertava Alston, gairebé sempre s'acompanyen d'una **reducció de la despesa en polítiques socials, una caiguda dels seus beneficiaris i un esvaïment de la idea que l'Estat ha de retre comptes a la ciutadania** (i no només a l'inrevés). Quatre anys després els seus advertiments continuen vigents.

Alguns dels algorismes més comuns avui al sector públic estan pensats per **detectar el frau en les ajudes socials, perfilar persones treballadores aturades o predir el comportament de col·lectius** històricament discriminats. Aquest tipus d'eines s'han estès des de fa més d'una dècada entre administracions de tot el món, també a **Catalunya** i a l'Estat espanyol, gairebé sempre sota un mantell d'opacitat intencionat.

Per descomptat, aquests **no són els únics algorismes presents avui al sector públic**. Hi ha també

sistemes pensats **per expandir drets i no per vulnerar-los**. En els darrers anys les iniciatives que promouen **un ús responsable, just i transparent de la intel·ligència artificial (IA)** i altres sistemes s'han multiplicat. Aquesta guia precisament pretén recollir recomanacions que continuïn aquest camí mantenint la consciència de quin és l'estat actual de les coses quan ens enfoquem en la digitalització de l'acció pública.

En primer lloc, és urgent **enfrontar el paper protagonista que les empreses privades han acaparat** en el disseny, el funcionament i fins i tot la supervisió d'aquest tipus de sistemes. Avui dia són escasses les administracions públiques amb capacitat (i intenció) de desenvolupar algorismes propis.

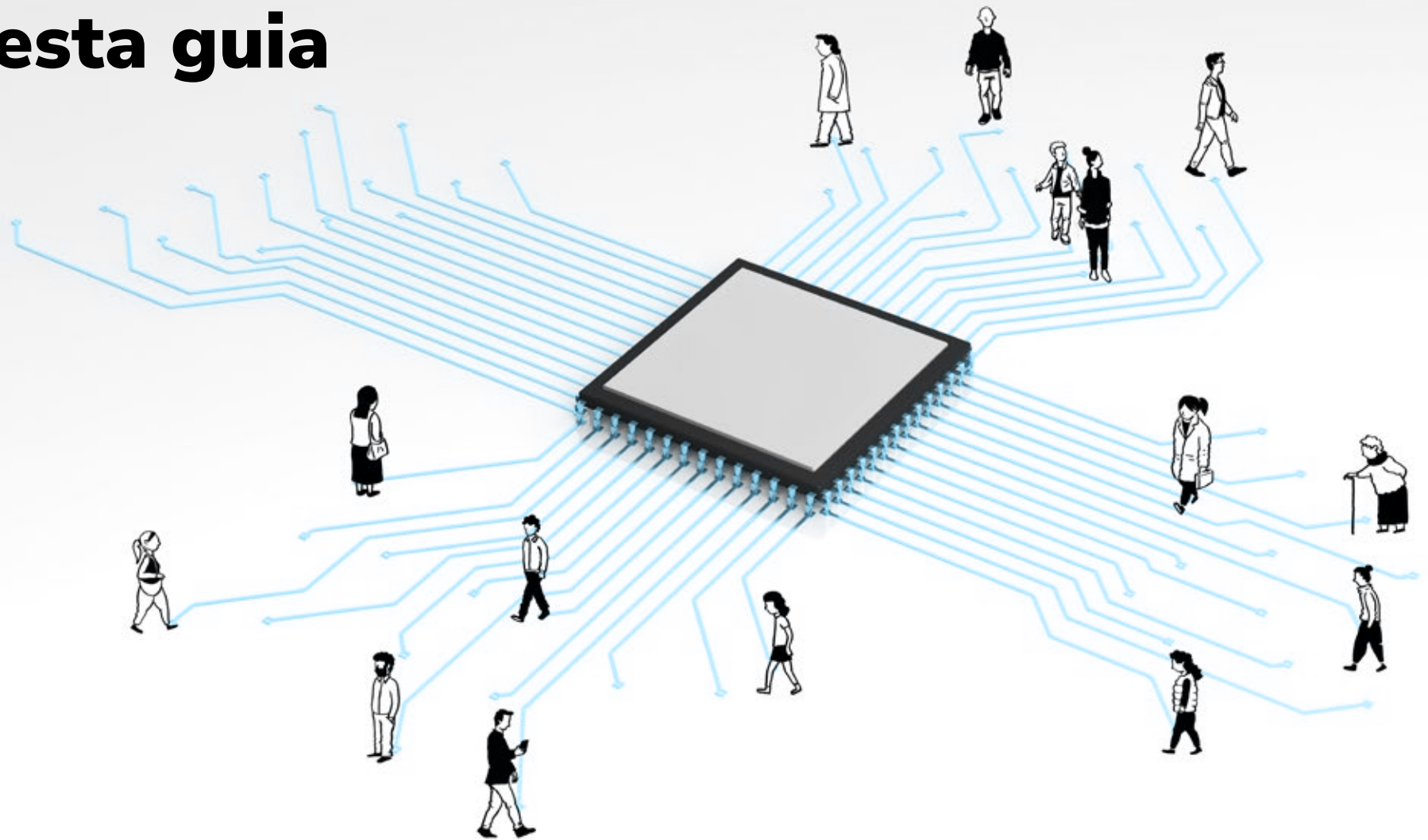
A més, les necessitats d'aquestes tecnologies (un gran poder computacional i grans volums de dades), unides a la lògica del lliure mercat, han provocat una **concentració del poder d'aquesta indústria a poques mans**. Aquesta situació provoca una evident pèrdua de sobirania de les administracions, els estats i la societat civil en conjunt.

Alhora cal acabar d'una vegada per totes amb la idea que digitalitzar o automatitzar és sempre una bona idea. Més aviat al contrari: l'ús d'una IA o un algorisme complex és útil en poques ocasions. En canvi, la seva capacitat per reproduir o agreujar problemes socials ja existents està prou demostrada, per exem-

ple a l'hora de discriminar comunitats senceres o augmentar l'opacitat sobre l'acció del govern.

En definitiva, amb aquesta guia volem **obrir nous relats sobre com les administracions i les professionals que hi treballen poden ajudar-se de la tecnologia en la recerca del bé comú**. Sempre sota la idea, com assenyalava Alston fa gairebé un lustre, de *"millorar el nivell de vida a les persones vulnerables i desfavorides i concebre noves maneres de cuidar els qui s'han quedat enrere"*.

2—Paraules per entendre aquesta guia



A vegades, al llarg d'aquesta guia farem servir termes tècnics o especialitzats. Per fer la lectura més comprensible, en el següent glossari trobareu els significats d'algunes d'aquestes paraules:

Algorisme / Algoritme

Conjunt d'instruccions que permeten fer un càlcul matemàtic o resoldre un problema. A través de la informàtica els algorismes permeten automatitzar digitalment un procés que abans es feia de manera manual.

Intel·ligència artificial (IA)

Disciplina que s'ocupa de crear màquines capaces de simular algunes habilitats humanes i fer tasques autònomament a través d'aquestes. A partir de l'anàlisi de moltes dades aquests sistemes informàtics generen patrons, els identifiquen i, d'acord amb aquests, produeixen nova informació. Tot i el seu nom, la IA no és intel·ligent en el sentit humà del terme, ja que no té capacitats com l'empatia, el judici crític o la contextualització.

