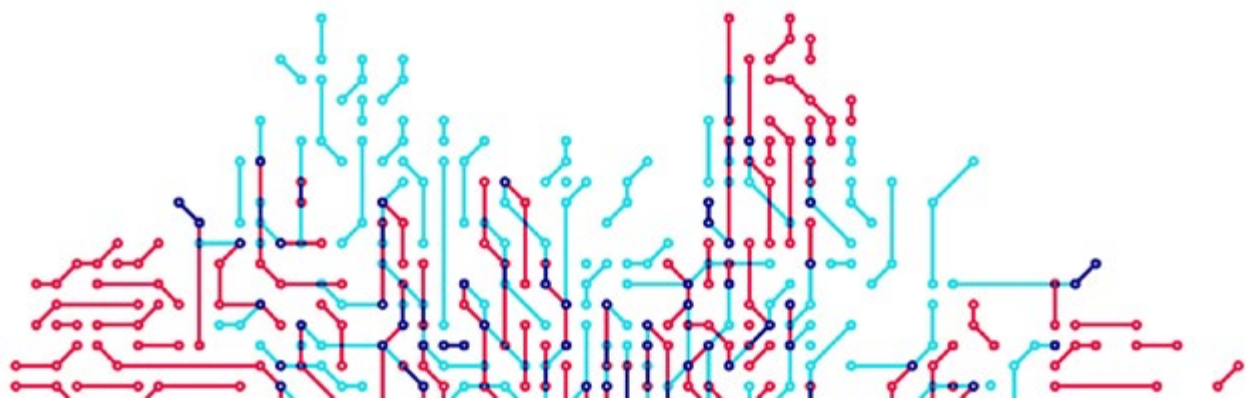


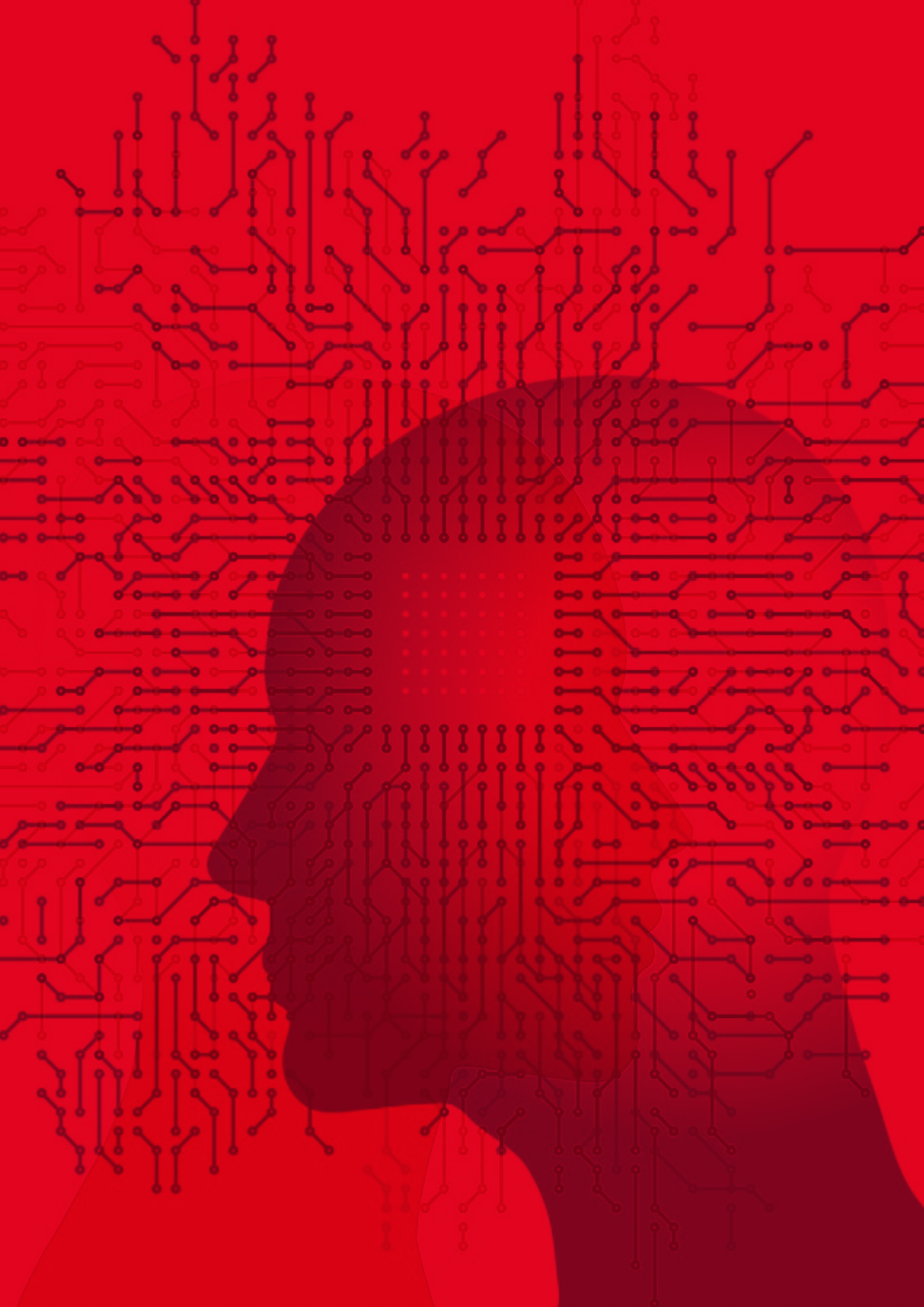
INTEL·LIGÈNCIA ARTIFICIAL

Decisions Automatitzades a Catalunya





Preàmbul	5
1. Protecció de dades i ètica en les decisions automatitzades	11
2. Ada en acció: treball de recerca a Catalunya	19
2.1. Com la intel·ligència artificial està canviant el món	20
2.2. Riscos dels algorismes de decisió automatitzada (ADA)	24
2.3. De decisió automatitzada o de suport a la decisió?	30
2.4. On s'apliquen els ADA a Catalunya? Més de 50 exemples per entendre...	33
2.4.1. Salut	34
2.4.2. Sistema judicial	42
2.4.3. Educació	46
2.4.4. Banca	51
2.4.5. Comerç	56
2.4.6. Social	62
2.4.7. Treball	67
2.4.8. Ciberseguretat	71
2.4.9. Comunicació	74
2.4.10. Visió per computador	77
2.5. L'ètica de la intel·ligència artificial	82
2.6. Dilemes pendents	111
2.7. Lectures recomanades per aprofundir-ne més	113
2.8. Agraïments	119
3. Protecció de dades i la IA	123
Recollida i tractament de dades de forma il·legítima	123
Dret a no ser objecte de decisions automatitzades	124
El principi de transparència	125
El principi de lleialtat	126
Principi de limitació de la finalitat	127
Anàlisi d'aplicacions dels ADA	129
Avaluació del risc de violència	129
Sistemes de <i>scoring</i>	130
Detecció de frau en les targetes de crèdit	131
Detecció de malalties	131
4. Recomanacions finals	133
Recomanacions per a les persones	133
Recomanacions per a les autoritats supervidores	136
5. Glossari	139





PREÀMBUL

Cap a una societat intel·ligent

Per M. Àngels Barbarà,
directora de l'Autoritat Catalana de Protecció de Dades

Els darrers deu anys s'ha produït un evolució tecnològica tan àmplia, profunda i a tanta velocitat que els éssers humans no som capaços de preveure com aquestes tecnologies s'incorporaran a les nostres vides.¹

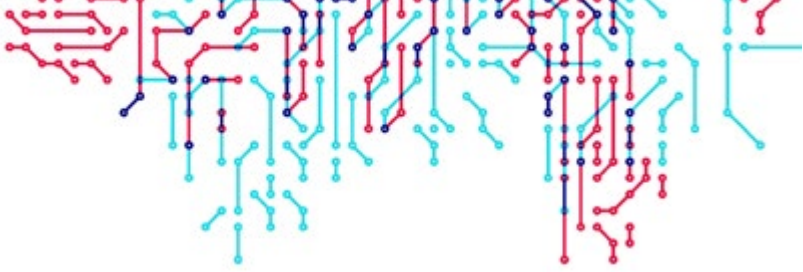
La innovació i el desenvolupament tecnològic fonamenten el seu creixement en les dades, personals i no personals. La convergència en el temps de diverses tecnologies –integrades per utilitzar de forma massiva les dades i generar valor– amplifica l'impacte en la vida de les persones. A més, l'explotació de la informació no esgota el seu valor en un ús, sinó que n'hi poden haver múltiples, de manera que es genera una cadena de valor que atorga un poder immens a aquells que governen la tecnologia i les dades.

Em refereixo a tecnologies com ara la computació en el núvol, que permet accés des de diferents llocs i gran capacitat d'emmagatzematge a baix cost; la Internet de les coses (IoT, en anglès), que recull informació de tots els entorns, en qualsevol format i moment; l'anàlisi de dades massives (*big data analytics*), que permet generar nou coneixement en salut, transport, educació, medi ambient, etc.; i, finalment, la intel·ligència artificial (IA), que engloba aquells sistemes que manifesten un comportament intel·ligent, capaços d'analitzar el seu entorn i actuar –amb cert grau d'autonomia– amb la finalitat d'assolir objectius específics.²

La IA entesa en un sentit ampli, la podem veure com una eina de suport a la presa de decisions en qualsevol àmbit. Aquest suport en la presa de decisions afecta, i de manera molt rellevant, les persones, ja que les dades i la tecnologia dibuixen perfils i patrons que serveixen per decidir i influir-hi directament.

1 “Hemos perdido la capacidad de censurar, de gobernar o de siquiera negociar con la tecnología, que es vista como una consecuencia lógica de nuestro progreso intelectual” (Joaquín David Rodríguez Álvarez. *La civilización ausente*. 2016, p.11).

2 Comunicació de la comissió al Parlament Europeu, al Consell Europeu, al consell, al Comitè Econòmic i Social Europeu i al Comitè de les regions: Intel·ligència artificial per Europa. COM/2018/237 final.



Canvis invisibles

Ens endinsem en un context digital que ens exposa a noves formes de vulnerabilitat a través del tractament de les nostres dades personals. Aquestes dades són utilitzades per les tecnologies emergents per generar valor i modifiquen, en gran mesura, la manera en què les persones i les organitzacions entenen la realitat que les envolta.

El canvi s'està produint d'una manera quasi invisible, entrant en la nostra vida personal, professional, social, política amb aparença de normalitat, de comoditat, de millora de les nostres vides sense impacte, sense conseqüències.

Aquest és un dels riscos més importants que es deriven de les tecnologies emergents: la seva intrusió silenciosa en els nostres drets i llibertats, transformant la societat en què vivim, d'una manera que sembla que no causi cap impacte. Aquesta societat modulada per les tecnologies, la que anomenem *societat digital*, serà diferent de la que hem conegut fins ara.

6 En la la societat digital –que es desenvolupa dins de la quarta revolució industrial– les tecnologies es fusionen mitjançant els mons físic, digital i biològic.³ La tecnologia s'integra en el nostre cos i es desenvolupa a partir de les noves dades que generem. Està dissenyada per aprofundir en el coneixement dels nostres impulsos, amb l'objectiu d'incidir –de manera directa, invisible i en temps real– en la presa de decisions de les persones i sobre les persones.

La seva invisibilitat i opacitat fa que no siguem conscients del risc que generen per als valors i principis que han sustentat la nostra societat fins ara: la dignitat, la llibertat, el lliure desenvolupament de la personalitat, la igualtat, reformulant aspectes bàsics de la democràcia i de l'estat de dret.

Nou model de societat

S'està gestant un nou model de societat sense que aquells que la integren se n'adonin. Per aquest motiu, des de l'Autoritat de Protecció de Dades –amb la mateixa orientació que el conjunt d'autoritats de control europees i de les principals institucions de garantia dels drets humans–, volem aportar claredat i comprensió en aquest sentit al context de la societat catalana. Igualment, volem ajudar les persones a defensar de manera més activa els seus drets i llibertats des del coneixement i la informació.

3 Klaus Schwab. *La cuarta revolución industrial*. Barcelona: Debate, 2016.

En aquesta ocasió, ens hem centrat en la intel·ligència artificial (IA) a Catalunya i, en concret, en els algorismes de decisió automatitzada (ADA) pel seu potencial d'impacte en la privacitat i en el conjunt dels drets i les llibertats de les persones.

Com no pot ser d'una altra manera, ho fem des de la regulació en matèria de protecció de dades. Des del maig del 2018 és de plena aplicació el Reglament General de Protecció de Dades (RGPD). Aquesta norma europea –d'aplicació directa al conjunt dels estats membres de la Unió Europea– dona forma jurídica a la voluntat de protegir els drets i les llibertats de les persones en el context del canvi de model de societat cap a la societat digital. I és un impuls europeu a les tecnologies emergents com a eina per sortir de la crisi conjuntural en què estava immersa Europa.

L'RGPD estableix un nou model de garantia dels drets i les llibertats de les persones basat en la responsabilitat i el compromís de les entitats amb els drets de les persones, i en un enfocament sobre el risc en què cada entitat ha de valorar com afecta els seus usuaris el tractament de dades concret que duu a terme.

Aquest nou marc europeu es basa en els principis tradicionals de protecció de les dades (lleialtat, licitud, transparència, limitació de la finalitat, minimització, exactitud, limitació dels terminis de conservació, seguretat) i en uns requisits de legitimació dels tractaments molt similars als existents, però canvia radicalment la forma d'afrontar-ne el compliment, justament, sobre la base dels criteris, abans esmentats, de responsabilitat proactiva i enfocament sobre el risc.

El nou model obliga les organitzacions a plantejar-se quina és la manera més adequada de fer els tractaments per reduir-ne l'impacte sobre els drets de les persones i, alhora, adaptar-se de manera integrada a la idiosincràsia de cada organització.

Dimensió ètica i social

La dimensió ètica i social ha pres, en els darrers anys, una rellevància especial en el debat entorn de l'evolució del dret a la protecció de dades i de com aquesta dimensió ha de regir la innovació tecnològica i protegir els valors, els principis i els drets que són inherents a la nostra societat i que defineixen la tradició europea.

El context de la IA i dels ADA –d'innovació i disrupció continua– exigeix a les entitats que volen utilitzar aquestes eines fer un pas més en el compliment. Ja no hi ha prou amb complir les obligacions d'una manera mecànica i formal. Cal dissenyar, desenvolupar i utilitzar aquestes eines incorporant-hi una dimensió ètica i social. L'ús intensiu de les dades –majoritàriament, dades personals– obliga, des d'una perspectiva jurídica i ètica. La IA no pot ser un fi en si mateix, sinó un mitjà per assolir una transformació de la



societat que se centri en la persona. Però complir aquesta premissa no és fàcil, ja que la IA s'incrusta en les tecnologies que fem servir de forma rutinària i invisible.

La IA serveix a finalitats molt diverses i té el potencial de millorar el benestar de les persones. Com indica l'OCDE –en una recomanació recent–⁴ la IA té el potencial de contribuir a una activitat econòmica global i sostenible, d'augmentar la innovació i la productivitat.

En aquest entorn, el dret a la protecció de dades personals s'ha convertit en un element clau per protegir les persones davant del potencial ús abusiu de la informació. Vivim en una societat en què la identitat digital de cada individu, respongui o no a la realitat, és la que es pren com a base per a la personalització de productes i serveis, per influir en les decisions de les persones i per decidir sobre elles. Però, tal com remarca el grup d'experts en ètica creat pel supervisor europeu, la protecció de dades no és un tema tècnic o legal, és una qüestió profundament humana.

La tecnologia millora la nostra qualitat de vida, ens la fa més còmode, però no pot ser a canvi que els nostres drets es vegin limitats sense mesura; per exemple, mitjançant la manipulació i distorsió de la nostra percepció de la realitat o “vigilant les nostres emocions”.⁵ Com exposa l'informe *Automating Society*, elaborat per Algorithm Watch, cal fer-se altres preguntes, com ara qui decideix quan s'ha d'utilitzar IA, amb quines finalitats, i/o qui i com la desenvolupa.

Són qüestions bàsiques, i el Parlament Europeu així ho indicava quan deia que en IA cal invertir no només en tecnologia i innovació, sinó també en els àmbits socials, ètics i de responsabilitat. Hi ha elements que cal integrar en el disseny, desenvolupament i aplicació de les tecnologies. Aquest elements han d'anar més enllà de les obligacions que les normes regulen, han de tenir una projecció universal. És a dir, han de ser comprensibles i aplicables de manera global.

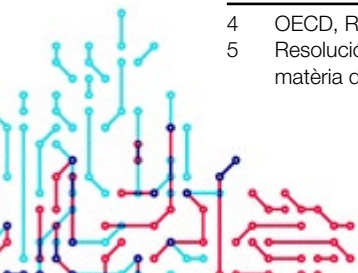
IA segura i fiable

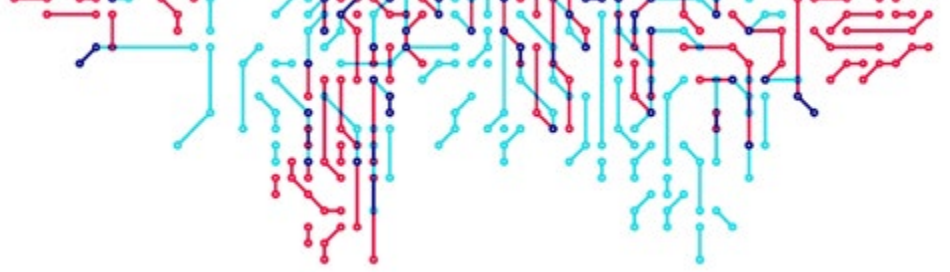
Davant d'aquestes enormes capacitats de les tecnologies emergents, hem de reflexionar sobre si el camí que han pres la innovació, el seu disseny i el seu ús, es pot fer d'una manera menys intrusiva per a la vida de les persones.

Hem de promoure un disseny que no limiti la llibertat de l'individu per prendre decisions i que no permeti que aquells que ostenten el poder sobre la tecnologia puguin controlar ni dirigir el comportament de les persones. I això només ho aconseguirem si la societat entén com funcionen els algorismes i pot fer plantejaments crítics respecte de les

4 OECD, Recommendation of the Council on Artificial Intelligence (OECD/Legal/0449), adopted May 21, 2019.

5 Resolució 2018/2088(INI) del Parlament Europeu, de 12 de febrer de 2019, sobre una política industrial global europea en matèria d'intel·ligència artificial i robòtica.





tecnologies que s'utilitzen. Cal dissenyar una IA segura i fiable, en què s'integrin els principis de la transparència, l'explicabilitat, la seguretat, l'auditabilitat i la responsabilitat.

L'octubre del 2018, en la 40a Conferència Internacional d'Autoritats de Protecció de Dades, es va aprovar la Declaració sobre ètica i protecció de dades en intel·ligència artificial. L'abril del 2019, el grup independent d'experts d'alt nivell sobre intel·ligència artificial de la Comissió Europea va fer pública les seves directrius ètiques per a una IA fiable. I el maig del 2019, l'OCDE publicava les seves recomanacions sobre IA. Finalment, en el decurs de la 41a Conferència Internacional d'Autoritats de Protecció de Dades, el Consell d'Europa anunciava l'inici dels treballs per a l'elaboració d'un marc legal en IA.

En aquest context, des de l'Autoritat Catalana de Protecció de Dades, conscients que Catalunya és un important motor de desenvolupament tecnològic, i partint del treball que hem realitzat en aquests darrers anys en matèria d'ètica i protecció de dades, hem iniciat un projecte específic destinat a identificar com s'està utilitzant la IA a Catalunya, i com s'incorpora el vessant ètic en aquests entorns.

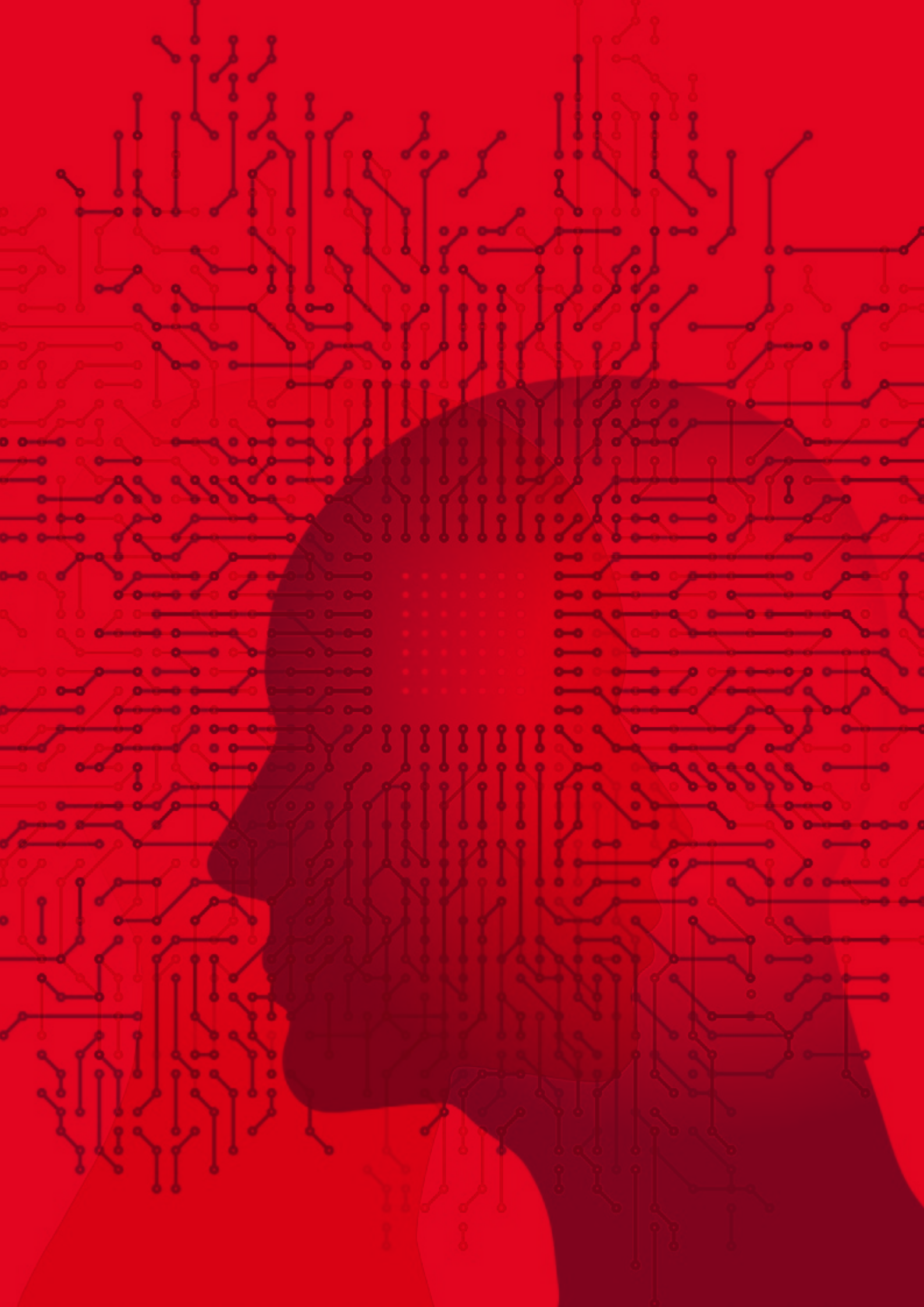
Aquest projecte s'ha focalitzat en un ús concret de la IA: els algorismes de decisió automatitzada (ADA) i com afecten els valors que defineixen la tradició europea. Forma part d'aquest informe un recull d'aplicacions dels ADA fet per la periodista Karma Peiró, especialitzada en tecnologies de la informació i la comunicació (TIC). Vull assenyalar que, tot i que tot l'informe té un caràcter divulgatiu, aquest caràcter és especialment marcat en la descripció de les aplicacions dels ADA. Amb una voluntat de fer aquestes tecnologies més entenedores, defugim dels detalls més tècnics del seu disseny o la seva aplicació.

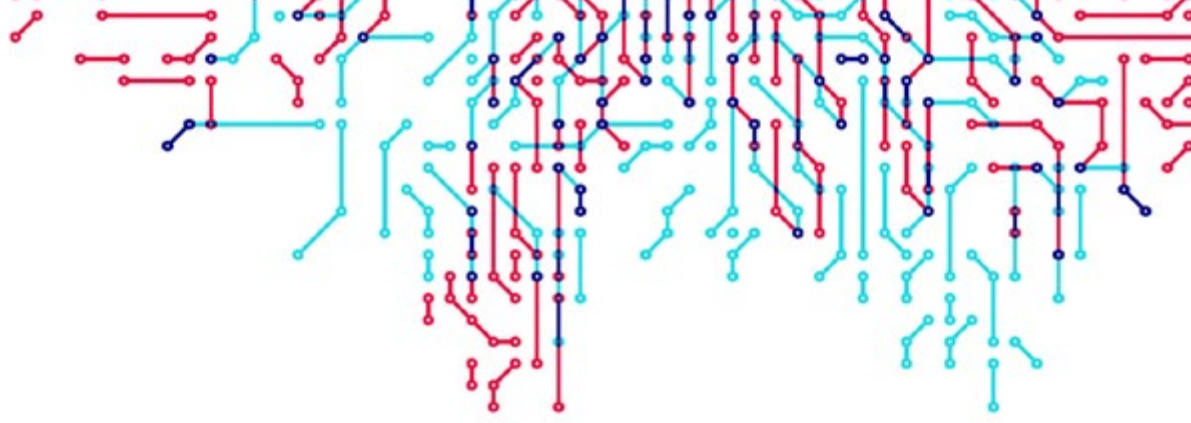
9

També cal identificar les possibles desviacions i discriminacions derivades de l'ús de les tecnologies emergents per evitar que es perpetuïn discriminacions "històriques" basades en la raça o el sexe, o que se'n generin de noves, no perceptibles, a partir de la creació de grups sobre la base de patrons i correlacions algorísmiques.

L'enfocament ètic d'Europa respecte de la IA reforça la confiança de les persones i ha de permetre generar un avantatge competitiu per a les empreses europees d'IA.⁶ Des de l'Autoritat Catalana de Protecció de Dades, hi volem contribuir amb l'elaboració dels principis que han de governar el disseny, el desenvolupament i l'ús de la IA. Aquest informe que es presenta només és el començament del llarg camí que ens queda per recórrer en la garantia dels drets i les llibertats de les persones en un món en constant evolució.

⁶ Comunicació de la comissió al Parlament Europeu, al Consell Europeu, al consell, al Comitè Econòmic i Social Europeu i al Comitè de les regions: Generar confiança en la intel·ligència artificial centrada en l'ésser humà. COM/2019/168 final.



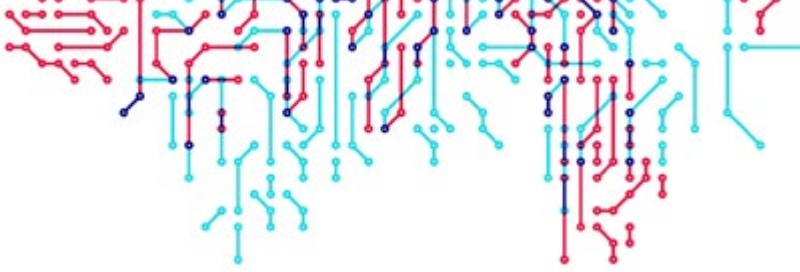


1. PROTECCIÓ DE DADES I ÈTICA EN LES DECISIONS AUTOMATITZADES

Les nostres vides estan cada cop més condicionades per la interacció que tenim amb la tecnologia. En el centre d'aquest ecosistema hi ha les nostres dades. Instagram té les nostres fotos, Google el nostre correu electrònic i documents, Facebook coneix el nostre cercle social i Twitter els nostres interessos. En aquests casos, som nosaltres mateixos que generem les dades i les posem a disposició de multitud d'organitzacions. Altres vegades, les dades es generen i recullen sense que en siguem conscients. Per exemple, el rastre que deixem amb els dispositius mòbils (amb sensors de tota mena), amb la navegació a Internet o amb les compres amb targeta.

Aquesta gran quantitat de dades personals disponibles accentua la incidència de la tecnologia sobre la vida de les persones. En altres paraules, permet un altíssim grau de personalització en els serveis que ens ofereixen. Això és veu clarament en el cas de les notícies. Abans, podíem triar un diari o un telenotícies en funció de la seva línia editorial. Ara hi ha aplicacions de mòbil, recomanadors de notícies o xarxes socials (Google News, Facebook, Twitter) que adapten la informació al perfil de cada persona. L'objectiu no és altre que millorar l'experiència dels usuaris mostrant-los només aquella informació que és del seu interès, per tal de fidelitzar-los.

Aquest altíssim grau de personalització pot tenir molts avantatges, però també riscos pel que fa als drets i llibertats. Els riscos directament relacionats amb el tractament de dades personals són força obvis: un incident de seguretat pot posar en perill la confidencialitat de les nostres dades, es pot fer un ús inadequat de la informació, etc. Un altre tipus de riscos, no tan obvis, són els que provenen de l'aplicació concreta que es



fa de les dades. En el cas de les notícies personalitzades, hi ha el risc que la informació que ens mostren estigui massa restringida, cosa que limita el nostre camp de visió de la realitat. És l'anomenat *filtre bombolla*. A més, cal tenir present que aquell que controla el que veus, pot controlar el que penses, sents i fas.

En aquest context, la capacitat dels algorismes de prendre decisions de manera automatitzada ha esdevingut essencial. Un portal de notícies podria personalitzar manualment el contingut per a un petit nombre de persones, però aquesta tasca és inabordable quan els usuaris es compten per milers. Avui en dia, són els algorismes que analitzen el perfil d'un usuari i decideixen quins són els continguts més adients per a ell.

Hem posat l'exemple del cas d'un portal de notícies, però aquest és només un exemple d'entre els milers d'aplicacions que els algorismes de decisió automatitzada tenen en multitud de camps. Com a part d'aquest informe s'inclou un estudi sobre l'ús dels algorismes de decisió automatitzada on s'expliquen multitud d'aplicacions en sanitat, assumptes socials, treball, banca, etc.

Aquestes aplicacions deixen clar que els algorismes de decisió automatitzada poden tenir una incidència gran sobre la ciutadania: diagnosi incorrecta d'una malaltia, denegació d'un ajut social, ser rebutjat en un procés de selecció de personal, denegació d'un crèdit per part d'un banc, etc.

12

Aquest informe

L'objectiu d'aquest informe és conscienciar la ciutadania i dotar-la del coneixement necessari perquè faci un ús responsable de les seves dades personals. Això s'aconsegueix en dos fases.

La primera fase es materialitza en un estudi de recerca sobre l'ús dels algorismes de decisió automatitzada a Catalunya, fet per la periodista Karma Peiró, especialitzada en tecnologies de la informació i la comunicació. L'objectiu és informar la ciutadania, en to divulgatiu, sobre els avantatges i els riscos que comporten.

En la segona fase, entenent que la major part d'aquestes d'aplicacions fan ús de dades personals, reflexionem sobre les principals friccions entre els algorismes de decisió automatitzada (i, per extensió, de la intel·ligència artificial), a més dels principis de protecció de dades personals. Fruit de les reflexions anteriors, proposem una sèrie de recomanacions finals.

Algorismes de decisió automatitzada en l'RGPD

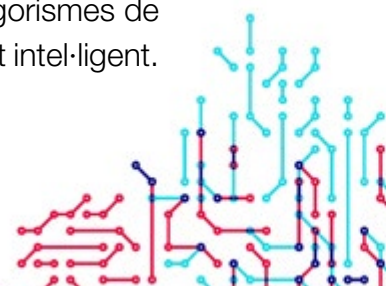
Des del maig del 2018, el Reglament general de protecció de dades (RGPD) és la norma europea bàsica que regula el tractament de dades personals. La norma anterior tenia un caire massa formalista i no va poder donar resposta als importants reptes que han acompanyat el desenvolupament tecnològic. Un enfocament basat en el risc, d'una banda, i el principi de responsabilitat proactiva, de l'altra, són dues novetats importants de l'RGPD; es busca que s'adaptin a l'RGPD perquè pugui encarar els reptes derivats del desenvolupament tecnològic amb prou garanties. Amb l'enfocament basat en el risc, es vol que les mesures que pren el responsable del tractament per protegir els drets i les llibertats de les persones siguin proporcionals als riscos del tractament. La responsabilitat proactiva exigeix que el responsable del tractament sigui capaç de demostrar que es fa d'acord amb l'RGPD. És a dir, no n'hi ha prou amb complir amb l'RGPD, cal poder demostrar-ho.

L'RGPD reforça els drets que tenim les persones respecte al tractament de les nostres dades; l'anomenada autodeterminació informativa. Ara bé, perquè les persones puguem exercir els drets que tenim sobre les nostres dades d'una manera apropiada, primer cal que coneguem quins són aquests drets i, després, que siguem conscient de la necessitat de protegir-los. Per exemple, un dels principals errors que cometem és que donem el consentiment per a l'ús de les nostres dades de manera massa generosa o, fins i tot, despreocupada. Donant sempre el consentiment per a l'ús de les nostres dades reduïm sensiblement la protecció que l'RGPD ens proporciona.

La redacció de l'RGPD busca ser al més general possible. És per això que l'RGPD no esmenta tractaments de dades concrets: es basa en uns principis generals. L'única referència a un tipus concret de tractament és en l'article 22, que parla de la presa de decisions automatitzades. La gran rellevància que han adquirit aquest tipus d'algorismes i les fortes implicacions que poden tenir sobre les persones fan que sigui necessari donar proteccions específiques.

Algorismes de decisió automatitzada i intel·ligència artificial

Els algorismes de decisió automatitzada i la intel·ligència artificial són conceptes fortament relacionats. És cert que és possible dissenyar un algorisme que prengui decisions automatitzades de manera poc (o gens) intel·ligent. Per exemple, un algorisme per triar el candidat més adient per a un lloc de treball podria decidir de manera aleatòria (sense avaluar cap dels candidats). Ara bé, hi hauria una gran probabilitat que la decisió presa per un algorisme d'aquestes característiques no fos òptima. Perquè els algorismes de decisió automatitzada siguin realment útils cal que tinguin un comportament intel·ligent.



La IA engloba els algorismes la finalitat dels quals és dotar els ordinadors d'un comportament intel·ligent. La definició de *comportament intel·ligent* no és clara del tot. Avui, la intel·ligència artificial ha superat les capacitats humanes en moltes tasques específiques, però encara no pot fer moltes de les funcions dels humans. És a dir, és capaç de guanyar al millor jugador d'escacs del món, però seria incapaç de dur a terme les tasques ordinàries d'una persona de mitjana edat, com ara tenir una reunió amb la mestra dels fills. Les persones ens movem en diferents contextos i ambients i decidim sobre multitud de temes cada dia perquè hem acumulat un coneixement previ que les màquines encara no tenen. Això també es pot explicar de la manera següent: els humans fem moltes coses amb poca precisió; les màquines, molt poques amb molta precisió.

Hi ha diversos nivells de comportament intel·ligent: des dels sistemes més bàsics, que es limiten a aplicar un conjunt de regles que han estat prefixades, fins a sistemes complexos d'aprenentatge profund, que estan inspirats en les xarxes neuronals biològiques. Fixar quin és el límit a partir del qual un algorisme és intel·ligent és una qüestió controvertida i que no té una rellevància especial per a la finalitat dels algorismes de decisió automatitzada —és a dir, prendre decisions sense intervenció humana. Per tant, en aquest informe farem servir els termes *intel·ligència artificial* i *algorismes de decisió automatitzada* de manera intercanviable. L'única exigència que posem per qualificar com a intel·ligència artificial és que l'algorisme analitzi l'escenari actual i prengui la decisió en funció d'aquesta anàlisi.

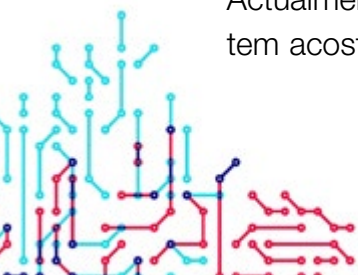
14

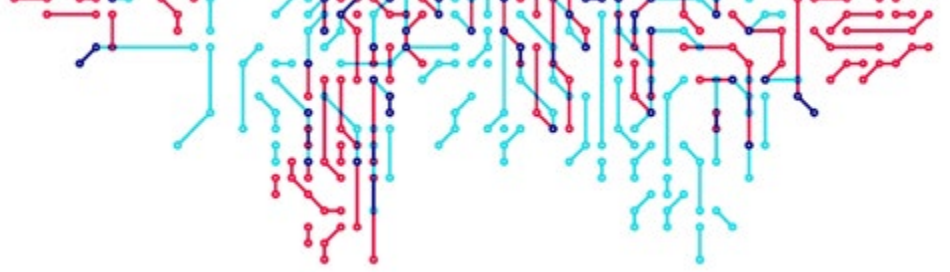
Decidir els límits i fixar què és ètic

Donades les fortes implicacions que els algorismes de decisió automatitzada i la intel·ligència artificial tenen sobre les persones, cal fixar el límit del que és raonable. Òbviament, la llei és un instrument bàsic (particularment, l'RGPD com a principal regulació europea en protecció de dades). Ara bé, no és l'únic instrument. L'ètica té també un paper essencial. Per exemple, a l'hora de definir la llei i de guiar la seva aplicació.

La determinació del que és ètic i del que no ho és no és una qüestió que pugui resoldre l'Autoritat Catalana de Protecció de Dades (APDCAT) o un grup d'acadèmics. L'ètica és una contracte social, i és la societat qui ha de fixar què és ètic. Perquè aquest debat es produeixi d'una manera satisfactòria, cal dotar la societat dels coneixements bàsics necessaris. En aquesta tasca, aquest informe pot ser de gran utilitat: donant a conèixer les aplicacions dels algorismes de decisió automatitzada a Catalunya i analitzant-ne les implicacions.

Actualment, la societat té una doble visió respecte a la IA i els ADA. D'una banda, estem acostumats a les millores que comporten els avenços tecnològics, i els adoptem





ràpidament; per això es pot dir que no fer ús de les capacitats de la intel·ligència artificial en la detecció de malalties, per exemple, no seria ètic. D'altra banda, cal no ser ingenus i dir que tot s'hi val; ser massa permissiu pot donar lloc a usos que interfereixin d'una manera excessiva amb els drets i les llibertats de les persones.

El gran potencial de la intel·ligència artificial és un condicionant a l'hora de fixar-ne els límits. Es considera una tecnologia disruptiva capaç de canviar les relacions de poder entre els estats, i això fa que ningú no es vulgui quedar enrere. Ja es parla de la nova carrera espacial entre superpotències, aquest cop centrada en el desenvolupament de la IA. Els Estats Units –fent valer la seva posició com a líder en desenvolupament tecnològic– encara té una posició dominant. La Xina s'hi ha incorporat més recentment, amb un pla d'inversions multimilionari que, en una dècada, vol, primer, igualar en coneixement als Estats Units i, després, superar-lo. El Japó és líder en robòtica. Alguns països europeus busquen una parcel·la pròpia en el món de la intel·ligència artificial, però amb pressupostos més modestos. És en el desenvolupament d'una IA ètica (també des del punt de vista de la protecció de dades) on Europa té la davantera respecte als altres actors.

Protecció de dades i IA

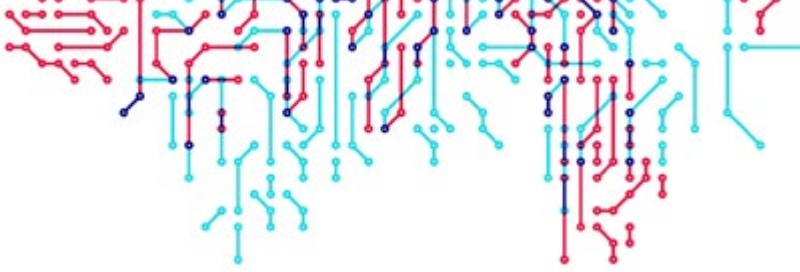
15

La gran quantitat de dades disponibles ha estat un dels pilars fonamentals per al desenvolupament de la IA que hem viscut en la darrera dècada. És per això que l'efecte de la protecció de dades sobre la intel·ligència artificial és un tema controvertit.

Alguns sectors critiquen les garanties que ofereix l'RGPD, perquè entenen que són un fre per al desenvolupament de la IA a Europa. Aquests actors advoquen per una regulació més laxa, que faciliti la reutilització de dades.