“La IA no raona sobre el coneixement. No se li pot atorgar la capacitat de raonar, però sí la de fer càlculs ràpidament, molt més que els humans. Per això li atorguem aquesta capacitat d'intel·ligència. Però així, una calculadora també podria ser considerada intel·ligent”
[AlgoRace]

"Lluny de la mitificació que avui envolta la IA, no parlem d'una intel·ligència sobrehumana, autònoma i que pot prendre decisions de forma racional i objectiva, sinó de sistemes construïts i programats per persones, entrenats amb bases de dades que poden contenir biaixos"
[Paula Boet]

Sistema automatitzat de decisió (SAD)¹

Qualsevol procés en què es fan servir grans volums de dades, algorismes i, en algunes ocasions, tècniques d'entrenament d'IA per optimitzar un servei o una decisió pública. Per exemple, mitjançant l'automatització de prediccions, recomanacions o càlculs. Un SAD cal entendre'l com un sistema complex on el que és tècnic només és una part i el context juga un rol fonamental.

Els SAD al sector públic en general no automatitzen completament una decisió, sinó que la complementen i, com a mínim, sobre el paper, ajuden les treballadores públiques a prendre decisions per si mateixes.

Cicle de vida algorísmic

L'automatització d'un servei o una decisió és un procés laboriós que requereix diferents passos, entre els quals s'inclouen:

- (a) la definició del problema que volem resoldre i de les dades que farem servir
- (b) la recollida i el processament d'aquestes dades
- (c) la construcció d'un model matemàtic a partir de fórmules o usant tècniques estadístiques que el generin per si mateix a partir de les dades
- (d) desenvolupament, implementació i monitoratge del sistema
- (e) seguiment a la seva evolució i impactes

Biaix algorítmic

Hi ha molts tipus de biaixos en els algorismes. Un sistema pot estar esbiaixat perquè està mal construït, pel fet que les dades utilitzades no són les correctes o les variables no són les adequades per a la finalitat que es busca. Però com a eines creades per humans dins d'un context sistèmic determinat, els algorismes incorporen també biaixos. Aquests, solen ser més difícils de detectar quan estan relacionats amb desigualtats històriques sobre certs col·lectius pel seu origen, gènere o classe social.

Discriminació algorísmica

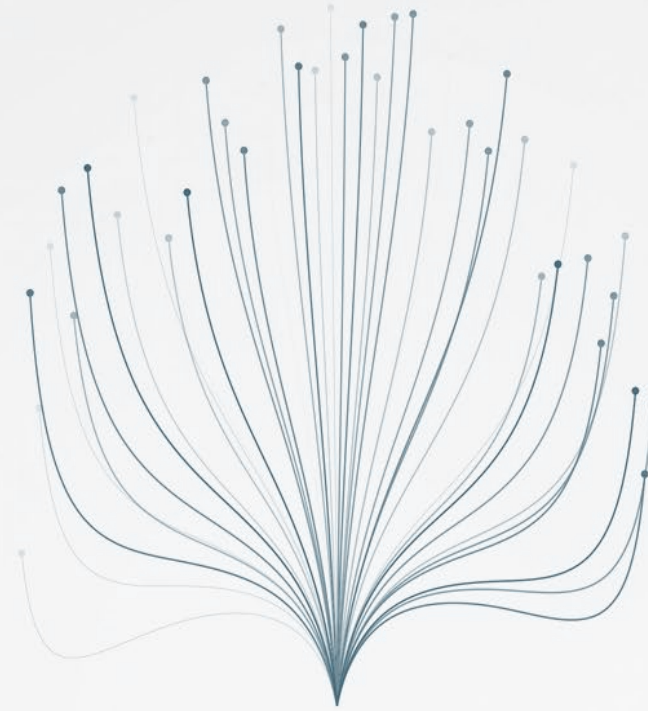
Es considera que un algorisme discrimina quan tracta de manera sistemàtica un col·lectiu de manera diferent d'un altre. És habitual que aquestes discriminacions s'acarnissen negativament amb col·lectius vulnerabilitzats, a partir de biaixos intencionats o no. La reproducció i la perpetuació d'aquestes discriminacions és un dels grans riscos de l'automatització. També hi pot haver una discriminació del sistema positiva i intencionada, per tal de corregir una injustícia històrica.

Drets digitals

Els algorismes i la IA, com altres tecnologies, afecten drets humans ja guanyats i reconeguts. Avui cal protegir aquests drets en un context com el digital que és nou, canviant i cada cop més interconnectat amb l'analògic.

1
El terme SAD és una adaptació de l'anglès ADMS (Automated decision-making system), una paraula que s'ha estès els darrers anys a l'estudi de sistemes algorítmics. En aquesta guia el farem servir per referir-nos als sistemes algorítmics i d'IA usats al sector públic, que inclouen també una part política, social, econòmica i mediambiental.

3—Automatitzar què i per a què? Per on començar?



Els SAD existents avui dia al sector públic estan **disenyats majoritàriament per complementar processos**. És a dir, no prenen decisions per ells mateixos, sinó que estan pensats per donar suport, influir o determinar d'alguna manera la feina d'un funcionari. És important, així i tot, fer diversos matisos.

Encara que no estiguin ideats per decidir de manera autònoma, està demostrat que **els humans solem fiar-nos massa de les màquines**. Tenim tendència a donar una credibilitat excessiva als resultats que ofereixen. Això és un problema si l'algorisme en qüestió s'està equivocant o està prenent una decisió injusta.

D'altra banda, el concepte mateix **automatitzar** pot donar la idea que simplement traslladem a un sistema informàtic un procés que abans ho feien persones, sota un paraigua de suposada imparcialitat o invariabilitat. Però la realitat moltes vegades és diferent. **"L'ordinador ni funciona com una persona ni utilitza informació qualitativa, que és un tipus d'informació molt utilitzada per exemple en els serveis socials"**, explica Sánchez Monedero.

Per exemple, un sistema per filtrar sol·licituds d'accés a una escola d'idiomes basat en les rendes de les famílies sí que estaria automatitzant digitalment un procés, ja que el resultat de la màquina seria sempre el mateix que el d'una persona, en fixar-se només a la renda. Podríem assumir que l'algorisme ho faria d'una manera més ràpida i barata.

En canvi, si parlem de sistemes d'avaluació individual de riscos com **RisCanvi**², un SAD usat a les presons de Catalunya des de fa més d'una dècada, la realitat és una altra.

"Aquests models estadístics capturen la dinàmica poblacional en el sentit estadístic, però es perd totalment la valoració individual. I, de vegades, això és acceptable i altres vegades no", assenyala Sánchez Monedero. **És realment possible que un algorisme prengui una decisió justa sense tenir en compte aquest context individual?**

Precaucions basades en riscos

Amb tot, **els àmbits d'actuació dels SAD i els seus objectius varien molt**. Actualment, **a Europa**, Catalunya i **a l'Estat espanyol** hi ha en ús sistemes de puntuació de risc de frau sobre la població que rep algun tipus de subsidi públic. Però també hi ha algorismes en àrees molt diferents. Per exemple, recomanadors de llibres en biblioteques públiques que basen els seus resultats a l'historial de préstecs de l'usuari.

Els riscos potencials sobre els drets de la ciutadania d'un sistema i un altre, com és obvi, no tenen res a veure. I, per tant, tampoc les precaucions i protocols que els han d'acompanyar, apunta Paula Boet. **"No hem de mirar de la mateixa manera els sistemes**

que no tenen un risc d'impacte en els drets fonamentals de la ciutadania que els que sí que en tenen".

2

Els algorismes de RisCanvi assignen un nivell de risc a cada pres de les presons de Catalunya en funció de desenes de factors relacionats amb el perfil social i delictiu de l'intern i el seu comportament a la presó. La nota assignada tracta de mesurar, entre altres qüestions, la probabilitat que el pres cometi de nou un delictiu si surt de presó. Aquesta informació es trasllada tant als funcionaris de presons com als jutges que han de decidir, per exemple, sobre una petició de tercer grau.

Les propostes actuals de governança pública d'aquestes tecnologies caminen precisament cap a aquest enfocament basat en els riscos. És el cas de la proposta de regulació europea de la IA, que divideix els SAD **en funció de si representen un risc inacceptable, alt, limitat o mínim**³. La nova norma, a la qual s'hauran d'acollir tots els estats membres i els seus diferents nivells administratius, entrarà en vigor el 2024.

Cal tenir en compte que el risc potencial d'un SAD no només té a veure amb la seva complexitat tècnica: el disseny i l'àmbit d'actuació de vegades és determinant per estimar-ne el risc. “Molts dels sistemes que s'estan utilitzant per a temes sensibles com la gestió de recursos públics, etc., no són en realitat res complexos tècnicament”, analitza Sara Suárez-Gonzalo. “Però no per ser més simples són menys perjudicials”, afegeix Boet.

D'on ve la iniciativa?

En els darrers anys el sector públic s'ha convertit en una oportunitat de negoci molt lucrativa per a les empreses que desenvolupen o comercialitzen algorismes i IA. Sota promeses genèriques com optimitzar processos o estalviar costos les companyies del sector estan venent “solucions digitals” a organismes públics, impulsades per potents campanyes de màrqueting.

Aquestes compres de vegades responen a una necessitat real -o almenys a la recerca de millores davant de problemes com la manca de personal o les noves demandes de la ciutadania-, però en altres són fruit de la gran publicitat i foment de les expectatives que avui envolta aquestes tecnologies. El **fals mite segons el qual digitalitzar o automatitzar un procés sempre és bona idea** continua movent molts alts funcionaris a signar contractes milionaris per adquirir productes tecnològics.

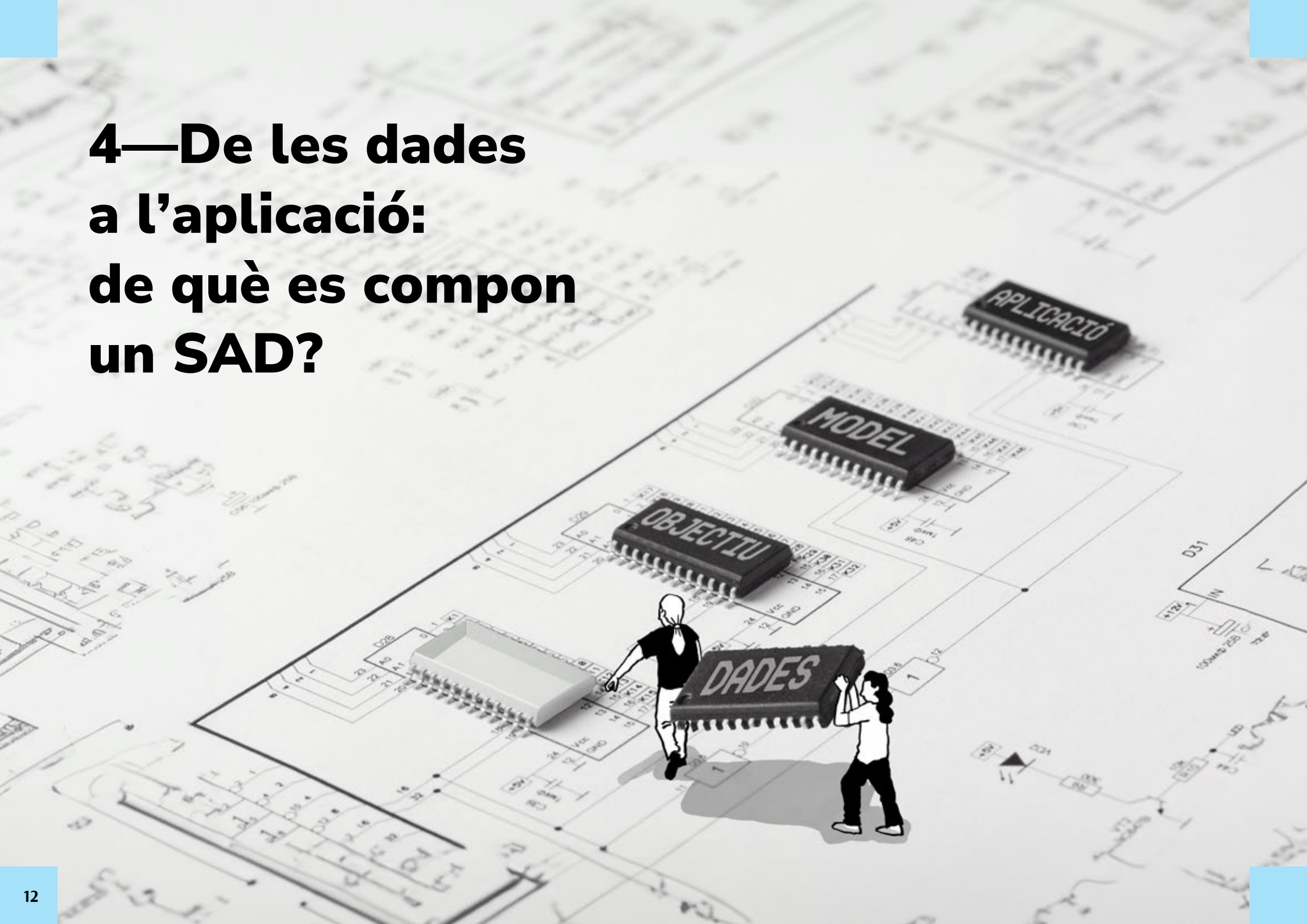
Aquest problema es veu agreujat per **la manca d'evidències que, moltes vegades, envolten aquestes eines**. És freqüent que els governs adquireixin productes i serveis que no han demostrat una utilitat real per tal que suposadament cobreixen. Així ho assenyala **un informe recent de la UNESCO**, que critica la influència creixent de la gran indústria tecnològica a les polítiques educatives de mig món, inclòs l'Estat espanyol.

L'organisme internacional adverteix que amb prou feines s'estan avaluant els resultats d'implantar aquestes eines digitals a l'educació pública. I assenyala la manca de “proves sòlides” que demostrin el suposat valor afegit que representen per a la formació dels menors.

3

La **nova llei europea d'intel·ligència artificial (AI Act)** és la primera norma integral que tracta de regular els usos d'aquestes tecnologies tant al sector privat com al públic. La norma incorpora unes exigències de transparència i control especials sobre els sistemes considerats de més risc.

4—De les dades a l'aplicació: de què es compon un SAD?



Per parlar d'automatització és important conèixer quins són els elements clau d'un SAD. L'ONG alemanya **AlgorithmWatch** proposa la divisió següent: **les dades, l'objectiu, el model i la seva aplicació.**

En aquest repàs comencem a dibuixar unes primeres recomanacions per a les administracions sobre com pensar, desenvolupar i implementar aquests sistemes des dels principis de la justícia global.

a) LES DADES

Les **dades** són la pedra angular sobre la qual es construeix i desplega qualsevol algorisme. Per tant, definiran el funcionament futur del SAD que volem desenvolupar.