També hi ha la visió contrària, des de la qual les limitacions que imposa l'RGPD són un incentiu per al desenvolupament d'una intel·ligència artificial que faci un ús més raonable i més efectiu de les dades disponibles; per exemple, fomentant el desenvolupament d'algorismes més sofisticats que aprenen a partir de conjunts limitats de dades, com ho fan les persones. Els escacs (tot i no tractar amb dades personals) en són un cas molt clar: els algorismes més sofisticats s'han entrenat jugant milers de milions de partides, mentre que els grans mestres humans n'han jugat a tot estirar uns quants milers.

El punt de vista de l'APDCAT, com no pot ser d'una altra manera, és que l'ús de la intel·ligència artificial ha de ser compatible amb l'RGPD.



La societat que volem construir

Sense restar importància a les crítiques a l'RGPD, mirar el passat ens pot ajudar a reflexionar a l'hora de decidir el tipus de societat que volem construir. La primera menció al dret a la privacitat, en el context de protegir les persones de les intromissions dels avenços tecnològics de l'època, data del 1980:⁷

“Les fotografies instantànies i l'empresa periodística han envaït els recintes sagrats de la vida privada i domèstica; i nombrosos dispositius mecànics amenacen de fer bona la predicció que el que s'ha parlat a cau d'orella en la cambra més retirada, ho pregonaran des dels terrats.”

Salvant les distàncies, estem davant d'un cas semblant: decidir si deixem que la tecnologia faci un ús indiscriminat de les dades personals o si hi posem límits. Indubtablement, les noves tecnologies són molt més intrusives que una fotografia, en el sentit que tenen la capacitat de generar un perfil de la persona molt més extens. Si fer una fotografia d'una persona en el seu àmbit privat pot violar el dret a la intimitat, no hauríem de limitar també l'ús de les dades que generem? La Declaració Universal dels Drets Humans mateixa, en l'article 12,⁸ reconeix el dret a la privacitat.

16

Al món actual, podem veure diferents sensibilitats pel que fa a la relació entre intel·ligència artificial i el respecte als drets i les llibertats de les persones. En particular, respecte a la protecció de dades personals tenim aquests tres escenaris:

Europa

Disposa de la legislació de protecció de dades personals més avançada. Llevat d'algunes excepcions (com ara obligacions legals, interès públic, interès legítim), la legislació europea promou l'autodeterminació informativa; és a dir, el dret de cada persona a decidir l'ús que s'ha de fer de les seves dades.

Com hem comentat anteriorment, la normativa de protecció de dades europea presenta algunes friccions amb la intel·ligència artificial. (*Vegeu el capítol “Intel·ligència artificial i protecció de dades”.*)

⁷ Warren i Brandeis. “The Right to Privacy”. *Harvard Law Review* (1980).

⁸ Declaració Universal dels Drets Humans (https://www.ohchr.org/en/udhr/documents/udhr_translations/cln.pdf).

Xina

La protecció de les dades personal és un tema de poc desenvolupat a la Xina. 'Hi ha un estàndard de protecció de dades personals, però no és 'de compliment obligat.

La Xina fa un ús de la intel·ligència artificial molt invasiu. Per exemple, el sistema de crèdits socials, actualment en implantació, utilitza vigilància electrònica massiva per controlar el comportament dels ciutadans, assignar una puntuació a cada individu, i premiar-los o penalitzar-los. Algunes de les implicacions que pot tenir aquest sistema per als ciutadans xinesos són la prohibició de viatjar, l'exclusió de certes escoles, la repressió a minories religioses i la humiliació pública.

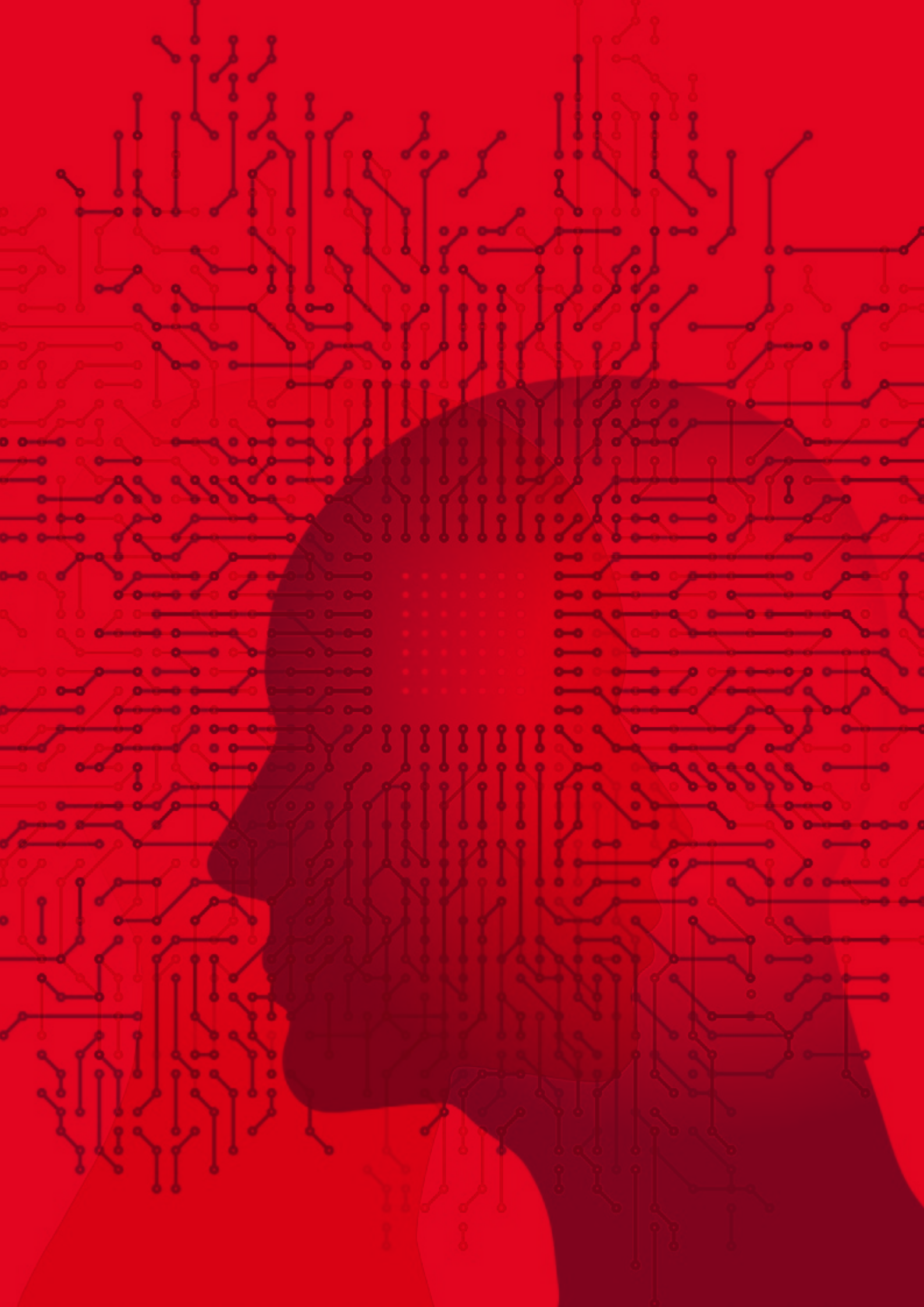
Estats Units

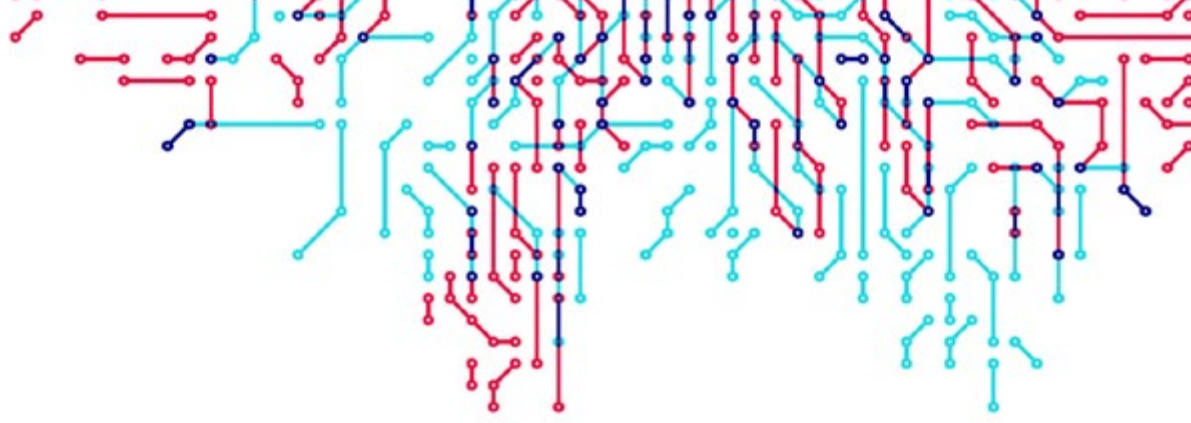
Als Estats Units no hi ha una llei de protecció de dades. El seu marc legal està format per centenars de lleis (estatals i federals) que en tracten aspectes específics.

Una diferència remarcable entre Europa i els Estats Units té a veure amb el consentiment per a l'ús de dades personals *opt-in* contra *opt-out*. Això vol dir que a Europa el consentiment s'ha de recollir de forma explícita, informada i per a una finalitat determinada. Als Estats Units, el consentiment és implícit: si una persona no vol que una organització tracti les seves dades, ho ha de demanar explícitament.

Podem dir que els Estats Units fan un ús pragmàtic de la intel·ligència artificial, aprofitant tota la informació que està disponible.

Per tant, la relació entre protecció de dades personals i intel·ligència artificial és un tema que pot enfocar-se de diverses maneres. Correspon a cada societat decidir-ne els límits i fixar què és ètic. Ara bé, això no és possible si la societat desconeix els usos que se'n fan i les implicacions que tenen sobre les persones. El contingut dels capítols següents busca donar resposta a aquestes qüestions.





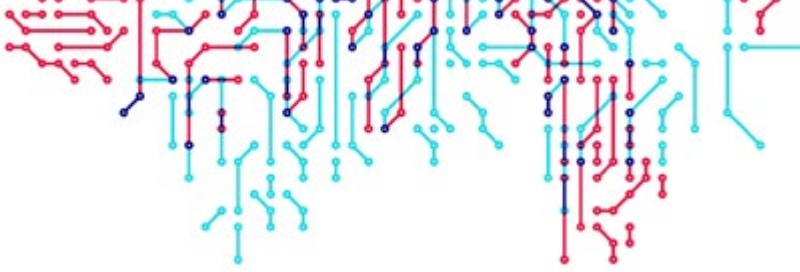
2. ADA EN ACCIÓ: TREBALL DE RECERCA A CATALUNYA

Per Karma Peiró

19

Karma Peiró és periodista especialitzada en les tecnologies de la informació i la comunicació (TIC) des del 1995. Conferenciant en seminaris i debats relacionats amb les dades, la transparència de la informació i l'ètica en la intel·ligència artificial. Ha col·laborat en l'informe *Automating Society*, d'Algorithm Watch, una radiografia sobre l'aplicació dels algorismes de decisió automatitzada a Europa.

És membre del Consell Assessor de Transparència de l'Ajuntament de Barcelona; del Consell Assessor de la Iniciativa Barcelona Open Data; i del Consell per a la Governança de la Informació i els Arxius. Més informació a: <http://karmapeiro.cat>.



2.1. Com la intel·ligència artificial està canviant el món

El 2011, el popular concurs de televisió nord-americà *Jeopardy* –basat en preguntes i respostes de cultura general– acumulava un pot de cinc milions de dòlars. Optaven al gran premi dos dels millors jugadors en dècades: Ken Jennings (guanyador en 74 ocasions) i Brad Rutter (que acumulava uns guanys de més de tres milions de dòlars). Tots dos van acceptar el repte de lluitar, aquesta vegada, amb un oponent molt atípic que seria en una habitació contigua al plató, a causa del soroll del sistema de ventilació que el feia funcionar. Es tractava del sistema intel·ligent d'IBM Watson,⁹ una màquina capaç d'entendre preguntes complexes, respondre-les en temps real i aprendre de cada interacció. S'imaginem qui va guanyar?

De manera molt simple, es podria dir que la intel·ligència artificial (IA) és una part de la informàtica que potencia que les màquines funcionin i reaccionin com els humans. És a dir, que puguin raonar, aprendre i actuar de manera intel·ligent. Per aconseguir aquest objectiu, la IA desenvolupa algorismes que prediuen i prenen decisions de manera automatitzada.

20

La idea no és nova. El filòsof i escriptor Ramon Llull¹⁰ (1232-1316) va dedicar la seva vida a l'*ars machina*, una màquina que servia per fer proves lògiques i facilitar el raonament. Més recentment, el matemàtic Alan Turing va crear el *test de Turing*¹¹ (1950), que plantejava la possibilitat que un ordinador pogués “pensar”. Cosa que quedava demostrada en el moment que un persona no podia distingir si qui li parlava era una màquina o un humà. A Turing també se li concedeix el mèrit d'haver salvat milions de vides gràcies a trencar els codis matemàtics de la màquina d'encriptació de missatges Enigma¹² dels nazis. I així va escurçar la Segona Guerra Mundial en dos anys.

Hi ha moltes altres fites històriques que han anat apropant la intel·ligència artificial al present, tot i que, fins avui, només eren casos excepcionals. És ara, que sense adonar-nos-en, la tenim més a prop que mai.

Algorismes que ens acompanyen

Els algorismes són el conjunt d'ordres que cal seguir per resoldre un problema o una tasca. Ja van ser aplicats en les primeres civilitzacions de la humanitat. I nosaltres mateixos, en el nostre cervell, quan reaccionem a una situació imprevista, mentalment em-

9 “Watson” ([https://ca.wikipedia.org/wiki/Watson_\(intel%2%B7lig%C3%A8ncia_artificial\)](https://ca.wikipedia.org/wiki/Watson_(intel%2%B7lig%C3%A8ncia_artificial)))

10 Ramon Llull (<https://www.cccb.org/es/exposiciones/ficha/la-maquina-de-pensar/223672>).

11 “Test de Turing” (https://ca.wikipedia.org/wiki/Test_de_Turing).

12 “Enigma” (https://ca.wikipedia.org/wiki/M%C3%A0quina_Enigma).

prem algorismes per resoldre-la. El mateix fem quan cuinem un plat, seguint els minuts de cocció de cada ingredient que marca la recepta.

Els algorismes que trobem als ordinadors ens fan la vida més fàcil i senzilla. Quan busquem la millor ruta per arribar a un lloc, quan demanem un taxi des d'una *app* de mòbil, quan rebem altres recomanacions després de comprar un llibre, quan veiem la nostra sèrie preferida en una plataforma en línia, quan cerquem habitatge en un barri nou, quan accedim a una oferta de treball, quan demanem un crèdit al banc, quan entrem en una llista d'espera per a una operació quirúrgica, quan encarreguem una pizza a domicili o quan contractem l'assegurança de la llar, els algorismes decideixen i ens donen les millors solucions o opcions. Què faríem sense ells? Prescindir-ne seria una enorme pèrdua de temps i d'eficiència!

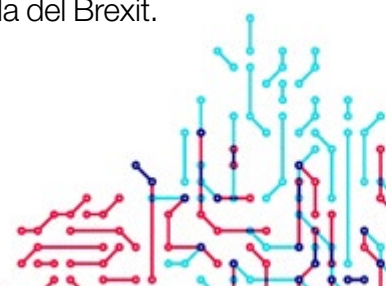
Els algorismes més sofisticats fan servir aprenentatge automàtic (*machine learning*), una branca de la intel·ligència artificial que aconseguix que les màquines millorin amb l'experiència. És bo per establir patrons i relacions ingents, i també per agilitzar processos. Dins del *machine learning* hi ha l'aprenentatge profund (*deep learning*), una tècnica de processament de dades basada en xarxes neuronals artificials amb moltes capes. La tècnica està inspirada en el funcionament bàsic de les neurones del cervell. Existeix fa més de cinquanta anys, però ara tenim prou volum de dades i capacitat de computació per aplicar-la en una multitud de casos pràctics.

21

Fabricants incansables de dades

Els quatre mil milions i mig de persones connectades avui a Internet (58% de la població mundial) generem un nombre gegantesc de dades. Cada moviment que fem ja és, pràcticament, una dada. Fabriquem dades les 24 hores al dia, els 365 dies de l'any. Inclús quan dormim estem creant dades: les del descans, la no-activitat. Sense dades, el món hiperconnectat s'alentiria tant que semblaria aturat.

Les nostres accions als ordinadors i als mòbils –que portem encaixats permanentment a les mans– produeixen dades que descriuen els nostres gustos, estat d'ànim, reaccions a estímuls, pors i ambicions. Les dades més valuoses per a les empreses, els partits polítics o els governs són les que defineixen el comportament de grups de persones. Després d'analitzar-les i classificar-les, configuren perfils o patrons de consum per fer-nos arribar ofertes de productes, serveis o missatges tan personalitzats (i tan al nostre gust) que gairebé no podem rebutjar-los. La consultora Cambridge Analytica va utilitzar aquestes tècniques d'anàlisi del rendiment per llançar desinformació (*fake news*) a través de Facebook en campanyes polítiques com la del president Donald Trump o la del Brexit.



Hi ha altres dades, com ara les empremtes digitals, la cara, l'iris o el genoma, que ens identifiquen com a éssers únics i irrepetibles. També hi ha dades acumulades, derivades de les nostres accions passades, que ens classifiquen com a bons o mal pagadors, o que ens obren i tanquen les portes d'accés a nous llocs de treballs en funció del lloc de residència, el gènere, l'edat o el color de la pell.

Avui les dades ho decideixen tot sobre les nostres vides en el present, i en el futur. És com si, de sobte, ens haguéssim posat unes ulleres d'augment i haguéssim descobert una infinitat de matisos del nostre entorn, mai apreciats abans. L'editor tecnològic de la revista *The Economist*, Kenneth Cukier,¹³ explica –en el seu llibre *Big Data: A Revolution that will transform how we work, live and think*– que “el valor de la informació avui resideix en la manera de correlacionar les dades per descobrir patrons que ni tan sols s'havien imaginat”.

Confiarem per evolucionar

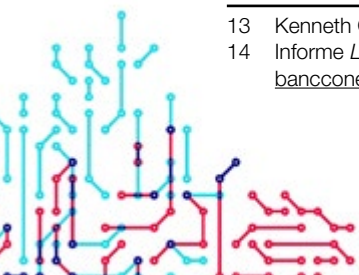
22

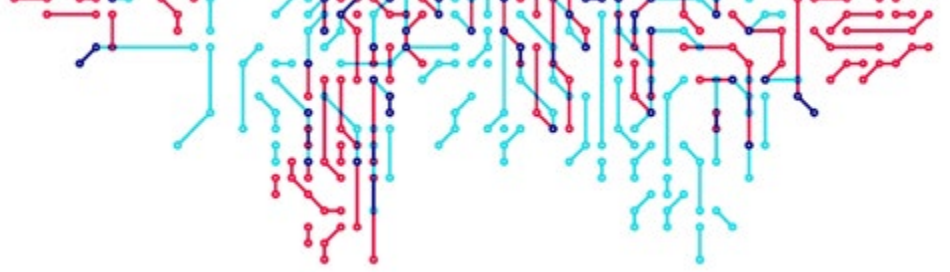
Ja hem normalitzat que una màquina –o assistent virtual– ens desitgi bon dia cada matí, tenir un aparell que dona voltes pel terra de casa netejant o que un sistema automatitzat ens suggereixi la sèrie que ens ve de gust veure segons el moment del dia. Aviat acceptarem que la nevera faci la compra de la setmana, que la rentadora s'ocupi de la bugada de manera autònoma o que la calefacció de casa truqui al tècnic quan falli la caldera. Tots aquests objectes produeixen (o produiran) dades relacionades amb el consum, les interaccions i l'ús. Es coneix com la Internet de les coses (IoT, en anglès); és a dir, la interconnexió digital d'objectes per comunicar-se i interactuar entre ells. Es calcula que el 2025, hi haurà al món 21.500 milions d'objectes connectats.¹⁴ També disposarem robots domèstics que tindran cura de les persones grans i dels petits a la llar, els alumnes a classe estudiaran les matèries més aptes per a la seva personalitat i confiarem en els cotxes autònoms com els vehicles més segurs per als nostres desplaçaments.

Confiarem en la tecnologia perquè sabem que, en cas contrari, no evolucionarem com a societat. Confiarem perquè la història ens ha demostrat que la tecnologia sempre ens ha fet avançar com a éssers humans.

13 Kenneth Cukier (https://www.ted.com/speakers/kenneth_cukier).

14 Informe *La Internet de les coses a Catalunya*. Acció. Generalitat de Catalunya (<https://www.accio.gencat.cat/web/.content/bancconeixement/documents/pindoles/iot-cat.pdf>).





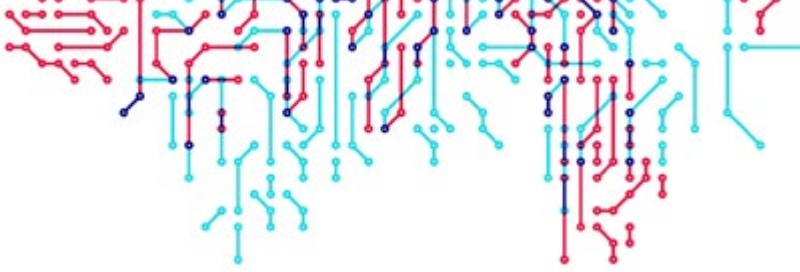
La complexitat s'entén amb exemples

Però, per confiar-hi, s'han de conèixer els beneficis i els riscos que implica la tecnologia que ens envolta i la que se'ns promet. Només així podem ser crítics i valorar en cada moment quina dosi de tecnologia volem a les nostres vides.

L'objectiu d'aquesta recerca és sacsejar la curiositat individual, explicar –d'una manera senzilla– per a què serveixen els algorismes de decisió automatitzada (ADA) que s'apliquen ja a Catalunya. Per explicar-ne la complexitat, hem recopilat més de cinquanta exemples en l'àmbit social, laboral, de la salut, de l'educació, la banca, la comunicació o la ciberseguretat, entre d'altres.

Per elaborar aquest treball hem comptat amb la col·laboració desinteressada d'una trentena d'experts del nostre país –màxims referents de la investigació acadèmica en intel·ligència artificial–, i d'emprenedors i professionals relacionats amb la IA, que han aportat coneixement i reflexió sobre els avenços actuals i els riscos que cal tenir en compte. De ben segur que hem deixat de mencionar moltes investigacions i iniciatives interessantíssimes en els sectors esmentats o en d'altres. Aquesta és només una mostra per entendre i decidir.

Perquè d'això es tracta: de decidir a títol individual què fer si un algorisme de decisió automatitzada ens afecta de manera negativa o discriminatòria. De saber quins organismes o legislacions ens emparen. De conèixer els procediments que cal seguir per tal que es revisin els criteris que s'han fet servir per dissenyar l'algorisme, així com per detectar i mitigar els biaixos de les dades que han provocat una desviació desfavorable per a una persona o un col·lectiu.



2.2. Riscos dels ADA

Imagini's que està cercant feina en una empresa de gran prestigi, amb unes condicions de treball i un sou més que acceptables. La companyia a la qual aspira rep moltes demandes, i no fa entrevistes personalitzades ni tests perquè considera que les habilitats que es destaquen en els currículums són sovint exagerades. A canvi, demana la contrasenya del correu electrònic del candidat perquè un algorisme rastregi els missatges personals i decideixi si és el candidat que cerca l'empresa. Donaria vostè el consentiment perquè un sistema d'intel·ligència artificial revisés la seva bústia a canvi d'optar a la feina de la seva vida?

L'exemple és real i s'explica en l'informe elaborat per l'ONG Algorithm Watch *Automating society: Taking stock of Automated Decision-Making in the EU*.¹⁵ L'agència de recol·locació de treball finlandesa DigitalMinds té una vintena de grans corporacions com a clients. Com que rep milers de currículums cada mes, només pot fer-ne la selecció amb algorismes de decisions automatitzades (ADA). DigitalMinds també els utilitza per escodrinyar el Facebook i el Twitter personal dels aspirants. El sistema analitza l'activitat del candidat i com reacciona. Amb els resultats es pot saber si una persona és introvertida o altres aspectes de la personalitat. De nou la pregunta: donaria vostè el seu consentiment? Com que la tecnologia ja existeix, no s'hauria de descartar el fet que aviat es comencin a aplicar mètodes similars a Catalunya.

24

Prenen decisions pels humans

Els algorismes avui ja són capaços d'assumir algunes funcions dels humans, i això els fa més imprescindibles cada dia. Poden fer reconeixement facial o d'imatges, interactuar (assistents virtuals), decidir automàticament (recomanadors de sèries, llibres o altres productes), tenir un impacte social (robots que s'ocupen de malalts a hospitals), i aprendre de moltes situacions en temps real (els cotxes autònoms que aviat tindrem).

“Hi ha tres tipus d'algorismes de decisió automatitzada –explica Jordi Vitrià, catedràtic de Llenguatges i Sistemes Informàtics de la Universitat de Barcelona (UB)–: Els que prediuen un nombre (fabricaré 33 cotxes); els que prediuen una classe (cert o fals, sí/no, malalt/sa); i els recomanadors –que entre un conjunt molt gran de possibilitats– asseguren que compraràs un determinat producte. Aquests tres tipus prèviament fan la predicció i, segons el resultat, prenen la decisió (automatitzada o no).”

Elisabeth Golobardes, doctora en enginyeria informàtica, explica el següent: “Els algorismes són de predicció i de decisió automatitzada. Poden predir potencials clients en

¹⁵ Informe elaborat per Algorithm Watch, *Automating society: Taking stock of Automated Decision-Making in the EU* (<https://algorithmwatch.org/en/automating-society/>). Karma Peiró és l'autora del capítol “Spain”.

un àmbit comercial o un possible accident en una central nuclear. I també hi ha els que prenen decisions com ara la d'evacuar la població en risc cent quilòmetres fora del nucli de la central nuclear per una possible explosió. O decidir fer fora cent treballadors d'una empresa per possible fallida. Són decisions molt complicades i confio que, al darrere, sempre hi hagi un professional que valori la situació i se'n faci responsable en cas que s'hi apliqui la decisió de la màquina de manera automatitzada”.

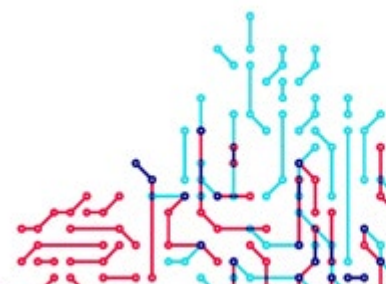
Tradicionalment, els bancs han emprat molts algorismes abans de concedir o denegar un crèdit. L'única diferència respecte al moment actual és que abans era el director de l'oficina bancària qui ho decidia, en funció d'unes variables marcades per l'experiència d'anys, i ara ho fan les màquines de manera automatitzada. El mateix passa amb les assegurances de cotxes o d'habitatge: el que abans feia un professional, el *comercial de tota la vida*, ara ho fa una màquina. Tot consisteix a tenir una gran base de dades amb milers o milions de casos del passat i entrenar l'algorisme. “Però s'ha d'anar amb compte amb els biaixos!”, alerta Jordi Vitrià. “Si agafo una base de dades d'un banc i durant trenta anys hi ha hagut una visió masculista de la societat, per la qual no concedien crèdits a les dones, l'algorisme perpetuarà aquest biaix i avui tampoc no els en donaran”. Els algorismes no són ni bons ni dolents, però tampoc no són neutres. S'alimenten de dades, i les dades tenen biaixos. “Sempre han existit. El que cal fer és detectar-los i mitigar-los”, afegeix Vitrià.

25

Maleïts biaixos!

En els anys vuitanta, el vicedegà de l'Escola de Medicina de l'Hospital St. George de Londres, Geoffrey Franglen, havia d'avaluar unes dues mil cinc-cents sol·licituds cada any. Per automatitzar el procés, va escriure un algorisme que l'ajudés a revisar-les, basant-se en el comportament d'avaluació de sol·licituds anteriors. Aquell any, els candidats es van sotmetre a una doble prova abans de ser admesos: la de l'algorisme i la dels professors. Franglen va adonar-se que les qualificacions coincidien en un 90-95%, la qual cosa demostrava que la IA podia reemplaçar els humans en aquesta fase tan tediosa. Però, quatre anys més tard, la direcció del centre va adonar-se de la poca diversitat dels candidats. I la Comissió per a la Igualtat Racial del Regne Unit va denunciar l'escola per discriminació xenòfoba i de gènere. Va resultar que cada any, l'algorisme havia deixat fora de la selecció unes 60 persones: els discriminava perquè tenien cognoms no europeus o perquè eren dones. S'estaven perpetuant biaixos.

“Hi ha tres tipus de biaixos clàssics: l'estadístic, el cultural i el cognitiu”, explica Ricardo Baeza-Yates, catedràtic d'Informàtica de la Universitat Pompeu Fabra i de la Northeastern



University.¹⁶ “L'estadístic procedeix de com obtenim les dades, d'errors de mesuratge o similars. Per exemple, si la policia està més present en alguns barris que en altres, no serà estrany que la taxa de criminalitat sigui més alta on tingui major presència. El biaix cultural és aquell que deriva de la societat, del llenguatge que parlem o de tot el que hem après al llarg de la vida. Els estereotips de les persones d'un país en són un exemple clar. Per últim, el biaix cognitiu és aquell que ens identifica i que depèn de la nostra personalitat, dels gustos i de les pors. Si llegim una notícia que està alineada amb el que pensem, la nostra tendència serà validar-la encara que sigui falsa.”

Aquesta última desviació es diu també *biaix de confirmació*. Bona part de les notícies falses (*fake news*) s'alimenten d'aquest raonament per difondre's més ràpidament. Per aquest motiu, si no ens qüestionem el que llegim o veiem, correm el risc d'avançar cap a una involució humana. L'historiador Yuval Noah Harari¹⁷ alerta en el seu darrer llibre, *21 Lliçons per al segle XXI*, que “amb la tecnologia actual, és molt fàcil manipular les masses”. I si seguim el que pensa la majoria de la gent, què passa quan moralment la massa està equivocada?

26

I encara més biaixos...

Ricardo Baeza-Yates explica¹⁸ que existeix el biaix d'ordre (rànkning) quan fem una cerca a la Web, ja que les persones tendeixen a fer clic a les primeres posicions i el buscador podria interpretar que aquestes respostes són millors que les següents.

Els biaixos de presentació són els que trobem a les recomanacions en l'àmbit del comerç electrònic. Només allò que el cercador mostra a l'usuari podrà tenir clics. Tot el que no surt en la pàgina de resultats queda fora de la consulta. És un cercle viciós, com el de l'ou i la gallina. I l'única manera de trencar-lo és mostrar l'univers total de resultats. “Això es coneix com a ‘filtre bombolla’: el sistema mostra únicament allò que t'agrada. Com es basa en les accions del passat, no és possible veure allò que es desconeix”, afegeix Ricardo Baeza-Yates.

En cas que el món continuï funcionant d'aquesta manera, arribarà un moment que ens sentirem com el personatge principal de la pel·lícula *El show de Truman*,¹⁹ que un bon dia es va adonar que tot el seu món era un engany i que s'havia perdut què hi ha més enllà de l'horitzó. Les xarxes socials funcionen així, i a les multinacionals ja els està bé. Potencien el que es coneix amb el nom d'economia de la dopamina. La dopamina és

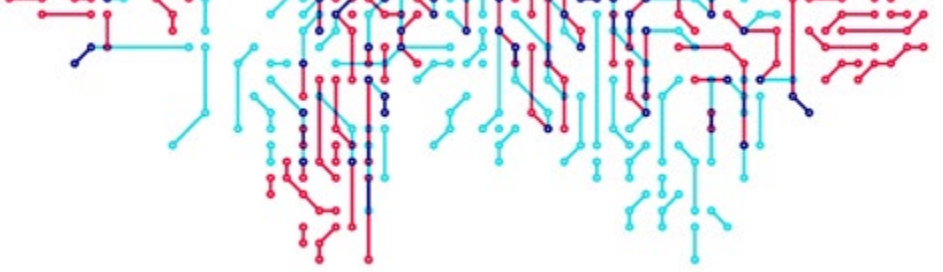
16 Ricardo Baeza-Yates, Karma Peiró. “És possible acabar amb els biaixos dels algorismes (1a i 2a part)”. 2019 (<https://www.karmapeiro.com/2019/06/17/es-possible-acabar-amb-els-biaixos-dels-algoritmes-1a-part/>)

17 Yuval Noah Harari (<https://www.ynharari.com/>).

18 És possible acabar amb els biaixos dels algorismes (1a i 2a part), Ricardo Baeza-Yates i Karma Peiró (2019) (<https://www.karmapeiro.com/2019/06/17/es-possible-acabar-amb-els-biaixos-dels-algoritmes-1a-part/>)

19 Film *El show de Truman* (https://ca.wikipedia.org/wiki/The_Truman_Show).





un neurotransmissor que tenim al cervell i que proporciona sensació de plaer, sentiments de goig o reforç per motivar les persones a dur a terme accions. Estar en el filtre bombolla d'una xarxa social pot arribar a crear una addicció, que ens incita a passar moltes hores interactuant.

Els que discriminen les dones

El 2014, Amazon va estrenar un algorisme per reclutar nous treballadors per als seus magatzems.²⁰ L'eina puntuava d'una a cinc estrelles els millors candidats. Tot semblava ideal: la IA estalviaria hores al departament de recursos humans. Però, un any més tard, la multinacional es va adonar que en els llocs tècnics, com el de desenvolupador de programari, no s'havia contractat cap dona. Podria ser que no hi hagués cap candidata amb aptituds per aquella feina?

L'exemple ja és un clàssic dels errors que ha provocat recentment la IA. Les dades massives que van servir per alimentar l'algorisme de selecció de personal es van basar en currículums rebuts en la dècada anterior, en què la majoria dels programadors eren homes. Quan el sistema automàtic detectava la paraula *dona* –o un sinònim–, directament penalitzava el currículum posant-li menys puntuació.

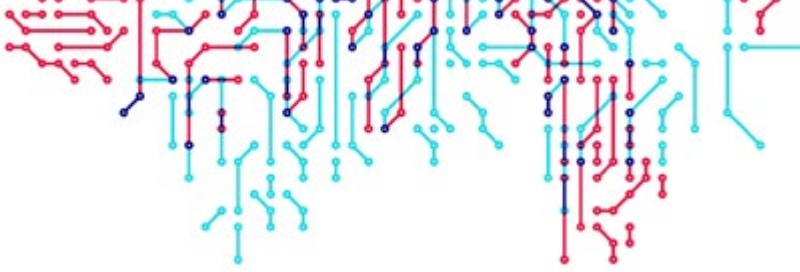
27

Un altre exemple de discriminació de gènere va passar el 2016. Investigadors de Microsoft Research i de la Universitat de Boston van fer servir una col·lecció massiva de notícies de Google per entrenar algorismes sobre els estereotips femení/masculí que sortien a la premsa. Els resultats donaven que els homes eren programadors informàtics i les dones mestresses de casa. Ells eren metges i elles, infermeres. I té la seva lògica, perquè als Estats Units la majoria dels periodistes que redactaven les notícies sobre les quals es va basar l'entrenament de l'algorisme eren homes. Per tant, Google només reflectia el biaix de gènere existent en la realitat.

De biaixos, n'hi ha per donar i per vendre. N'hi ha desenes de culturals i encara més de cognitius. S'han arribat a classificar fins a un centenar, però els més importants són uns vint-i-cinc. En l'article "The Ultimate List of Cognitive Biases: Why Humans Make Irrational Decisions",²¹ s'enumeren uns 49 biaixos cognitius. Aquests són els més perillosos, perquè estan arrelats a cada persona. L'única manera de resoldre'ls és fer canviar cada persona, cosa la qual, d'entrada ja sembla una proesa impossible. Cal donar la raó a l'historiador Harari quan diu que "és molt fàcil manipular les persones i molt complicat eliminar els biaixos".

20 "Por qué la inteligencia artificial discrimina a las mujeres?". Ricardo Baeza-Yates, Karma Peiró (2019) (<https://medium.com/think-by-shifta/por-qu%C3%A9-la-inteligencia-artificial-discrimina-a-las-mujeres-18b123ecca4c>).

21 *The Ultimate List of Cognitive Biases: Why Humans Make Irrational Decisions* (<https://humanhow.com/en/list-of-cognitive-biases-with-examples/>).



Però els sistemes intel·ligents no són autònoms, ni actuen per si mateixos. I ho explica molt bé la investigadora Elisabeth Golobardes: “L’algorisme com a tal no té cap biaix. Són les dades que s’introdueixen i l’objectiu pel qual ha estat dissenyat el que discrimina”.

Es pot ser just amb tothom?

Tampoc no tots els biaixos són perjudicials. “Per exemple, que hi hagi més infermeres que infermers pot ser positiu per les seves qualitats empàtiques en el tracte dels pacients. Però que els polítics siguin majoria no ho és, perquè un punt de vista de la població (el femení) no està igualment representat”, diu Baeza-Yates. Els resultats dels algorismes poden discriminar per raons de gènere, raça, edat o classe social, per mencionar els més importants.

Si sabem que els algorismes tenen biaixos i que poden discriminar..., per què s’apliquen? “Una resposta podria ser que el benefici o encert dels resultats és considerablement superior (més d’un 90% en la majoria dels casos) que el perjudici o error. És això just pels que surten perjudicats? En aquest punt es podria encetar una llarga discussió sobre què és o no és just a la vida”, continua Baeza-Yates.

28

Resulta molt difícil ser just amb tothom. Un algorisme pot ser just amb un col·lectiu de dones, però discriminar un home. En aquest sentit, l’investigador Andrew Selbst,²² del Data&Society Research Institute,²³ explica que decidir la discriminació en intel·ligència artificial és ben complicat: “És un procés en constant evolució, igual que qualsevol aspecte de la societat”.

Per ser rigorosos, també cal dir que els algorismes ben dissenyats –tot i tenir biaixos– són justos d’acord amb els paràmetres que se’ls han donat. A diferència dels humans –que poden variar la seva decisió en funció d’un estat d’ànim o de cansament físic i mental– els algorismes sempre funcionen igual.

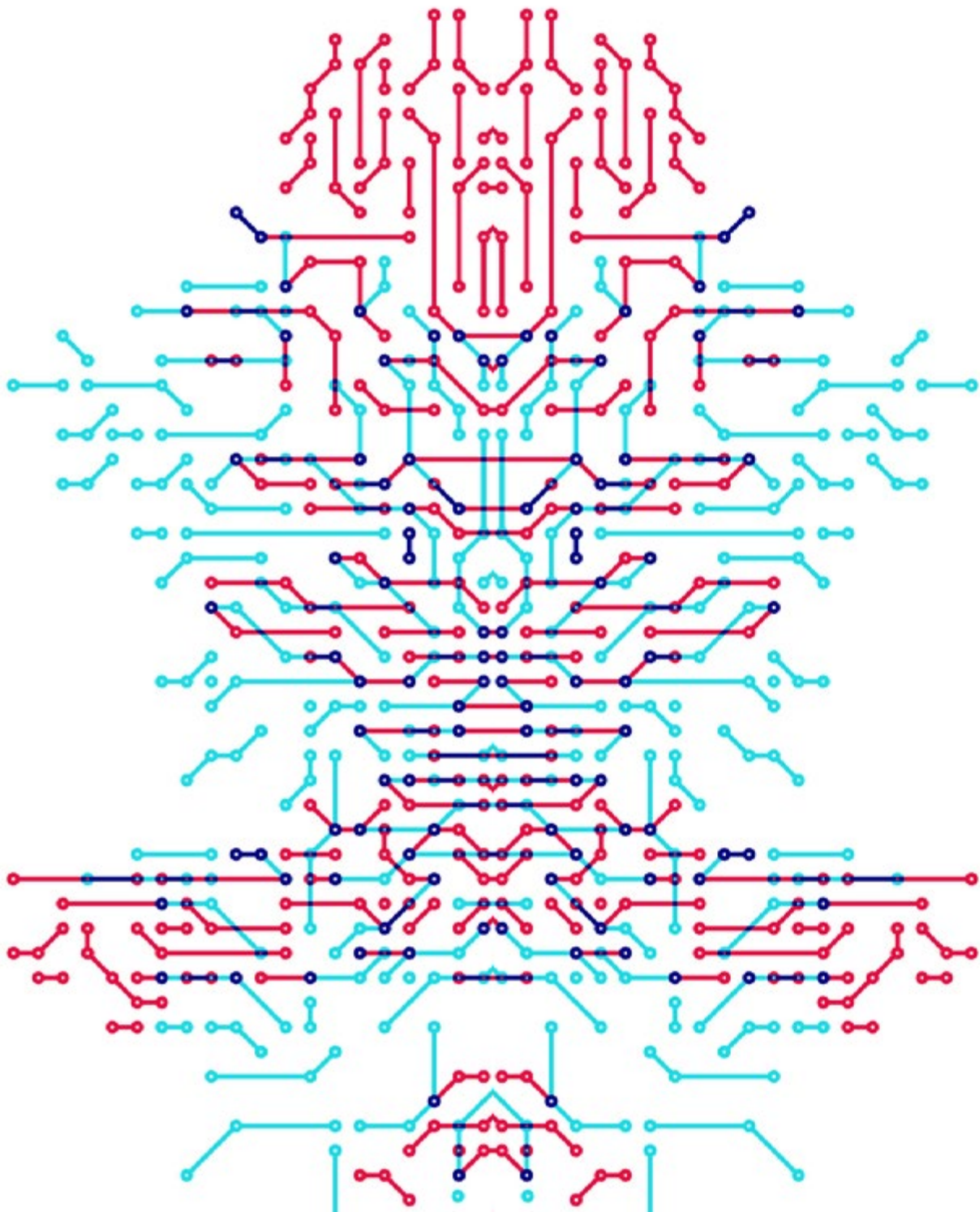
Com els prejudicis

Els biaixos són semblants als prejudicis: tots en tenim, en un grau o un altre. Molts els heretem del nostre entorn social o familiar sense adonar-nos-en. El biaix més gran és creure que no tenim cap prejudici. Però atenció! Si els biaixos no es corregeixen, hi ha el risc d’habitar un futur en què cada cop sigui més difícil el progrés social perquè els prejudicis s’hauran perpetuat.