Un primer pas és **avaluar quin tipus de dades ens calen i comprovar si efectivament comptem amb elles o no.** També la manera com estan disponibles: és habitual que les dades en mans d'una administració no estiguin correctament digitalitzades o estructurades, cosa que pot ser un impediment per construir algorismes sobre la base d'ells.

Un altre punt a tenir en compte és la **privadesa**. És freqüent que les dades personals en mans del sector públic es puguin utilitzar únicament per a finalitats molt específiques. És per això que a vegades cal recórrer a un altre tipus de fonts. "Les administracions

han d'establir **protocols interns de revisió de les bases de dades que fan servir, posant especial atenció en si aquests són suficients per a la finalitat que es busca, si són de qualitat, etc.**", comenta Sara Suárez-Gonzalo.

També cal ser conscients que les dades que seleccionem per construir un SAD mai no seran una foto perfecta de la realitat que volem analitzar. Més aviat es tracta d'una imatge fixa que inclourà certs biaixos i assumpcions sobre el problema segons el context momentani en el que s'hagin recollit i incorporat.

Així, com ja hem esmentat abans, aquests sistemes acostumen a fer servir dades quantitatives (valors numèrics) per als seus càlculs. Però **moltes vegades no és fàcil reduir la realitat que pretenem analitzar dades. I si ho fem, és probable que deixem fos realitats o matisos importants.** Tal com comenta Suárez-Gonzalo, si un sistema vol predir qüestions complexes com la pobresa o el risc d'abandonament escolar, és complicat definir quines categories de dades farem servir per fer-ho. "Quin tipus de dades tens en compte per predir si un xaval deixarà el col·legi? Ni tan sols en l'àmbit sociològic estan clars quins factors ho defineixen. Són coses delicades", adverteix.

Preguntes similars poden sorgir en pensar en un sistema predictiu com RisCanvi. **Fins a quin punt la informació del passat ens pot ajudar o no a**

"Les administracions han d'establir protocols interns de revisió de les bases de dades que fan servir", comenta Sara Suárez-Gonzalo

endevinar el futur? Entre les variables utilitzades per aquest algorisme usat a les presons Catalunya hi ha l'històric criminal de la persona interna, la seva xarxa social i familiar fora de presó i possibles addicions, entre altres factors delicats.

ONGs i professionals del dret han alertat com Ris-Canvi, pel fet d'emprar dades molt sensibles de caràcter social, pot contribuir a la perpetuació de les discriminacions ja existents a les presons catalanes. Aquest fet és especialment problemàtic, ja que Ris-Canvi pot derivar en decisions injustes dels jutges sobre qüestions tan vitals com una llibertat condicional. De fet, **l'ús de SAD predictius en àrees especialment sensibles des del punt de vista dels drets, com ara el món penitenciari i l'acció policial, és rebutjat per múltiples experts.**

La **infrarepresentació o, al contrari, la sobre representació de certs col·lectius a les dades** és un altre problema habitual. Això pot venir per una selecció basada en criteris racistes, masclistes o d'altres tipus, o, en altres ocasions, per la manca de dades de qualitat sobre certs grups. "Per exemple, s'acostuma a treballar amb menys dades de la població envellida, que de la gent jove, de les dones respecte dels homes o de les persones d'ingressos més baixos, en comparació a les de persones riques", comenta Suárez-Gonzalo.

Un SAD entrenat només amb dades d'un col·lectiu majoritari oferirà resultats esbiaixats en aplicar-lo

sobre altres col·lectius. Un exemple molt conegut d'aquest tipus de discriminació és el provocat per les **tecnologies de reconeixement facial**. Aquests sistemes cometen errors principalment sobre persones racialitzades o d'altres col·lectius històricament discriminats⁴, per això cada vegada més institucions estan plantejant un veto al seu ús.

b) L'OBJECTIU

Abans d'automatitzar qualsevol procés o decisió a l'administració és important definir quin és el nostre **objectiu**. És a dir, **identificar el problema que volem resoldre o el servei públic que volem millorar** i després avaluar si un SAD ens pot ajudar realment per a la finalitat perseguida. "L'administració ha de definir què significa que el sistema funcioni i com ho comprovarà. Això, a vegades, també s'oblida", adverteix Sánchez Monedero.

Els algorismes usats al sector públic, com qualsevol altra decisió dins de l'administració, tenen una intenció política determinada. En els darrers anys, com comentàvem a la Introducció d'aquesta guia, la **digitalització de molts serveis públics ha seguit els principis propis del capitalisme neoliberal**: austeritat, reducció de la despesa pública i de les polítiques socials i una mal anomenada "eficiència" en la provisió de serveis.

L'automatització de l'activitat policial ha estat una altra àrea d'interès creixent per a molts governs. N'és un exemple la **polícia predictiva**, un terme que agrupa SAD basats en dades relacionades amb delictes del passat (hora, ubicació, tipus de crim, etc.) que tracten de predir on hi haurà més crims en el futur, per, d'aquesta manera, enviar més o menys patrulles a una zona determinada.

4

Aquest tipus de biaixos no són els únics problemes associats al reconeixement facial i als sistemes de recollida de dades biomètriques. Un dels més greus és el seu ús per la policia en contextos com els controls fronterers, on la vulneració de drets fonamentals ja és una realitat que es pot veure agreujada per l'ús d'aquestes tecnologies.

Múltiples estudis han acreditat que aquests sistemes perpetuen la discriminació cap als barris o zones geogràfiques més pobres i exclosos. A més, a països com la **Xina** i els **Estats Units** diverses administracions n'han cancel·lat l'ús, després de fer inversions milionàries, després de no poder demostrar la seva efectivitat en la reducció de la criminalitat.

Per no caure en aquest tipus de polítiques públiques fallides, **les administracions han d'abandonar la idea que un algorisme o una IA ajudarà en tots els casos a resoldre un problema.** I, en canvi, ha de ser molt conscient de les limitacions i dels riscos que el seu ús pot implicar.

Si després d'aquesta reflexió considerem útil l'automatització d'un procés, és important **mantenir aquesta prudència en els passos següents del cicle de vida algorísmic:** desenvolupament i construcció del model, implementació, supervisió, etc.

c) EL MODEL

Els models algorítmics es poden construir de dues maneres. Podem escriure directament unes regles i normes específiques a través de llenguatges de programació; o, de forma indirecta, mitjançant tècniques d'entrenament de la IA que extreuen informació de les dades que hem introduït al sistema.

En tots dos casos, un **SAD analitza un segment de la realitat a partir d'uns criteris determinats i limitats**, en teoria suficients (segons el criteri dels seus dissenyadors) per a resoldre el problema a què ens enfrontem.

Durant l'estiu del 2020, enmig de l'emergència de la Covid-19, l'agència pública que regula les proves d'accés a la universitat al **Regne Unit (Ofqual)** va decidir suspendre els exàmens i crear un programari per assignar les notes. Quan els alumnes van rebre la seva qualificació van observar que el programa havia baixat majoritàriament les dels estudiants de centres públics, mentre que havia mantingut o pujat les dels privats⁵.