²² Andrew Selbst (<https://andrewselbst.com/>).

²³ Data & Society Institute (<https://datasociety.net/>).

Com podem estar segurs, doncs, que totes les dades que introduïm en l'algorisme representen l'univers que volem predir i que no discriminaran ningú? No en podem estar segurs. I aquí apareixen els aspectes ètics de la intel·ligència artificial que cal tenir molt presents. (Vegeu el capítol 6, "Aspectes ètics de la IA amb les opinions d'investigadors").



2.3. De decisió automatitzada o de suport a la decisió?

La matemàtica Cathy O'Neil²⁴ –autora del llibre *Armes de destrucció matemàtica*, dedicat a evidenciar els algorismes de decisió automatitzada (ADA) als Estats Units–, en una de les seves visites a Barcelona, em va plantejar aquest dilema: “Series capaç de decidir si un menor corre perill a casa perquè està a punt de ser maltractat o abusat sexualment per algun dels seus progenitors?”. Davant la meua cara de preocupació, ella va afegir: “Als EUA ho estan decidint les màquines. L'algorisme Aura²⁵ va ser creat per detectar els menors que podien ser víctimes potencials d'abusos abans de produir-se els fets, amb la bona intenció d'evitar un trauma a l'infant”. Però... com pot saber del cert una màquina que s'ha d'allunyar el nen dels seus pares abans que passi res?, vaig pregar preguntar perplexa i bastant amoïnada. I la resposta d'O'Neil va ser: “No ho pot saber”. L'algorisme Aura va ser una prova pilot que, finalment, no es va arribar a aplicar. No obstant això, hi ha altres sistemes automatitzats, com el COMPASS²⁶ –que prediu la criminalitat a les presons–, que sí que han estat molt polèmics. El mitjà independent *Pro Publica* va realitzar una investigació periodística²⁷ que demostrava els biaixos que tenia sobre les persones negres.

30

A Catalunya no hem arribat a aquests extrems. Estem en una fase inicial de la intel·ligència artificial i molts dels exemples que expliquem encara formen part d'investigacions o de proves pilot. D'altra banda, tots els experts que hem consultat per a aquest informe deixen clar que les decisions automatitzades s'apliquen únicament en els sectors que no comporten un risc per a la vida d'una persona. Per tant, en l'àmbit de la salut, en el judicial (i en certs casos d'ensenyament), la decisió de l'algorisme és un suport al professional i no s'aplica de manera automatitzada. L'última paraula sempre la té el metge, el jutge o el mestre responsable.

Qüestió d'escala

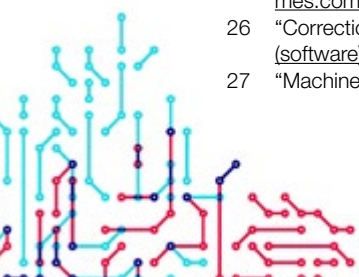
En sectors com el de reclutament de llocs de treball, on s'ha d'atendre una demanda de centenars o de milers de persones, l'algorisme fa una primera selecció. Per atorgar un ajut social, amb molts sol·licitants en les mateixes condicions, el sistema automatitzat decideix. I d'alguna manera, també pot comportar canvis importants i repercussions desfavorables per a la vida de les persones.

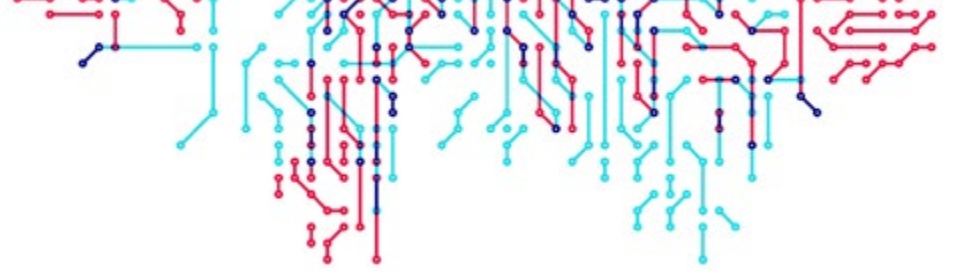
24 Cathy O'Neil (<https://mathbabe.org/about/>).

25 “Cathy O'Neil, author of ‘Weapons of Math Destruction,’ on the dark side of big data”. *Los Angeles Times* (<https://www.latimes.com/books/jacketcopy/la-ca-jc-cathy-oneil-20161229-story.html>).

26 “Correctional Offender Management Profiling for Alternative Sanctions (COMPASS)” ([https://en.wikipedia.org/wiki/COMPASS_\(software\)](https://en.wikipedia.org/wiki/COMPASS_(software))).

27 “Machine Bias”, *ProPublica* (<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>).





“Automatitzar la selecció de personal en departaments de recursos humans és un exemple molt comú. Si reps mil currículums cada mes, o cada setmana, no te’ls pots mirar tots i decidir. Els algorismes fan una preselecció. I els responsables es miren només aquells que han passat el filtre dels ADA. Però les màquines no veuen coses que els humans sí que veuen. I llavors es quan venen les discriminacions”, explica Ramón López de Mántaras, professor investigador del Consell Superior d’Investigacions Científiques (CSIC).

“En la medicina és diferent. Els metges fan servir els ADA només com a suport a la decisió. La Unitat de Cures Intensives va plena d’algorismes, i és clar que algun pot disparar una alarma, però un metge (o personal sanitari) mira què passa i pren la decisió final. No hi ha cap sistema automatitzat que t’administri un medicament sense la supervisió d’un professional. I espero que sempre sigui així”, afegeix. “Perquè el metge veu coses que l’algorisme no veu. I l’algorisme troba patrons que l’ull humà no veu. L’algorisme mira els arbres, i el metge el bosc. Per això, quan treballen junts disminueix molt l’error. En la detecció del càncer de mama, hi ha estudis que demostren que el millor metge té un error del 5% o 6% amb les mamografies i l’ADA, del 6% o 7%, però que, treballant plegats, l’error és només del 0,5%.”

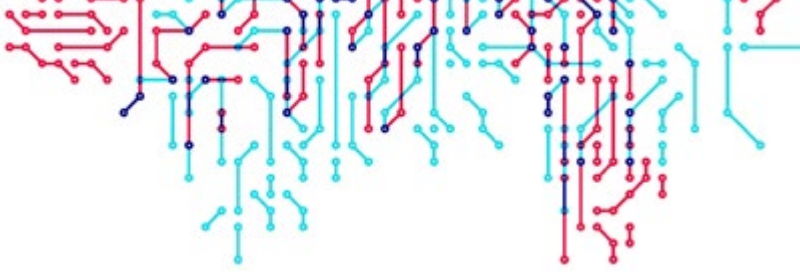
31

Les “caixes negres”

Recapitem: la intel·ligència artificial fa servir algorismes que aprenen de dades massives. Aquestes dades tenen biaixos que poden discriminar. Cal detectar-los i mitigar-los. Els algorismes prediuen i després, alguns, prenen decisions automatitzades. Bona part d’aquests sistemes són d’aprenentatge automàtic, és a dir, que aprenen abans de fer-los servir, i alguns altres són d’aprenentatge profund (*deep learning*).

Qualsevol sistema que es consideri intel·ligent ha de tenir l’habilitat d’aprendre, és a dir, de millorar amb l’experiència. El que no s’ha aconseguit explicar encara és com ha arribat l’algorisme a la predicció o decisió. D’això se’n diu “caixes negres”. Les aplicacions dedicades al processament del llenguatge natural com ara els traductors, la diagnòsi mèdica, la bioinformàtica o la detecció del frau financer són caixes negres.

Per tant, si un sistema automatitzat prediu alguna acció que em discrimina, podem preguntar com ha arribat a aquella decisió? Ho podem preguntar però no obtindrem, per ara, una resposta. I aquest és, actualment, un dels grans dilemes de la IA.



L'explicabilitat o transparència algorísmica

Per entendre com va fer la predicció o va prendre la decisió l'algorisme es parla de l'*explicabilitat* (*explainable artificial intelligence*).²⁸ “En els darrers dos anys s’ha treballat molt en aquest problema dels algorismes d’aprenentatge automàtic. En particular, en els més opacs, que són els d’aprenentatge profund”, explica Ricardo Baeza-Yates, “Però encara falten respostes”.

La investigadora de l’Oxford Institute Sandra Wachter²⁹ considera que hauríem de tenir el dret legal de saber per què els algorismes prenen les decisions que ens afecten. Explica que els propietaris dels algorismes –multinacionals, empreses de totes les mides, bancs, però també governs, administracions públiques o la policia– posaran molts inconvenients a fer transparents les fórmules dels seus sistemes per defensar la propietat intel·lectual o la seguretat ciutadana. Per això proposa que es donin “explicacions contrafàctiques”. És a dir, si t’han denegat una hipoteca, poder preguntar al banc: “Si guanyés 10.000 euros més a l’any, me l’haguessin concedit?”. Si no has passat la selecció d’un lloc de treball, poder preguntar: “Si hagués tingut un màster, me l’haurien donat?”.

32

La comunitat científica treballa perquè els algorismes donin autoexplicacions de les seves decisions. I podem pensar que molt aviat es trobarà la manera. Estem en els inicis d’un canvi social que ha de magnificar-se en els pròxims anys. I la ciutadania té un rol molt important. La tecnologia té parts brillants i parts fosques. Considerem com a parts fosques els errors involuntaris, però també el mal ús intencionat que es pot tenir respecte a una persona o a un col·lectiu en funció dels biaixos de les dades, o dels criteris sota els quals s’hagi creat el sistema intel·ligent. Cal estar alerta a les anomalies i les discriminacions per aconseguir –amb la legislació vigent i els organismes públics responsables– que es continuï garantint la privacitat, la confidencialitat i la llibertat individual.

28 “Explainable artificial intelligence” (https://en.wikipedia.org/wiki/Explainable_artificial_intelligence)

29 Sandra Wachter. “How to make algorithms fair when you don’t know what they’re doing”. *Wired*. (<https://www.wired.co.uk/article/ai-bias-black-box-sandra-wachter>).

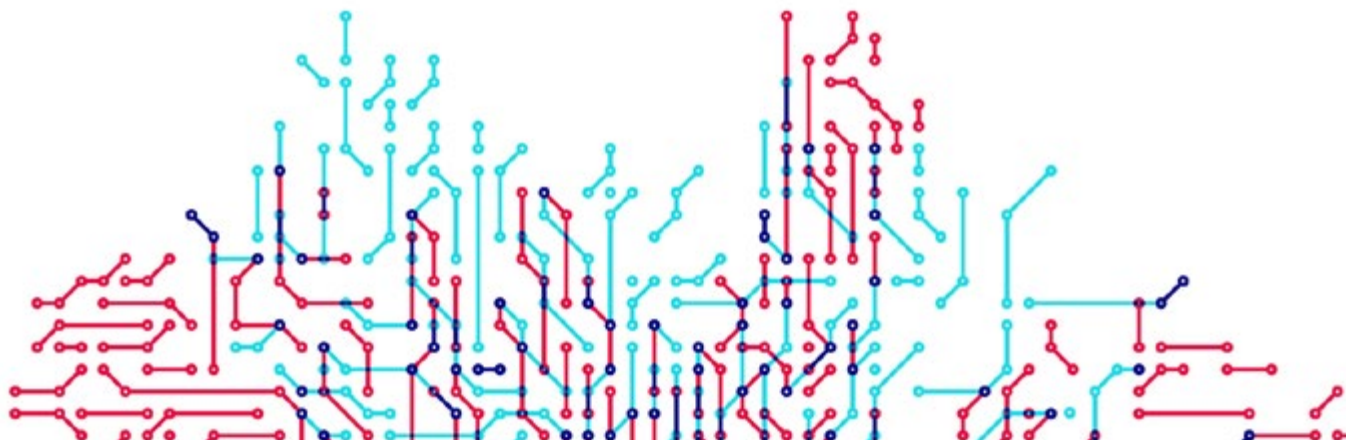
ADA EN ACCIÓ

2.4. On s'apliquen els ADA a Catalunya? Més de 50 exemples per entendre...

La millor manera d'entendre conceptes tecnològics és amb exemples. Quan ens posem en una situació determinada, podem imaginar amb més claredat què faria la màquina en qüestió o com actuaria el sistema intel·ligent. Només així podem valorar la importància, els riscos i les conseqüències de la seva aplicació.

Per aquest motiu s'ha considerat important incloure en aquest informe un capítol dedicat a explicar on es troben els algorismes de decisió automatitzada (ADA) en diferents sectors: salut, sistema judicial, educació, àmbit social, comerç, ciberseguretat, banca, àmbit laboral, comunicació, etc. No és, ni de bon tros, el recopilatori més exhaustiu. Som conscients que ens deixem infinitat d'exemples. Però sí que considerem que és una mostra per fer-se a la idea de com d'invisibles són els ADA, i per entendre tant el valor de la seva aplicació com els riscos que comporta.

De manera generalitzada, es pot dir que mentre a les empreses i el sector privat la seva implementació està molt avançada, a l'administració pública encara li queda un llarg camí per recórrer. És destacable també que en els sectors de la salut i judicial, a partir de col·laboracions entre hospitals locals i internacionals, universitats, ens privats i *start-ups*, o amb desenvolupaments propis, fa anys que s'exploren les possibilitats que ofereix la intel·ligència artificial per avançar en un món més eficient i just.



2.4.1. Salut

La intel·ligència artificial és un gran suport en l'àmbit de la salut. Els algorismes ajuden a entendre o extreure conclusions sobre malalties en molt menys temps, a suggerir un diagnòstic i medicació pertinent, a fer una millor gestió dels centres hospitalaris o a llegir històries clíniques a gran escala.

En aquest cas, les decisions automatitzades –a diferència d'altres sectors– no s'apliquen sense la supervisió d'un doctor o especialista en la matèria. I aquest és un fet important. Què passaria si l'algorisme no detectés un càncer en un pacient i el doctor no donés un tractament? A Catalunya –almenys quan es redacta aquest informe– la IA sempre és un suport a la decisió del metge especialista.

La intel·ligència artificial té un potencial enorme però sempre en col·laboració, i com a complement del potencial i coneixement que tenen els professionals de la salut. Com s'ha dit en l'apartat anterior, quan la decisió automatitzada implica un risc per a la vida de les persones, l'última paraula la té el professional especialista.

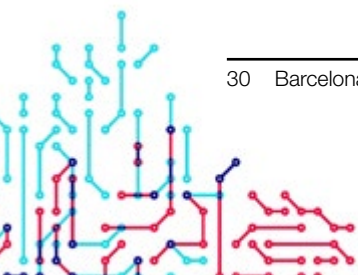
34

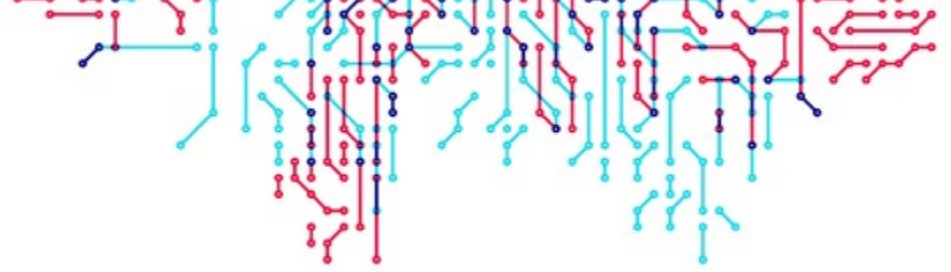
“El cervell humà és molt lent, comparat amb el que avui poden fer les màquines”, apunta la investigadora Itziar de Lecuona, professora del Departament de Medicina i sotsdirectora de l'Observatori de Bioètica i Dret de la Universitat de Barcelona (UB). “Però no per això hem de confiar cegament en la tecnologia. Tots els models s'han de validar, s'han de fer proves de concepte i preguntar-se si l'algorisme funciona per a l'objectiu inicial”, afegeix.

“Els algorismes sempre han existit en medicina, però ara hem fet un pas més: anem cap a la predicció, cap a l'experiència més personalitzada. Amb la tecnologia actual (i les seves capacitats) seria molt poc eficient no aplicar-la.” Lecuona insisteix que en l'àmbit de la salut tothom està pensant a fer el bé. “Però també s'ha de tenir en compte l'ètica. Perquè els algorismes es nodreixen de dades que formen part de la intimitat de les persones.”

D'altra banda, l'experta en seguretat i resiliència de l'Agència Europea de Seguretat de Xarxes i Informació (ENISA) Dimitra Liveri va parlar en el Barcelona Cybersecurity Congress³⁰ dels hospitals intel·ligents, que molt aviat deixaran de ser excepcionals a tot arreu. “Un hospital intel·ligent és un ecosistema de màquines interconnectades que prenen decisions automatitzades”, va explicar. Què passaria si una bomba d'insulina intel·ligent cometés un error? El pacient podria morir. L'experta d'ENISA va recordar en la seva presentació que encara queden reptes per resoldre. “1. La disponibilitat de l'aparell: si estic malalt vull que el meu dispositiu funcioni tot el temps, sense aturar-se. 2. La integritat: vull que m'injecti les dosis correctes de glucosa i que mai no s'equivoqui”.

³⁰ Barcelona Cybersecurity Congress (<https://www.barcelonacybersecuritycongress.com/es/front-page/>).





• Una píndola que grava els intestins

El càncer de còlon a Catalunya és una malaltia que afecta majoritàriament a població de més de 50 anys. Es calcula que es diagnostica cada any a unes cinc mil persones i està lligat a l'estil de vida i l'alimentació de cadascú. Des de fa anys, la Generalitat de Catalunya s'ocupa de fer arribar a domicili uns tests per detectar-lo. A partir d'una prova senzilla, es pot agafar a temps en cas d'anomalia. Quan el test dona positiu, es fa una prova endoscòpica o colonoscòpia, que, en alguns casos, requereix ingrés hospitalari, sedació i resulta costosa per a la sanitat pública.

Als EUA i a alguns països d'Europa s'està practicant una nova tècnica que es podria implementar a Catalunya aviat. Investigadors de l'Hospital de la Vall d'Hebron, juntament amb el Departament de Matemàtiques i Informàtica de la Universitat de Barcelona (UB) experimenten amb una càpsula endoscòpica³¹ per estudiar l'estat dels intestins. Els pacients s'han d'empassar una petita píndola que va dotada d'una càmera, quatre leds³² i una bateria. La píndola viatja pel cos del pacient i enregistra imatges que poden ser de gran ajuda mèdica. Les imatges s'envien per wifi en connexió amb un dispositiu que porta la persona, agafat a un cinturó. L'enregistrament dura entre vuit i dotze hores.

La píndola no té cap contraindicació i si dona alguna alerta es procedeix a fer la colonoscòpia. L'invent sembla perfecte, però llavors sorgeix la pregunta de si els metges tenen temps de visionar tantes hores d'enregistrament de cada pacient? "Impossible", respon Jordi Vitrià, director del Departament de Matemàtiques i Informàtica de la UB. "O fas un sistema automatitzat, que analitzi i detecti si hi veu alguna anomalia o no té sentit. Amb el que l'algorisme detecti, el doctor prendrà la decisió de si es fa l'endoscòpia o altre tipus de proves", afegeix Vitrià.

35

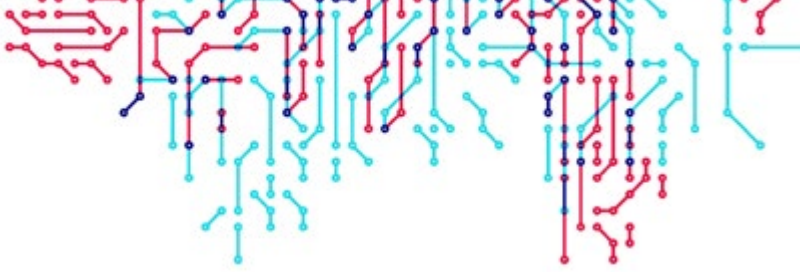
• Detector del dolor exagerat

Imaginem que un treballador ha patit un accident o una lesió en un braç, en una cama o a l'esquena i que demana la baixa de mitjana/llarga durada per aturar la seva activitat professional. La durada i la quantia de la prestació va en funció del temps que hagi d'estar de baixa. Per norma, aquests casos els gestionen les asseguradores laborals amb un reconeixement físic, durant el qual fan passar el treballador lesionat per diferents màquines que posen a prova les articulacions. Però sempre hi ha un percentatge de casos en què les asseguradores no poden detectar realment la magnitud del dolor.

"S'ha constatat que en un percentatge alt les persones són magnificadores del dolor", explica Jordi Vitrià. És a dir, que aquella part del cos no els fa tant mal com diuen.

31 "Càpsula endoscòpica" (<https://hospital.vallhebron.com/en/diagnostic-tests/endoscopy-endoscopic-capsule>).

32 "LEDs, dispositius miniaturitzats que emeten llum" (https://ca.wikipedia.org/wiki/D%C3%ADode_emissor_de_llum).



Quan això es detecta, la Seguretat Social denega la baixa. “Hi ha molts diners públics en joc. Per això, van demanar a la UB crear un model algorísmic, d’acord amb tot l’història de proves i diagnòstics acumulats durant anys. Ara veiem que, en molts casos, els pacients tenen raó, però en d’altres, no. Es pot detectar l’exageració del dolor de manera automatitzada”, puntualitza l’investigador de la UB. Després d’implementar-se a Barcelona, s’ha implementat a Sevilla, amb els mateixos criteris, de manera que els nivells de permissivitat del dolor i les decisions que es prenguin siguin semblants en totes dues ciutats.

• **Telèfons intel·ligents per als avis**

La població de més de 80 anys ha augmentat a Europa en les darreres dècades. Una esperança de vida més alta, juntament amb la davallada de la natalitat, hauria de servir per plantejar-se com s’inverteixen els recursos de manera més eficient. Una vida més llarga no significa necessàriament una salut millor. Què cal fer per mantenir el benestar dels avis? Quins ajustos calen en les polítiques socials, econòmiques i de salut pública?

36

El projecte europeu NESTORE³³ es presenta com un entrenador virtual que acompanya les persones grans, amb bona salut, i els dona consells personalitzats sobre els hàbits alimentaris o els exercicis diaris que han de fer. El seguiment es fa a través d’objectes connectats i sensors repartits per la casa, i d’una aplicació de mòbil que monitora tots els seus moviments.

“Se’ls entrega un telèfon intel·ligent i han de fotografiar el que tenen a la nevera, què mengen, etc. Amb algorismes de visió per computació es fa la predicció del que han d’ingerir d’acord amb el seu estat de salut”, explica Itziar de Lecuona, una de les investigadores del NESTORE. “A l’habitatge també tenen un aparell (assistent virtual) que els recorda que han de sortir a caminar, les passes mínimes o altre tipus d’exercici físic. S’acaba convertint en un acompanyant molt íntim perquè ho saben tot de tu: les malalties i la medicació, el menjar, les hores de descans, si t’has mogut o no, si has estat en contacte amb la família, si has rebut visites, etc.”, afegeix.

• **Encertar la medicació després d’un trasplantament**

L’Hospital del Mar de Barcelona té una de les unitats de trasplantament de ronyó més grans de tot Espanya. Va començar el 1973, i avui dia ja ha fet més de 1.400 trasplan-

³³ Projecte europeu NESTORE (<https://nestore-coach.eu/home>).

taments.³⁴ A Catalunya, gairebé la meitat dels nous pacients sotmesos a teràpia renal substitutiva tenen més de 70 anys, i bona part dels donants renals superen els 60.

Un dels problemes principals després d'un trasplantament és encertar la dosi de medicació que es dona de per vida al receptor del ronyó. Per això, aquest pas és molt sensible. "Si et passes, pot tenir efectes secundaris i si la medicació és poca, l'òrgan pot ser rebutjat pel cos. És molt complicat", explica Fernando Cucchietti, director d'Anàlisi de dades i visualització del Barcelona Supercomputing Center (BSC). "Vam proposar a l'Hospital del Mar utilitzar algorismes per decidir si una persona rep o no un ronyó amb tots els condicionants que actualment es tenen en compte entre el donant i el receptor, però també per ajudar-los a precisar el protocol de medicació exacte", afegeix. Per ara, el projecte encara no s'ha posat en marxa.

• Detectar a temps la cirrosi

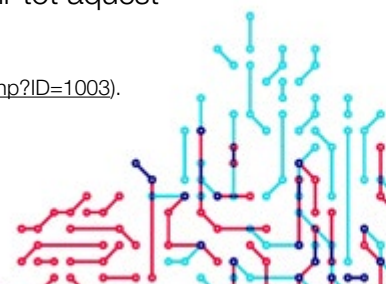
Cada cop que el fetge pateix una lesió, ja sigui per malaltia, consum excessiu d'alcohol o una altra causa, l'òrgan intenta reparar-se a si mateix. En el procés, es forma un teixit de cicatrització. A mesura que la cirrosi avança, es formen més teixits de cicatrització i fan que el fetge funcioni amb dificultat (cirrosi descompensada). La cirrosi avançada és potencialment mortal. Però si es diagnostica a temps i se'n tracta la causa, es pot limitar el mal i –a vegades– es pot revertir.

L'Hospital Clínic de Barcelona té una de les millors unitats de tractament del fetge, però quan els malalts hi acudeixen ja és massa tard. Per això és tan important desenvolupar tecnologies i sistemes de detecció primerenca, explica Fernando Cucchietti. "Ara formem part d'un projecte europeu amb el Clínic, que ens permetrà tenir els símptomes preliminars a una cirrosi de fetge. La prova es farà amb 40.000 persones de tot Europa, durant cinc anys. I es coordina des de Barcelona", afegeix l'investigador del BSC. Segons explica, la idea és instal·lar un aparell en centres d'atenció primària i hospitals, de manera que els algorismes facin el diagnòstic i els metges decideixin si el pacient s'ha de fer una biòpsia o un trasplantament. "Es volen detectar els símptomes cinc o deu anys abans que sigui massa tard, perquè en aquesta fase primerenca només cal un canvi d'estil de vida i molt poca medicació", conclou.

• Diagnòstic amb electrocardiogrames digitals

Una empresa del sector de la salut vascular fabrica aparells que fan electrocardiogrames en format digital. Això permet processar-los i associar-los a patologies. "Tenir tot aquest

34 "40 anys de trasplantament renal a l'Hospital del Mar" (<https://www.parcdesalutmar.cat/es/noticies/view.php?ID=1003>).



coneixement encapsulat és com tenir un comitè de metges que, al llarg del temps, ha diagnosticat una sèrie de malalties, amb molts centenars de milers d'electrocardiogrames”, explica Jordi Navarro, expert en dades i CEO d'una empresa catalana dedicada a la IA.

“L'algorisme fa un diagnòstic similar al del metge, a partir de trobar patrons. Amb la diferència que, si hi hagués vint cardiòlegs, un s'hauria equivocat en el diagnòstic i 19 l'haurien encertat, però l'algorisme sempre agafa el resultat de la majoria. En aquest cas, el bo. En aquest sentit, podem dir que els algorismes són democràtics.”

• **Reconeixement facial per detectar el TDAH**

38

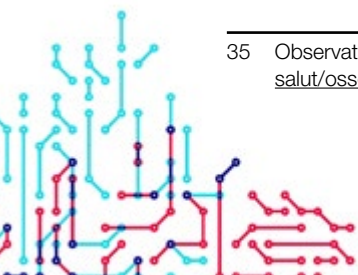
El trastorn per dèficit d'atenció amb hiperactivitat (TDAH) és d'origen neurobiològic i es caracteritza per la impossibilitat de mantenir l'atenció, hiperactivitat i impulsivitat, i influeix considerablement en la vida d'un menor. Els símptomes del TDAH es manifesten a l'escola o en qualsevol altre ambient social. Aquest trastorn s'inicia en la infància, però pot continuar en l'adolescència i l'edat adulta. A escala mundial, s'estima que entre el 3% i el 7% dels nens poden estar-hi afectats. A Catalunya, l'1,3% de les nenes i el 3,7% dels nens de 6 a 17 anys residents a Catalunya van consumir algun fàrmac per al TDAH durant el 2015. Aquesta proporció és d'1,9% i de 5,5% en les nenes i els nens, respectivament, de 12 a 15 anys, que és el grup amb més consum.³⁵

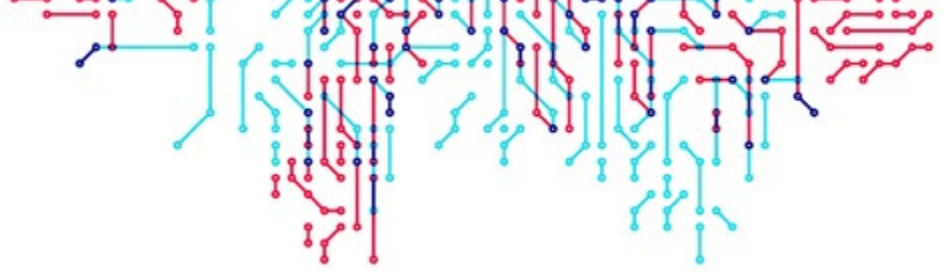
Ara, amb el reconeixement facial, es pot diagnosticar més encertadament la malaltia. El Centre de Visió per Computació (CVC) de la Universitat Autònoma de Barcelona (UAB) ha posat en marxa un sistema intel·ligent que analitza els gestos i les expressions facials del menor. Amb els resultats es pot donar un grau de certesa sobre els símptomes. Aquest diagnòstic serveix al metge/psicòleg que tracta el menor per fer el protocol d'actuació. Amb el mateix algorisme s'intenta diagnosticar també altres malalties de caire depressiu entre el jovent. “Tot just s'està començant”, explica Meritxell Bassolas, directora de Coneixement i Transferència de Tecnologia del CVC. “A més dels resultats de l'algorisme, s'analitza el comportament del menor a les xarxes socials. Amb tot, el professional té més informació per aplicar un tractament”, afegeix. *(Vegeu altres exemples amb visió per computació en l'apartat 2.4.10.)*

• **El malalt no va a l'hospital que li pertoca**

El CatSalut –organisme que gestiona la sanitat pública de Catalunya– distribueix els recursos sanitaris entre les diferents comarques catalanes, que s'agrupen en nou re-

³⁵ Observatori de la Salut de la Generalitat de Catalunya (http://observatorisalut.gencat.cat/web/.content/minisite/observatori-salut/osscc_central_resultats/informes/fitxers_estatics/MONOGRAFIC_26_TDAH_CdR.pdf).





gions sanitàries, i en subregions de diverses categories. Tothom té assignat un centre d'atenció primària, a més d'un hospital per a les urgències o les estades més llargues. No obstant això, no sempre s'acudeix al recurs que es té assignat, de manera que els recursos –previstos inicialment per a un nombre determinat de persones– poden ser escassos o sobrants. Quins són els motius que fan que els malalts no acudeixin allà on els pertoca per zona de residència? Les raons poden ser múltiples: problemes logístics, falta de professionals en el recurs, esperes massa llargues o horaris del transport públic per arribar-hi insuficients.

Fins ara, aquesta pregunta es responia amb la minuciosa tasca dels tècnics que cercaven les anomalies després de molta observació de funcionament. Ara ho pot fer un programa automàticament:³⁶ l'algorisme agafa les variables de comportament dels pacients i les mostra en verd o vermell en funció de si estan anant al lloc que els pertoca o no. Amb aquesta informació extreu patrons anòmals de cada centre de salut. De manera que són més ràpids a organitzar el territori i minimitzar un problema de recursos públics.

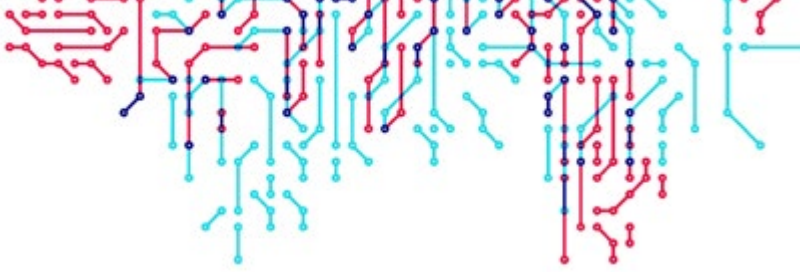
• Predir que el malalt tornarà a l'hospital

Les persones que pateixen malalties cròniques del cor o del ronyó tornen, cada cert temps, a l'hospital. A vegades, per petits imprevistos desestabilitzadors, com és ara un dia de calor intensa o córrer per agafar el bus. L'indicador de readmissió hospitalària –reconegut internacionalment com a prova de qualitat–³⁷ marca en trenta dies la taxa òptima perquè un pacient torni a l'hospital. “Si el pacient torna abans d'un mes, alguna cosa falla”, explica Ricard Gavaldà, professor i coordinador del laboratori de recerca de la Universitat Politècnica de Catalunya (UPC). “Potser l'hospital necessitava llits i ha enviat el pacient a casa abans del previst, potser no hi ha hagut comunicació amb el metge de capçalera en el seguiment de la medicació administrada, o simplement perquè la persona necessitava més atenció”, afegeix.

“A Catalunya, el CatSalut paga el 100% de la despesa hospitalària. Però si el pacient torna als vint-i-nou dies d'haver marxat, l'hospital només rep el 40%. D'aquesta manera, es persegueix que hi hagi una atenció millor”, explica Gavaldà. Amb la informació bàsica de milers de casos d'un hospital, s'ha fet un programa que prediu quin és el risc que el malalt torni abans de trenta dies. “Amb un risc elevat, l'hospital el pot tenir un parell de dies més al llit. Amb un risc baix, es pot fer un seguiment diari a casa per saber si tot va bé, però queda lliure una plaça per a una altra urgència”, continua Gavaldà. “Així s'estalvien recursos públics i es dona benestar al pacient si no cal que estigui hospitalitzat”.

36 Amalfi Analytics (<https://www.amalfianalytics.com/>).

37 “Readmission rate as an indicator of hospital performance. The case of Spain” (https://www.researchgate.net/publication/8267981_Readmission_rate_as_an_indicator_of_hospital_performance_The_case_of_Spain).



Sobre les decisions automatitzades que els algorismes ja poden prendre en l'àmbit de la salut, per a qualsevol diagnòstic de malalties, el professor i coordinador del laboratori de recerca a la UPC ho té clar: “No es tracta de si la màquina ho fa millor que el metge o no, sinó que el metge ho faci millor amb la màquina que sense ella”, sentència Gavaldà.

• Agilitzar el triatge a urgències

Quan una persona acudeix d'urgències a un hospital –sobretot a les grans ciutats– suporta hores d'espera, a la sala d'entrada, en un passadís o al box (reservat). Normalment, després d'una llarga estona és visitada pel metge, que l'envia a fer-se proves. Després d'hores, a vegades tot un dia i una nit, el metge decideix si li assigna un llit.

Què passaria si des del primer moment, en el triatge d'entrada, se sol·licités un llit per als casos més recurrents? Amb els milers d'entrades acumulades a urgències i la seva resolució, un algorisme pot decidir de manera automatitzada quines persones el necessitaran. Entrarien en llista d'espera de llit des del moment del triatge, i no després d'acumular hores d'espera.

Si la màquina encerta, el malalt només ha d'esperar a urgències el temps que triguen les proves encarregades per al diagnòstic. D'aquesta manera, l'hospital guanya en eficiència. Si l'algorisme s'equivoca i el pacient no necessita el llit, ràpidament s'assignarà a un altre malalt.

• Traduir les històries clíniques

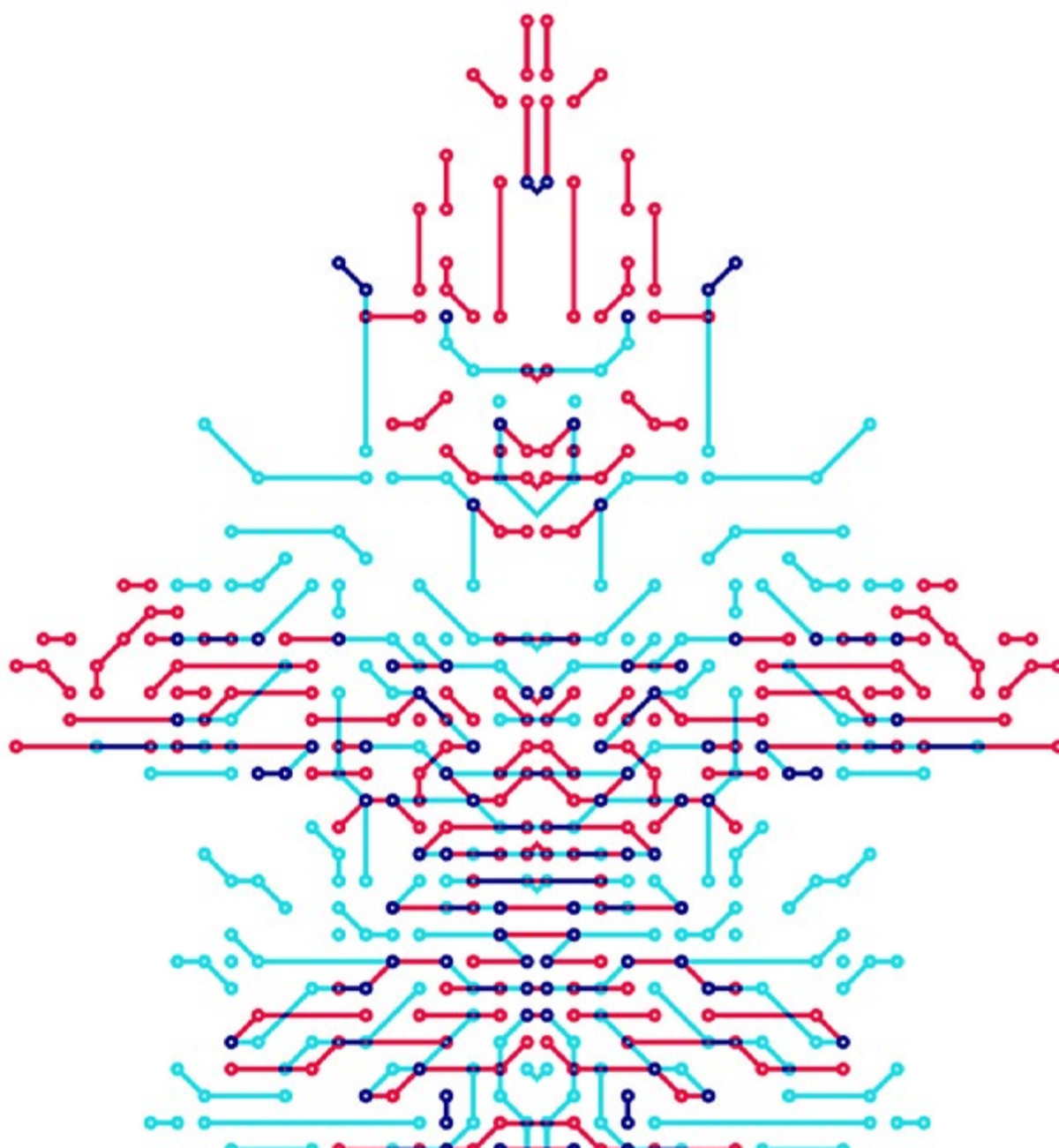
Quan un investigador vol fer una recerca es troba amb el gran problema que el 80% o el 85% dels historials estan en el que s'anomena *format lliure*. És a dir, el pacient explica què li fa mal i el metge ho afegeix al seu historial. El diagnòstic de la malaltia, el tractament, els medicaments subministrats, etc., tot queda en un format incompreensible per a les màquines. I amb la informació en aquest estat no es pot fer cap investigació a gran escala.