El model utilitzat per Ofqual feia servir una quantitat molt petita d'informació per al seu càlcul i, a més, prioritzava l'historial acadèmic de l'escola, **assignant poc pes als èxits de l'alumne en qüestió.** D'aquesta manera es perpetuava el privilegi de classe dels alumnes de centres privats davant dels públics. Aquest exemple demostra que **decidir quins factors es consideren rellevants per solucionar el problema que afrontem és una feina delicada.** Encara més quan davant tenim qüestions d'alta complexitat.

Per tant, **a l'hora de desenvolupar un model dins un SAD, és important ser conscient de les oportunitats que ofereix, però també de les limita-**

cions. Especialment, si aquestes darreres poden impactar en col·lectius discriminats i vulnerar drets fonamentals.

Un altre impacte habitualment ignorat d'aquestes tecnologies és la petjada ecològica. Lluny del relat segons el qual aquestes eines són innòcues per al planeta, **els algorismes i la IA –especialment en sistemes avançats– requereixen l'ús de molts recursos naturals per al seu desenvolupament.** Aquests recursos els trobem en tot el cicle de vida algorítmic, des de les matèries primeres necessàries per al muntatge d'un ordinador fins a l'aigua usada per refrigerar els centres de dades on s'entrenen aquests sistemes.

5

Després de la mobilització de l'alumnat, el govern va descartar aquest sistema amb un clar biaix de classe i va fer que els professors avaluessin, sota el seu propi criteri, les notes que havien estat rebaixades pel programari.

d) L'APLICACIÓ

Per entendre l'impacte de l'automatització sobre els drets i evitar mirades reduccionistes cal entendre els SAD com a sistemes no només tècnics. És a dir, com a **iniciatives que inclouen una part tècnica, però que s'apliquen en un context social, cultural, polític i normatiu determinat**. Cal entendre'ls, doncs, com a sistemes sociotècnics.

Al sector públic aquest context es pot reconèixer en múltiples qüestions. Per exemple, si les persones funcionàries que interactuen amb el sistema no tenen la formació i els coneixements adequats sobre el funcionament del sistema i les possibles implicacions de la seva utilització, difícilment s'aconseguiran els objectius marcats. O si l'àrea on s'aplica el SAD no té els recursos necessaris per donar resposta al problema a resoldre, la precisió tècnica de l'algorisme no servirà de res.

També és important **establir clarament quins protocols o passos se segueixen dins d'un SAD**. Per exemple, amb relació a com el funcionariat públic interpreta la informació generada per un algorisme.

El sistema **VioGén**⁶ està integrat al protocol de seguiment de casos de violència de gènere del Ministeri de l'Interior espanyol. Les investigacions sobre aquest SAD mostren com la manca de suport legal i psicològic a les dones denunciants és un problema

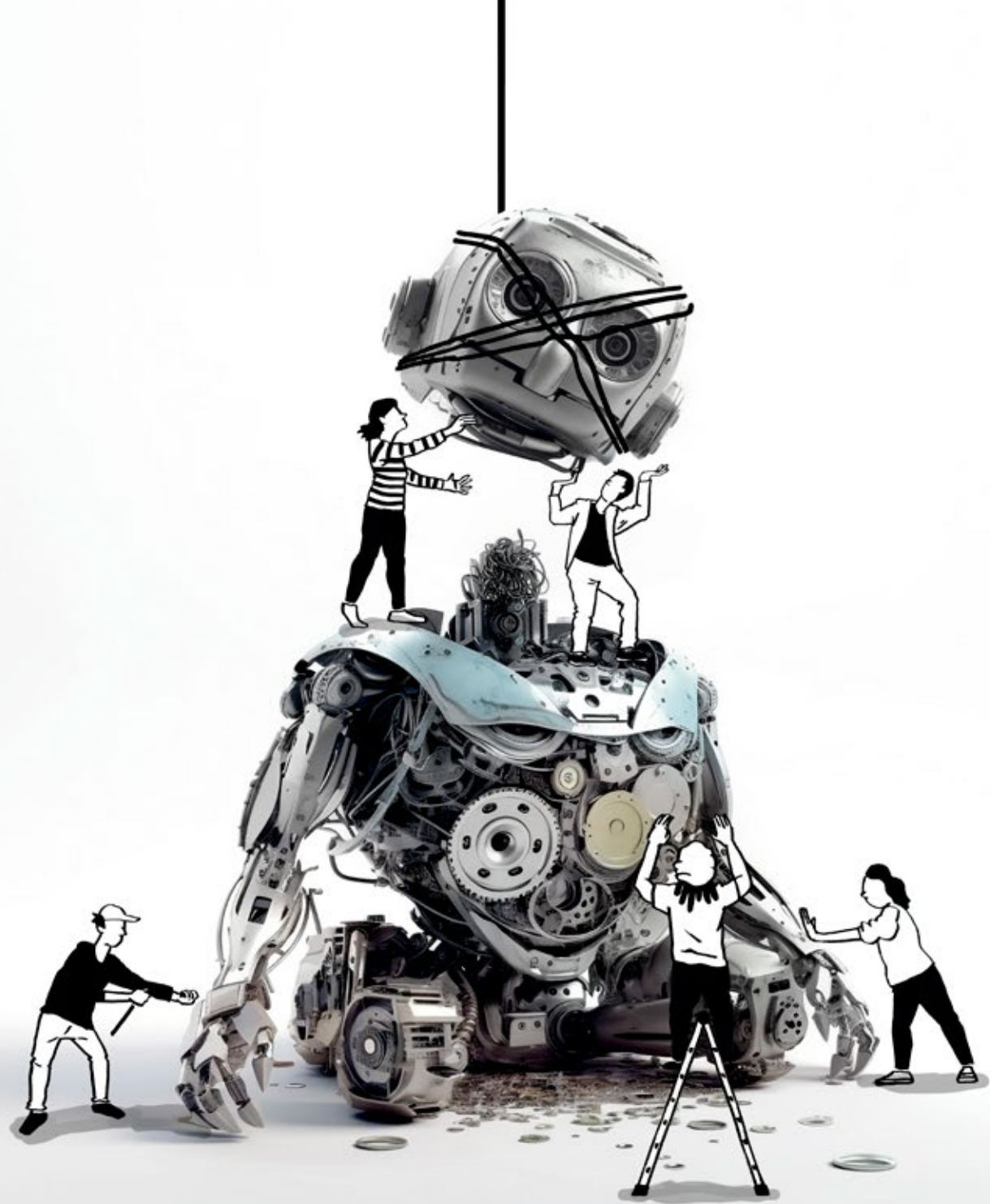
a l'hora de demanar les dades que alimenten l'algorisme. Un altre factor que problematitza el seu rendiment és l'escassa formació en violència de gènere dels cossos policials, fet que compromet la bona recollida i registre de les dades.

Així, per construir sistemes responsables és fonamental **fixar unes regles clares sobre el seu ús i un pla que garanteixi els recursos necessaris per a un bon funcionament**. A més, quan parlem de sistemes que poden afectar drets bàsics sempre hi ha d'haver una comprovació humana en totes les fases del cicle de vida algorímic. Perquè les decisions i els judicis són cosa d'humans, no d'algorismes.

6

Igual que RisCanvi, VioGén és un mesurador de risc, que en aquest cas intenta estimar quin risc corren les dones que han denunciat un cas de violència masclista. Aquesta estimació es calcula a partir d'informació sobre l'agressor, la víctima i la relació entre tots dos. La puntuació assignada per VioGén serveix la policia per decidir les mesures de protecció sobre la víctima.

5—Principis aplicats a la pràctica: com ho fem?



Portar a la pràctica principis com la **no-discriminació, la transparència, la justícia, la rendició de comptes i el bé comú** en l'àmbit de l'automatització del sector públic no és una tasca senzilla. Sánchez Monedero apunta que encara estem en una fase inicial d'aquest repte, un repte que sens dubte portarà temps. "Quant hem trigat al camp de la medicina a assolir una governança democràtica i política en l'àmbit global? Doncs tot el segle XX".

"En aquest moment d'explorar solucions allunyades de l'afany de lucre de les empreses i d'una visió punitiva i reduïda del paper de l'Estat, calen noves estratègies", afegeix Boet. **"Ens falten relats i imaginaris per pensar que la IA no és només per fer policia predictiva o per detectar el frau, que són usos que no haurien ni de plantejar-se"**, afirma.

Amb aquesta idea al cap a continuació llistem una sèrie de pautes que podria posar en pràctica qualsevol administració pública. La següent no és una llista exhaustiva i rígida de recomanacions, sinó un full de ruta pràctica adaptable a cada context particular per avançar cap a uns algoritmes més democràtics en el sector públic.

a) Posar les comunitats afectades al centre i fer-les partícips

Històricament, **el desenvolupament de projectes tecnològics al sector públic ha estat en mans de perfils tècnics** -generalment contractats a l'empresa privada. Això suposa un problema, entre altres motius, perquè ofereix una visió molt limitada del potencial impacte dels SAD sobre els drets de la població, especialment entre els col·lectius vulnerabilitzats. Simplificant-ho molt: és complicat que un programador informàtic conegui la realitat d'una dona que ha migrat sola a Catalunya i tingui en compte les seves necessitats en construir un algorisme per a un servei públic.

Igualment, la manera de prevenir i combatre els riscos vinculats a l'automatització no pot ser definida únicament per persones allunyades d'aquests riscos per la seva situació de privilegi, poder o seguretat institucional.

Per tant, és fonamental **ampliar el nombre d'actors implicats en el disseny, el desenvolupament i l'aplicació d'aquests sistemes** i incloure-hi les veus de les comunitats directament afectades i de la societat civil organitzada, especialment les defensores de Drets Humans. Cal pensar també, sobretot quan estem davant de col·lectius vulnerabilitzats i/o en situacions precaritzades, que per fomentar la seva participació cal destinar-hi recursos i apostar per mesures flexibles i de conciliació

Per construir un SAD de forma més democràtica i horitzontal és convenient **consensuar amb un ampli ventall de veus qüestions com ara quin és l'objectiu del projecte, com pretenem aconseguir-ho, quin tipus de seguiment farem al sistema i com actuarem si aquest falla.**

Aquí algunes pistes que poden ser útils:

- 1** Desagregar les característiques de la població potencialment impactada per un SAD ens permetrà identificar veus representants d'aquestes comunitats.
- 2** Crear grups de treball diversos en què participin aquests col·lectius, a més de perfils tècnics, autoritats i altres representants de la societat civil.
- 3** Involucrar tant com sigui possible aquests grups de treball en totes les fases del cicle de vida algorítmic facilitarà un seguiment adequat del projecte.
- 4** Un cop el SAD estigui en funcionament, crear vies de denúncia i participació àgils per a altres veus fora de l'administració. Per exemple persones afectades per l'algorisme no tingudes en compte inicialment o investigadors independents que en puguin revisar el funcionament.

Amb aquesta finalitat, **plataformes digitals per a la participació ciutadana com Decidim.org poden resultar de gran utilitat**. Tot i això, si es fan servir aquestes eines, cal ser conscients de les esclatxes digitals que afecten moltes veïnes i veïns i que poden provocar una barrera per a la seva participació efectiva.

Una altra via d'acció és l'impuls de projectes més ambiciosos, com les assemblees o consells de ciutadania que proposa l'**Ada Lovelace Institute** en el seu projecte **Citizens Biometric Council**, al Regne Unit.

Aquesta iniciativa va reunir un grup de ciutadanes de diferents condicions per debatre sobre les tecnologies que recullen i processen dades biomètriques. El marc creat per aquesta organització britànica **per involucrar, mitjançant processos participatius, la ciutadania en l'ús de les dades personals** també pot servir d'inspiració.

Eines com les avaluacions d'impacte algorítmic serveixen per estudiar l'ús d'un SAD en un context determinat i entendre, categoritzar i respondre als danys o riscos potencials que el seu ús implica

b) Avaluar els impactes i adaptar els sistemes a cada context

Tal com hem vist, un SAD pot tenir un impacte directe a la vida de la ciutadania. Per tant, **és responsabilitat de les administracions conèixer i avaluar aquests efectes**, especialment si parlem de col·lectius que no tenen altres vies de participació en les decisions públiques (per exemple, les persones migrants sense dret al vot).

Un algoritme pot funcionar molt bé en el paper o en entorn de desenvolupament i, en canvi, resultar un desastre quan s'implementa al món real. Per tant, per conèixer els impactes reals **cal fugir de les anàlisis únicament centrades en el funcionament tècnic**.

A aquest efecte, en els darrers anys s'han popularitzat eines com les **avaluacions d'impacte algorítmic**. Un mecanisme que, tot i les limitacions, ens pot donar informació de gran utilitat per descartar, corregir o adaptar un projecte. Aquestes avaluacions serveixen per **estudiar l'ús d'un SAD en un context determinat i entendre, categoritzar i respondre als danys o riscos potencials que el seu ús implica**. El més ideal és portar-les a terme abans del desplegament del sistema i no quan ja està en ús, encara que el millor és fer avaluacions periòdiques d'aquest tipus durant tot el cicle de vida del sistema.

Tot i la seva relativa novetat en el camp de l'automatització, processos similars són habituals en altres àrees de l'acció pública com la mediambiental -com a requisit per aprovar un projecte urbanístic- o els drets humans -com a anàlisi prèvia a un projecte amb potencial impacte en aquest terreny-.

Igual que en el disseny i desenvolupament d'un SAD, per treure el màxim partit a aquestes eines és recomanable **implicar-hi les comunitats directament afectades pel seu ús**, encara que actualment sigui una pràctica poc usual. També es recomana fer públics els resultats d'aquestes avaluacions: se'n poden treure lliçons molt útils tant per a altres administracions com per a la societat civil en conjunt (investigadors, organitzacions, tecnòlegs, periodistes, etc.).

Una altra eina d'ús creixent són les **auditories algorísmiques**, un procés mitjançant el qual s'avalua un SDA segons una sèrie de criteris específics que poden relacionar-se amb els objectius marcats pel desenvolupador o amb qüestions com els biaixos i

discriminacions, la privadesa, la transparència i la rendició de comptes, el compliment normatiu o la seva empremta ecològica.

Tenir en compte **qui realitza aquestes auditories és clau per valorar-ne la utilitat**. Per exemple, les grans tecnològiques compten avui dia amb equips propis per auditar els sistemes que elles mateixes desenvolupen, mentre que **moltes consultores privades (Deloitte, Accenture, etc.) ofereixen aquests serveis previ pagament de tarifes molt elevades**. L'evident conflicte d'interessos generat o la cerca del lucre, respectivament, resten credibilitat i valor a aquestes anàlisis.