L'Hospital de la Vall d'Hebron fa servir un programa³⁸ que automatitza i converteix en dades estructurades les històries clíniques per una anàlisi posterior. Aquesta tasca a mà és molt lenta, i sempre va amb uns anys de retard. “Utilitzem processament de llenguatge natural. És a dir, hem ensenyat a l'ordinador a llegir històries clíniques, a partir de les variables demanades per l'investigador. L'algorisme fa una lectura automatitzada i torna les dades estructurades en un full de càlcul”, explica Gabriel Maeztu, doctor i

38 “IOMED i l'Hospital Vall d'Hebron desenvolupen nova tecnologia per a la gestió d'historials clínics” (<https://www.europapress.es/catalunya/noticia-iomed-vall-dhebron-desarrollan-nueva-tecnologia-gestion-historiales-clinicos-20180211115250.html>).

científic de dades del projecte. Maeztu vol deixar molt clar que “en cap cas, les dades dels pacients surten de l’entorn virtual de l’hospital. Sempre es respecta tota la cadena de la privacitat”.

Una vegada la informació està estructurada es pot predir, per exemple, si un pacient tindrà complicacions o no en passar pel quiròfan. “Per exemple, el departament de traumatologia del Vall d’Hebron fa servir IA per predir què passarà en una operació de túnel carpià. Això es pot saber a partir de les dades estructurades del seu historial clínic, però també amb la d’altres pacients que han passat per la mateixa operació en els darrers anys”, afegeix Maeztu. Amb tota aquesta informació, es fa un pronòstic que, finalment, el metge ha de valorar per decidir si opera o no.



2.4.2. Sistema judicial

L'aplicació de la intel·ligència artificial en l'àmbit judicial és un tema molt delicat perquè –com en l'àmbit de la salut– les decisions que prengui l'algorisme afecten directament la vida de les persones. El ressò d'experiències similars als Estats Units –on el programa Correctional Offender Management Profiling for Alternative Sanctions (COMPASS) decidia la reincidència criminal– tampoc no ajuden a tenir confiança en els resultats de la màquina. El mitjà independent *ProPublica* va delatar³⁹ que el COMPASS –elaborat per l'empresa Equivant–⁴⁰ tenia biaixos que marcaven una probabilitat més alta de cometre crims per als acusats negres que no pas per als blancs.

No obstant això, els sistemes automatitzats han evolucionat molt en els darrers anys. I tot i tenir present els biaixos, també s'hauria de pensar en els prejudicis dels jutges (per motius de raça, gènere, religió, etc.) i com poden influenciar-los a l'hora de prendre una decisió. Per tant, seria de rigor preguntar-nos: qui pot ser més just, un jutge o una màquina?

*Humans decisions and machine predictions*⁴¹ és un interessant estudi nord-americà que demostra com –en decidir atorgar una fiança en un procés judicial– l'aprenentatge automàtic pot funcionar millor que les decisions d'un jutge, fins i tot quan podrien discriminar els negres o els hispans. Els resultats van donar que, quan era molt evident que el risc de reincidència del pres era molt baix, els jutges i l'algorisme coincidien a alliberar-lo sota fiança abans del judici, però que la màquina era més justa que el jutge a l'hora de predir casos amb un risc de reincidència criminal més alt. I això és perquè les màquines són sistemàtiques, fins i tot quan són tan racistes com els jutges.⁴²

A Catalunya, fa uns deu anys que s'apliquen programes similars al COMPASS per detectar la reincidència criminal en adults i en joves. Fins a la data, cap investigació no ha demostrat que hi hagi biaixos perjudicials per als interns. L'investigador Carlos Castillo⁴³ –director del grup d'investigació de Ciència de la Web i Computació Social de la Universitat Pompeu Fabra (UPF)– ha realitzat diverses investigacions sobre aquests sistemes intel·ligents i, segons el seu criteri, funcionen prou bé. “Els tècnics que en fan ús, en última instància, valoren individualment els resultats que ha donat la màquina i decideixen la mesura que cal aplicar”, explica Castillo.

42

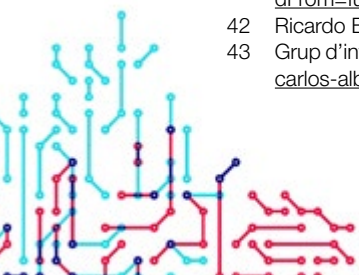
39 “How we analyzed the Compass recidivism algorithm”, *ProPublica* (2016) (<https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm>).

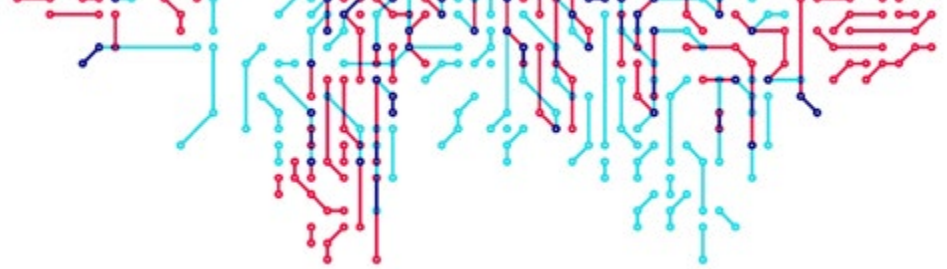
40 Equivant (<https://www.equivant.com/>).

41 *Humans decisions and machine predictions* (<https://academic.oup.com/qje/article-abstract/133/1/237/4095198?redirectedFrom=fulltext>).

42 Ricardo Baeza-Yates, Karma Peiró. “És possible acabar amb els biaixos dels algorismes?”.

43 Grup d'investigació de Ciència de la Web i Computació Social (<https://www.upf.edu/es/web/etic/entry/-/-/24095/adscriccion/carlos-alberto-alejandra-castillo>).





• Predir la reincidència criminal

Els permisos de sortida s'utilitzen per a la reinserció i la rehabilitació dels interns. El fet de poder pronosticar, amb l'eficàcia predictiva més gran possible, la probabilitat de trencament futur d'un permís representa una gran ajuda per als tècnics penitenciaris. El RisCanvi⁴⁴ és un protocol (o eina de valoració del risc) que es va posar en marxa el 2009 a totes les presons de Catalunya, per estimar les possibilitats que una persona torni a delinquir una vegada ha sortit de la presó. El Departament de Justícia de Catalunya va fer un encàrrec al Grup d'Estudis Avançats en Violència⁴⁵ (GEAV) de la Universitat de Barcelona (UB) i, en el temps que porta en funcionament, ja s'ha aplicat a uns vint mil presos.

La predicció de la criminalitat que fan els algorismes és individualitzada i personalitzada. “A les presons –en un moment o altre– els tècnics, directors, psicòlegs o juristes han de prendre la decisió sobre què pot passar en un permís d'un intern o quan aquest està a punt de sortir al carrer per sempre”, explica Antonio Andrés Pueyo,⁴⁶ investigador principal del GEAV i catedràtic de Psicologia de la UB. “La preocupació és si tornarà a cometre un delictes.”

Pueyo explica que tradicionalment es feia una anàlisi, a partir d'uns paràmetres i si l'intern complia certs requisits, i decidien l'actuació en funció dels resultats. “De deu anys ençà es fa de manera automatitzada amb intel·ligència artificial. A partir de 43 variables –que un sistema matemàtic combina a la perfecció–, el tècnic pot prendre la decisió més encertada”, afegeix Pueyo. “Si un pres demana un permís per visitar la família, i la resposta del RisCanvi és d'alta probabilitat de delinquir, alerta al tècnic amb un senyal vermell. El pres surt igualment, però s'activa un seguiment diari, una polsera electrònica o el contacte amb un familiar”, explica el director del GEAV.

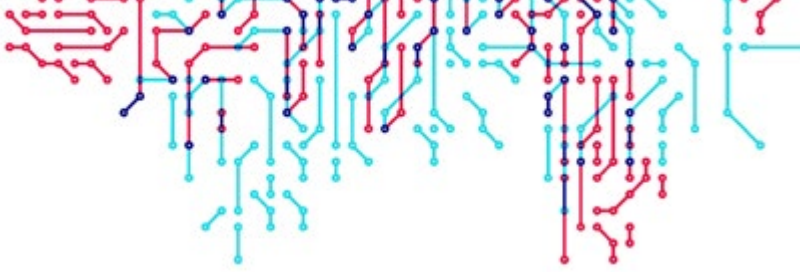
El programa també acumula el comportament violent de l'intern a la presó: si ha agredit algú altre, si ha intentat autolesionar-se, suïcidar-se, etc. És una tasca individual per a cada intern i habitual en el sistema penitenciar, però el tècnic ja no ha de retenir tota aquesta informació, ho fa el sistema automatitzat, que li aporta molt coneixement quan ha de prendre una decisió.

El RisCanvi ja va per la versió 3.0. Cada tres anys s'actualitza i s'hi incorporen millores perquè sigui més precís. Davant de possibles dilemes ètics, Pueyo puntualitza que “la resposta de l'algorisme es validada sempre per la Junta de Tractament, la qual pot: a) mantenir el mateix nivell de risc (baix, alt) de l'algorisme, b) augmentar-lo i/o c) dis-

44 RisCanvi (<http://cejfe.gencat.cat/es/recerca/cataleg/crono/2017/eficacia-del-riscanvi-2017/>).

45 GEAV (<http://www.ub.edu/geav/>).

46 Antonio Andrés Pueyo (http://www.ub.edu/personal/docencia/profes99_2000/pueyoficha.htm).



minuir-lo. En els casos b i c, ha de justificar aquest canvi amb les evidències que el sustenten”.

• Algorismes que detecten la reincidència juvenil

El programa Structured Assessment of Violence Risk in Youth (SAVRY)⁴⁷ funciona amb la mateixa lògica que el RisCanvi, però va ser elaborat el 2003 als Estats Units i no pel GEAV. “S'utilitza a molts països del món: Canadà, EUA i molts altres d'Europa com ara Holanda”, explica Antonio Andrés Pueyo. “Té menys factors de valoració, només 26, i igualment s'aplica de manera molt individualitzada. És més manual que el RisCanvi i la valoració final depèn més del tècnic. Perquè, en el cas dels joves, els canvis de comportament són molt bruscos o van més de pressa que els dels adults. Per tant, ja està bé que sigui així”, afegeix el director del GEAV. “Automatitzar molt més el SAVRY no seria una bona estratègia.”

Quan la valoració del risc de reincidència és alta –tant en el RisCanvi com en el SAVRY–, els equips de tractament decideixen les actuacions o les mesures que cal prendre per evitar que passi el que indica la màquina.

44

• Estadística predictiva per als advocats

Hi ha eines jurisprudencials⁴⁸ que ajuden també els advocats en l'àrdua tasca de llegir sentències i combinar-les per extreure'n noves conclusions. Els models matemàtics poden operar amb informació qualitativa de l'anàlisi de milions de casos, dels jutges que han dictat sentència, de l'aplicació d'articles i lleis, data, lloc, etc. Els algorismes combinen tota la informació procedent d'ordres i instàncies jurisdiccionals de tot l'Estat espanyol i defineixen l'estratègia processal més idònia per a cada cas.

• Orientació per a l'extradició, o no, de migrants

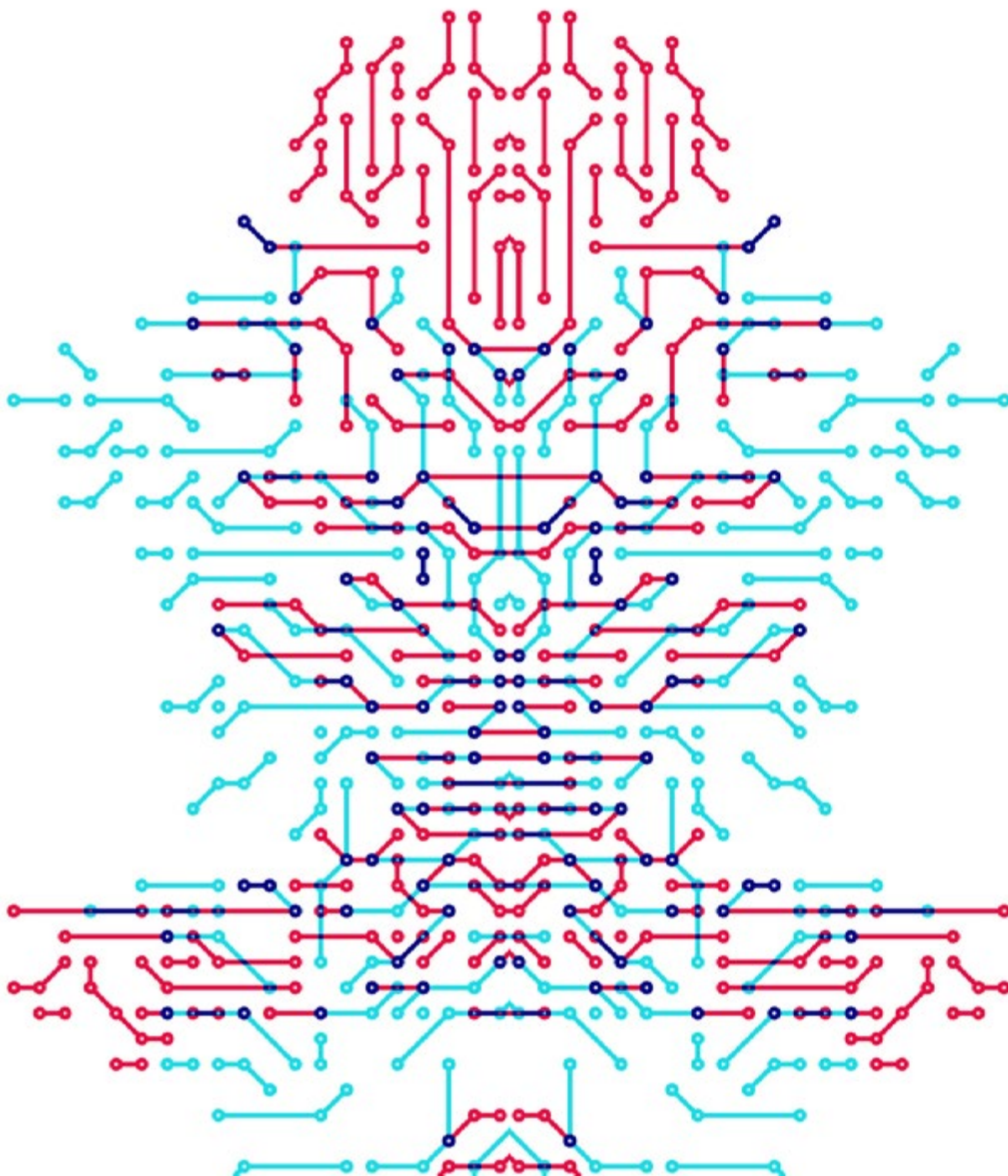
Quan detenen un migrant il·legal i el porten davant del jutge, hi ha certs criteris pels quals aquest pot decidir la no devolució al seu país: risc de mort, risc de contraure malaltia greu, una epidèmia, o risc que sigui sotmès a tortura o un càstig cruel, inhumà o degradant.⁴⁹ El jutge posa una puntuació a aquests criteris i, a partir d'aquí, decideix.

47 SAVRY (https://treballiaferssocials.gencat.cat/web/.content/03ambits_tematics/07infanciaiadolescencia/temes_relacionats/jornades_treball_dgaia_2012/docs_3_maig/valoracio_risc_reincidencia.pdf).

48 Jurimetria (<https://jurimetria.wolterskluwer.es/content/Inicio.aspx>).

49 Migration HR and Governance (https://www.ohchr.org/Documents/Publications/MigrationHR_and_Governance_HR_PUB_15_3_SP.pdf).

El Departament de Matemàtiques i Informàtica de la UB va elaborar un projecte amb tots els casos que havien passat pel Tribunal Suprem i va crear un model matemàtic. Jordi Vitrià, investigador de la UB, explica que l'eina dona a l'advocat una informació molt valuosa per a les associacions d'ajuda als migrants. "Les dades són totes de casos reals, i l'eina facilita una referència sobre per on podria anar l'extradició", afegix Vitrià. "És una referència per a l'advocat i ho podria ser també per al jutge, perquè el model està creat en funció dels casos del Suprem. Però està clar que el jutge té l'última decisió." Per ara no s'ha posat en pràctica a Catalunya.



2.4.3. Educació

La intel·ligència artificial ha revolucionat totes les indústries, i l'educació és un sector que no es queda al marge. El passat mes de maig, la UNESCO va reunir 50 ministres a Pequín⁵⁰ per consensuar el document: *Artificial Intelligence in Education. Challenges and Opportunities for Sustainable Development*.⁵¹ Entre els punts principals hi ha el missatge que la IA té el potencial de transformar profundament l'educació, sempre respectant els drets humans i els valors socials. I que les polítiques educatives s'haurien de planificar amb aquest enfocament, de manera que el jovent es pugui apoderar per al seu futur. S'ha acabat estudiar només de petit. L'aprenentatge és un fet que s'allargarà tota la vida.

La IA ajudarà moltíssim en l'àmbit de l'educació perquè –segons comenten els experts– es passarà de l'aprenentatge en tasques a l'aprenentatge basat en la col·laboració. D'altra banda, el tractament i l'anàlisi que es pot fer avui dia de les avaluacions, de plataformes d'aprenentatge a distància, de la interacció a classe amb el mòbil o de la navegació per webs educatius permet imaginar nous mètodes d'estudi. Els algorismes ja poden establir-hi relacions i adaptar-se a cada estudiant predient la millor manera d'adquirir coneixement.

A les aules d'estudis primaris i secundaris, el sistema automatitzat pot ser una guia per al professor. I en l'àmbit universitari pot servir per gestionar la recerca i per fer més eficients els recursos de la facultat.

46

• Aprenentatge actiu

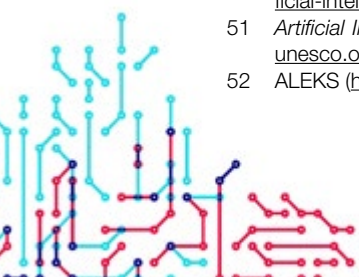
El programa Assessment and Learning in Knowledge Spaces (ALEKS)⁵² s'ha provat als Estats Units durant més de dotze anys amb milions d'estudiants, i ara arriba a les escoles catalanes. “No s'ha de tenir por a introduir la IA a les aules, pot convertir-se en una eina que millori el procés d'aprenentatge”, explica Carles Sierra, director de l'Institut d'Investigació en Intel·ligència Artificial (IIIA).

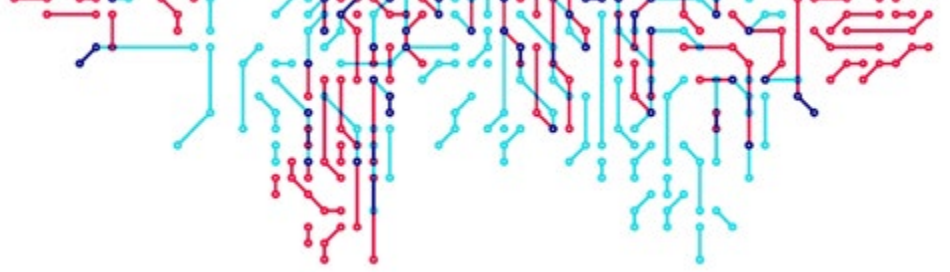
ALEKS utilitza preguntes adaptatives per determinar, d'una manera ràpida i precisa, el que un estudiant sap o no sap sobre una matèria. També aconsella a l'estudiant sobre els temes per als quals està més preparat. Durant tot el curs, el programa va reavaluant periòdicament l'estudiant. Serveix per planificar un aprenentatge personalitzat a cada alumne, o per a aquells que tinguin més problemes per aprendre una matèria determinada.

50 Internacional Conference on Artificial Intelligence and Education (<https://en.unesco.org/events/international-conference-artificial-intelligence-and-education>).

51 *Artificial Intelligence in Education. Challenges and Opportunities for Sustainable Development*. UNESCO (<https://unesdoc.unesco.org/ark:/48223/pf0000366994>).

52 ALEKS (<https://www.aleks.com/>).





“El sistema fa un model de cada estudiant, i sap quin és el seu estat de coneixement”, afegeix Sierra. El sistema aporta un gràfic al professor perquè faci seguiment de l’adquisició dels conceptes que assoleix l’alumne. “S’ha de tenir en compte sempre que és un suport al professor a l’aula. Estudis d’avaluació han demostrat que, amb aquesta tecnologia, el grau d’abandonament és menor a la universitat i que les notes a la secundària milloren”, conclou Sierra.

• L’algorisme fa els grups de classe

En una classe amb 30 alumnes que han de dur a terme una tasca per equips, un sistema automatitzat agrupa els estudiants en funció de les seves personalitats. “El professor defineix el nombre d’estudiants per equip i les competències que calen per a aquesta tasca. La IA agrupa i assigna una responsabilitat a cada estudiant”, explica Carles Sierra.

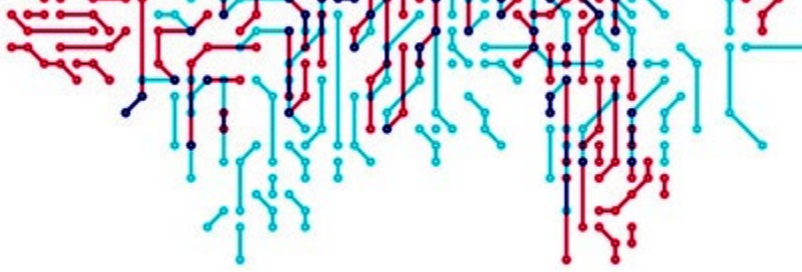
Però com pot una màquina fer els equips per dur a terme una tasca sense equivocar-se en l’agrupació? “Primer s’avaluen les personalitats a partir de 20 preguntes, i el sistema intel·ligent els classifica: introvertit-extravertit; assenyat-intuïtiu; analític-emotiu; resolutiu-reflexiu”, respon Carles Sierra. “En funció del resultat, el sistema automatitzat els agruparà de la manera més diversa i eficient per resoldre l’exercici proposat pel professor. El resultat de l’algorisme podria servir per reforçar, o no, l’opinió que el professor té dels alumnes. El que s’intenta, segons el director de l’IIIA, és buscar equips equilibrats i que tots siguin prou bons. “Aquesta és la part difícil. Trobar un equip bo és fàcil, però que tots siguin mínimament bons és molt complicat. I això és el que fa l’algorisme.”

Diversos instituts de Catalunya ja han provat aquesta tecnologia, i la millora del rendiment va oscil·lar entre un 25% i un 30%, segons explica Carles Sierra. Èticament, podrien sorgir dubtes d’aquestes aplicacions. No hi ha el risc que la màquina s’equivocui i que un alumne en surti perjudicat? “Els resultats que tenim ara és que, de mitjana, funciona bé. Això no vol dir que no hagi errors en algun cas en concret. Però cal tenir en compte que els professors tampoc no estan lliures de biaixos i també es poden equivocar en l’agrupació dels alumnes”, respon l’investigador.

• Correcció d’exàmens

Una altra tècnica que l’Institut d’Investigació en Intel·ligència Artificial ha dut a la pràctica –amb alumnes de secundària– és el que es coneix com a *avaluació entre parells*.

Una eina té definida una rúbrica de com s’ha d’avaluar els alumnes. El professor només avalua uns quants estudiants, però finalment tota la classe té la seva nota personalitzada. Com pot ser?



“Cada alumne té un grup de companys als quals ha d’avaluar. Què fem? Comparem la manera d’avaluar dels alumnes, i això dona un nivell de similitud en l’avaluació”, continua explicant Carles Sierra.⁵³ Com pot un alumne tenir prou criteri per avaluar a un company? “El sistema intel·ligent crea similituds entre la manera d’avaluar del professor i la dels alumnes. Els algorismes creen un nivell de confiança entre les valoracions del professor i les dels estudiants avaluadors.”

Aquesta experiment s’ha fet en una classe d’anglès i els resultats no s’allunyaven més d’un 10% dels resultats que hauria donat el professor.

• Superar la dislèxia

Des de fa un parell d’anys, unes setanta escoles de Catalunya disposen d’un sistema d’intel·ligència artificial per ajudar els alumnes amb dislèxia o dificultats de lectura. En total, se’n beneficien uns tres mil nens i nenes –d’entre 5 i 12 anys– i uns dos-cents més que, a títol particular, en fan un seguiment intensiu.

UBinding⁵⁴ és una plataforma d’aprenentatge basada en el desenvolupament cognitiu de cada infant, la família i l’entorn escolar. Un grup d’investigadors de la Universitat de Barcelona (UB) va desenvolupar-la el 2007. Segons expliquen els seus promotors, el mètode UBinding aconsegueix ajudar el 90% dels alumnes a millorar la fluïdesa lectora en una mitjana d’uns 7-8 mesos.

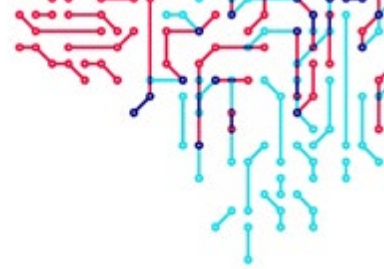
“Es tracta d’un algorisme de suport a la decisió. Un equip format per psicòlegs, logopedes i matemàtics especialitzats en els trastorns de l’aprenentatge fan un seguiment diari”, explica Jorge López, responsable d’àrea d’investigació i desenvolupament. Igualment, també es podria aplicar a nens amb dèficit d’atenció perquè fallen en funcions executives, o per reforçar el repàs d’algunes assignatures.

“L’entrenament permet adaptar-se a cada nen o nena en funció de la velocitat a la qual assimila la lectura”, afegeix la matemàtica Adina Nedelea. “Tenint milers d’alumnes, resulta molt complicat fer el seguiment individualitzat sense l’ajuda de la intel·ligència artificial. L’algorisme mai no decideix, però suggereix com serà la sessió següent de l’estudiant, i el professional valora si la fa o no la fa.”

L’equip d’UBinding disposa del *feedback* dels pares/mares, i del mateix nen o nena que ha fet els exercicis de lectura. “Sense els algorismes no hauríem pogut arribar a milers d’alumnes”, conclou Nedelea.

53 Carles Sierra. “Intel·ligència artificial i educació” (vídeo) (<https://www.youtube.com/watch?v=Xd5ZoKdl33A>).

54 UBinding (<http://www.ubinding.cat/>).



• Assistents virtuals a l'aula

Els programes més avançats en aprenentatge automàtic i profund poden crear un mètode d'estudi personalitzat per a cada alumne, a partir de les dades generades per la seva participació en l'aula o a distància. L'algorisme recomana el tram d'estudi d'acord amb el ritme, les aptituds i els objectius de cada alumne. També pot suggerir els millors companys per fer un treball, a partir de l'anàlisi de les compatibilitats i les habilitats de cada membre del grup. “En pocs anys, l'aula es transformarà: el professor disposarà d'un assistent virtual que l'ajudarà en les seves tasques, els alumnes tindran un programa personalitzat i l'assistència a classe serà presencial o virtual”, explica Ostrowicz,⁵⁵ expert en intel·ligència artificial aplicada al sector educatiu.

Estudiar tots els alumnes amb el mateix llibre de text, ja és antic. Segons Ostrowicz, els algorismes recomanaran continguts per a cada alumne d'acord amb el seu nivell; quadres de comandament que ajudaran el mestre a verificar l'avenç de cada alumne i assistents virtuals perquè els alumnes puguin resoldre dubtes en tot moment. Igualment, tindran un recomanador dels millors estudiants amb els quals es pot aprendre, i els temaris adaptats a cadascú. Però el més singular de totes aquestes promeses és que un algorisme serà capaç d'alertar l'alumne quan estigui a punt d'oblidar el coneixement après. El sistema de retenció adaptativa mesura la velocitat de l'oblit i recomana una revisió de la lliçó just abans que es perdi el que s'havia memoritzat.

49

• Reconeixement facial per entrar a l'institut⁵⁶

Aquest exemple no és una aplicació de la IA per millorar l'aprenentatge, sinó un sistema de reconeixement facial que controla la presència dels alumnes en un centre educatiu.

Des de feia set anys que funcionava en un institut de secundària de Barcelona i la majoria de famílies ho veien amb bons ulls. “El primer any em va xocar, però va molt bé perquè si el teu fill no passa pel sistema de reconeixement facial, envien un SMS als pares”, explica una de les mares en una entrevista de TV3. A través de diferents càmeres instal·lades al centre, els estudiants havien de “fitxar” en arribar a l'escola; si no, s'advertia de la falta del menor als pares. El sistema no controlava l'assistència a classe, tot i que si l'alumne era a l'institut segurament aniria a classe.

Segons la professora de Dret i Ciència Política Mònica Vilasau Solana,⁵⁷ les dades biomètriques i les dades de menors estan marcades com a especialment sensibles.

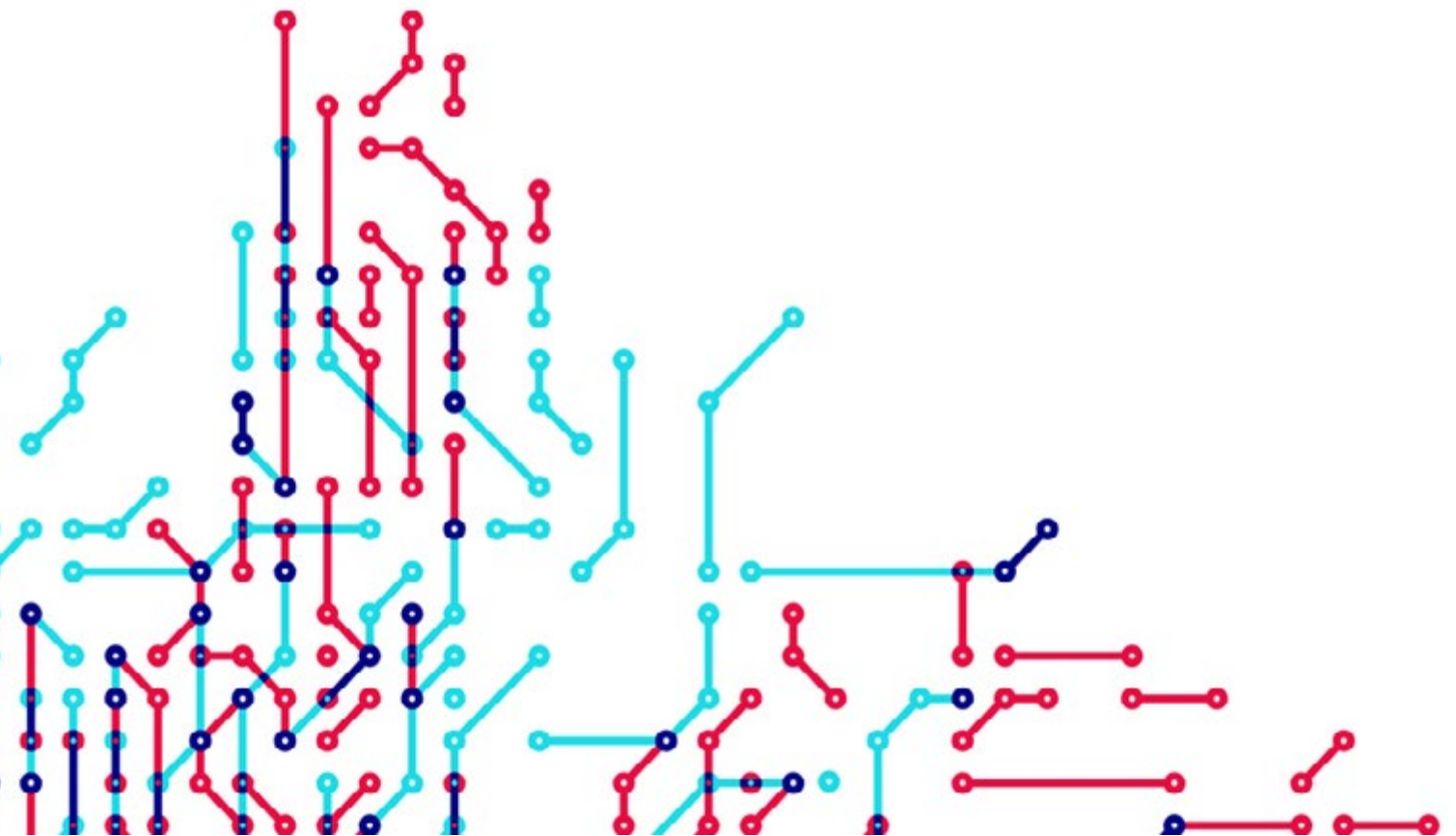
55 Ivan Ostrowicz (https://www.linkedin.com/in/ivanostrowicz/?locale=fr_FR).

56 Notícia publicada a 324.cat (5/10/19) (<https://www.ccma.cat/324/reconeixement-facial-per-passar-llista-en-un-institut-de-barcelona/noticia/2952712/>).

57 Mònica Vilasau (<https://www.uoc.edu/portal/es/news/kit-premsa/guia-experts/directori/monica-vilasau.html>).

Per aquest motiu, des de l'Autoritat Catalana de Protecció de Dades se li va obrir una investigació. “No hi ha prou amb el consentiment dels pares, sinó que s’ha d’avaluar l’impacte que pot tenir fer-les servir i mirar si no hi ha cap altra alternativa disponible”, explica Vilasau.

Altres centres educatius catalans havien implementat el reconeixement facial dels alumnes, però amb el Reglament de Protecció de Dades europeu van deshabilitar la instal·lació. Aquest institut de Badalona també el va acabar retirant arran de la investigació oberta per l'APDCAT. L'exemple xoca amb el vessant més ètic de la IA: en aquests casos es feia servir la tecnologia per controlar i no per millorar el rendiment en la formació dels joves.





2.4.4. Banca

Els bancs sempre han estat els primers a aplicar els algorismes de decisió automatitzada. Ja des dels anys setanta, amb operacions molt més simples que les actuals. Ara són xarxes neuronals avançades que ofereixen una varietat enorme de productes financers en funció de les dades que s'hi introdueixen: promocionar o no un servei a través de la bústia de correu electrònic d'un client o recomanar-li productes quan visita el web de l'entitat financera són accions habituals. La intel·ligència artificial també s'utilitza per a tasques internes per guanyar eficiència i rendiment en una entitat bancària, per gestionar les trucades dels clients, per aconseguir uns temps d'espera telefònica acceptables, per dissenyar el tracte preferencial dels clients prèmium o per decidir a quina part de la ciutat ubicar noves oficines.

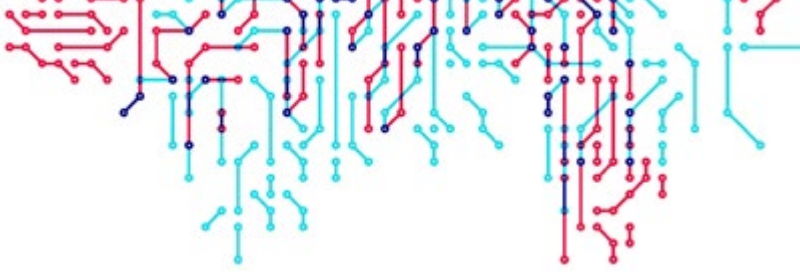
Els bancs estan molt controlats respecte a la privacitat i anonimització de les dades que nodreixen els seus algorismes. Per tant, d'entrada, impera la lògica de no perjudicar cap client amb possibles discriminacions generades pels biaixos. “Alhora s'ha d'entendre que un banc viu de discriminar, tot i que el verb té una connotació negativa”, explica Marco Bressan, ex-científic de dades en cap d'una entitat bancària d'àmbit estatal. “L'entitat ha de saber si recuperarà els diners d'una hipoteca o d'un préstec. La utilització d'algorismes per decidir a qui els concedeixen o no és bàsica. Per la seva banda, el banc ha de detectar i mitigar els possibles biaixos (de gènere, nivell de renda, origen del client, etc.) que hi pugui haver.”

Bressan està convençut que els algorismes –si estan ben dissenyats– són més precisos que les persones. “Imaginem un director d'un banc misogin: segurament prendrà decisions molt equivocades pels prejudicis personals que pot tenir. Els sistemes automatitzats responen a diferents variables i la decisió que prenen és més justa”, afegeix. “Les dades amb les quals s'alimenten els algorismes són el més important. Si d'entrada hi ha una discriminació, la perpetuaran en el temps i afectarà molta més gent.”

D'altra banda, els humans –segons Bressan– ja no arribem a calcular la quantitat d'operacions en temps real que poden fer les màquines amb dades massives. “No obstant això, com que actuen a gran escala, només que l'algorisme falli una vegada perjudica milers o milions de persones de cop. En canvi, el director del banc quan s'equivocava afectava com a molt cent persones.”

• El crèdit, per a qui tingui millor puntuació

Les accions més regulades d'una entitat financera són les d'accés a crèdit. L'algorisme que decideix a quin tipus de clients se li concedeix o se li denega, a priori, està validat i regulat pel Banc Central de cada país. Però el client desconeix aquesta valoració.



Com decideix un algorisme concedir o no un crèdit? Quan un banc presta diners s'assegura que els recuperarà. Abans es recopilava informació del client, i el director de l'entitat decidia. Amb tecnologies més potents, la predicció dels diners que es recuperaran o no és més encertada. Se sap la quantitat de diners que la persona guanya, i l'estalvi que pot fer cada mes després de pagar les despeses. Es classifica els clients en grups (o *clusters*) i se'ls assigna una puntuació (o *scoring*) que no està necessàriament relacionada amb el que guanyen. Una persona amb un salari d'uns 1.500 euros mensuals pot estalviar-ne 100 i tenir una puntuació més alta que una altra que guanyi 4.500 euros i no sigui capaç ni d'estalviar 50.

Com s'entrenen aquests algorismes perquè siguin eficaços? Posem que tenim les dades de milers de clients als quals se'ls ha concedit crèdits en el passat. En funció d'unes variables, es programa l'algorisme perquè predigui si aquestes persones són prou solvents per tornar-lo o no. Com que són dades del passat, són fàcilment comparables per saber si el sistema automatitzat està funcionant de manera correcta. Ja ha quedat demostrat que la precisió sempre és més gran amb un sistema automatitzat que fent-lo manualment.

52

• Contractació d'assegurances

Els sistemes de *scoring* (puntuació) són freqüents en la contractació d'assegurances (de vehicles, salut, llar, empresa, etc.). Per lògica, la probabilitat que un jove tingui un accident de trànsit és més alta que la d'una persona de 55 anys, i, per tant, el cost de l'assegurança també serà més elevat.

Podríem preguntar si aquests sistemes automatitzats no s'equivoquen mai: la resposta és que l'error o l'encert dependrà de la informació prèvia (dades massives) que s'introdueixi en l'algorisme. Quan les prediccions sobre grups s'equivoquen, en surt perjudicada més gent. Es pot donar el cas que un jove sigui més prudent que un home de més edat, i que els joves sempre acabin pagant més. Els algorismes de les asseguradores tenen en compte altres variables per decidir la prima, com ara la renda. S'ha demostrat que les persones amb una renda més baixa –normalment– tenen més sinistralitat que les de renda alta. L'orografia també es té en compte en el preu que han de pagar perquè condiciona molt els accidents.

• Predictors d'hipoteques

A mesura que la intel·ligència artificial és més potent, els algorismes són més matemàtics (i menys estadístics). "Això ha fet que passem de l'anàlisi descriptiva a la predictiva", explica Pier Paolo Rossi, director d'anàlisi i màrqueting d'una entitat bancària cata-

lana. “S’estudia el passat per predir el futur. L’algorisme analitza, per exemple, tots els clients que han demanat hipoteca en els darrers cinc anys i, a partir d’unes variables, prediu els que potencialment demanaran una hipoteca aviat.”

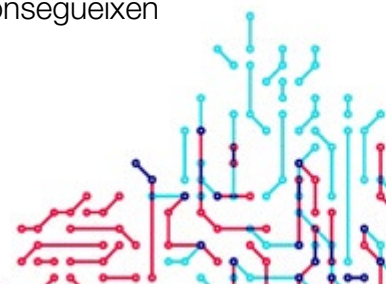
Les variables poden ser de tipus sociodemogràfic (edat, gènere, fills, etc.), de productes financers contractats (compte corrent, hipoteques, préstecs, etc.), de com interactua amb el banc (si té o no targeta de crèdit, si la fa servir molt o poc, si sempre treu diners del caixer o va a l’oficina, etc.). La relació amb els bancs als pobles és més propera que a les ciutats, i aquesta és una altra variable. Amb les dades obtingudes s’entrenen els algorismes que faran les prediccions, amb una probabilitat molt alta d’encert.