Per això és recomanable que aquestes auditories les facin tercers: des d'investigadors independents a òrgans reguladors amb veritable autonomia. La llista aquí pot incloure també activistes, organitzacions civils i periodistes -actors que de fet han destapat molts dels casos recents de mal ús de SAD al nostre entorn.

c) Una rendició de comptes real i efectiva

La transparència en un sentit ampli és un pilar bàsic per construir algorismes responsables i democràtics. **Només amb transparència no aconseguirem un ús respectuós amb els drets d'aquestes tecnologies, però aquesta sí que és una condició necessària per assolir aquest objectiu.**

No saber l'origen i el raonament darrere d'una decisió on ha intervingut un SAD pot dificultar o directament impedir la restitució dels drets de les persones afectades. El **dret a la impugnació d'una decisió pública s'ha de protegir**, també quan parlem d'algorismes i IA.

Per tant, les administracions han d'establir mecanismes efectius per rendir comptes cap a la ciutadania sobre els SAD que utilitza. Especialment si ho fa en àrees sensibles. "Especialment, quan està afectant drets bàsics, la informació sobre aquests sistemes ha d'estar disponible. **No només cal conèixer-ne**

l'existència i les característiques, sinó que la ciutadania ha de tenir una certa capacitat de discussió, de monitoratge i d'influència significativa", argumenta Sara Suárez-Gonzalo.

Malauradament, encara és freqüent que les autoritats no ofereixin ni tan sols informació essencial d'on, com i per què s'està automatitzant un procés. Per posar fi a aquesta anomalia un mecanisme cada cop més en ús són els **registres públics d'algoritmes**. Repositoris d'informació, generalment en format web, on es publiquen qüestions com el funcionament intern, els objectius, les dades usades i qui l'ha desenvolupat.⁷

Un model interessant en què fixar-se és **la plantilla creada per la xarxa europea de ciutats Eurocities**, on ja apareixen predefinits els camps d'informació bàsics que han de tenir aquests registres.

D'altra banda, **ser transparent no sols és oferir informació sinó també comunicar de la manera més clara possible**. Sánchez Monedero incideix en la necessitat de fer-ho de manera que la gent pugui "desafiar els resultats del sistema". "No pot fer falta un tècnic o professional especialitzat en dades per entendre aquesta informació", assenyala. Aquest investigador apunta també que cal democratitzar la visualització de dades perquè resulti accessible a tothom.

d) Enfrontar les desigualtats entre el Nord i el Sud global

Les administracions han de ser conscients que **el desenvolupament d'algoritmes no és aliè a les relacions de poder entre el Sud i el Nord global**. Les discriminacions generades per aquests sistemes s'entenen a partir de les desigualtats històriques patides per la població originària de països del sud. Així ho demostren els biaixos cap a la població racialitzada o d'origen migrant descrits en aquesta guia.

Però els desequilibris entre Nord i Sud van molt més enllà. El disseny i desenvolupament de **la tecnologia usada per aquests sistemes es fa principalment als Estats Units, la Xina i, en grau més baix, alguns països europeus**. Un desequilibri que ajuda a entendre perquè els patrons històrics que han definit les relacions entre les dues regions globals s'estan reproduint en aquest terreny.

Per això, cal canviar l'enfocament i incorporar els plantejaments de governança algorítmica que parteixen del Sud Global. Aquests incideixen, per exemple, en la urgència de **combatre l'extractivisme digital que avui practica la gran indústria tecnològica**, subcontractant una **força de treball precària (i invisibilitzada)** a països d'Àfrica, Àsia i Llatinoamèrica. Els moderadors de contingut i etiquetadores de dades entrenen eines d'intel·ligència artificial que es comercialitzen al Nord, com en el cas del mundial-

7

Encara que el seu potencial per oferir informació d'utilitat a la societat civil és gran, els registres oberts fins ara inclouen principalment sistemes innocus o ofereixen informació molt limitada sobre aquests. Deixen fora, per exemple, els resultats de les avaluacions d'impacte, una informació clau per conèixer els seus efectes sobre col·lectius vulnerables.

ment conegut ChatGPT, per millorar-ne el rendiment o evitar que difonguin discursos d'odi.

Tot i el seu desconeixement entre el gran públic, **aquestes treballadores són una part fonamental dels recents avenços tecnològics en aquest camp.** Si aspirem a un desenvolupament digital més just, aquestes innovacions s'han de fer amb condicions laborals dignes a tota la cadena de producció.

Un altre canvi de pes ha de venir per la via medi-ambiental. En les anàlisis sobre els **impactes de l'automatització, se sol deixar de banda la seva empremta ecològica.** Els models més recents d'IA requereixen unes enormes capacitats de computació per al seu entrenament i ús. Aquests processos no més es poden fer en **enormes centres de dades que consumeixen grans quantitats de recursos naturals per al seu bon funcionament.**

Aquestes infraestructures digitals tradicionalment s'han instal·lat als països del Nord, més a prop de les seues físiques de les grans empreses del sector i en llocs amb més abundància d'aigua. Però els darrers anys cada vegada més companyies les estan traslladant a països del Sud, on les sequeres i l'acceleració del canvi climàtic ja estan tenint efectes encara més catastròfics.

Quan les administracions desenvolupen o contracten un producte tecnològic han de tenir en compte

també aquesta problemàtica. Cal tenir presents els costos per al planeta que impliquen aquest tipus de projectes digitals i conèixer quines infraestructures i tecnologies fa servir l'empresa proveïdora i/o calcular la despesa d'aquells serveis desenvolupats en el si de l'administració.

Les discriminacions generades en el desenvolupament d'algorismes s'entenen a partir de les desigualtats històriques patides per la població originària de països del Sud global

e) Dotar-se dels recursos necessaris per al control i bon ús dels algorismes

L'externalització del disseny i desenvolupament no és l'única cessió al sector privat que les administracions han fet els darrers anys. Com dèiem, el treball **d'auditoria i control d'aquestes tecnologies també s'està deixant sovint a les mans de firmes privades**, a través, moltes vegades, de consultores amb interessos particulars allunyats del bé comú.

"No té sentit que l'administració estigui comprant sistemes fora, implementant-los a casa i demanant-los a altres, que també tenen interessos privats, que avaluin si funcionen bé o no", exemplifica Sara Suárez-Gonzalo.

Fins i tot els organismes públics creats per a aquesta vigilància, com l'**Agència Espanyola de Supervisió de la IA (AESIA), s'estan dissenyant així.** "Estem externalitzant-ho tot. Això és una cosa que no pot passar. **L'administració ha de tenir capacitat per vigilar aquests sistemes.** O com a mínim tenir internament el coneixement i el criteri per fer-ho", apunta Paula Boet.

Per recuperar part d'aquest terreny perdut és urgent que el sector públic, d'una banda, aprofiti les habilitats dels professionals que hi ha a dins. **Les administracions compten amb personal funcionari amb coneixements i experiència de gran utilitat, que no és usat per a aquest fi.**

Deixar en mans privades infraestructures rellevants per al funcionament d'una administració és un problema per l'amenaça a la seguretat que suposa aquesta pèrdua de sobirania

Però cal anar més enllà. Els responsables públics han d'invertir per dotar-se dels recursos necessaris per a un bon ús i control dels SAD. D'una banda, és clau formar els treballadors públics en el funcionament d'aquests sistemes i en com supervisar-los. Igualment, cal **impulsar òrgans realment independents que puguin vigilar** que s'estan complint tots els requisits necessaris. Una via que permetria, a més, involucrar més veus en aquests processos, assenyala Suárez-Gonzalo. "Si la supervisió d'aquests sistemes les portessin agències públiques independents, igual que hi ha per exemple els Consells de l'Audiovisual, es podria canalitzar una certa capacitat de la ciutadania per exercir un control i influir-ne l'ús".

f) Emancipació i empoderament davant del sector privat

Deixar en mans privades infraestructures cada cop més rellevants per al funcionament d'una administració, i per tant per a les vides i els drets de la ciutadania, és un problema per diversos motius. Un és **l'amenaça a la seguretat que suposa aquesta pèrdua de sobirania**.

Si el sistema falla, les eines per arreglar el desgavell no estan en mans de les administracions, sinó sota control privat. Una cosa semblant passa amb els protocols de supervisió i control esmentats al punt anterior.

Davant d'aquestes amenaces els organismes públics s'han de fer amb eines per augmentar-ne el control sobre els SAD. A hores d'ara, una iniciativa interessant d'algunes administracions és el **desenvolupament de clàusules estàndard de contractació que s'apliquen quan liciten un producte d'aquest tipus**. "Si les administracions són tan exigents amb tot el que compren, no pot ser que amb això no ho siguin", apunta Suárez-Gonzalo.

Boet proposa establir unes clàusules estàndard de contractació que incloguin salvaguardes específiques i més exigents per a les empreses. **Aquesta proposta feta per la Comissió Europea, Eurocities i diverses ciutats europees** més recull diverses recomanacions útils per a una contractació més responsable.

g) Restaurar els drets de les víctimes de l'automatització

La manera com les administracions poden restaurar els drets de persones afectades per l'automatització encara és un terreny poc explorat. Fins ara, **la reparació de les víctimes en alguns dels casos més sonats s'ha centrat en la compensació econòmica**⁸. Però aquesta via té evidents limitacions, relacionades amb el dany causat en l'àmbit emocional i moral a les víctimes.

Quan parlem de discriminacions (no només en l'àmbit de l'automatització), a l'Estat espanyol s'ofereix com a resposta principal a les víctimes l'àmbit penal. Perquè aquests mecanismes siguin realment útils, **les víctimes han de ser escoltades i tingudes en compte durant el procés penal, per definir com reparar el dany** que els han causat.

D'altra banda, els principis jurídics davant els quals ha de respondre un SAD quan provoca una vulneració de drets encara no estan clarament definits. Encara que ja hi ha algunes línies a seguir. Per exemple, el **Reglament General de Protecció de Dades (GDPR)** europeu protegeix el dret d'una ciutadana a saber en què es basa un algorisme per prendre una decisió que l'afecta i el dret a no ser objecte d'una decisió basada només en un procés automatitzat, excepte en casos puntuals.

La *Carta Catalana pels drets i les responsabilitats digitals*, per la seva banda, incideix també en el nostre dret a entendre aquests sistemes i a conèixer els paràmetres i les dades utilitzades en la seva construcció.

Per a una restauració completa de drets en aquest camp, pot ser interessant seguir el marc conceptual de la **Justícia Restaurativa (JR)**⁹. Una filosofia que en el context europeu s'ha fet servir fins ara com a complement a la justícia penal, organitzant per exemple trobades per al diàleg entre les víctimes i l'ofensor -en el cas que ens ocupa, l'Estat o l'administració de torn-.

Seguint en aquesta línia, les administracions podrien organitzar **trobades entre associacions de veïns o grups comunitaris i els responsables polítics (i tècnics) d'un SAD que hagi resultat perjudicial**. Amb el doble objectiu d'abordar les causes estructurals del dany causat i de transformació real; però també d'assumir responsabilitats per part de les institucions, creant garanties perquè un cas semblant no es repeteixi

8

Així ha passat per exemple amb l'escàndol de Robodebt a Austràlia, en què un sistema automatitzat de reclamació de deutes va assenyalar erròniament centenars de milers de receptors d'una ajuda pública. I amb el sistema contra el frau en les prestacions per cura de fills als Països Baixos (toeslagenaffaire), que va acusar falsament milers de famílies del país, la majoria d'origen migrant.

9

Aquesta filosofia s'allunya de la preocupació de la justícia tradicional a l'incompliment de les lleis, la culpa i el càstig associat; i se centra, en canvi, al dany causat a les víctimes, les necessitats creades i les obligacions que es generen a partir d'aquí.

Recursos útils

Definició de metodologies de treball i protocols per a la implementació de sistemes algorítmics. Ajuntament de Barcelona. 2023

https://ajuntament.barcelona.cat/digital/sites/default/files/def_metodologia_sist_algoritmics_cat_0.pdf

From Principles to Practice. An interdisciplinary framework to operationalise AI ethics. AI Ethics Impact Group (AEIG). 2020

https://www.bertelsmann-stiftung.de/fileadmin/files/BSt/Publikationen/GrauePublikationen/WKIO_2020_final.pdf

Data Justice in Practice: A Guide for Policymakers. Global Partnership on AI. 2021

<https://advancingdatajustice.org/data-justice-in-practice-guides/>

Automated Decision-Making Systems in the Public Sector – An Impact Assessment Tool for Public Authorities. AlgorithmWatch. 2021

<https://algorithmwatch.org/en/adms-impact-assessment-public-sector-algorithmwatch/>

Algorithmic impact assessment: a case study in healthcare. Ada Lovelace Institute. 2022

<https://www.adalovelaceinstitute.org/report/algorithmic-impact-assessment-case-study-healthcare/>

Fundamental Rights and Algorithms Impact Assessment (FRAIA). Government of the Netherlands. 2021.

<https://www.government.nl/documents/reports/2021/07/31/impact-assessment-fundamental-rights-and-algorithms>

Llistat de recursos utilitzats per a l'elaboració d'aquesta guia

1 Digital Rights Guide (Guia de Drets Digitals). UN-Habitat

Enllaç: [Digital Rights Guide](#)

2 Informe: Una introducción a la IA y la discriminación algorítmica para movimientos sociales

Autor: Algorace

Enllaç: [Introducción a la IA y Discriminación Algorítmica](#)

3 Informe: Auditoría externa e independiente del sistema Viogen

Autor: Fundación Éticas

Enllaç: [Auditoría Éticas Viogen](#)

4 Podcast: Capítulo 4: Viogen, el software contra la violencia machista

Autor: Digital Future Society

Enllaç: [Viogen: Software contra la violencia machista](#)

5 Informe: Algorithmic Accountability in the Public Sector

Autor: Open Government Partnership

Enllaç: [Algorithmic Accountability Public Sector](#)

6 Informe: Impact Assessment Public Sector

Autor: AlgorithmWatch

Enllaç: [ADMS Impact Assessment Public Sector](#)

7 Informe: Missing Links in Ai Governance

Autor: Mila i UNESCO

Enllaç: [Mila UNESCO](#)

8 Informe: Justicia Restaurativa en Casos de Odio y Discriminación

Autor: Institut de Drets Humans de Catalunya

Enllaç: [Justicia Restaurativa](#)

9 Govern Obert - Transparència algorítmica al sector públic

Autor: Pablo Jiménez Arandia - Generalitat de Catalunya

Enllaç: [Govern Obert 9](#)



www.lafede.cat

 [Lafede.cat](https://www.facebook.com/Lafede.cat)

 [Lafede_cat](https://twitter.com/Lafede_cat)

 [lafedecat](https://www.instagram.com/lafedecat)