L’analítica prescriptiva també comença a implementar-se. Els algorismes identifiquen les necessitats dels clients en funció del seu cycle de vida. No és el mateix una parella de 25 anys, que viu en un pis de lloguer, sense fills; que la mateixa parella deu anys més tard. Si tot els ha anat bé, segurament tindran un sou més alt, voldran un pis de compra, tindran fills i les necessitats seran diferents. “Amb aquesta analítica, hi guanyen el banc i el client, perquè abans s’oferien hipoteques a tothom i ara només a aquelles persones que les necessiten”, afegix Pier Paolo Rossi. Les dades no fallen.

• **Recomanadors financers**

Com ja ha quedat clar en els exemples anteriors, els bancs juguen amb avantatge respecte a altres empreses perquè tenen molta informació dels seus clients. Les dades són diners. Per saber quins productes financers els interessaran, només n’han de fer una *anàlisi descriptiva*: és a dir, estudiar com són, quants clients arriben econòmicament bé a final de mes, quants són de la zona alta de la ciutat, de quines edats són, quin ritme de vida porten, etc. Els algorismes ajuden els clients a prendre decisions financeres, a comprar a crèdit, a gestionar les despeses de la llar o la vida familiar, a contractar assegurances, etc.

El negoci dels bancs ha canviat considerablement en els darrers anys i ja no consisteix només a prestar i guardar diners, sinó a oferir serveis que els diferenciïn. De quina manera? Amb la informació relacionada dels hàbits de vida de milers de persones: des de la zona de la ciutat on viu, fins al tipus de feina, el sou, els membres de la família, els menors al seu càrrec, etc. Però també amb dades de consum: de la llum, aigua, gas o qualsevol altra despesa domiciliada, les compres diàries, despeses de viatges, estudis, activitats extraescolars, vacances, reserves d’avions i hotels de feina o d’oci. Amb totes aquestes dades, els algorismes fan una radiografia integral de la vida de cada persona, i els ofereixen els serveis que *els poden interessar*. D’aquesta manera, aconsegueixen fidelitzar al màxim cada client.



• **Coaching financier**

Com que les dades de consum són les més valuoses, amb elles es fan anàlisis del rendiment anonimitzats, de manera que es classifica el client en un grup determinat. Així es detecten productes que tots ells comprarien. En l'exemple anterior ja hem vist les variables de nivell econòmic i despesa mensual que es poden tenir en compte.

Els bancs envien i/o fan ofertes molt personalitzades de compra que, amb molt poca probabilitat el client rebutjarà, perquè s'estalviarà diners en un servei que utilitza o serà un avantatge familiar en poc temps. És a dir, si una persona de la mateixa edat que jo, amb el mateix sou i amb la mateixa despesa té un wifi contractat a casa molt més econòmic que jo, el banc m'oferirà que canviï de companyia i que contracti un de nou mitjançant l'entitat financera. Aquestes ofertes personalitzades van destinades a centenars o milers de clients.

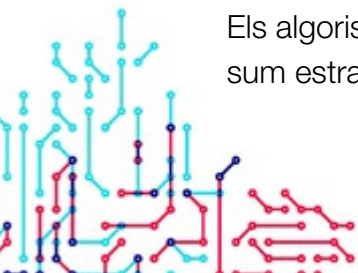
El recomanador financier automatitzat substitueix el gestor tradicional del banc, aquella persona en la qual es confiava per obrir un pla de pensions o fer un estalvi a llarg termini. Posem que una persona té 10.000 euros que vol guardar per a la seva jubilació o invertir per treure'n un rendiment. L'algorisme el classificarà segons si té un perfil de risc "agressiu" (arriscarà molt) o si és més aviat "conservador" (poc risc), i aconseguirà que la proposta del banc s'adapti a les seves preferències. Aquests tipus d'assistents sense intervenció humana ajuden tant al personal del banc com al client: alerten el client si té risc de perdre els seus diners i, encara que en aquesta operació el banc no guanyi res, el fidelitza.

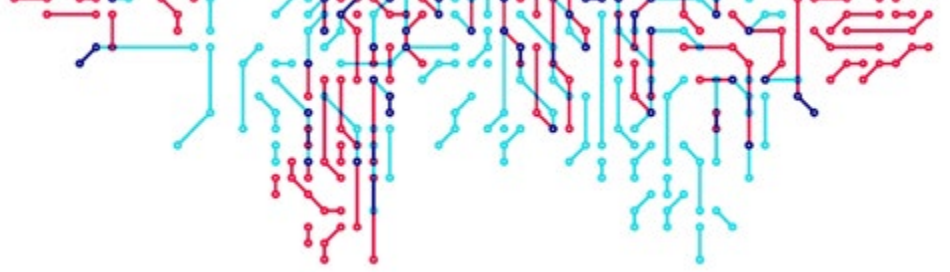
Són accions destinades a recuperar la confiança en els bancs, molt malmesa després de la crisi econòmica. Una altra operació que fidelitza és avisar el client que es quedarà en números vermells en molt poc temps. Com que el sistema automàtic monitora els comptes, el pot alertar d'una situació incòmoda que tothom es voldria estalviar. I aconsella frenar certes despeses de restaurants, oci, compres secundàries, etc.

• **Frau en targeta de crèdit**

Alguna vegada li ha passat que han rebutjat la seva targeta de crèdit en una botiga on no havia comprat mai? O que se li ha bloquejat perquè havia comprat més de l'habitual? Encara que no en som conscients, això és més habitual del que podríem imaginar: cada dia es bloquegen milions de targetes a tot el món.

L'ús de l'aprenentatge automàtic per detectar fraus financers es fa servir des dels anys noranta del segle passat i avui ja està molt avançat, però encara no és del tot perfecte. Els algorismes estan entrenats per bloquejar targetes de crèdit quan detecten un consum estrany o exagerat. I ho fan monitorant milions de transaccions diàries.





Però què passa si el sistema automatitzat s'equivoca en considerar estranya una operació que no ho és? Això es coneix com a *fals positiu*; és a dir, errar en la predicció i en la decisió de bloquejar la targeta. Un estudi realitzat el 2015 per la consultora Javelin Strategy Research⁵⁸ va estimar que només una de cada cinc prediccions de frau és correcta, i que els errors al banc li podien costar milers de milions en ingressos perduts perquè les persones que han vist la seva targeta bloquejada ja no la tornen a fer servir. Tot i que la detecció del frau automatitzat està molt avançada, encara presenta algunes limitacions, ja que, per detectar casos fraudulents reals, cal que els algorismes s'equivoquin moltes vegades.

Ara, per evitar més frustracions als clients i maldecaps als bancs, investigadors del Massachusetts Institute of Technology (MIT) fan servir una tècnica⁵⁹ –que ja s'aplica també en alguns bancs ubicats a Catalunya– per reduir gairebé a la meitat la possibilitat d'error. Aquesta tècnica es fixa en altres característiques de la compra que fins ara no s'havien tingut en compte, com ara la distància entre dos comerços i l'hora en què s'ha produït la compra, a més de si s'han fet les dues de manera presencial o en línia.

• Algorismes que interactuen amb el client

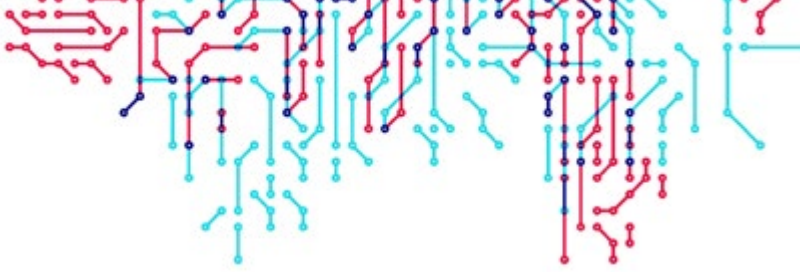
55

Els *assistents virtuals en línia* o *xatbots* són algorismes que interactuen directament amb els clients, simulant un gestor de banc. Es recull la conversa escrita en forma de xat, com els moviments que fa per les diferents opcions de la pàgina web. Les galetes o *cookies* (programes que segueixen el rastre) permeten al banc saber per on s'ha desplaçat una persona en l'estona que ha visitat el web.

“En funció del moviment, de les preguntes que ha fet al *xatbot*, de les respostes que ha donat, l'algorisme va aprenent i pot personalitzar a la pantalla del client, en aquell moment, el que necessita. Així sabem si està interessat en un préstec, si és client o no, quina quantitat necessita, per a quin ús (viatge, casa, estudis, cotxe), i l'algorisme personalitza la resposta al moment”, explica Pier Paolo Rossi. Aquest sistema és el que fa servir també Amazon o qualsevol botiga en línia. “Si coneixes el teu client, tens un avantatge respecte als altres. I generes una confiança que abans no tenies”, conclou.

58 2015 Data Breach Fraud Impact Report (<https://www.javelinstrategy.com/coverage-area/2015-data-breach-fraud-impact-report>).

59 Solving the false positives problem in fraud prediction using automated feature engineering (<http://www.ecmlpkdd2018.org/wp-content/uploads/2018/09/567.pdf>).



2.4.5. Comerç

Sense adonar-nos-en, estem envoltats d'algorismes que ens suggereixen productes, recomanen una sèrie, ens lloguen un pis i ens ofereixen un cotxe de segona mà, etc. I sempre encerten els nostres gustos.

D'igual manera, hem introduït a les nostres vides petits aparells que ens saluden a primera hora del matí, o amb els quals interactuem per demanar-los música, que encenguin el televisor mentre cuinem, o que ens diguin la previsió del temps. Són els assistents virtuals, microintel·ligències que entenen el llenguatge natural. Els trobem també en format *xatbot*, sovint en pàgines webs i interpreten el que un client o un ciutadà els pregunta. Amb la informació que reben, decideixen la resposta. Són sistemes que estalvien temps al servei d'atenció al client d'una empresa, però també substitueixen treballadors que, en un passat no tan llunyà, atendien trucades i resolien dubtes per telèfon o per correu electrònic. Però el que tenim avui són intel·ligències artificials molt prematures. No hem vist res del que està per arribar.

56

A Catalunya, les multinacionals tecnològiques copen el mercat, i ser competitiu en intel·ligència artificial no és fàcil, explica Jordi Navarro⁶⁰ –CEO d'una empresa dedicada a l'anàlisi predictiva i el *machine learning* des de fa quatre anys. Ell considera que el sector turístic encara té camí per recórrer. “A partir d'informació històrica, podem entendre el passat i extreure'n patrons. Això ens permet operar en el sector hotelier, en les cancel·lacions d'habitacions d'hotel, per exemple. Amb la IA, d'alguna manera, és com si tinguéssim una bola de vidre, que ens permet actuar amb un grau prou alt d'encert. El comportament humà és molt previsible. Si no, per què triomfa el líder mundial del comerç en línia?”, pregunta.

A Jordi Mas –membre de Softcatalà i pioner de la Internet catalana–, li preocupa la part ètica dels sistemes d'aprenentatge profund que es fan servir en àmbits comercials, sobretot perquè no es pot explicar com una màquina pren certes decisions. “Em preocupa que no hi hagi un codi deontològic respectat per tothom. Per exemple, cap professional no hauria de treballar en un projecte comercial que no respecti les convencions de drets humans internacionals. També em preocupa que la regulació vagi per darrere de la tecnologia, i que siguem molt reactius. Com el que està passant ara amb el reconeixement facial, que no som conscients que les dades biomètriques tenen molt valor. Els algorismes trien a cada moment el contingut que llegim, mirem o escoltem, els productes que acabem comprant, filtren allò que no ens interessa, i mil coses més. Hi ha el perill de crear bombolles reduccionistes, en un món que és complex i amb moltes opcions.

60 Jordi Navarro (<https://www.linkedin.com/in/jordi-navarro-perez/>).

Em preocupen les empreses poc ètiques, però també, i molt, els governs. Només cal pensar que molt del que tenim ens arriba dels EUA o de la Xina.”

• Cancel·lació d'habitacions

El 2017, un passatger de l' aerolínia United Airlines⁶¹ va acaparar totes les mirades en ser expulsat, per la força, d'un avió. Hi havia sobrerreserva o *overbooking* (la venda de passatges superava la capacitat de l'aparell) i algú havia de quedar-se a terra. Li va tocar al client que menys milles (o vols) acumulava. Un altre passatger va gravar l'incident, i el vídeo va fer la volta al món. La companyia va reemborsar el preu dels bitllets a tots els que viatjaven en el desafortunat vol 3411.

Aquesta situació no és nova. L'*overbooking* sempre ha existit. Abans, es feia a ull, per intuïció o coneixement del sector, i ara es fa amb les matemàtiques i la IA. Jordi Navarro explica que la mateixa estratègia s'empra en el sector hotelier, amb les reserves i les cancel·lacions d'habitacions. “Ara ens equivoquem menys. Tot està basat en estadística a partir del perfil de la persona que ha fet la reserva, el comportament en anteriors ocasions i altres variables. L'algorisme és capaç de predir si cancel·larà a última hora o no aquella habitació. L'amo de l'hotel pot oferir places, però no perd diners”, afegeix Navarro.

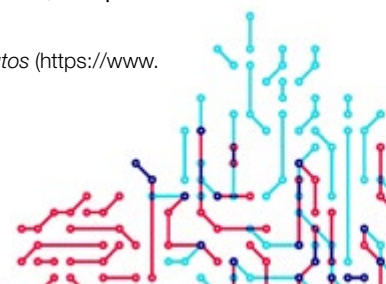
57

• Un clic no vol dir un client

Sovint, quan cerquem un allotjament a una destinació turística, accedim als hotels des d'un portal o un web de preus intermediari. Introduïm al cercador la ciutat, la quantitat de nits i ens dona un resultat. Aquests webs intermediaris són majoristes que ofereixen tot el que els establiments tenen, com les agències de viatges d'abans. Quin és el negoci del majorista? Els anuncis dels hotels. Per exemple, si algú busca passar dues nits a Menorca, el web majorista oferirà una llista d'allotjaments amb els diferents preus. Només pel fet que es cliqui en alguna de les ofertes, el propietari de l'establiment ja paga al majorista. Si després no es reserva l'habitació, ha perdut diners. I això passa milers de vegades cada dia. Per tant, l'hotelier ha d'estar molt convençut que el que ofereix al majorista és prou atractiu perquè els clics suposin reserves d'habitacions.

Ara, amb la IA, es pot encertar molt més –en funció de la ciutat, de l'hora, de si demanen una o dues places, de les dates, etc.– i assignar una probabilitat alta de reserva. Així, l'hotelier només ofereix al majorista una habitació a Menorca quan sap gairebé del cert que el client se la quedarà. Això es podria fer sense algorismes? “No”, respon

61 “United Airlines reemborsará los billetes a todos los clientes del vuelo del que expulsó un pasajero”. *20 Minutos* (<https://www.20minutos.es/noticia/3011324/0/united-airlines-devolvera-billetes-pasajeros/>).



Jordi Vitrià, investigador del Departament de Matemàtiques i Informàtica de la UB, des del qual han col·laborat en aquest projecte. “És una qüestió d’escala. Abans, l’agència de viatges de tota la vida tenia contactes amb hotels, empreses de lloguer de cotxe, avions, etc., i es quedava un marge per tot allò que oferia i venia. Aquest seria el paper del majorista. Ara, en temps real, les 24 hores del dia, és impossible fer-ho personalment.”

És ètic que, si et connectes des del barri més ric de la ciutat, el preu d’aquella habitació et costi el doble? “*Business is business*”, afegeix Vitrià. “Tot i que no jugues amb unes regles clares, perquè ningú no t’ha explicat que t’han agafat les dades de geolocalització, i que això pot influir en el preu final que pagues”.

• La ITV més assequible

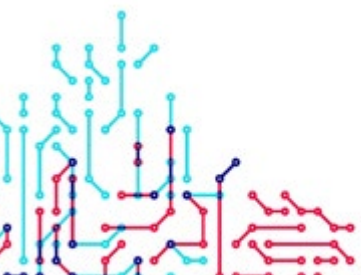
58

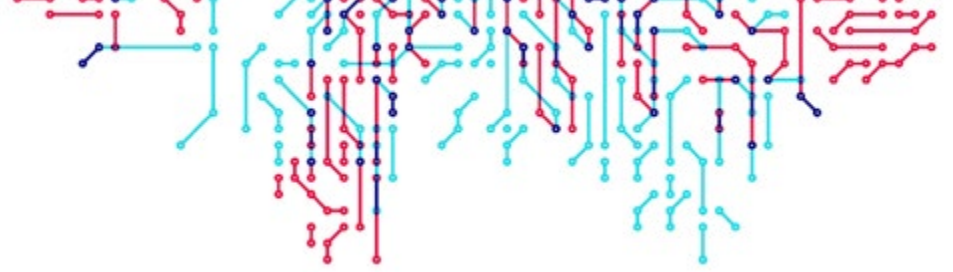
A França, el preu que es paga per passar la inspecció tècnica dels vehicles està liberalitzat. Com que hi ha molta competència entre les franquícies que revisen els cotxes, la Universitat de Barcelona va col·laborar en un projecte per establir preus automàtics en funció del lloc on estigués ubicat el centre de la revisió. “No és el mateix si tens set centres d’ITV de la competència, que si ets l’únic a la zona”, explica Jordi Vitrià. “Vam fer un algorisme que, a partir de totes les dades d’inspeccions de vehicles de França dels darrers anys, automàticament proposava al gestor de la franquícia d’ITV el preu que el podia fer competitiu, en funció de les circumstàncies que l’envolten, o del dia del mes, etc.”

• Factors externs per encertar en la producció

Els grans concessionaris de cotxes necessiten tenir un control molt estricte sobre l’aprovisionament de materials. Algunes peces s’han de demanar amb temps perquè procedeixen de països asiàtics. La central de la marca de vehicles fa previsions, però sovint no s’ajusten a la realitat perquè no tenen en compte altres factors externs a la fabricació. Això provoca que hi hagi grans estocs als magatzems; la qual cosa vol dir, diners aturats.

Les prediccions de producció es poden millorar amb un model matemàtic, diu l’investigador de la UB Jordi Vitrià. “Amb l’històric d’aprovisionaments, però també amb la previsió meteorològica, amb els dies festius i de vacances del lloc de cada magatzem i amb dades de factors socioeconòmics, els resultats són molt diferents. És informació que mai no hauries tingut en compte per a la fabricació de cotxes, però que hi influeixen molt.”





• Com es poden explotar al màxim els casaments

El *customer journey* o mapa de l'experiència del client és el seguiment del client que es fa des del moment que entra en contacte amb una empresa (per sol·licitar informació, per exemple) fins que l'abandona (s'esborra d'un servei). A partir de les interaccions es pot predir quan es perdrà aquell client. Si ens donem d'alta en un servei de lloguer de motos per la ciutat, per exemple, i durant un temps no en contractem cap, l'empresa ens enviarà correus i missatges al mòbil recordant ofertes, descomptes, oportunitats, rutes a fer fins que lloguem de nou una moto. I si no reaccionem, ens donarà per perduts.

Amb aquesta mateixa estratègia operen la major part de les empreses avui. A Sant Cugat hi ha una que ofereix tot tipus de serveis i productes per a un casament. És el gran cercador dels casaments: des de maquilladores fins a restaurants, lloguers de cotxes especials, vestits, etc. Hi treballen més d'un centenar de persones que resolen dubtes i consultes en qualsevol idioma. Està ubicada a Catalunya, però el mercat és global.

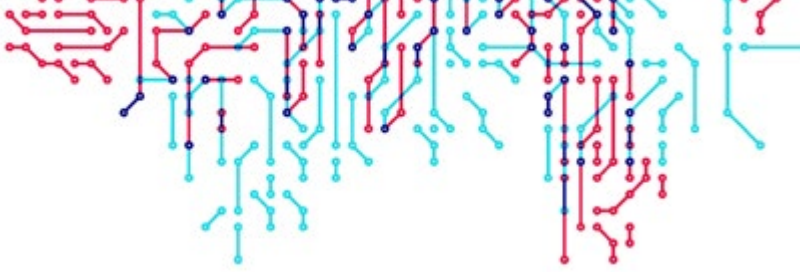
El negoci funciona igual que el de les reserves d'habitacions d'hotel. L'empresa guanya de cada clic que fan en un servei, i li paga al restaurant que s'hi anuncia. Aquest paga per aparèixer dels primers en el seu cercador. Si surts dels primers, tens moltes probabilitats que facin reserves de casament. “Ni els naixements, tot i ser un moment important de la majoria de persones, causen tanta expectació, ni mouen tants diners com els casaments”, explica Jordi Vitrià de la UB. “Els algorismes ajuden a millorar el *customer journey* d'aquesta empresa i l'alerten quan hi ha perill que marxi el client, quan s'hauria de fer una acció presencial o per correu electrònic”, afegeix.

59

• Algorismes per ser més ràpids

Les empreses de repartidors amb bicicleta o amb moto també fan servir algorismes. En el repartiment d'última milla (distàncies curtes) es demanen coses petites i molt diverses. Aquestes empreses confien als algorismes l'eficiència de la gestió de les comandes i les entregues. Per exemple, assignen el repartidor que complirà abans amb la petició del client i li indiquen on ha de comprar el que demana, en funció de la distància que s'ha de recórrer i el temps establert. Si la distància és molt gran, l'aplicació traspasarà la comanda a un repartidor que vagi amb moto. Això també suposa un benefici per al repartidor, que fa esperar menys temps al client, i pot rebre una qualificació millor en acabar el servei.

Els algorismes assignen una puntuació als repartidors en funció de la reputació que tinguin, basada en criteris com les hores dedicades al repartiment, si ha estat en caps de setmana, si ha estat en les hores de més demanda, si no tenen queixes dels clients, etc. Amb aprenentatge automàtic, el sistema també pot entendre què demanen els clients i com ho demanen. “Per exemple, si es demanen tres aigües, què vol dir? Tres garrafes de



cinc litres d'aigua o tres ampolles d'un litre? La plataforma disposa d'una base de dades amb tota la informació acumulada i la processa perquè l'entrega sigui al més acurada possible", explica el responsable de comunicació d'una empresa de repartiment a domicili.

• Les dades de comportament, les més valuoses

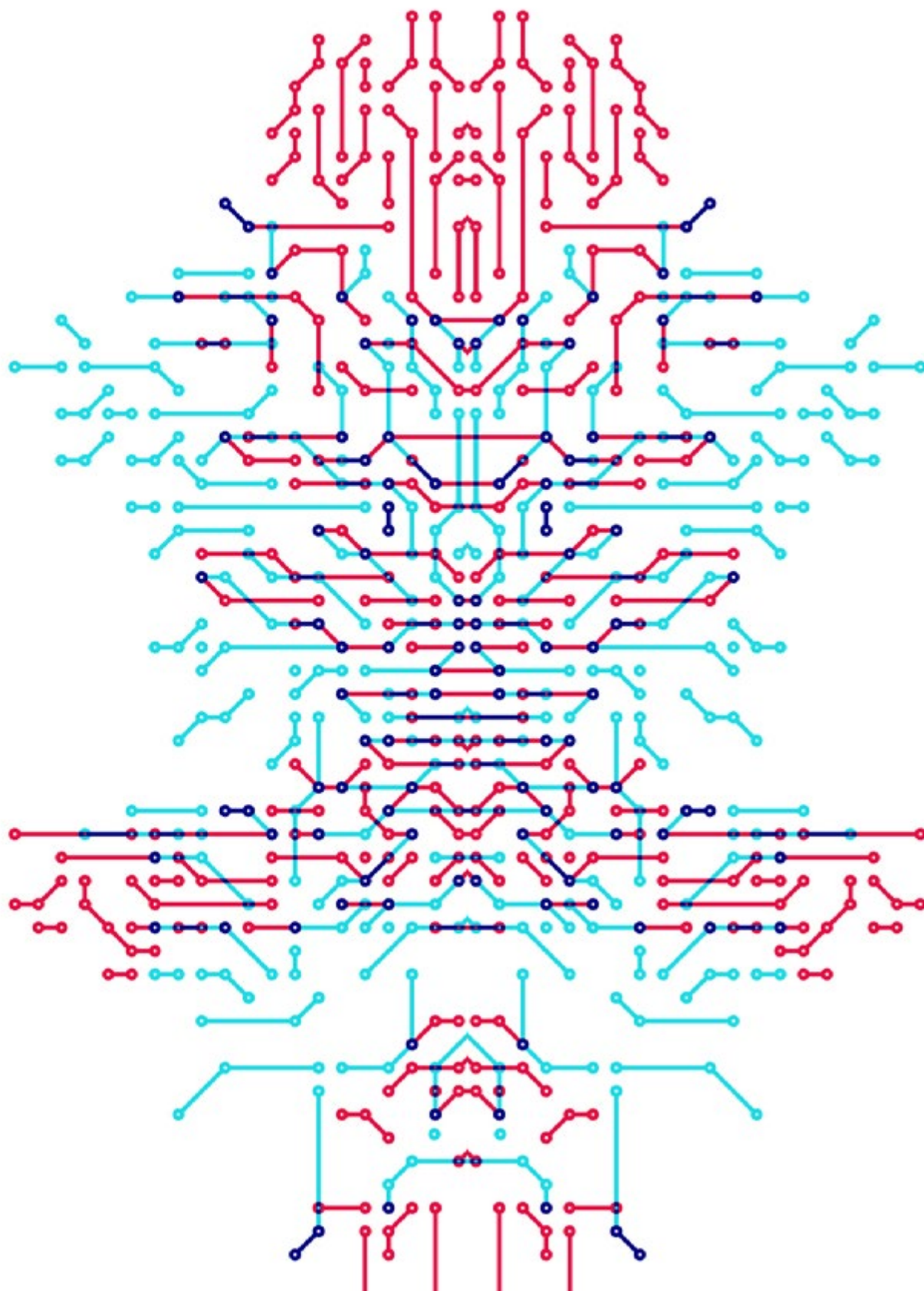
Tots visitem portals amb anuncis de compra i lloguer d'habitatge, de cotxes de segona mà, d'ofertes de treball o per veure una sèrie. Però com funcionen els algorismes que tenen al darrere? L'objectiu principal és tenir l'usuari el màxim temps enganxat (*user engagement*) a la plataforma o al web. Aquest és el negoci: com més temps s'inverteixi en aquests portals, més possibilitat de fer diners amb els anuncis, productes o serveis que s'hi ofereixen. Amb aprenentatge automàtic es fan recomanacions que encerten gairebé sempre, a més de detectar si algú hi posa anuncis fraudulents (que n'hi ha, i molts), i poden interactuar amb l'usuari amb respostes intel·ligents (*smart replies*).

60

"Els algorismes de recomanació se'ls van inventar fa vint anys, no són res nou. Aprenen del comportament de l'usuari", explica el responsable d'un mercat web (*marketplace*), una empresa internacional ubicada a Barcelona i propietària de diversos portals d'anuncis. "Però tenim dos problemes: 1) Introduir nous productes o serveis. Si tens un *portfolio* de set mil productes, que ja funcionen sols, però n'hi ha un que no el coneix ningú perquè és nou, com fas perquè es vegi? S'entrena l'algorisme perquè ho recomani fins que aconseguix atraure l'atenció de l'usuari. 2) El que es coneix com a *cold start* (usuari nou). No sabem res d'ell, no tenim dades de comportament. Potser ha navegat per vint pàgines, però no ha comprat res. L'algorisme no sap què li agrada, per tant, va a cegues en les recomanacions. Si aconseguix que compri o fiqui algun producte a la cistella, això ja és moltíssima informació! Normalment, una persona s'hi està vint minuts al portal, que li mostra coses molt genèriques", explica aquest professional.

Totes les empreses de comerç electrònic funcionen igual. Quan una persona arriba al web d'un portal d'anuncis, se li injecta una galeta (*cookie*: petit fitxer que l'identifica amb una numeració única). Les *cookies* perduren en el temps, i durant mesos o anys ofereixen molta informació. Quines dades es recullen? Les categories visitades, les cerques, les pàgines vistes, el temps invertit en cada pàgina, les compres, etc. "Si tinc un usuari que cada cop que entra al portal va a la secció de llibres, fa cerques de Java, Python, HTML, m'ha visitat mil vegades però no s'ha donat d'alta i no m'ha comprat res, igualment tinc moltes dades!", afegeix el responsable del *marketplace*. "Analitzem tota aquesta informació, i cada cop que un usuari és identificat, es posen en marxa els recomanadors. Fem milions de recomanacions diàries per a tots els portals que tenim. Són de tipus: "Si has vist aquesta bicicleta, segur que t'interessa aquest casc". Els

anònims també compten. Potser entren avui i no tornen fins d'aquí a sis mesos, però encara tenen la *cookie*, perquè no expira", conclou.



2.4.6. Social

La intel·ligència artificial en el sector social resulta molt útil sobretot per a processos a gran escala. Els algorismes poden concedir o no concedir ajuts econòmics, assistència d'algun tipus i derivar a altres serveis especialitzats en funció de la problemàtica personal. Avui, ja són de gran ajuda, sobretot en grans ciutats quan hi ha milers de ciutadans sol·liciten serveis.

Els algorismes –com ja s'ha explicat– s'alimenten o entrenen amb dades massives del passat. I aquestes dades és molt probable que tinguin biaixos, perquè les societats evolucionen i els hàbits, els costums, les maneres de viure, d'educar-se o d'accedir al món laboral han variat. No és el mateix la societat catalana d'avui que la de fa cinquanta anys.

Per exemple, les dones no accedíem tant a llocs de treball, pràcticament no hi havia famílies monoparentals, tampoc no s'adoptava tant com avui, etc. Podem tenir problemes similars (econòmics, d'immigració, de violència, etc.), però les realitats són diferents. Per això, parlant de l'aplicació de la IA en el sector social és tan important detectar i mitigar –com més aviat millor– les possibles discriminacions que vagin adossades a les dades massives que es fan servir.

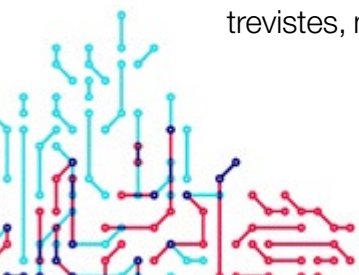
62

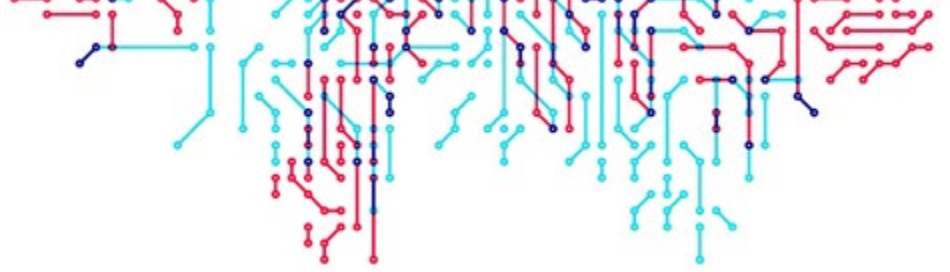
A Catalunya, l'Administració pública tot just comença a posar-se les piles pel que fa a utilitzar la IA per redistribuir ajuts socials, però sobretot s'està contemplant –de moment– per ser més eficients en els tràmits i gestionar millor els recursos.

• Intel·ligència col·lectiva per als ajuts socials

L'Àrea de Drets Socials de l'Ajuntament de Barcelona pot atendre una mitjana de cinquanta mil primeres visites a l'any. Les persones que acudeixen als 40 centres de serveis socials repartits per la ciutat tenen problemes econòmics, de dependència, per malaltia mental, d'alcoholisme, poden necessitar ajuda psicològica, d'adaptació, poden patir una situació de violència de gènere, etc. Problemàtiques molt diverses que són ateses per una plantilla de més de set-cents professionals, entre treballadors socials, psicòlegs i educadors socials.

Quan la persona arriba al centre, se l'atén en unes cabines privades. El treballador social enregistra la conversa i, en acabar, transcriu la problemàtica, així com l'ajuda o servei a la qual ha estat derivada. En el sistema intern es descriu amb tres lletres: demanda (D), problema (P), recurs (R). Actualment, l'Ajuntament disposa de centenars de milers d'entrevistes, moltes de les quals acaben sent repetitives perquè els problemes s'assemblen.





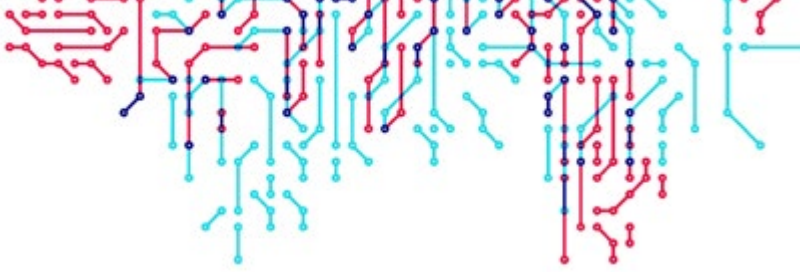
“Vam entrar en un repositori tres-centes mil entrevistes i vam dotar-lo de tècniques d’aprenentatge automàtic”, explica Lluís Torrens, director d’Innovació Social de l’Àrea de Drets Socials, Justícia Global, Feminismes i LGTBI de l’Ajuntament de Barcelona. “La màquina va llegir tots els comentaris anotats pels treballadors socials per les D, les P i les R. Ara suggereix els recursos a partir del que ha après. Classifica les demandes i les respostes possibles.”

Els resultats s’han aplicat ja a tres centres. Segons explica Torrens, les recomanacions són més acurades que les dels professionals, perquè eviten la dispersió. I el nivell de satisfacció és alt. “Si tens set-cents professionals és molt fàcil que no tots destinin els recursos de la mateixa manera. Bé perquè el professor universitari els explicava les matèries a la seva manera, bé perquè el professional ha investigat més en un tipus de problemàtica, etc. La màquina homogeneïtza les respostes, i dona la llibertat al professional que acabi decidint. M’agrada dir que no és un sistema d’intel·ligència artificial, sinó col·lectiva”, apunta el director d’Innovació Social de l’Àrea de Drets Socials de l’Ajuntament de Barcelona.

• Potenciar les avaluacions

Es calcula que unes vint-i-cinc mil famílies de Barcelona viuen en situació de pobresa cronificada. Per estrany que pugui semblar, mai fins ara no s’han fet avaluacions de les ajudes socials que l’ajuntament mateix hi destina. No se sap si els ha anat bé o no a les persones o les famílies a les quals se’ls ha concedit una prestació o un servei de recuperació. “Un metge quan et dona una pastilla sap en quin percentatge et curarà la malaltia. Els professionals socials no saben si funcionarà o no el tractament que apliquen, perquè en les persones influeixen un munt de factors”, explica Lluís Torrens. “D’entrada, si els que han demanat l’ajut ja no venen més al servei, pot ser que la seva situació s’hagi resolt, però també que hagin desistit de la nostra ajuda o marxat de la ciutat”, afegeix.

“La IA ens permet fer avaluacions acurades i comparar el que hauria decidit el professional amb el que decideix la màquina. Ara volem crear un sistema integral de dades massives, que consisteix a agafar tota la informació que podem obtenir de registres administratius d’una família demandant d’ajuts econòmics, saber quins són els seus ingressos a través de l’Agència Tributària, si cobra una pensió, si ha demanat altres ajudes a altres administracions, per aconseguir una evolució temporal i saber si està millorant o no”, continua el director d’Innovació Social de l’Àrea de Drets Socials de l’Ajuntament de Barcelona.



Les avaluacions s'han posat en pràctica en el projecte europeu B-Mincome,⁶² en què durant dos anys s'ha fet un seguiment exhaustiu a nou-centes cinquanta famílies dels deu barris del Besós. “Els hem injectant una renda, amb polítiques actives de reinserció. Tot un projecte molt controlat, amb l'agència pública l'avalua al darrere. Aquest projecte, que ara s'acaba, ha permès tenir molt coneixement de com una acció molt avaluada està afectant les famílies. Les intervencions públiques s'han d'avaluar, i els algorismes ens ajuden a fer-ho”, afegeix Torrens.

• Detectar els biaixos

Després de posar en marxa algunes proves pilot amb intel·ligència artificial, la màxima preocupació del director d'Innovació Social de l'Àrea de Drets Socials de l'Ajuntament de Barcelona, Lluís Torrens, és que els algorismes amb els quals opera tinguin el mínim de biaixos possible. Per a això, han encarregat una auditoria a una empresa externa. “És molt probable que, per l'origen de les famílies, per gènere o per edat –de manera involuntària– tinguem resultats que no pertocuen. I que hi hagi alguna discriminació per l'algorisme mateix, que no n'hagi generat cap sense que nosaltres en siguem conscients. És el més normal, ja que les respostes socials han anat canviant al llarg del temps. Però hem d'estar alertes als biaixos i mitigar-los”, explica Torrens.

64

“Igualment, ens agradaria detectar amb l'algorisme una problemàtica que el professional social no hagi estat capaç de captar. Com ara un fet possible de violència en un llar (sigui vers a la dona, els fills o la gent gran). Això es faria correlacionant informació d'altres serveis socials, per exemple, dels serveis d'acompanyament a les dones amb problemàtica de violència masclista. Connectant els serveis es podria fer un programa d'alerta precoç i detectar una situació complicada a casa. Encara ho estem treballant i ho explorarem més. Però el professional assistencial podria treballar coordinadament amb els Mossos d'Esquadra, amb la Guàrdia Urbana i amb d'altres serveis per ajudar les famílies”, conclou el responsable del consistori de la capital catalana.

62 Projecte B-Mincome (<http://ajuntament.barcelona.cat/bmincome/ca/pressupost-ajudes-barcelona>).

• **Estudiar l'envelliment actiu**

La directora del grup consolidat de recerca Machine Learning and Computer Vision de la Universitat de Barcelona (UB), Petia Radeva, lidera el projecte Lifelogging,⁶³ que consisteix a capturar imatges d'una persona al llarg de la seva vida amb una càmera portàtil.

Va ser finançat per La Marató de TV3 i es desenvolupa juntament amb l'equip de la doctora Maite Garolera, del Consorci Sanitari de Terrassa. L'interès és monitorar la vida diària per estudiar l'envelliment actiu de les persones grans. Les càmeres de resolució temporal llarga (LTR) són excel·lents per a aquest propòsit.

Com menja, es mou, dorm, va de compres, cuina, neteja, socialitza, etc., la gent gran estan relacionades amb el deteriorament cognitiu. I se sap que els que mantenen habilitats altes pel fet de fer diferents activitats cada dia són menys fràgils físicament i cognitivament. Amb tècniques de visió per computació i aprenentatge profund s'estan analitzant aquestes accions a partir de les imatges capturades amb càmeres LTR. També serveixen com a àlbum de records per millorar la memòria.

• **Assistent intel·ligent de dietes personalitzades**

La Universitat Politècnica de Catalunya (UPC) participa en el projecte Diet for You, juntament amb l'Institut Hospital del Mar d'Investigacions Mèdiques (IMIM), per potenciar un estil de vida saludable i l'adherència a la dieta. "L'algorisme agafa totes les dades del pacient –de salut, d'estil de vida, de la dieta que ha fet o està fent, de l'exercici que fa; eventualment es podria agafar informació genòmica, etc.– i fa un perfil de la persona", explica Karina Gibert, investigadora principal del projecte i membre de l'Intelligent Data Science and Artificial Intelligent Research Center (IDEAI-UPC). Així el sistema té coneixement per saber els seus patrons de dieta", afegeix.

Automàticament, crea els menús per a tres o cinc mesos, amb tots els plats preparats, que s'adapten a la prescripció nutricional més estàndard. "Amb la gràcia –continua Gibert– que el sistema es configura amb el context de què agrada i no agrada a la persona, a quins ingredients no pot accedir perquè són massa cars, com és l'estil de vida allà on viu (si el dinar és més llarg o més breu, si es menja de pressa un sandvitx o fa tres àpats forts, si pren te o cafè, etc.), a més de saber totes les restriccions de tipus al·lèrgic o alimentari, com, per exemple, si la persona és diabètica", explica. Sobre aquesta base de coneixement es creen els menús descompostos nutricionalment. Finalment, el nutricionista dietista revisarà la recomanació automàtica i activarà les restriccions que

⁶³ Lifelogging (<http://www.ub.edu/cvub/egocentric-vision/>).

corresponguin o ajustarà el suggeriment de l'algorisme. El projecte està finançat pel Ministeri de Ciència, Innovació i Universitats dins del Programa Retos.

• El robot que fa companyia

La robòtica és encara molt experimental al nostre país, però cada dia més ens acostumarem a tenir-la ben visible en l'assistència sanitària, així com en residències de gent gran, llars d'infants i escoles.

A Catalunya, el robot Pepper es va passejar per sales de centres hospitalaris el 2018. L'incloem en aquest apartat social (i no en el de salut) perquè les principals funcions de la màquina eren informar o acompanyar pacients. Es desplaça silenciosament i té aspecte d'humanoide, a més d'interactuar amb persones en 21 idiomes. Encara està en fase de desenvolupament, però ja s'ha pensat perquè expliqui a la gent gran com s'han de fer les cures o com han de tractar la seva malaltia. Seria també d'ajuda per a l'acompanyament a menors immunodeprimits després d'una operació quirúrgica que resideixen aïllats.

66

El robot utilitza tècniques d'aprenentatge automàtic, i pot reconèixer persones de manera molt precisa memoritzant-ne els trets facials. També pot identificar l'estat d'ànim dels pacients, a partir de l'expressió de la cara i del to de veu. En el desenvolupament del projecte s'han ajuntat diferents centres mèdics com els hospitals de Sant Joan de Déu i el Clínic, el Departament de Robòtica de la Universitat La Salle i l'empresa YASYT.⁶⁴

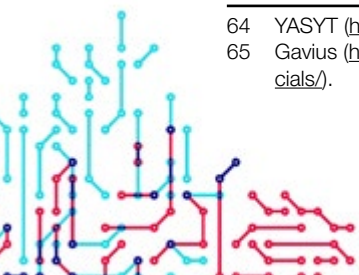
• Gavius, per ser més socials

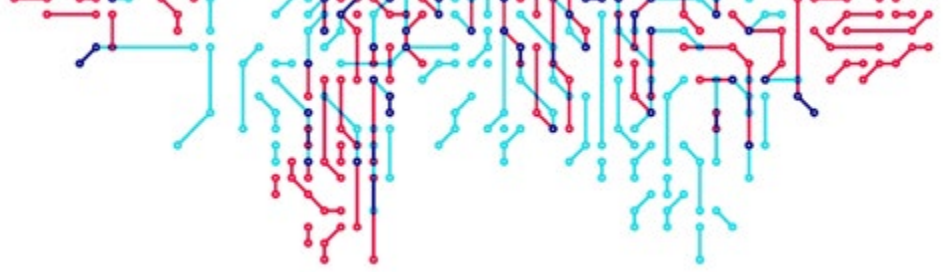
Un projecte que s'inicia aquest 2020 a l'Ajuntament de Gavà és el Gavius⁶⁵. Un assistent virtual comunicarà a la ciutadania quins ajuts socials tenen a l'abast, com es tramiten, com s'atorguen i com es poden percebre de manera còmoda, ràpida, senzilla i a través del mòbil.

Amb finançament del fons europeu Urban Innovation Actions –dotat de 4,3 milions d'euros–, és un projecte en combinació entre Administració (Consorci AOC, ajuntaments de Gavà i de Mataró), empresa (GFI i EY), ciutadania (Xnet) i recerca (UPC i CIMNE).

64 YASYT (<https://yasyt.com/ca>).

65 Gavius (<https://www.aoc.cat/2019/1000265639/laoc-collabora-amb-gavius-un-projecte-innovador-en-lambit-dels-ajuts-socials/>).





2.4.7. Treball

Les innovacions tecnològiques en el sector del reclutament de personal no són cap novetat: els primers formularis en format test van aparèixer a la dècada dels anys quaranta del segle passat; als anys noranta ja es feien servir tècniques digitals i, ara, és el torn de la intel·ligència artificial. I es pot arribar fins a límits no imaginats abans. Per exemple, l'agència de recol·locació de treball finlandesa DigitalMinds té una vintena de grans corporacions com a clients. Com que rep centenars de demandes de candidats, vol estalviar temps a l'hora de seleccionar el millor treballador per als seus clients, i per això no fa entrevistes personalitzades: demana la contrasenya de correu electrònic (i perfils a les xarxes socials) de l'aspirant, i un algorisme de decisió automatitzada escodrinya tota la informació personal (missatges enviats i rebuts, interaccions, etc.), a més de decidir si es mereix el lloc de treball.

A Catalunya també trobem aplicacions de la IA en el sector laboral. Afortunadament, aquestes no són tan intrusives com al cas finlandès. Ja es fa servir per a l'automatització de la selecció de personal o per predir les probabilitats que una persona desocupada té de trobar una nova feina.

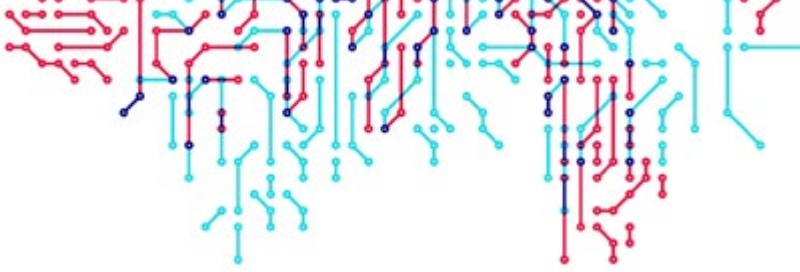
67

• Selecció de personal pels gestos de la cara

El Centre de Visió per Computació va presentar una demo d'un recomanador de recursos humans en el Mobile World Congress 2019. Hi ha un grup de recerca que a partir de l'anàlisi comportamental i gestual de les persones pot dir quin tipus de perfils ocupacionals i quines aptituds tenen més o menys emfatitzades els candidats.

Ens podem situar en l'escena d'una persona que acudeix a una entrevista de treball. El director de recursos humans li fa l'entrevista, però hi ha una càmera que està enregistrant i un programari amb algorismes que està captant els trets de la personalitat.⁶⁶ “El responsable de recursos humans es basarà en la seva impressió personal però també en l'anàlisi del sistema intel·ligent per decidir si és el candidat que busquen o no”, explica Meritxell Bassolas, directora de Coneixement i Transferència de Tecnologia al CVC. “La màquina pot detectar que tenia una actitud nerviosa, confiada, menys artística, més reflexiva, etc. I ho fa tant pels gestos com per les emocions i les microexpressions facials”, afegeix Bassolas. La investigadora considera, però, que mai no tindrem entrevistes únicament amb una càmera i les valoracions dels algorismes. “Recordem que aquests sistemes han de ser sempre de suport a la decisió del professional. Els experts

66 CVC at Mobile World Congress 2019 (<http://www.cvc.uab.es/outreach/?p=1780>).



en recursos humans tenen molt coneixement que també han de posar sobre la taula en reclutar gent. Saben quins paràmetres han d'avaluar en funció del càrrec que cerquen o les competències necessàries. La màquina no arriba fins aquí.”

• Els biaixos de les plataformes professionals

El Grup d'Investigació de Ciència de la Web i Computació Social de la UPF, juntament amb la Universitat Tècnica de Berlín i el Centre Tecnològic Eurecat, va crear un algorisme que detecta i mitiga biaixos d'altres algorismes. El sistema l'han anomenat FA*IR.⁶⁷ “Vam estudiar dades d'ofertes de treball, de reincidència de presos i rànquings d'admissió a les universitats per detectar patrons de discriminació en directoris que puguin afavorir o relegar certs col·lectius, per gènere, edat o raça”, explica Carlos Castillo, director del grup d'investigació de Ciència de la Web i Computació Social de la UPF.

Una de les seves alumnes de doctorat, Meike Zehlike, va investigar com eren classificats els homes i les dones en les plataformes de professionals LinkedIn i Viadeo. “Si hi ha cent perfils d'homes i dones igualment qualificats i en els primers resultats del cercador només apareixen homes, tenim un problema.”

68

FA*IR detecta aquesta mena de discriminació i la corregeix incorporant-hi un mecanisme d'acció positiva per reorganitzar els resultats i evitar la discriminació sense afectar la validesa del resultat.

• Predir les baixes laborals

A partir de les assistències o absències del personal d'un centre hospitalari, un algorisme pot predir en cada servei, cada perfil laboral, quantes baixes poden esperar-se cada dia. El sistema dona una estimació de quantes persones no es presentaran al seu lloc de treball.

“No filem treballadors mai, ho fem per servei”, explica Ricard Gavaldà, coordinador del laboratori de recerca a la Universitat Politècnica de Catalunya (UPC). Ara es fan contractacions de personal de salut per dies, o caps de setmana, i van a urgències o a planta, però no tenen l'experiència d'haver treballat regularment en aquell centre. Això provoca distorsions, no hi ha temps de formar els professionals, no coneixen el context de l'hospital, etc. “No es pot predir el motiu pel qual s'agafarà la baixa, però sí que el servei es quedarà coix de personal durant certs dies”, argumenta Gavaldà.

⁶⁷ M. Zehlike, F. Bonchi, C. Castillo, S. Hajian, M. Megahed, R. Baeza-Yates. “FA*IR: A Fair Top-k Ranking Algorithm” (<https://arxiv.org/pdf/1706.06368.pdf>).

• Automatitzar per gestionar en temps rècord

Una de les tasques del Servei Públic d'Ocupació Català (SOC) és la formació de les persones treballadores o desocupades. Quan gestionen les subvencions destinades a aquesta formació, això els genera un problema de temps i recursos enorme, perquè són moltes les entitats públiques i privades al servei públic que opten a fer la formació i a rebre diners de l'Administració.

“Parlem de quantitats de diners molt importants, que poden pujar a cinquanta milions d'euros, i als quals opten unes tres-centes o quatre-centes entitats de formació. Cada expedient és un projecte, i cada projecte representa tres o quatre cursos”, expliquen responsables de la Secretaria Tècnica del SOC. “Abans d'atorgar les subvencions, el personal intern havia d'introduir totes les variables manualment per avaluar-les, d'acord amb uns criteris establerts prèviament. I així, expedient per expedient”, afegixen. Com que era un volum enorme, el procés trigava mesos abans no se sabés el resultat de les entitats que farien la formació.

El 2014, es va decidir automatitzar aquest tràmit, i ara es fa en només unes hores. En aquest cas, no hi ha aprenentatge automàtic, però sí que hi ha algorismes d'ordres lògiques que s'apliquen a una màquina. Hi ha una seqüència lògica –més o menys complexa, perquè afecta molts paràmetres– que resol una tasca feixuga i complicada.

69

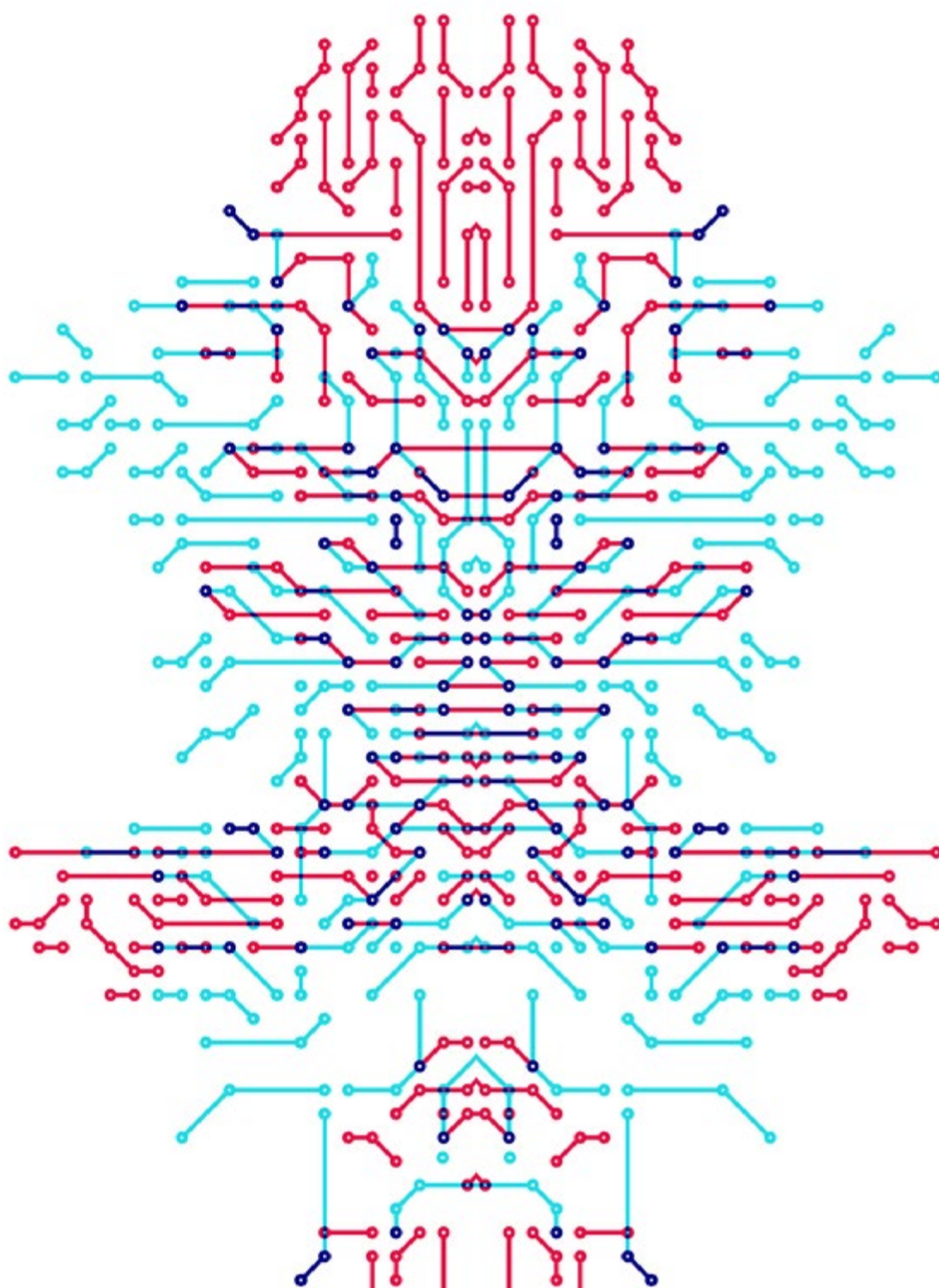
• Automatitzar per planificar millor

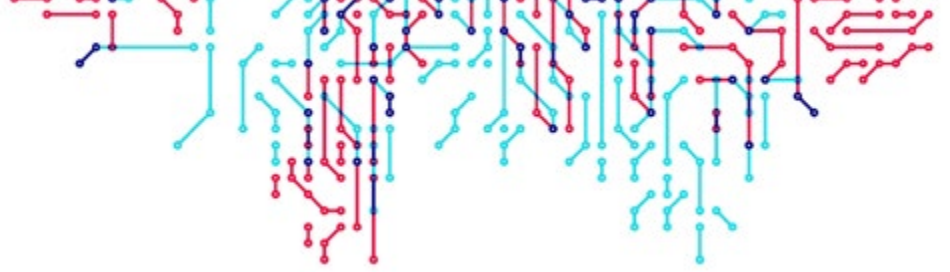
El Centre Tecnològic Eurecat ha dissenyat un algorisme que analitza les ofertes dels portals Infojobs i Feina Activa⁶⁸ (els més consultats pels ocupadors). L'objectiu d'aquest projecte pilot és evidenciar que la intel·ligència artificial pot ser molt útil per a la classificació de l'oferta formativa, tant en l'àmbit territorial com en el sectorial. “Podem saber les mancances, les coincidències i també ajuntar les competències que demanen els empresaris amb la formació dels departaments competents, el Servei Públic d'Ocupació de Catalunya i el Departament d'Educació”, expliquen des de la Secretaria Tècnica del SOC.

Una de les potencialitats és donar suport a l'orientador. En funció de les característiques de la persona i de la seva experiència professional l'algorisme podria predir quina probabilitat d'inserció té. “Per exemple, si ets una dona amb estudis universitaris, que ha passat per diferents llocs de responsabilitat, se't classifica en un *cluster* determinat. Amb les dades massives de tota la població catalana –prèviament introduïdes–, l'algorisme podrà predir el percentatge de probabilitat de trobar feina.” A continuació,

68 Portal Feina Activa de la Generalitat (<https://feinaactiva.gencat.cat/web/guest/home>).

l'orientador del SOC –amb altres recursos d'orientació professional– donarà les opcions que s'adeqüin millor al seu perfil per millorar la seva ocupabilitat. “És una eina més que té l'orientador, mai no es farà una decisió automatitzada per concedir o no formació, o qualsevol altre servei, perquè es podria caure en discriminacions”, afegixen. Tot just s'ha estrenat aquest 2020.





2.4.8. Ciberseguretat

Els atacs informàtics a les empreses –grans i petites–, així com a governs i organitzacions que manegen grans volums d'informació privada o financera obliguen a disposar d'una ciberseguretat cada cop més sofisticada. “Les empreses dedicades a la ciberseguretat utilitzen algorismes per caracteritzar el funcionament normal de tots els ordinadors i interaccions que es produeixen en l'entorn de l'organització, a partir d'informació trobada en diferents fonts de dades; les correlacionen i, així, detecten desviacions de la normalitat, indicadores d'alguna possible anomalia”, explica Manel Medina,⁶⁹ catedràtic de la Universitat Politècnica de Catalunya i fundador i director de l'esCERT-UPC, l'equip espanyol de seguretat en xarxa.

Però el cibercrim està en expansió i tal com van reconèixer la majoria dels principals ponents al Barcelona Cybersecurity Congress –celebrat l'octubre del 2019–, els que volen fer mal ho tenen més fàcil que els que lluiten per la defensa i la protecció. “Les ciberamenaces són una tendència en augment arreu del món, que afecta totes les indústries”, explica l'informe *La ciberseguretat a Catalunya*,⁷⁰ elaborat per la Generalitat.

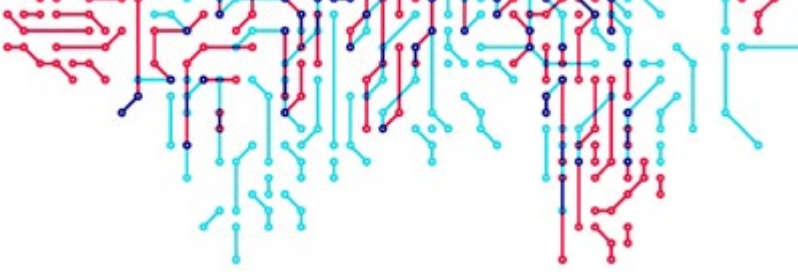
El motiu principal és la transformació digital que s'ha experimentat en tots els sectors de la societat la darrera dècada. “Els mateixos avenços tecnològics que han impulsat la productivitat i l'eficiència dels negocis són els que han fet les organitzacions més vulnerables als ciberatacs”, afegeix l'informe. I assenjala que cada cop que una empresa pateix un atac es veu obligada “a interrompre les seves operacions una mitjana de 17 hores a l'any. Altres efectes negatius poden ser: l'aturada completa de les operacions, la disminució de la facturació; el desvetllament d'informació confidencial; les implicacions legals; la pèrdua de qualitat dels productes; els danys a la propietat física i fins i tot a la vida humana.”

El pitjor és que la complexitat de les amenaces augmenta ràpidament i de manera constant. Això obliga les empreses a estar en alerta permanent. Manel Medina explica que els sistemes automatitzats actuals poden detectar si et connectes a adreces IP de països inusuals per a les activitats de l'organització, o si en algun ordinador de l'empresa hi ha un tràfic de dades que no es correspon amb l'habitual, per detectar possibles fugites d'informació. “El problema ve amb l'amenaça permanent avançada (*advanced permanent threats* –APT–);⁷¹ és a dir, programes espies que s'autoinstal·len en els ordinadors de les empreses i queden latents durant llargs períodes de temps. Durant aquest temps, van generant un degoteig de dades, amb molt poca intensitat, perquè els programes de detecció d'anomalies no percebin desviaments significatius del comportament normal.”

69 Manel Medina (<https://inlab.fib.upc.edu/es/persones/manel-medina>) (<https://www.linkedin.com/in/manelmedina/?originalSubdomain=es>).

70 *La ciberseguretat a Catalunya* (https://www.accio.gencat.cat/web/.content/bancconeixement/documents/informes_sectorials/ciberseguretat-informe-tecnologic.pdf)

71 “APT” (https://en.wikipedia.org/wiki/Advanced_persistent_threat).



Els APT són un conjunt de processos sigil·losos i continus amb la intenció de saltar-se la seguretat informàtica d'una empresa o organització, normalment per motius de negoci o polítics. I es programen per romandre actius un període llarg de temps. “Fa quatre anys hi va haver un robatori de dades personals a l'oficina de gestió de personal (OPM) del govern dels Estats Units que va afectar fins i tot les dels que estaven jubilats. Tres mesos abans, el govern s'havia protegit amb un programa de detecció i prevenció d'intrusions”.⁷² No s'explicaven com podia haver passat, i el motiu era que la fuga d'informació ja s'estava produint quan el programa de detecció va aprendre el que es podia considerar un comportament normal. “Són sistemes d'aprenentatge automàtic (*machine learning*)⁷³ que aprenen els paràmetres normals d'un conjunt de dades suposadament normals i quan els dones una mostra que no s'ajusta a aquesta normalitat, t'avisen. Però si la mostra “normal” ja conté tràfic generat pel programa espia, aquest és considerat part de la normalitat, i no generarà cap avís”, explica Medina.

• El nou desafiament: robatori de dades biomètriques

72

Avui les identificacions humanes es fan habitualment amb empremtes digitals, escaneig d'iris, reconeixement facial, de veu o ADN. Cada cop és més comú que les empreses obliguin als treballadors a triar un d'aquests sistemes per assegurar-se que uns no fitxen en nom dels altres.

Les dades biomètriques són úniques i intransferibles. No hi ha dues cares iguals, ni dues empremtes digitals iguals. Un article de la revista *Forbes*⁷⁴ recordava fa uns mesos que “l'ús de la tecnologia biomètrica i l'autenticació humana suposarà un gran impacte en la societat. Si bé són una gran solució per a la identificació, plantegen problemes seriosos de seguretat. Els països no estan preparats per assegurar els patrons de detecció biomètrics. D'altra banda, els riscos per al rendiment laboral, la precisió, la privacitat, la interoperabilitat i els possibles riscos per a la salut –problemes de visió pels escàners de retina– han de ser gestionats de manera efectiva”, explicava la periodista Jayshree Pandya a l'article.

El director de l'esCERT-UPC també alerta d'aquests riscos de seguretat, que a priori es venen com un avenç tecnològic. “S'hauria de conscienciar les empreses que si recullen dades biomètriques, les protegeixin molt!”

72 “Intrusion Prevention System (IPS)” (https://es.wikipedia.org/wiki/Sistema_de_prevenci%C3%B3n_de_intrusos) (https://en.wikipedia.org/wiki/Office_of_Personnel_Management_data_breach) (<https://www.csoonline.com/article/3130682/the-opm-breach-report-a-long-time-coming.html>).

73 “Aprenentatge automàtic” (https://ca.wikipedia.org/wiki/Aprenentatge_autom%C3%A0tic)

74 “Hacking Our Identity: The Emerging Threats From Biometric Technology” (<https://www.forbes.com/sites/cognitiveworld/2019/03/09/hacking-our-identity-the-emerging-threats-from-biometric-technology/#353ed3505682>).

• Algorismes per protegir els clients

Internet ha canviat la manera com es compren béns i serveis, i això ha provocat una ràpida transformació també en l'organització interna de les empreses. Si no s'inverteix en seguretat, el cost davant d'un atac informàtic de grans dimensions pot repercutir en greus pèrdues econòmiques i de reputació.

El risc de rebre atacs cibernètics –cada cop més sofisticats– és molt alt. Malauradament, els petits negocis i comerços sempre van tard en la transformació digital, i la ciberseguretat només la tenen en compte després d'haver patit un atac. “Els cibercriminals tenen tàctiques, tècniques i procediments per comprometre les empreses i monetitzar les dades robades”, expliquen en l'informe *Cyberthreat Intelligence for Retail & E-Commerce*,⁷⁵ publicat per BlueLiv (una de les *start-ups* de ciberseguretat creada a Barcelona, i que ja té seus a San Francisco i a Londres). “El risc mai no ha estat tan alt com ara: des de campanyes de suplantació d'identitat (*phishing*)⁷⁶ que enganyen els usuaris per compartir informació personal i financera, segresten comptes per fer frau, desenvolupen i despleguen *crimeware* i *malware* (programes maliciosos que s'autoinstal·len en els ordinadors), sistemes de pagament digital i bases de dades”, continua l'informe. Els algorismes de decisió automatitzada proporcionen notificacions en temps real de detecció de frau per a targetes de crèdit robades, però també per prevenir-lo, interceptant targetes abans que es revenguin en el mercat negre.

73

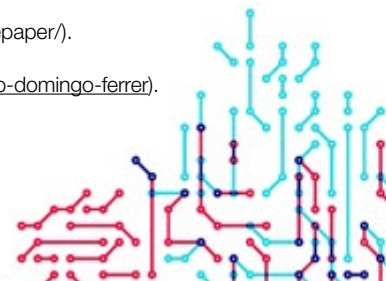
• Aprenentatge automàtic per detectar atacs

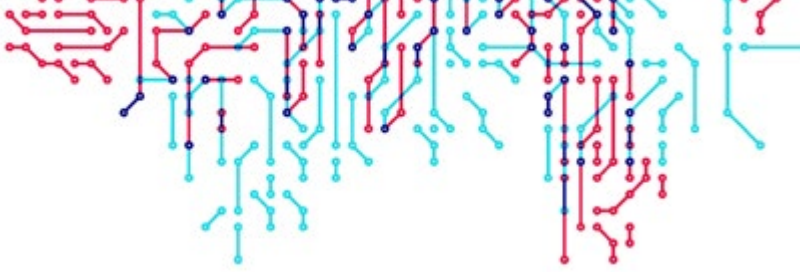
La intel·ligència artificial s'està utilitzant tant per a la defensa com per a l'atac. “És una cursa per veure qui es defensa o ataca millor”, explica Josep Domingo Ferrer,⁷⁷ director del Centre d'Investigació en Ciberseguretat de Catalunya (CYBERCAT) i professor catedràtic de la Universitat Rovira i Virgili. “Per detectar possibles amenaces, s'apliquen mètodes d'aprenentatge profund, a partir de l'entrenament amb dades del passat. Per exemple, a partir d'atacs previs, s'analitzen les característiques de qui els va provocar i s'alerta si es donen les mateixes condicions. Es fa el mateix per trobar el virus que ha causat un atac: s'analitzen casos anteriors. Els algorismes funcionen molt bé quan es disposa de grans bases de dades històriques.”

75 *Cyberthreat Intelligence for Retail & E-Commerce* (<https://www.blueliv.com/thanks-ecommerce-retail-whitepaper/>).

76 “Phishing” ([https://ca.wikipedia.org/wiki/Pesca_\(inform%C3%A0tica\)](https://ca.wikipedia.org/wiki/Pesca_(inform%C3%A0tica))).

77 Josep Domingo Ferrer (<http://www.urv.cat/es/universidad/conocer/personas/profesorado-destacado/2/josep-domingo-ferrer>).





2.4.9. Comunicació

En l'àmbit de la comunicació, els algorismes de decisió automatitzada fa més de deu anys que s'apliquen, sobretot amb tècniques de visió per computació, per exemple, per a interpretar la llengua dels signes.⁷⁸ Així, les persones amb sordesa poden mantenir una conversa i comunicar-se amb qui no entén la llengua dels signes, ja que fan una traducció de signe a paraula en temps real.

Més recentment, la IA s'ha posat en pràctica a la Viquipèdia –per detectar el vandalisme o les entrades incorrectes–, i també per generar notícies en mitjans o per valorar de quina manera es mostren sèries i pel·lícules a les plataformes de vídeo sota demanda, com ara Netflix.

Sorgeixen alguns dilemes ètics, línies vermelles que no s'haurien de traspassar i riscos de viure en filtres bombolla per interessos comercials. També sorgeixen preguntes que no es voldrien ni plantejar: les màquines substituiran els periodistes? Tecnològicament, ja es pot fer. O aquesta altra: podríem acabar perdent de vista la producció audiovisual local perquè un algorisme només m'ensenyia la nord-americana? Tecnològicament, ja es fa.

74

• L'algorisme periodista

“Una de les àrees més interessants i amb futur de la intel·ligència artificial és l'aprenentatge automàtic”, explica David Llorente, fundador d'una empresa que genera notícies amb IA. “La màquina és superexacta en la creació de continguts. No dirà que un altre jugador ha marcat el gol, ni s'equivocarà en els resultats electorals de tal partit, ni en la previsió del temps, perquè beu de les dades.”

L'empresa de Llorente treballa actualment per a 25 mitjans de comunicació i agències de notícies espanyoles. Un altre tema, comenta Llorente, és la percepció social. “És normal que els periodistes se sentin amenaçats professionalment. Però encara queda camí per recórrer. No s'acceptarà que una màquina faci una notícia en què s'acusi certes persones de frau”.

L'algorisme d'aquesta empresa redacta notícies només d'aquelles temàtiques que poden tenir dades objectives, és a dir, resultats electorals, esportius, dades econòmiques, loteries, previsió del temps, trànsit, etc. “El mitjà que ens contracta veu com doblem o tripliquem el volum de notícies que fan diàriament, i amb molta precisió. Això els pot ajudar a vendre més subscripcions o aconseguir més publicitat.”

78 “Creat un sistema visual per interpretar llengües de signes” (<https://www.uab.cat/web/noticies/detall-d-una-noticia/creat-un-sistema-visual-per-interpretar-llengues-de-signes-1090226434100.html?noticiaid=1275458325318>).

David Llorente explica que el sistema intel·ligent elabora la notícia en tres fases: en la primera, la màquina decideix les dades més rellevants per a la construcció de la notícia. “En resultats electorals, el més important és que s’expliqui qui ha guanyat. Però en un partit de futbol narrar un gol, quan han marcat vuit, potser no és tan important”, apunta Llorente. En la segona fase es decideix com escriure la informació: a partir d’un volum de notícies similars, la màquina aprèn l’estructura de la notícia i la forma. Aquí hi ha una revisió final de periodistes que treballen en l’empresa per validar la narrativa. Quan ja està entrenat, el sistema pren les dades definitives, selecciona les frases i redacta la notícia final. La màquina sempre s’entrena amb l’hemeroteca del mitjà pel qual està en funcionament, per agafar el to i l’estil d’escriptura. No només es guanya en velocitat, volum d’informació, sinó també en precisió i correcció ortogràfica.

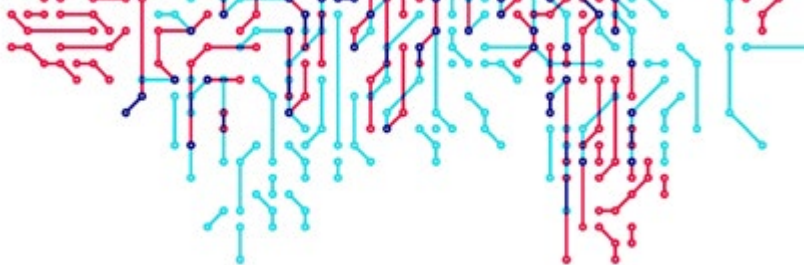
David Llorente creu que actualment els periodistes fan moltes tasques rutinàries, que no són pròpiament creatives, que podrien confiar-se als sistemes intel·ligents. “El que s’intenta és que els periodistes disposin de més temps per fer una informació més ben elaborada, de context, reportatges en profunditat, etc.” No obstant això, segons Llorente, s’hauria d’evitar que la màquina expressi opinió, tot i que tecnològicament ja pot fer-ho. “El tema de l’opinió és més delicat. L’opinió la generem els humans. S’ha d’anar molt amb compte. Són línies vermelles per a nosaltres. Perquè de seguida podrien acusar-nos de crear *fake news*, amb titulars enganyosos de grans temes com el canvi climàtic o temes polítics, que podrien canviar el sentit de la realitat. El tema ètic el tenim molt present. Hem rebut ofertes per fer aquesta mena de treballs i ens hem negat rotundament”, conclou David Llorente.

75

• El maleït filtre bombolla

El Consell de l’Audiovisual de Catalunya (CAC) va encarregar a l’investigador Carlos Castillo⁷⁹ un estudi per detectar com presentava l’algorisme vídeos en plataformes com ara Netflix, OrangeTV i d’altres. Europa obliga que hi hagi un 30% de producció audiovisual europea en cada plataforma. “Però una cosa és que hi hagi aquesta producció audiovisual en el catàleg i, una altra, que l’algorisme l’acabi mostrant a l’usuari”, explica Castillo. “La nova revisió de la Directiva de Serveis de Comunicació Audiovisual (DSCAV) explicita que garanteixin aquesta quantitat de producció europea als usuaris en els seus catàlegs. Si l’algorisme te l’ha mostrat els primers dies i no l’has seleccionat, potser ja no te la mostra mai més perquè interpreta que no t’interessa.”

79 Carlos Castillo, professor investigador distingit del Departament de Tecnologies de la Informació i les Comunicacions de la Universitat Pompeu Fabra. Informe per al CAC: “La oferta y la demanda del contenido audiovisual en la Era de los Datos Masivos”. 2018 (https://chato.cl/papers/castillo_2018_oferta_disponibilidad_contenido_audiovisual-ES.pdf).



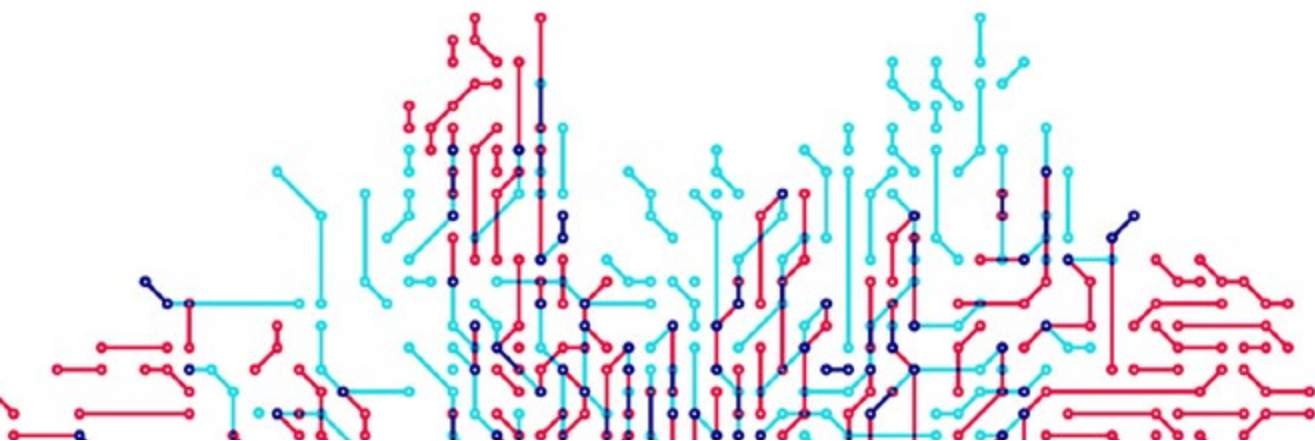
La primera cosa de què hem de ser conscients, explica Castillo, és que els usuaris tenim la sensació que en una plataforma de vídeo sota demanda estem triant la sèrie o la pel·lícula que volem veure, “però no és així en absolut. Potser té al davant quaranta títols, però aquesta és una part molt ínfima del catàleg total. I l’usuari no decideix el criteri pel qual s’han seleccionat aquells primers quaranta títols”.

L’informe que va encarregar el CAC parla dels riscos a l’autonomia i la diversitat, la bombolla informativa, viure en un filtre bombolla. “Si només veus pel·lícules d’acció, no et mostra més que això. Per tant, aquí no hi ha diversitat. Però també hi ha risc d’autonomia perquè hauries de poder explorar el catàleg sencer. Ara només hi ha una caixeta de cercar, cosa que fa que sigui impossible que hi accedeixis al volum total. Podrien haver-hi altres dissenys d’interfície que, per exemple, poguessis fer la cerca parlant. D’igual manera, podrien donar l’opció de decidir que el catàleg estigui ordenat o desordenat.”

• IA per a la Viquipèdia en català

76

Una de les preocupacions més importants dels projectes oberts Wikimedia és la revisió de contribucions potencialment perjudicials (“edicions”). Així com les contribucions que, sense voler, contenen errors. Des del 2018, Viquipèdia fa servir aprenentatge automàtic per fer les revisions d’articles a través d’ORES,⁸⁰ una eina que “marca les accions que són potencialment vandalismes, per exemple. D’aquesta manera, els voluntaris que revisen les modificacions recents a Viquipèdia poden fer la seva feina més fàcilment”, expliquen a la seva pàgina web. “Per poder habilitar ORES, la comunitat d’editors ha d’*ensenyar al sistema*, etiquetant edicions ja fetes, tot indicant si són correccions ortogràfiques, gramaticals, actualitzacions de continguts o –donat el cas– vandalisme.”



⁸⁰ “ORES” (<https://ca.m.wikipedia.org/wiki/Viquiprojecte:Viquirepte/ORES>).

2.4.10. Visió per computador

La visió per computador (VC) és una branca de la intel·ligència artificial que fa que els algorismes entenguin una imatge igual que ho fem les persones. Ja poden distingir objectes i persones, i també descriure'ls o dir què fan aquestes persones.

La VC no és un terreny nou. “Fa més de quaranta anys que es va començar a investigar. En els anys setanta del segle passat, el científic i pare de la intel·ligència artificial Marvin Minsky⁸¹ va experimentar amb ella”, explica Petia Radeva, la directora del grup de recerca Machine Learning and Computer Vision de la Universitat de Barcelona (UB). “Fa uns trenta anys que la VC va entrar a les fàbriques d’automoció per acoblar cotxes i altres funcions en entorns controlats. La novetat és l’aprenentatge automàtic, perquè és capaç de resoldre problemes més complicats, com podria ser la conducció de cotxes autònoms”, afegeix Radeva. “Ara la investigació va vint vegades més ràpid que fa una dècada.”

Hem decidit dedicar una secció a aquesta tecnologia perquè –tot i que els exemples que s’expliquen podrien haver-se encabint dins dels àmbits de la salut, social o comercial– és bo diferenciar el potencial de l’aprenentatge profund en la VC.

77

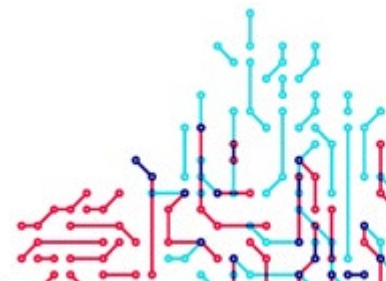
On trobem visió per computador?

Ja hi és a tot arreu. En els pàrquings, quan una màquina llegeix la matrícula del cotxe i s’obre la barrera de sortida. En el reconeixement facial, tant de l’iris com de l’empremta digital, que fan servir els sistemes de seguretat d’aeroports i de la policia, però també els nostres mòbils en substitució d’una contrasenya.

La VC està en els porters automàtics d’edificis, d’empreses i algunes cases. En algunes escoles i instituts han començat a aplicar-lo per a controlar els alumnes al centre. Els bancs l’ofereixen als seus clients per accedir al compte corrent des d’un caixer. Als hospitals, la VC es troba en l’anàlisi de tota mena d’imatges del cos: tomografies, radiografies, densitometries, ressonàncies, ecografies, mamografies, etc., per detectar i predir malalties.

Es fa servir també per comptabilitzar persones –com el nombre de manifestants en un espai públic–, o en accions d’esportistes, com ara les vegades que cada jugador té possessió de la pilota en un partit, els atacs, les faltes, etc. Aquests algorismes no són pròpiament de decisió automatitzada, si no que calculen, estimen i quantifiquen. “Els humans som millors en qualificar, les màquines, en quantificar”, explica la investigadora de la UB.

81 “Marvin Minsky” (https://ca.wikipedia.org/wiki/Marvin_Minsky).



En els mòbils és útil per afegir informació turística quan es fotografia un monument, sense importar l'angle o el tipus de llum del moment. També s'utilitza per al reconeixement de documents digitalitzats de segles passats o per dir quin és l'objecte trobat en un jaciment arqueològic. I la VC està present en els robots de la planta automatitzada que Amazon té a Catalunya, i en els que passen per passadissos d'hospitals (encara en proves) per dur menjar als malalts o fer-los companyia. I a l'edició del 2019 del Mobile World Congress⁸² de Barcelona s'entrava –literalment– per la cara: en lloc de mostrar l'acreditació, les barreres es podien obrir també amb el reconeixement facial.

Un últim exemple: gràcies al reconeixement de l'iris amb VC, el fotògraf Steve McCurry va localitzar Sharbat Gula, la dona afgana fotografiada el 1985, quan només tenia 12 anys i fugia del seu país en guerra. La seva cara es va fer mundialment famosa perquè va ser publicada a la portada del *National Geographic*.⁸³

A continuació, detallem alguns exemples que es troben en les fàbriques o en algunes indústries catalanes i que són fruit de la col·laboració de diferents centres d'investigació

78

• **Detecció de pizzes defectuoses**

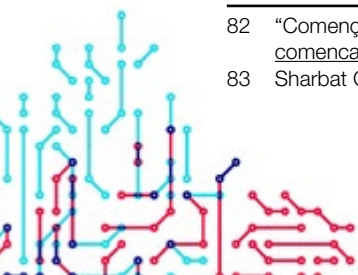
El Centre de Visió per Computació (CVC) de la Universitat Autònoma de Barcelona (UAB) fa més de quinze anys que ajuda les empreses catalanes a desenvolupar projectes digitals i processos industrials. “Quan parlem d'intel·ligència artificial, pensem en la mobilitat autònoma i una presa de decisió dels algorismes a gran escala. Però hi ha molts altres projectes interessants, a una escala menor, que estan impulsant grans avenços”, explica Mertixell Bassolas, la directora de Coneixement i Transferència de Tecnologia al CVC.

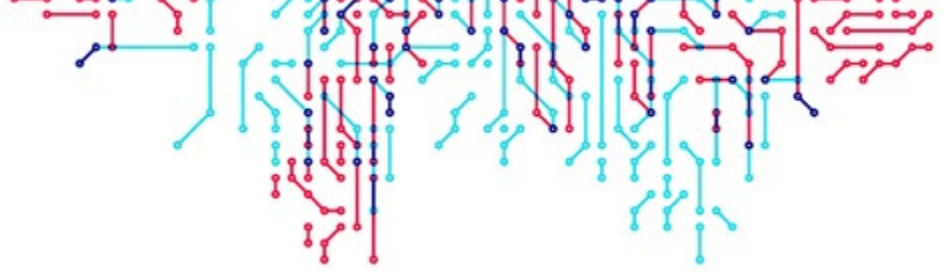
En l'àmbit de l'alimentació, i amb aprenentatge profund, podem detectar si l'estètica de les pizzes és correcta o no, si el repartiment dels ingredients en la base de pizza és l'establert o els materials intrusius que hagin pogut entrar en l'envasat”, explica Bassolas. “Per aconseguir això, l'algorisme necessita entrenar-se amb moltes imatges, milions si pot ser. Com més imatges, més encertat serà el resultat. I si pot ser, de pizzes en bon estat i en mal estat, perquè aprengui a diferenciar-les. I necessita anotacions per distingir què és cada ingredient.”

Segons Bassolas, el repte és aconseguir el que volem, però amb el mínim esforç inicial per part nostra”. Amb videojocs es generen imatges sintètiques de pizzes virtuals, o de

82 “Comença el Mobile World Congress amb la mirada posada en el 5G”. Betevé. Febrer 2019. (<https://beteve.cat/economia/comenca-mobile-world-congress-2019-5g/>).

83 Sharbat Gula (<https://www.nationalgeographic.com/news/2017/12/afghan-girl-home-afghanistan/>).





cotxes autònoms, tantes com necessita l'algorisme per ser entrenat. “Però això també dona molta feina”, afegeix. “I al final, aconseguim que un sistema intel·ligent, amb molt poca informació, sigui capaç de resoldre el que volem. I a més, que no s'oblidi del que ha après, perquè pugui ser funcional en un altre cas sense començar de zero.”

• Seguiment de l'engreix del porc

Sectors com l'agroalimentari s'estan digitalitzant a poc a poc. Processos que fins ara es feien de manera molt manual són automatitzats per extreure'n noves conclusions que ajudin a rendibilitzar més els recursos o millorar els productes. En el cas de les granges d'animals, ja es monitora tot el procés: des de l'engreix fins a l'arribada de l'animal a l'escorxador. “Controlem el porc durant tota la cadena de valor, i això permet comparar quin tipus de menjar és el millor, tant per escurçar temps com per incrementar la qualitat del producte al final”, diu Meritxell Bassolas.

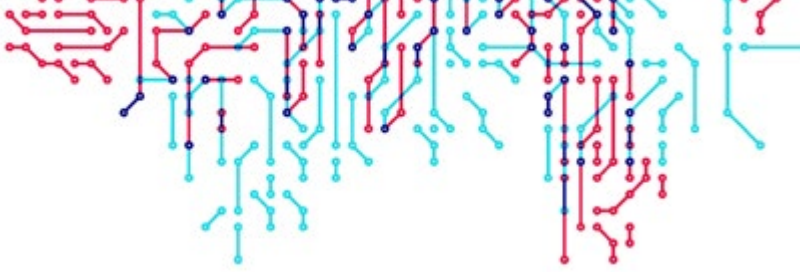
El Centre de Visió per Computació va fer una eina per detectar els porcs a l'escorxador. “Estem aplicant xarxes neuronals que avui ja són capaces de detectar qualsevol moviment en una escena.” A les granges d'engreix també s'implementa el mateix sistema intel·ligent per seguir l'evolució de l'animal, els pesos, els volums, l'estat del porc, i valorar com impacta en la qualitat del producte.

Anteriorment, aquest sistema s'havia utilitzat per classificar residus de manera selectiva en una planta de brossa, o per llegir els dígit dels comptadors de l'aigua.

• La complexitat dels cotxes autònoms

Hem sentit a parlar molt de les meravelles que faran els cotxes autònoms, però també hi ha molta incertesa respecte a les capacitats de les màquines en qualsevol imprevist. De qui serà la responsabilitat en cas que hi hagi un accident?

Entrenar els algorismes perquè s'apliquin a la mobilitat és un dels reptes més complicats que s'han plantejat els investigadors d'intel·ligència artificial fins ara. Perquè la màquina ha de saber tot el que hi ha a l'escena mentre circula. Ho ha de tenir prèviament identificat, i reconèixer-ho. Des d'objectes estàtics (edificis o mobiliari urbà) fins a tot el que es mou: persones i animals, però també l'herba, un full de paper, un estel o qualsevol objecte que surti volant o llenci algú, com ara una pilota. I, per acabar, el cotxe també ha de saber com reaccionar en cada situació. “És inviable entrenar l'algorisme en un entorn real –puntualitza Bassolas–, per tant, generem videojocs i els sistemes intel·ligents s'entrenen a partir de situacions virtuals.”



Però després de ser entrenada la màquina en un entorn virtual, s'adaptarà als carrers d'una ciutat o una carretera real? Fins fa molt poc no hi havia resposta a aquesta pregunta, explica Bassolas. Ara ja es parla del *domain adaptation*,⁸⁴ és a dir, la capacitat que l'algorisme té de canviar d'entorn i adaptar-se. “En un projecte internacional, hem creat un superentorn virtual en codi obert, en col·laboració amb Intel. Empreses com Toyota hi han mostrat interès. Ara hi ha moltes altres marques que aprofiten el que hem fet per adaptar diferents parts del procés.”

• **Reconeixement facial**

Un dels grans problemes que ha donat fins ara el reconeixement facial és que els algorismes han estat entrenats amb poques dades d'un determinat tipus de persones. Per exemple, si s'han posat en pràctica a Europa, potser se'ls han donat menys imatges de cara o de cos de persones negres, asiàtiques o hispanes. Si, després, aquest sistema intel·ligent s'ha d'adaptar a una situació concreta, pot donar problemes com el de l'assegador de mans d'un lavabo públic que només eixugava les de pell blanca perquè les negres no les reconeixia.

80 En els darrers anys s'han detectat i denunciat a bastament els biaixos dels algorismes de reconeixement facial perquè discriminen, especialment les cares no blanques i les femenines. El 2015, Google Photos va classificar algunes persones negres com a gorilles.⁸⁵ La solució que va donar és eliminar certes etiquetes, com ara el terme *goril·la*, en lloc d'arreglar l'algorisme, que continua esbiaixat. També, el 2018, el programa Rekognition d'Amazon va identificar erròniament 28 dels 435 membres del congrés dels Estats Units com a delinqüents.⁸⁶

A Catalunya, en l'àmbit judicial s'aplica el reconeixement facial tant per reconèixer un possible criminal com per saber si l'acusat està mentint, en situacions de custòdia de menors. “Si la persona menteix, el sistema ho detecta a partir de les microexpressions facials”, explica Bassolas. La pregunta és: l'algorisme de decisió automatitzada pot fallar? I la directora de Coneixement i Transferència de Tecnologia del Centre de Visió per Computació nega amb el cap. “Fa uns deu anys que s'utilitza i es perfecciona aquesta tècnica.”

84 “Domain Adaptation” (https://en.wikipedia.org/wiki/Domain_adaptation).

85 “Google apologizes after Photos app tags two black people as a gorillas” (<https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas>).

86 “MIT researchers: Amazon's Rekognition shows gender and ethnic bias (updated)” (<https://venturebeat.com/2019/01/24/amazon-rekognition-bias-mit/>).

• Fotografies satel·litàries

Les imatges satel·litàries poden ajudar a resoldre un munt de problemes socioeconòmics d'un país o d'una població. Des d'estalviar costos i recursos en la producció i distribució d'aliments fins a qüestions d'energia, per entendre què s'ha conreat en un país en un any, temes de pol·lució, desastres naturals, inundacions, etc. "El problema és que la tecnologia emprada fins ara té un cost molt elevat. Si vull tenir una foto satel·litària d'alta resolució de tot Catalunya, actualitzada cada mes, em podria costar uns quatre-cents milions de dòlars per foto", explica Marco Bressan, científic de dades i expert en imatge satel·litària.

"Perquè realment siguin útils hi ha d'haver una freqüència de captura, en temps real. Per exemple, si hi ha un tifó a Bangladesh, vull entendre l'impacte en els conreus d'arròs, tenir una informació pràcticament diària de l'efecte, quantes hectàrees han estat negades, quina mena de mesures he de prendre per ajudar la població, etc. Per obtenir totes aquestes dades, no he d'enviar un satèl·lit, sinó centenars. Però, a quatre-cents milions de dòlars el satèl·lit, és inviable", continua Bressan. Després de certes investigacions, l'empresa que lidera ha aconseguit mantenir la resolució de la imatge per menys d'un milió de dòlars.

L'objectiu és aconseguir una imatge de la Terra cada setmana. I, més endavant, una al dia. Actualment, aquestes fotografies són útils per als sectors agrícola, forestal, d'assegurances o energètic. I s'aplica sobretot per gestionar infraestructures. "Si tens un gasoducte que creua Sibèria o la Patagònia, verificar l'estat de la instal·lació per temes regulatoris, de seguretat, o perquè s'ha desbordat un riu, han caigut arbres, etc., és molt complicat i suposa un cost enorme. Però tenir una subscripció a imatges satel·litàries de la zona d'interès, pot reduir molt la despesa", explica Marco Bressan.

"Els algorismes analitzen en temps real tot el que diuen les imatges. Si hi ha cotxes afectats per una inundació, quants? En temps real, es pot saber quantes hectàrees hi ha sembrades en un país. Per exemple, un productor d'aliments que depengui d'aquesta matèria primera per a la producció, pot saber-ne l'inventari. Això era impossible abans. Abans, si hi havia una plaga que matava el cereal, et quedaves pagant una barbaritat. I ara es pot predir, etc."

Els governs ja utilitzen aquesta mena d'imatges per al seguiment del moviment i estimació dels camps de refugiats. O en defensa, per a la planificació d'atacs amb drons.

2.5. L'ètica de la intel·ligència artificial

L'interès actual per la intel·ligència artificial no té precedents. Molt aviat serà imprescindible per a la societat, perquè aporta molta eficiència, coneixement i creativitat. L'aprenentatge automàtic (*machine learning*) –i, en particular, l'aprenentatge profund (*deep learning*)– ha permès avançar com no es podria haver imaginat fa una dècada.

Però també cal pensar en l'impacte que tindrà sobre les persones. És el moment de fer-se preguntes sobre el significat del que està bé o malament, sobre la relació entre poder i abús, o entre biaix i distorsió. La guia ètica de la Comissió Europea per a una IA fiable,⁸⁷ la Declaració de Barcelona⁸⁸ i l'Estratègia IA per a Catalunya⁸⁹ –presentada aquest 2019 i que incorpora la proposta de creació d'un Observatori Ètic– són alguns exemples de la preocupació per assegurar els drets bàsics davant d'aquest avenç tecnològic.

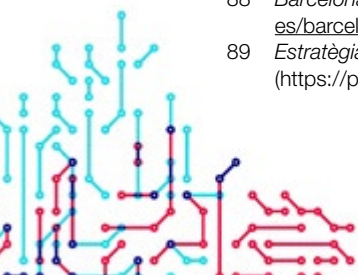
Els experts consultats per aquest informe han estat preguntats per les qüestions ètiques. Tots insisteixen a trobar la manera d'explicar com un algorisme ha pres una decisió o predit una situació, a més de tenir la possibilitat de qüestionar el raonament del sistema automatitzat, i plantegen el debat de la responsabilitat en el cas d'un mal funcionament o una discriminació de l'algorisme. Igualment, assenyalen la urgència de formar la ciutadania –des de la classe política fins als educadors, passant pels pares i mares i l'opinió pública en general–. Formar i/o conscienciar perquè sàpiguen quan han de reclamar els seus drets en cas que un sistema automatitzat redueixi (de manera parcial o total) la seva llibertat individual.

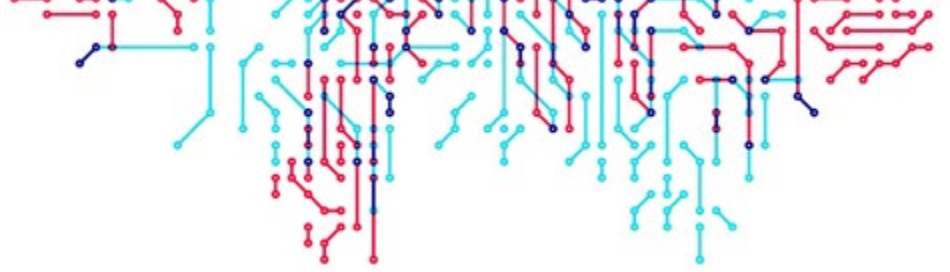
82

87 *EU artificial intelligence ethics checklist ready for testing as new policy recommendations are published* (<https://ec.europa.eu/digital-single-market/en/news/eu-artificial-intelligence-ethics-checklist-ready-testing-new-policy-recommendations-are>).

88 *Barcelona Declaration for the proper development and usage of artificial intelligence in Europe*. (2017) (<https://www.iiia.csic.es/barcelonadeclaration/>).

89 *Estratègia IA per a Catalunya*. Conselleria de Polítiques Digitals i Administració Pública de la Generalitat de Catalunya (2019). (<https://participa.gencat.cat/uploads/decidim/attachment/file/818/Document-Bases-Estrategia-IA-Catalunya.pdf>).





**Catedràtica d'Ètica i Filosofia del Dret moral
i Polític de la Universitat Autònoma de Barcelona (UAB)**

VICTÒRIA CAMPS:⁹⁰ “Amb la intel·ligència artificial s’ha perdut la privacitat”

Què és l’ètica? I per què és tan important ocupar-se de l’ètica en l’aplicació de la intel·ligència artificial?

“L’ètica es podria definir com un conjunt de principis, normes o valors que orienten la nostra conducta i que no poden ser anul·lats pels altres. Per exemple, el respecte a la dignitat és un valor que no pot ser eliminat perquè sigui beneficiós per a un altre”, explica Victòria Camps.

“Les religions no ens serveixen per definir l’ètica perquè és cert que tenim els Deu Manaments en la cristiana, que ens diuen: ‘No mataràs’ o ‘No robaràs’, però es podria qüestionar el de: ‘No desitjaràs la dona de l’altre’. L’ètica va més enllà, inclou una exigència d’universalitat –com la no-discriminació o el respecte vers les altres persones.”

“L’ètica té obligatorietat però no té la coacció del dret –que obliga i, si no compleixes, et penalitza–”, explica Camps. “La norma ètica obliga en consciència”, afegeix. “Amb la intel·ligència artificial hi ha molt poca consciència dels perills. S’ha perdut el que abans consideràvem un bé preuat: la privacitat. Els joves (però també gent de totes les edats) no tenen cap mena de pudor i ho donen tot: fotos, dades, reconeixement facial i després surten les violacions filmades, accions d’assetjament a les escoles, etc.”

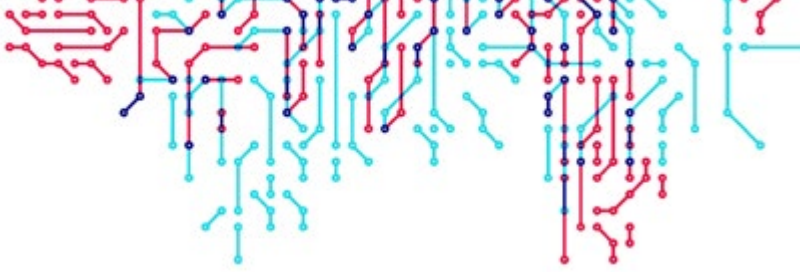
83

L’equitat no la resol només l’algorisme

“El més important en ètica és l’equitat”, continua Victòria Camps. “Un sistema equitatiu vol dir afavorir els que estan en pitjor situació. L’Estat té l’obligació de protegir les persones i la seva equitat. Això ho pot fer un algorisme? Sí, podria afavorir a qui han desnonat, als que no tenen feina i viuen amb fills menors a càrrec seu. Això podria acreditar que se li concedís una ajuda social. Però cada cas és individual, i això també s’ha de tenir en compte quan es programa l’algorisme. Perquè podria haver-hi qui es queixés, i l’Administració pública hauria de tenir respostes per a l’ús de la intel·ligència artificial. L’equitat no la resol només l’algorisme.”

“La IA no pot obviar les preguntes bàsiques de l’ètica que fan referència a la privacitat, a la confidencialitat. A l’informàtic que programa l’algorisme no li pots demanar que

⁹⁰ Victòria Camps (<https://www.uab.cat/web/el-departament/victoria-camps-cervera-1260171817458.html>).



pensi si s'utilitzarà bé o malament. La pregunta se li ha de fer a qui té la responsabilitat màxima, a qui encarrega i decideix el que ha de fer l'algorisme. Formo part del Comitè de Bioètica de Catalunya.⁹¹ Fa uns anys es va discutir molt al Parlament sobre la comercialització de les dades de salut de la gent amb el programa Visc++. Finalment, es va aprovar amb reserves i esmenes. La qüestió que cal plantejar-se és: tenir compartits els historials clínics és un avenç? Òbviament que sí. Però sense perdre el control de l'ús que es pugui fer d'aquestes dades." I què podria ser fer un mal ús? "Doncs que se sabés que un polític té una malaltia no explicada. La persona té dret que les seves dades de salut siguin privades i que si dona el consentiment de fer-les servir, només sigui per a allò que doni servei a la ciència i la medicina."

La responsabilitat de l'algorisme és un tema recurrent, sobretot imaginant possibles accidents en els cotxes autònoms (sense conductor) que han d'arribar. "Els problemes ètics en aquest cas, són de manual. Oblida't de la IA. Pensa en el conductor d'un tren: detecta que hi ha un suïcida a la via. Què fa? Una maniobra per esquivar-lo posant en perill la vida de molts més passatgers? Des del punt de vista ètic, no hi ha una resposta única. Sempre s'ha de procurar fer el mal menor. I amb la intel·ligència artificial s'ha de fer el mateix."

84

Diferència entre l'interès comercial i l'ètica

Victòria Camps considera que cal marcar una línia entre el que pot ser interès comercial i el que pot ser ètic. És ètic que una companyia segueixi el nostre rastre de navegació i ens bombardegi amb publicitat a la bústia de correu, al mòbil, als webs que consultem? És ètic que els bancs sàpiguen tot de nosaltres, hàbits de consum, despeses i preferències d'oci i ens vulguin fidelitzar amb ofertes de productes, en funció de l'anàlisi del rendiment que han fet de nosaltres? És ètic que l'habitació d'un hotel em costi més diners a mi que al meu veí perquè saben a quin barri visc i poden interpretar la meva renda?

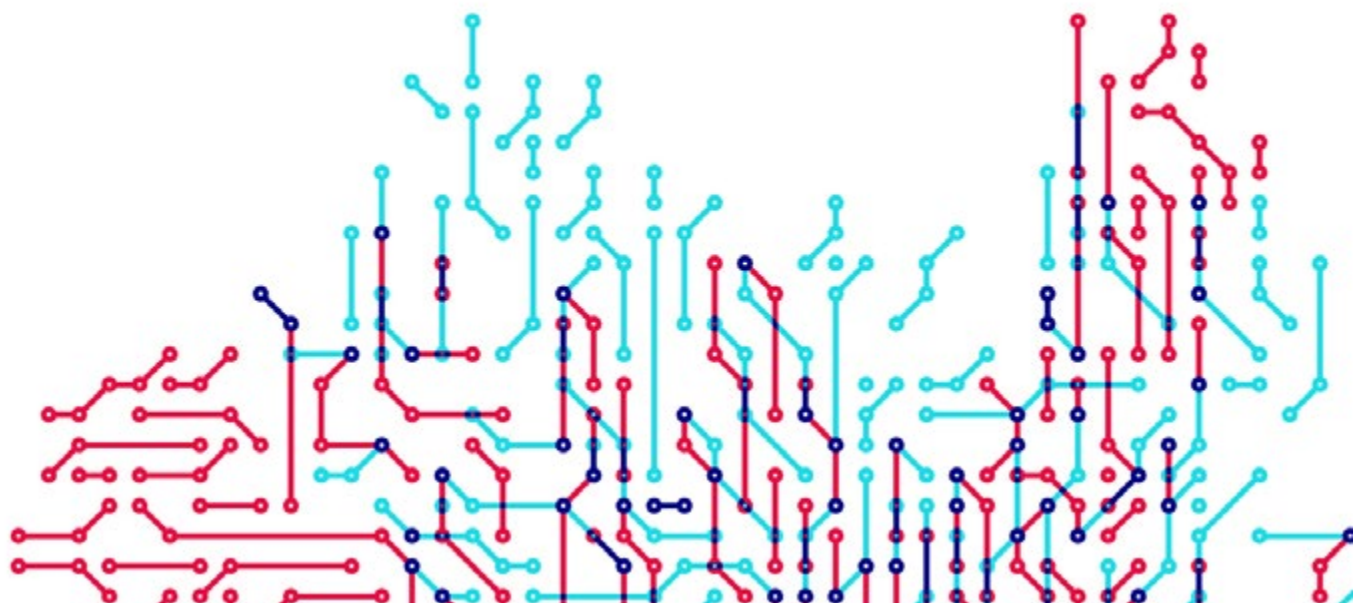
"Vivim en una societat de consum, basada en l'oferta i la demanda", respon la catedràtica d'ètica de la UAB. "I si compres una aspiradora, et pot incomodar tota la informació o publicitat que t'enviaran durant els mesos vinents, però és el món dels negocis. No l'acceptis. Una altra cosa és que amb les teves dades et vulguin perjudicar. Com, per exemple, que el banc sàpiga que ets alcohòlic i que n'informi l'asseguradora del cotxe perquè et discrimini en el pagament."

⁹¹ Comitè de Bioètica de Catalunya (<http://canalsalut.gencat.cat/ca/sistema-de-salut/comite-de-bioetica-de-catalunya/>).

La confiança no s'ha de perdre

“La vida en comú necessita la confiança vers l'altre. No s'ha de perdre mai”, explica Camps. “Tu renunciïs a part de les teves llibertats a canvi que uns representants públics que has elegit et protegeixin. Però si confies els teus diners a uns polítics i després es destapen casos de corrupció, hi perds la confiança.”

Sobre la confiança que dipositem en les xarxes socials com ara Facebook, i que ens fallen constantment, perquè el propietari ven les dades personals a consultores (Cambridge Analytica) que aprofiten per canviar l'orientació de vot en unes eleccions (EUA, Brexit, Brasil, etc.), la catedràtica d'ètica i filosofia argumenta que “aquestes plataformes són totalment prescindibles. No és l'algorisme de Twitter que et mostra la vida en una bombolla, ets tu, que acceptes aquesta situació. El que és una llàstima és que els polítics es comuniquin amb la població a través de Twitter perquè és un deteriorament de la política”. I insisteix que abans dels algorismes ens passava el mateix, triant el mitjà de comunicació amb el qual ens informàvem. “Però això no té gaire a veure amb l'ètica. El que èticament importa és preservar la llibertat individual, que ningú no t'acabi dient com has de viure.”



ALESSANDRO MANTELERO:⁹² “No hi ha una ètica global que es pugui aplicar a tots els països”

“L'ètica i la intel·ligència artificial (IA) és un debat molt vigent a Europa”, explica l'autor d'informes sobre regulació del *big data* i la IA.⁹³ Alessandro Mantelero vol puntualitzar que, quan parlem d'ètica, s'ha de deixar clar a què ens referim. “L'ètica es pot considerar com un nivell addicional i complementari del nivell de protecció de la llei. Ètica i dret estan connectats, però són diferents. Si l'ètica es transforma en llei, té sancions, i ja no és ètica”, puntualitza. I continua: “Quan parlem de protecció de dades, no parlem de riscos i beneficis, sinó de drets fonamentals. I cal tenir en compte que l'ètica a Espanya no és el mateix que a Rússia.” I llavors, quina preval? És possible una ètica global?

86

No hi ha una resposta concreta: “L'ètica mai no pot ser global, correspon a un entorn, a una comunitat. Per tant, és molt difícil que s'arribi a una ètica europea. Tampoc no serà el mateix en funció de les aplicacions de la intel·ligència artificial. L'ètica per a un algorisme que actui sobre malalts no serà la mateixa que per a un robot assistencial de gent gran. Ni per al robot que cuidi d'un nen i que li haurà de transmetre certs valors de vida. A més, a cada família predominen valors diferents”.

Davant aquest panorama complex, Mantelero conclou que tots els documents i els informes que s'estan redactant actualment no definiran un marc únic de l'ètica. “Estem al començament. Es parla molt d'ètica, però poc de societat. Si dic que invertiré molts diners per predir i controlar el crim, a partir d'una aplicació policial amb IA, la pregunta ètica i social podria ser: I per què no invertir aquests diners per crear escoles, amb un bon sistema educatiu, que intenti reduir la criminalitat i la diferència social. Així s'ataca l'origen del problema.”

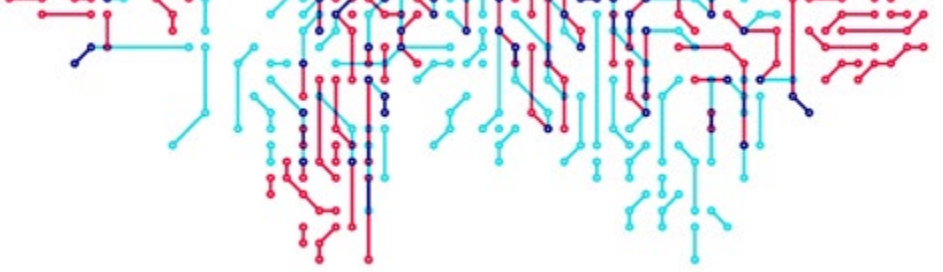
92 Alessandro Mantelero (<http://staff.polito.it/alessandro.mantelero/>).

93 *Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework* (<http://isiarticles.com/bundles/Article/pre/pdf/106203.pdf>).

Artificial Intelligence and Data Protection: Challenges and Possible Remedies (<https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6>).

També el *High Level Expert Group on Artificial Intelligence* de la Comissió Europea avança en aquestes qüestions. (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).





**Professora del Departament de Medicina de la Universitat de Barcelona (UB).
Subdirectora de l'Observatori de Bioètica i Dret de la UB**

ITZIAR DE LECUONA:⁹⁴ “Cal moderar l'entusiasme tecnològic que vivim”

L'experta en els aspectes ètics i legals de la biomedicina alerta que, en l'àmbit de la recerca biomèdica, bona part dels projectes de recerca que s'avaluen estan basats en la convergència de tecnologies, entre les quals es troben la intel·ligència artificial i l'aplicació de l'analítica de dades massives.

“La intel·ligència artificial ja és aquí, i tots estem contribuint amb els nostres conjunts de dades personals al seu desenvolupament”, explica. “Avui és possible explotar dades a escala massiva, per millorar la presa de decisions mitjançant el desenvolupament d'algorismes, que permetin –correlacionant les dades– permetin determinar patrons de comportament i predir conductes. I això ho volen fer tots: des de la iniciativa privada fins al sistema públic de salut. Per tenir sistemes de salut més eficients o perquè la ciutadania es beneficiï d'una medicina personalitzada. No fer-ho seria gravíssim des del punt de vista ètic”, afegeix De Lecuona.

87

El paper del ciutadà

Portem la intel·ligència artificial integrada en els aparells, com és el cas dels mòbils. Emetem dades contínuament des dels diferents dispositius digitals que utilitzem. També, hi ha sistemes més complexos en els quals s'exploten les nostres dades personals amb diversos objectius.

“El problema és que es prenguin decisions sobre mi, mitjançant dades que són meves, però que jo mai no sabré qui les té, ni per a què s'estan fent servir. Si l'ètica tracta de la felicitat i de societats més lliures, l'aplicació de la IA ens ha de fer reflexionar sobre el paper que hi tenen els individus, com a titulars d'aquestes dades personals. Qui en té el control i qui l'hauria de tenir? Es tracta de repensar la capacitat de controlar les nostres dades i la nostra intimitat en la societat digital”, argumenta Itziar de Lecuona.

Cal fomentar l'alfabetització digital, per poder prendre decisions lliures i informades com a ciutadans. En aquest sentit, l'experta en aspectes ètics assenyala que són nombrosos els professionals i les empreses que, en tant que tercers, col·laboren en la recerca científica. Per tant, tenen responsabilitats sobre la programació dels algorismes i el

94 Itziar de Lecuona (<http://www.bioeticayderecho.ub.edu/es/itziar-de-lecuona>).

tractament de les dades confidencials. Són, en la major part dels casos, professionals que tenen un coneixement expert, però no han rebut formació en ètica.

De Lecuona proposa revisar quins perfils –com ara els científics de dades– s’haurien d’incorporar en l’avaluació de la recerca biomèdica basada en intel·ligència artificial per part dels comitès d’ètica de la recerca. També planteja que s’analitzi quins coneixements i formació en ètica i protecció de la confidencialitat de les dades personals caldria que rebessin tots els agents que treballen en IA, per assegurar que s’hi compleixin els principis d’autonomia, beneficència, justícia i explicabilitat.⁹⁵

Mandra tecnològica

La subdirectora de l’Observatori de Bioètica i Dret de la UB destaca que a Catalunya tenim unes bases de dades de salut de qualitat, com ara la Història Clínica Compartida⁹⁶(HC3) o el SIDIAP.⁹⁷ El repte ara és que la informació que hi contenen sigui interoperable i es pugui reutilitzar. Però De Lecuona posa èmfasi en el fet que “cal moderar l’entusiasme tecnològic que tenim i evitar una confiança cega en la IA en salut. La decisió final –especialment en l’àmbit biomèdic– ha de romandre en el professional, no en la màquina. Els algorismes *per se* no poden tenir el poder de decisió”.

88

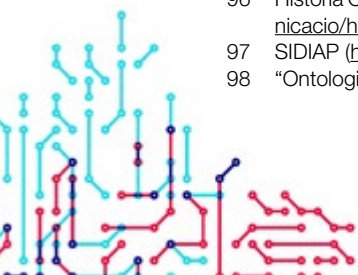
Des del punt de vista individual, Itziar de Lecuona proposa que ens espolem la mandra tecnològica que arrosseguem i que ens informem sobre com actuen els sistemes automatitzats que afecten la nostra vida quotidiana. Cal evitar que l’algorisme contribueixi a fomentar les desigualtats i les discriminacions, o a perpetuar-les. “No podem continuar pensant que nosaltres i els nostres conjunts de dades no són importants, i que ens és igual qui hi pugui accedir. Som rellevants, de tant en tant contribuïm a generar ontologies⁹⁸ per dissenyar i perfeccionar la intel·ligència artificial mitjançant els nostres conjunts de dades. Hem de ser curiosos i promoure una cultura de respecte de la intimitat. En la societat digital, protegir les dades personals és protegir les persones.”

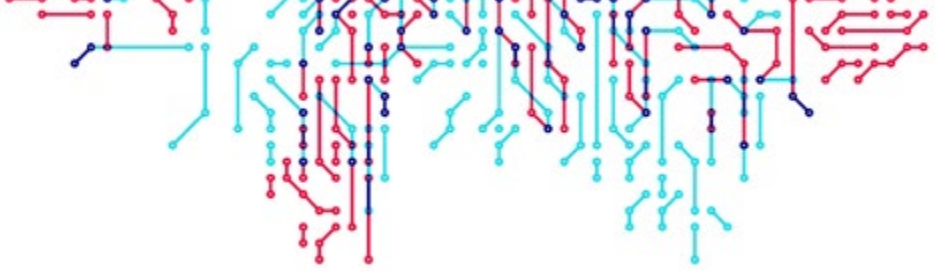
95 *Principis establerts pel High Level Expert Group sobre Intel·ligència Artificial de la Comissió Europea* (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

96 Història Clínica Compartida (http://salutweb.gencat.cat/ca/ambits_actuacio/linies_dactuacio/tecnologies_informacio_i_comunicacio/historia_clinica_compartida/).

97 SIDIAP (<https://www.sidiap.org/index.php/es>).

98 “Ontologia” ([https://ca.wikipedia.org/wiki/Ontologia_\(tecnologia_de_la_informaci%C3%B3\)](https://ca.wikipedia.org/wiki/Ontologia_(tecnologia_de_la_informaci%C3%B3))).





Professor investigador del Consell Superior d'Investigacions Científiques (CSIC). Exdirector de l'Institut d'Investigació d'Intel·ligència Artificial (IIIA)

RAMÓN LÓPEZ DE MÁNTARAS:⁹⁹ “S’hauria de poder certificar que els algorismes estan nets de biaixos”

Ramon López de Mántaras va coordinar l’elaboració de l’Estratègia Espanyola per a la Intel·ligència Artificial¹⁰⁰ –feta pública el març del 2019–, un pas importantíssim per definir les normes ètiques que corresponen a la seva implementació. Igualment, va promoure la Declaració de Barcelona,¹⁰¹ redactada fa dos anys per la comunitat científica i que marca molt bé les directrius ètiques que s’haurien de seguir.

Un dels punts prioritaris que la Declaració de Barcelona menciona és el del retiment de comptes (*accountability*). És a dir, que quan un algorisme prengui decisions, les persones afectades puguin rebre –en termes entenedors– una explicació de per què les ha pres i se’ls permeti qüestionar-les amb arguments raonats. Però dos anys més tard, aquest és un aspecte pendent de resoldre que preocupa molt els investigadors.

“Com a consumidor, ciutadà, tens dret a demanar explicacions,¹⁰² amb la Llei de protecció de dades europea. Pots preguntar per què t’han denegat un ajut, o no t’han donat un crèdit. El codi de l’algorisme no te’l donaran, però s’hauria de poder exigir que una entitat neutral avalués si l’algorisme és just. No existeix una entitat semblant a Catalunya. Algorithm Watch¹⁰³ o Ethical Tech Society¹⁰⁴ poden fer aquesta funció en l’àmbit internacional.

“Estaria bé que la Generalitat tingués un organisme que certifiqués si l’algorisme té un biaix o no. Això ja es fa amb els aliments i els medicaments, però no amb la IA. S’hauria de crear un segell de justícia (*fairness*). Tampoc no caldria certificar tots els algorismes, però sí aquells en què hi ha una decisió automatitzada que pot perjudicar significativament les persones. Això té un cost elevat per a l’Administració. Però els cotxes bé que passen la ITV. Si es fa per a altres coses...”

89

99 Ramon López de Mántaras (<http://www.iiia.csic.es/staff/ramon-l%C3%B3pez-de-m%C3%A1ntaras>).

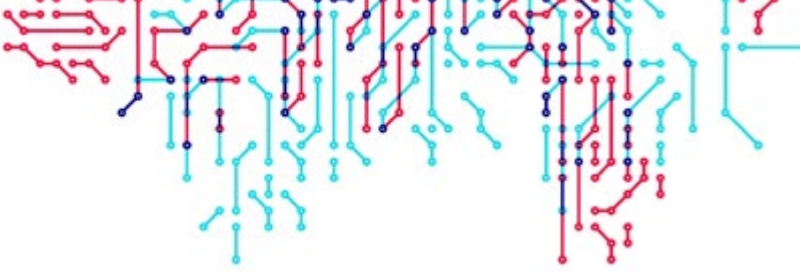
100 Estratègia espanyola de I+D+I en Intel·ligència Artificial (http://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf).

101 “Barcelona Declaration for the proper development and usage of artificial intelligence in Europe”. 2017 (<https://www.iiia.csic.es/barcelonadeclaration/>).

102 Ramón López de Mántaras. “La ética en la inteligencia artificial” (<https://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/el-multiverso-cuntico-711/tica-en-la-inteligencia-artificial-15492>).

103 Algorithm Watch (<https://algorithmwatch.org/en/>).

104 Lorena Jaume-Palusi (<https://twitter.com/lopalasi>).

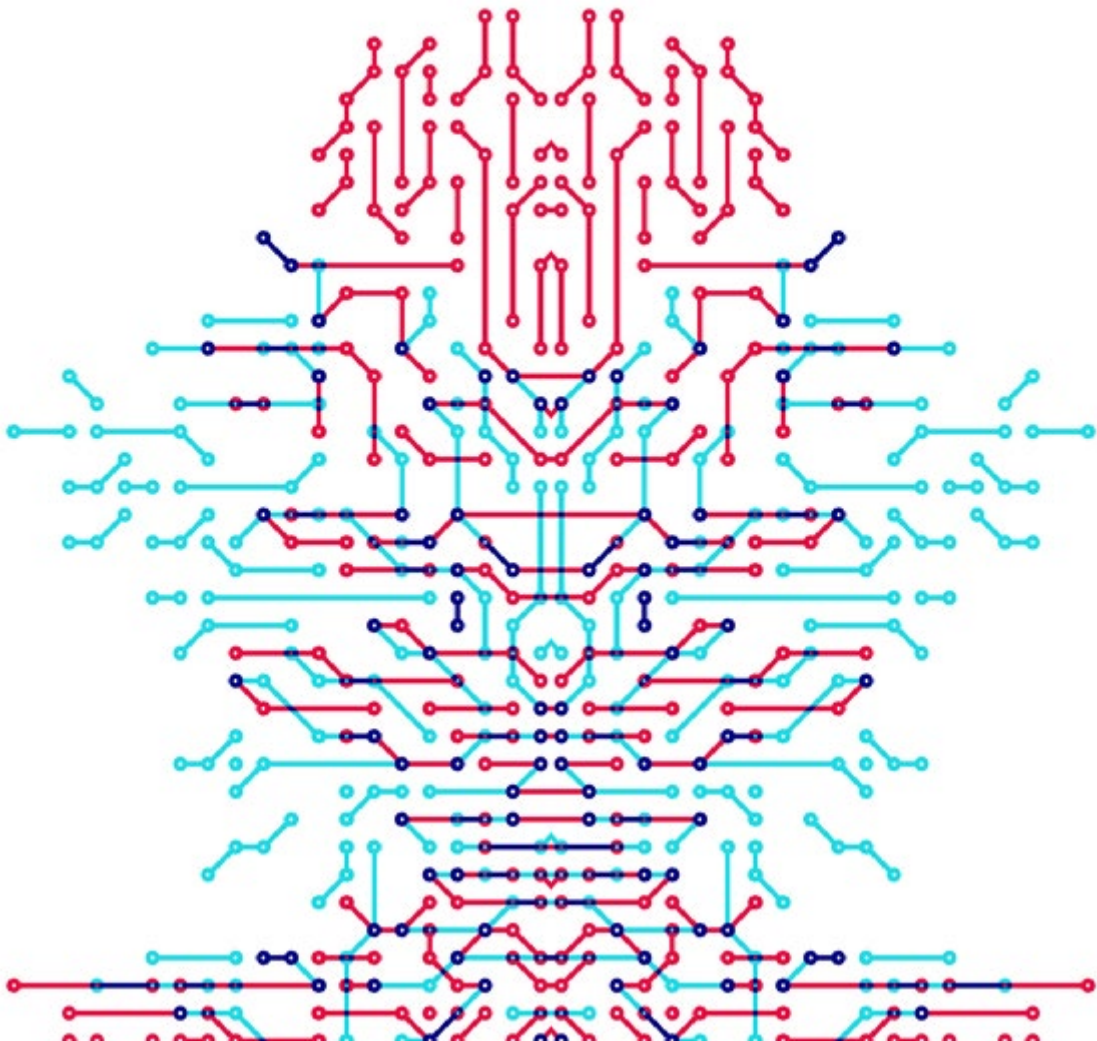


Perpetuar biaixos

“Sovint passa que les dades massives amb les quals entrenes l’algorisme estan esbiaixades, perquè són dades del passat, i en utilitzar-les de nou, perpetues i amplifiques biaixos del passat. Però les dades no te les pots inventar. Les has d’agafar d’algun lloc.”

Com detectar els biaixos? “Si en eliminar d’un algorisme la informació de raça, la decisió canvia, ja hi has detectat un biaix. El 100% de correcció no existeix, però, només que se’n poguessin treure els biaixos més importants, ja seria un avenç. Com el dispensador de sabó que no treu res quan s’hi acosta una mà negra, perquè el sensor només funciona amb la pell blanca. Fatal! I passa el mateix amb el reconeixement facial que confon les cares negres amb ximpanzés.”

“L’èmfasi en la importància dels aspectes ètics és la principal diferència que Europa pot abanderar respecte al que s’està fent en IA als Estats Units i la Xina. Per això, la UE exigeix a cada país que tingui una estratègia nacional per a la seva aplicació i que defineixi molt bé aquests criteris mínims.”



Doctora en informàtica i professora d'investigació a l'Institut de Robòtica i Informàtica Industrial (CSIC-UPC)

CARME TORRAS:¹⁰⁵ “Cal formar en criteris ètics a qui desenvolupa la tecnologia”

“Hi ha actualment una gran tendència a treballar amb aprenentatge profund (*deep learning*), un tipus de xarxa neuronal amb moltes capes que el que fa és associar entrades amb sortides. És un tipus d'aprenentatge anomenat de *caixa negra*, perquè no se sap amb exactitud què passa en aquest procés d'associació. Si les dades tenen biaixos, l'algorisme aprèn malament. Però qui el posa en marxa, sigui un banc per donar-te un crèdit o un organisme públic per avaluar un currículum o assignar una subvenció, no et podrà donar una explicació de com s'ha arribat a aquella decisió. Hi ha un projecte molt ambiciós de la Unió Europea¹⁰⁶ d'introduir el concepte d'*explicabilitat* (*explainability*) per obtenir explicacions de com ha processat la màquina les dades per arribar a una conclusió determinada. La comunitat informàtica està molt sensibilitzada per trobar una manera general d'argumentar les decisions que pren la màquina. A més, cal expressar els arguments en termes comprensibles per persones no-expertes en informàtica.”

Carme Torras –autora de novel·les de ciència-ficció des de les quals enfoca cap a on evolucionarem els humans i les màquines– reclama dos aspectes: “1. Regulació per assegurar els criteris mínims en l'aplicació de la intel·ligència artificial en la societat, en el moment actual. 2. Educació, començant per formar en criteris ètics als qui desenvolupen la tecnologia.”

91

Formació en tecnoètica

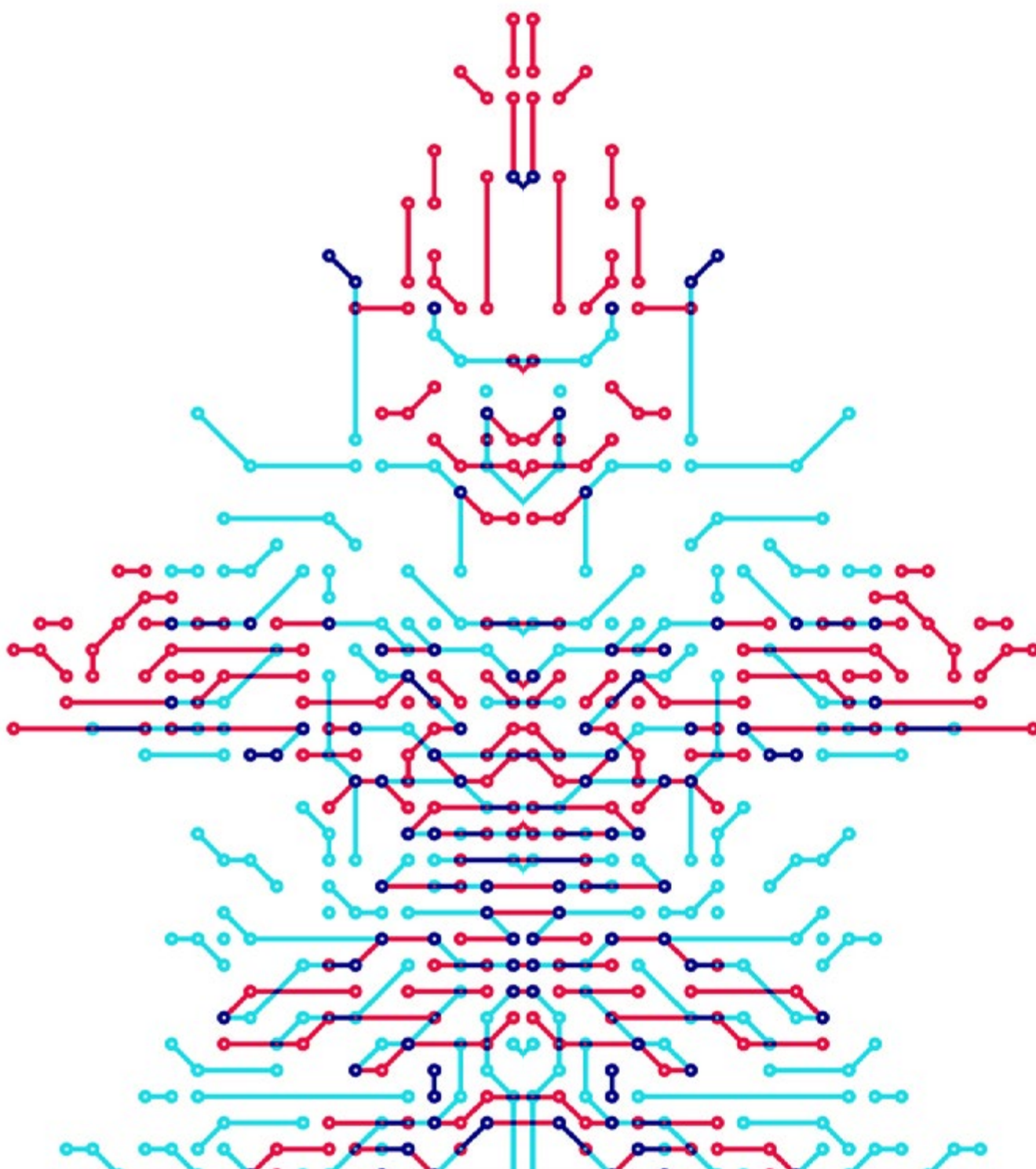
Segons Torras, “per ser partícip d'aquest futur, s'ha de conèixer com s'està creant”. Ella és una defensora de formar formadors, per tal que entre ells vagin disseminant els conceptes de *tecnoètica* o *roboètica*, que serien la fusió del coneixement de les tecnologies i/o robòtica amb una base de conceptes ètics.

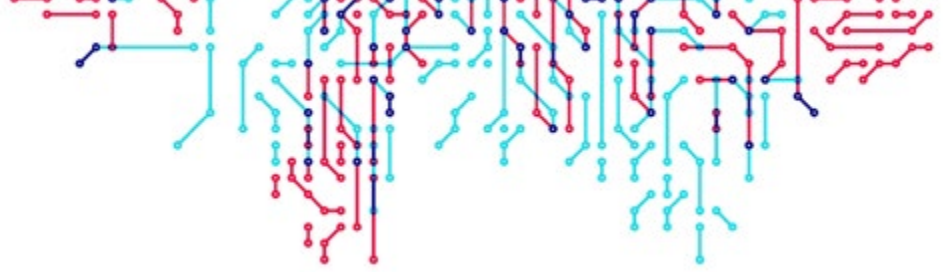
“Voldria arribar a secundària, als instituts, als joves de primària. Caldria començar ja, perquè no sigui massa tard. Aquesta generació creixerà en un món totalment tecnològic que estem creant ara. Hauria de ser una matèria transversal, per assignatures de llengües, de matemàtiques, de socials, de filosofia, de tecnologia. I es podria dir: ‘Ètica en la Societat Digital’.”

¹⁰⁵ Carme Torras (<https://www.iri.upc.edu/staff/torras>).

¹⁰⁶ *Ethics Guidelines for Trustworthy AI* (<https://ec.europa.eu/futurium/en/ai-alliance-consultation/guidelines/1>).

També proposa la creació d'un portal informatiu, en renovació i ampliació permanent, on es trobessin des de consells pràctics per preservar la privacitat al mòbil fins a regulacions i normatives, drets i dubtes digitals. "Crec que hi ha una franja molt gran de la població, gent molt diversa, amb curiositat per saber com manegar-se millor en aquest moment tecnològic, però no troba aquest servei mínim d'informació rigorosa. Aprofitem aquesta inquietud i vehiculem-la", afegix Carme Torras.





Cap de Tecnologia de NCENT. Catedràtic d'Informàtica de la Universitat Pompeu Fabra i de la Northeastern University

RICARDO BAEZA-YATES:¹⁰⁷ “La revolució tecnològica actual necessita ètica”

“Avui, les persones generen dades de moltes maneres” –explica Ricardo Baeza-Yates. “I aquestes dades es recullen, analitzen i tenen un valor més gran ara que en el passat. Poden fer-se servir amb motius legals, comercials o per manipular-nos electoralment. Les dades ja permeten predir comportaments o pràctiques de compra. Si tu em diguessis què preguntes a un cercador web en un dia, jo podria deduir si ets home/dona, jove o gran, localització, etc. Aquesta informació que donem voluntàriament, juntament amb el rastre dels nostres hàbits, pot ser utilitzada amb finalitats positives o negatives”, afegeix.¹⁰⁸

“La revolució actual necessita ètica, sí”, emfatitza el catedràtic d'informàtica. “Una part de regulació –no gaire estricta per no aturar el desenvolupament tecnològic– i unes mínimes normes ètiques com ara la IA explicable. És a dir, si un algorisme prediu o decideix automàticament, ha de poder donar-me una resposta de com ho ha fet. Perquè això passi, la gent ha de saber que hi ha biaixos en les dades que s'introdueixen en els algorismes”, explica el catedràtic d'Informàtica de la UPF. Baeza-Yates diu que la tecnologia sempre va més ràpida que la part social, i que només ens preocupem de l'ètica quan sorgeixen problemes, com ara mateix.

93

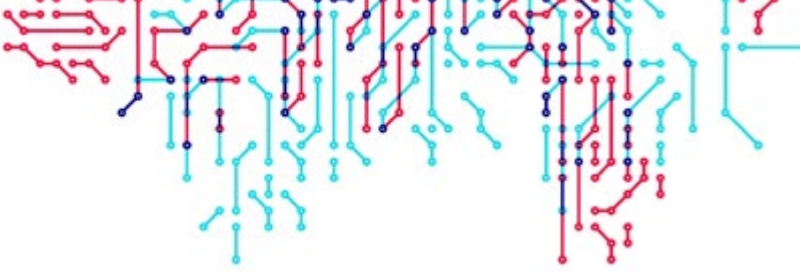
El Pep Consciències dels prejudicis

Si fem servir els algorismes sabent que tenen biaixos; per què deixem, doncs, que prenguin decisions? I Ricardo Baeza-Yates respon: “Per què tenim humans que prenen decisions si també s'equivoquen? Tots tenim prejudicis. Els sistemes automatitzats són de gran ajuda en situacions en què els biaixos no influeixen gaire, com, per exemple, en el control aeri: tenir persones moltes hores, en tensió, és més perillós que entrenar màquines per a aquesta tasca. No es cansen, estan programades, són més eficients.”

“De biaixos, n'hi ha un munt, el problema és que no ens adonem que existeixen fins que cometem un error. El mateix passa amb els nostres prejudicis, que no en som conscients la major part de les vegades. Les màquines poden ajudar-nos a adonar-nos-en i a crear

¹⁰⁷ Ricardo Baeza-Yates (<http://www.baeza.cl/spanish.html>).

¹⁰⁸ Participació en el documental xilè *Por la razón y la ciencia* *Por la razón* (https://www.youtube.com/watch?time_continue=1253&v=7PCC7tRyM2I).



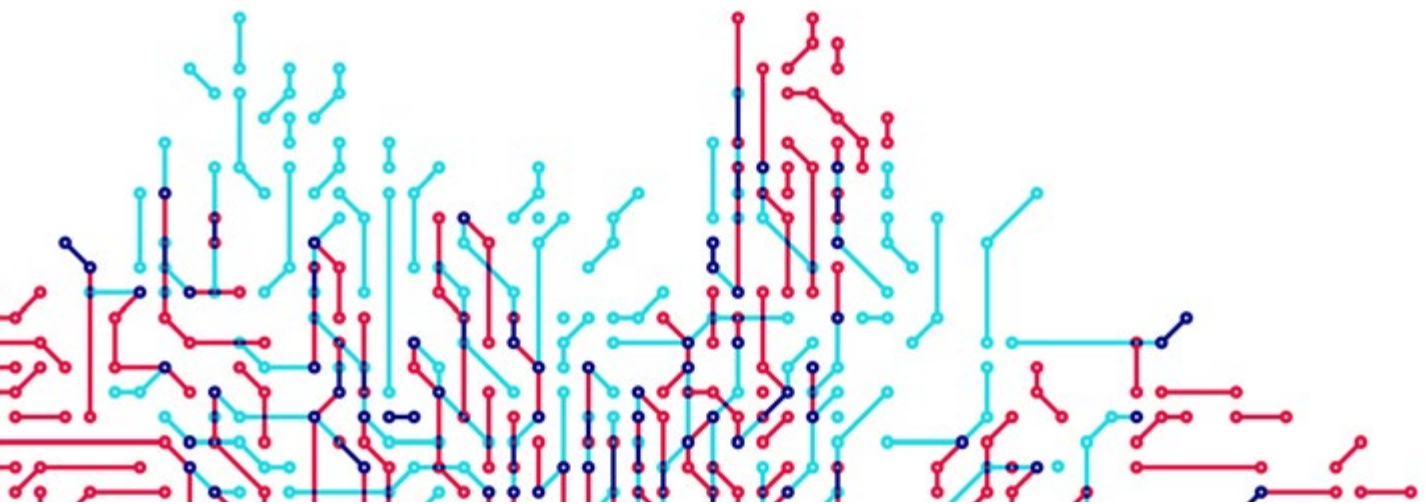
un món més just.” Sí, però com? “Es podria crear un Pep Consciències¹⁰⁹ (o assistent virtual) que alertés quan s’està cometent un prejudici, al parlar, actuar, jutjar, etc. O que ens avisés quan algú està tractant de manipular-nos.”

I sorgeix una nova contrapregunta: quants de nosaltres acceptaríem que una màquina (un mòbil o altre aparell) ens escoltés el que diem (en privat i en públic) i detectés els nostres prejudicis?

Millorar la societat

“En els pròxims anys, veurem com la detecció dels biaixos dels algorismes ajuda a millorar el món”, conclou Baeza-Yates per deixar un missatge optimista. Però alerta: “En aquest nou ecosistema de les dades, la tendència és tenir cada cop menys privacitat. Quan s’accepta una aplicació digital, cal llegir els termes d’ús i comprovar si el servei que se n’obindrà val la pena respecte a la privacitat que es perdrà”, afegeix. “I saber si legalment et poden demanar les dades que et demanen. Cal un canvi cultural enorme perquè tothom respecti la privacitat dels altres”. I recomana més educació, saber que fàcilment es pot ser manipulat, entendre la tecnologia actual i defensar els drets que assegurin la llibertat individual.

94



¹⁰⁹ Participació en el podcast “La inteligencia artificial tiene que ser nuestro Pepito Grillo”. BBVA (<https://www.bbva.com/es/podcast-la-inteligencia-artificial-tiene-que-ser-nuestro-pepito-grillo-ricardo-baeza-yates/>).

Director científic del grup d'Intel·ligència Artificial d'Alt Rendiment en el Centre de Supercomputació de Barcelona. Catedràtic d'Intel·ligència Artificial de la Universitat Politècnica de Catalunya

ULISES CORTÉS:¹¹⁰ “Cal ensenyar els límits de la tecnologia i pensament crític”

“Abans qui decidia donar-te un préstec o fer-te una operació de cor eren els algorismes que estaven al cap del banquer o del metge. Tu no sabies amb quins criteris ho decidien. Ara està passant el mateix, però amb màquines. Seria bo que una empresa pogués assegurar que la decisió és explicable si hi estàs en desacord. Però avui això és impossible, perquè ningú no els està dissenyant perquè siguin explicables.”

“Hi ha una petita porció de la població que té por dels automatismes pels errors comesos fins ara i, en concret, de l'ús de l'aprenentatge automàtic (*machine learning*). Però a la majoria de les persones, els importa ben poc el que es faci amb els algorismes. I els hauria d'importar, perquè les seves vides canvien constantment en funció del que decideixen un munt de màquines. S'han de produir casos molt flagrants, com el de Cambridge Analytica, per començar a prendre'n consciència. No és dolent fer servir Facebook, si se saben les conseqüències. El dolent és deixar a un menor que hi entri sense advertir-li els riscos que comporta”.

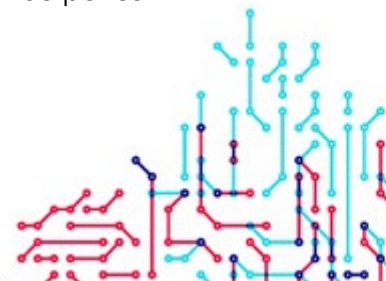
“Tinc clar que en una societat occidental, de passat cristià, hi ha un conjunt de valors comuns. Dels quals ja parlaven Plató i Aristòtil. Als EUA és d'utilitat plantejar-se: quin és el mal mínim que puc causar? És molt evident que el més just per a nosaltres no ho és per als de Malawi. Però ens ajuntem i decidim quin és el consens.”

L'ètica és cultura

“Faria tres preguntes a tothom: 1. Coneix la constitució del seu país? 2. Sap els manaments de la seva religió? 3. Quin és l'últim llibre d'ètica que ha llegit? Si les respostes són negatives, com podem demanar als enginyers que siguin ètics? L'ètica l'has de posar dins d'una cultura. I després demanar als programadors que facin algorismes que no perjudiquin els drets dels altres.”

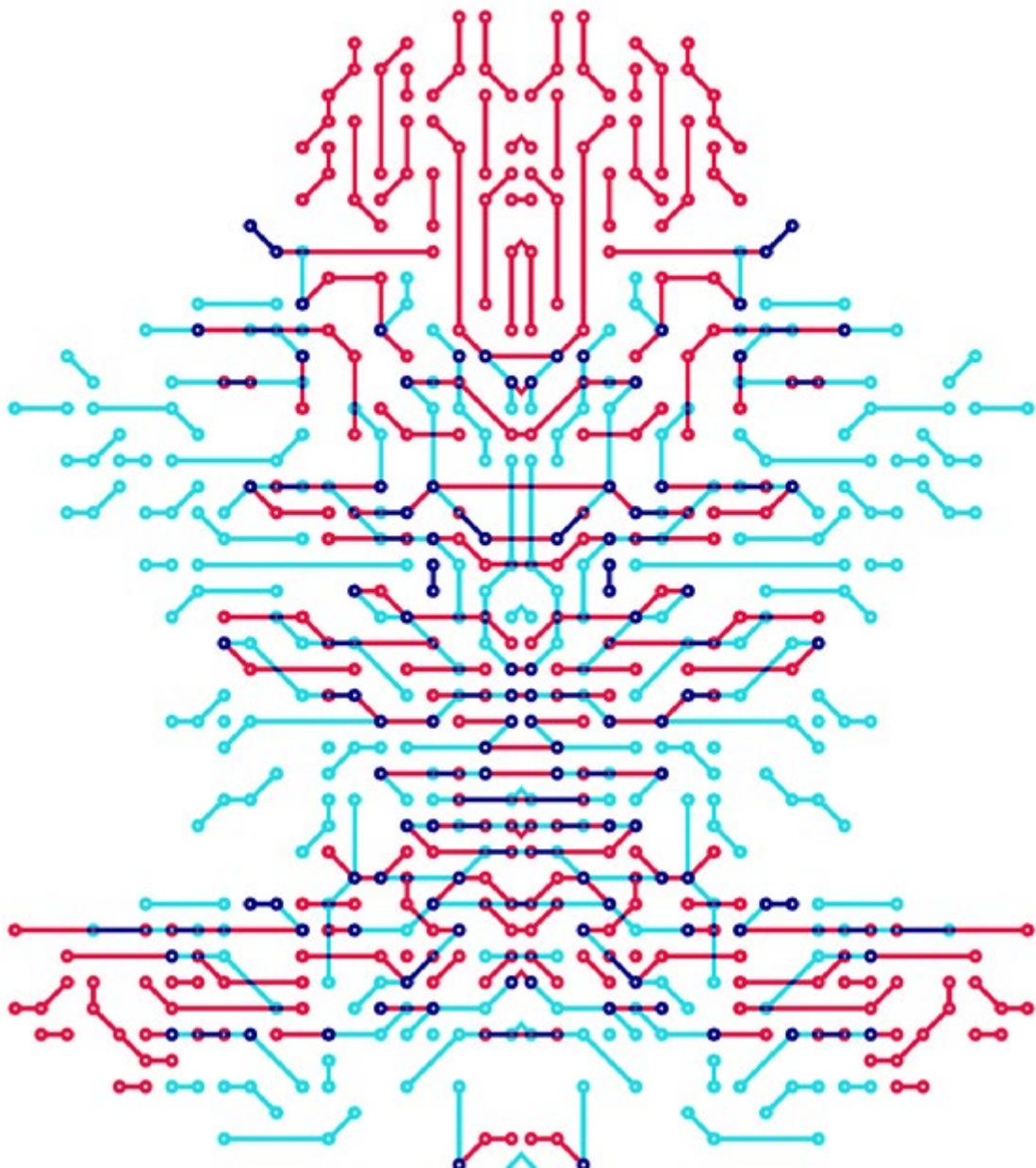
“Cap tecnologia no ha de fer por. El que fa por és la gent que la fa servir. De disseny, és innòcua, però pot tenir uns objectius il·lícits, bona part dels quals per guanyar diners. Mark Zuckerberg té una frase boníssima: ‘Nosaltres no necessàriament som ètics, però som legals’. Hem d'ensenyar a entendre els límits de la tecnologia. Però també pensa-

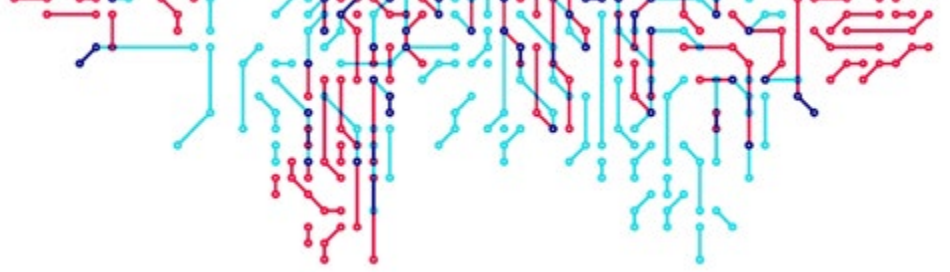
¹¹⁰ Ulises Cortés (<https://www.bsc.es/cortes-ulises>).



ment crític. A la gent no li interessa ser crítica, li interessa ser feliç. Viu en una distòpia increïble i es pregunta: 'per què hauria de preocupar-me de les meves dades? Si no tinc res que amagar!'."

"El problema és que ni tan sols ens plantegem aquestes qüestions. L'única manera de revertir aquesta situació seria que els polítics s'ocupessin dels problemes derivats de la tecnologia; i educar els educadors. Els pares han delegat la instrucció a l'estat. Però quan es tracta de l'educació sobre la tecnologia, no existeixen aquests pares. Si vols canviar el món, has d'aconseguir que els professors d'escoles, d'instituts i d'universitats sàpiguen de tecnologia. I que els programadors aprenguin més ètica. Però cada cop es veu més la televisió, es passa més temps a les xarxes socials, es fa menys esport i es llegeix menys."





Doctora en enginyeria informàtica i professora catedràtica d'Intel·ligència Artificial i Ciència de les Dades de La Salle - Universitat Ramon Llull.
Membre del grup de recerca en Data Science for the Digital Society (DS4DS)

ELISABET GOLOBARDES:¹¹¹ “S’estan resolent problemes que no sabem com s’han resolt”

“Tot canvia quan els telèfons intel·ligents superen en vendes els mòbils sense connexió a Internet. Democratitzem les dades i ara les rep i les genera qualsevol usuari. A l'Àfrica hauran passat de no tenir cap connexió a tenir, en breu, 5G perquè tothom tindrà un mòbil. Tothom vol saber. El poder ja no està en uns quants, el té la gent.”

“A les meves classes del grau d'informàtica i de robòtica dono conceptes d'IA-robòtica-ètica. És una qüestió d'educació. Hi ha molta polèmica també amb el reconeixement facial, amb la manera com lliurem –sense cap consciència– els nostres trets de la cara, que són únics. Els donem a qualsevol empresa que ens ofereix un entreteniment, i no sabem de quina manera els tractarà ni sota quina legislació. Seria molt més interessant que donéssim les dades a un servei de salut públic.”

“La tecnologia –per primera vegada– ha superat les demostracions matemàtiques que fan els humans. Aquest és l'únic punt fosc de la intel·ligència artificial actual. S'estan resolent problemes que no sabem explicar com s'han resolt, especialment en certes xarxes neuronals (*deep learning*).”

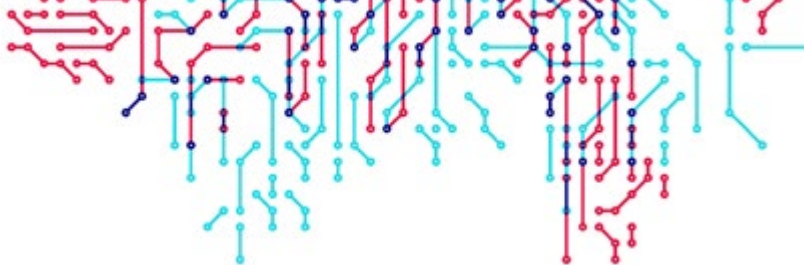
97

L'algorisme és neutre

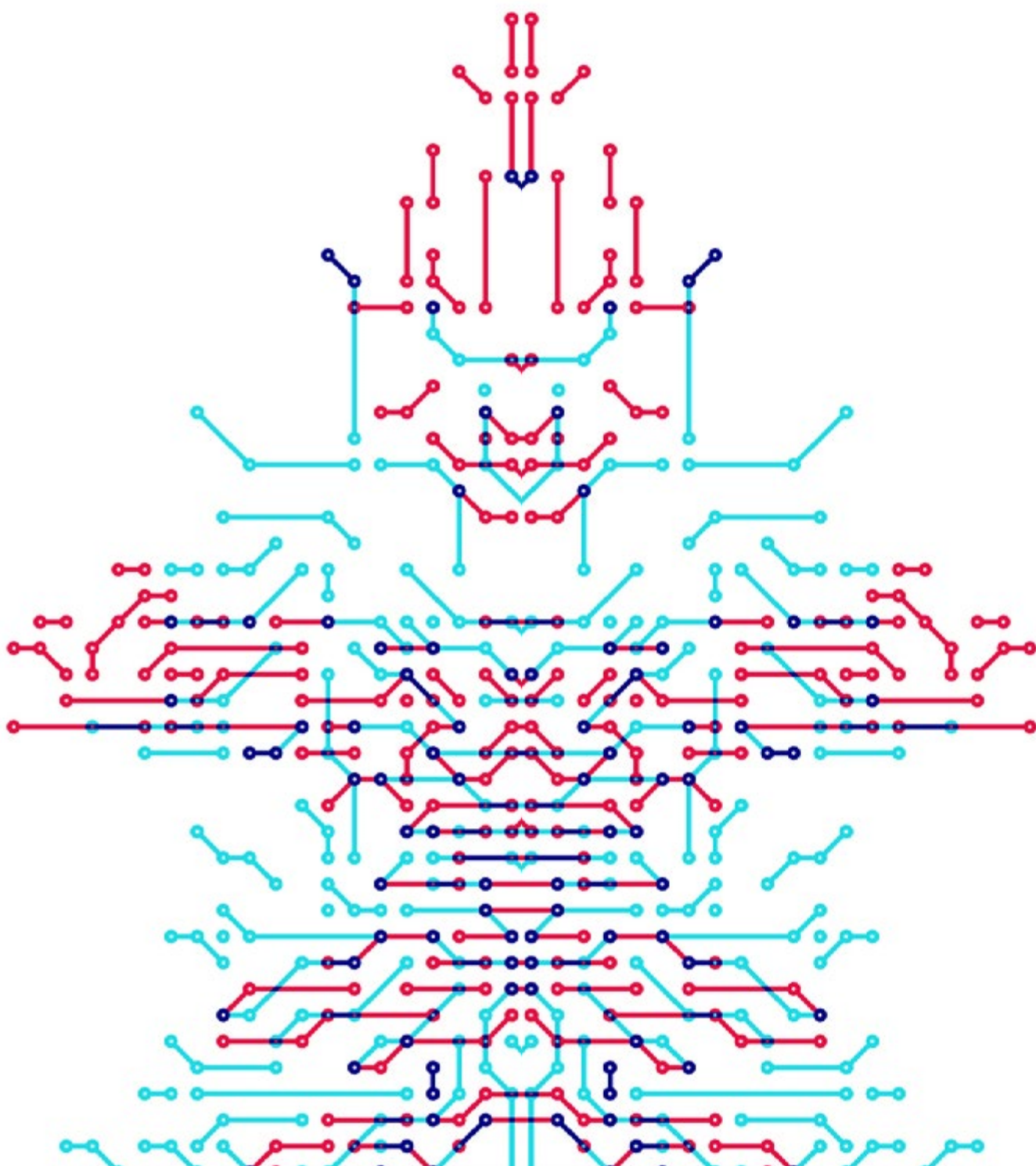
“Però m'agradaria desvincular l'algorisme de la part ètica. L'algorisme és un cargol d'un cotxe, només una peça. És neutre. El problema està en les dades massives que s'hi introdueixen per entrenar-lo i en l'objectiu de sortida. Els programadors tampoc no són els culpables: compleixen amb el que diu l'empresa o qui encarrega fer aquell algorisme. Caldrien polítiques públiques per obligar que les dades per entrenar-los fossin representatives de la realitat, plurals, ètiques i igualitàries. El problema és que qui mana crear l'algorisme –amb una finalitat determinada – pot tenir un objectiu econòmic o comercial concret.”

“Caldria una etiqueta de ‘Dades ètiques’, un segell de qualitat que assegurés que no hi ha discriminacions ni biaixos. Igualment, caldria assegurar que l'objectiu de l'algorisme té un propòsit ètic.”

¹¹¹ Elisabet Golobardes (<https://www.salleurl.edu/es/elisabet-golobardes-ribe>).



“Tot i això, em refio més de la intel·ligència artificial que d’alguns polítics: sota quins interessos treballen ells i decideixin el que em repercuteix com a ciutadana? No m’ho expliquen, tampoc. Em refiaria més d’una màquina que intentés ser justa a partir d’uns criteris definits que no de les decisions no raonades de molts líders del món. S’està posant massa el focus en la IA però no sé si és per despistar.”



JORDI VITRIÀ:¹¹² “Si no ens posem les piles, estem venuts”

“A mi no m’agrada la paraula *ètica*, perquè no s’entén. Prefereixo parlar de transparència, privacitat, diversitat i retiment de comptes (*accountability*). Si un hotel em diu que compartirà les meves dades amb altres, puc acceptar-ho o no. Però sé a què juguem. Que una companyia aèria em digui que les compartirà amb l’hotel per fer-me ofertes, puc acceptar-ho o no. Aquesta és la *transparència* que vull. Si un hospital o una administració m’assegura que anonimitzarà les dades i que en farà un ús responsable, jo puc acceptar-ho o no. Però exigeixo aquesta *privacitat*. El criteri de *diversitat* també hi ha de ser.

Quina empresa o govern em pot certificar que les dades no contenen biaixos que puguin provocar una discriminació? Avui, cap. Potser calen associacions ciutadanes que ho defensin. Però abans cal curiositat i entendre el funcionament de la IA. Un dels possibles negocis del futur serà el d’auditor/certificador d’algorismes, perquè els biaixos siguin els mínims. Per acabar, penso que el *retiment de comptes* hauria de ser també per a les empreses. Si rebo una oferta de wifi poc després de dir-li al meu germà per Whatsapp que tenia un problema amb el wifi actual. Em pot demostrar Whatsapp que no ha fet ús de les meves dades per generar aquesta oferta? Perquè a mi ningú no m’ha demanat consentiment per això. Hauríem de discutir sobre aquests termes. L’ètica, per definició, és voluntària, però el compliment d’aquests quatre punts hauria de ser obligatori.”

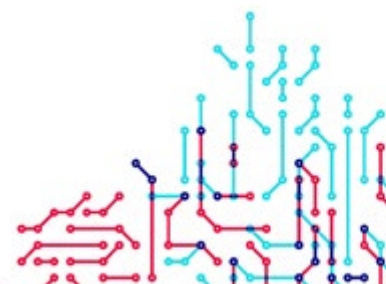
99

Els perills indirectes

“Existeixen altres perills per a la societat”, continua explicant Jordi Vitrià. “I per fer-hi front només pots conscienciar i educar la població. Un d’aquests perills és el de l’“Economia de la dopamina’.” I argumenta que un dels béns més preuats avui dia és l’atenció, el temps que cada persona inverteix en una xarxa social, un joc, un mitjà, navegant per una web de compravenda de productes, etc. “Els models de negoci en línia intenten maximitzar el temps d’atenció, oferint-te allò que et tindrà enganxat: una sèrie, els tuits dels teus seguidors, les fotografies dels teus amics, els productes que et temptaran, etc.”

D’altra banda, ens alertem molt del reconeixement facial que fan països com la Xina, o d’experiments com el que es va fer a Londres captant les imatges de tots els vianants

¹¹² Jordi Vitrià (http://www.ub.edu/dept_matinfo/professors/vitria-marca-jordi/).



sense avisar-los,¹¹³ “però..., enguany, a Barcelona, durant la celebració del Mobile World Congress, molts congressistes van entrar per la cara. Literalment! Animaven a entrar amb reconeixement facial. Si m’ho proposen, encara que sigui opcional, he de saber quin ús farà l’organització de totes aquestes cares. Perquè no ho han explicat, i espero transparència”, afegeix Vitrià. “Per a mi, educar la gent en tecnologia és això.”

“Des del món tecnològic s’acostuma a dir només part de la veritat. Qualsevol revolució tecnològica acaba generant més llocs de treball. Però el que no s’explica és que hi ha un període –més o menys llarg– durant el qual molta gent es queda sense feina perquè mai no serà capaç d’accedir a les noves professions. A un forner o a un taxista no el pots reconvertir en analistes de dades. Quant temps els dones, als taxistes? I als metges de capçalera? Els investigadors que fan recerca mèdica sempre hi seran, però i els cirurgians? Desapareixeran, sens dubte. Les infermeres no han de patir, perquè parlem d’empatia. Si volem ser més intel·ligents que en el passat, hauríem de dissenyar una estratègia que permetés passar el període de transició, amb mecanismes de regulació.”

Alfabetització de les dades (*data literacy*)

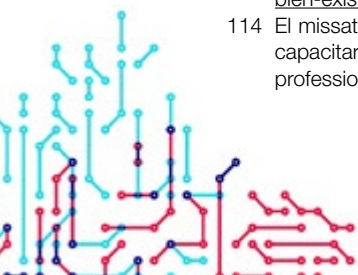
100

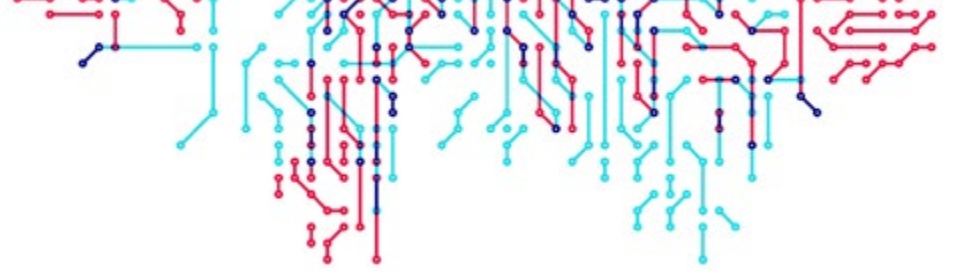
“El problema és que alegrement tothom pugi les seves cares a Facebook. La plataforma de Mark Zuckerberg té una voluntat de manipulació impressionant. Ens ho va demostrar en les eleccions nord-americanes. Encara ens cal molta formació.”

“Els comitès d’ètica dels hospitals han d’entendre de què parlem. Els bancs?, el mateix. I els polítics també haurien de formar-se en intel·ligència artificial perquè prendran moltes decisions que hi tindran a veure. L’ideal seria introduir un curs de *data literacy* a totes les carreres universitàries (de lletres, de ciències, polítiques, de dret o artístiques). La Universitat de Berkeley¹¹⁴ ho va implementar ja el 2017. Té lògica! Si no tens la base, pots equivocar-te molt! A Finlàndia, a Austràlia o a Nova Zelanda també s’estan implementant experiències similars. La Generalitat s’ho hauria de plantejar, des de l’escola, als instituts, a les universitats. Si no ens posem les piles, estem venuts. No podem regalar les dades com si no res. No ens podem creure tot el que ens venen.”

113 Karma Peiró, Ricardo Baeza-Yates. “Algoritmo, yo también existo” (<https://www.karmapeiro.com/2019/09/24/algoritmo-yo-tambien-existo/>).

114 El missatge de la Universitat de Berkeley és: “Vivim en un món envoltat de dades cada cop més complex. Aquest curs et capacitarà per respondre preguntes i explorar problemes amb els quals hauràs de lidiar en un futur immediat, en la teva vida professional o privada” (<https://data.berkeley.edu/news/fall-milestones-data-science-education>).





Professora del Departament de Matemàtiques i Informàtica de la Universitat de Barcelona (UB). Directora del grup consolidat de recerca Machine Learning and Computer Vision de la UB.

PETIA RADEVA:¹¹⁵ “Haurien d’existir els donants de dades”

Petia Radeva és una investigadora internacional que fa vint-i-sis anys que treballa en l'àmbit de la salut i de la visió per computació. És una defensora acèrrima de la tecnologia, dels algorismes d'aprenentatge automàtic, la visió per computador i la intel·ligència artificial.

“El món s’està digitalitzant. En els darrers cinc anys hem creat més informació que en els milers d’anys de civilització. Els algorismes són cada dia més importants per avançar científicament. Els necessitem per mesurar i comparar i per analitzar si hi ha una malaltia o no, si un òrgan està danyat o sa. Gràcies al *machine learning*, les màquines poden aprendre a partir d’un entrenament sobre un conjunt de dades. Per exemple, els podem mostrar mil casos d’òrgans malalts i la mateixa quantitat de sans. I així els algorismes poden construir les regles per decidir si un òrgan es sa.”

Radeva no veu un problema en el fet que els algorismes es comportin com a caixes negres, sinó que hi veu un repte. “Són difícils d’explicar, és cert. I això no és nou. Les xarxes neuronals van tenir el seu boom als anys setanta i ja en aquest moment ja se les va acusar de ser caixes negres. A la pràctica, el que va passar en aquell moment és que els ordinadors no eren tan potents com ara. Eren difícils de resoldre els problemes reals i més encara explicar-los. Ara són tremendament útils per la capacitat de computació que tenim i la quantitat de dades que utilitzem per entrenar-los. Qui s’atreveix a prescindir-ne? No només necessitem algorismes potents i eficients, també autoexplicables. Els metges no poden aplicar algorismes a la pràctica clínica si no els entenen. Per això, últimament, la comunitat científica està avançant molt a desenvolupar nous mètodes perquè els algorismes generin autoexplicacions.”

101

Donant de dades

“Avui hi ha possibilitat de recollir moltes dades. Ens hem de preguntar quin ús se’n fa? Com sempre, pot ser positiu o negatiu. Crec que el repte d’aquest moment tecnològic és trobar la manera que tothom s’animi a donar dades personals a la ciència. No conec cap multinacional que, després d’haver fet un ús il·lícit de les dades personals s’hagi enfonsat per la regulació europea (RGPD). En canvi, sí que l’RGPD posa obstacles al

¹¹⁵ Petia Radeva (<http://www.ub.edu/cvub/petiaradeva/?p=36>).

fet que els investigadors recol·lectin dades personals i puguin fer-ne ús per avançar en solucions a problemes de la societat i la salut de la gent.”

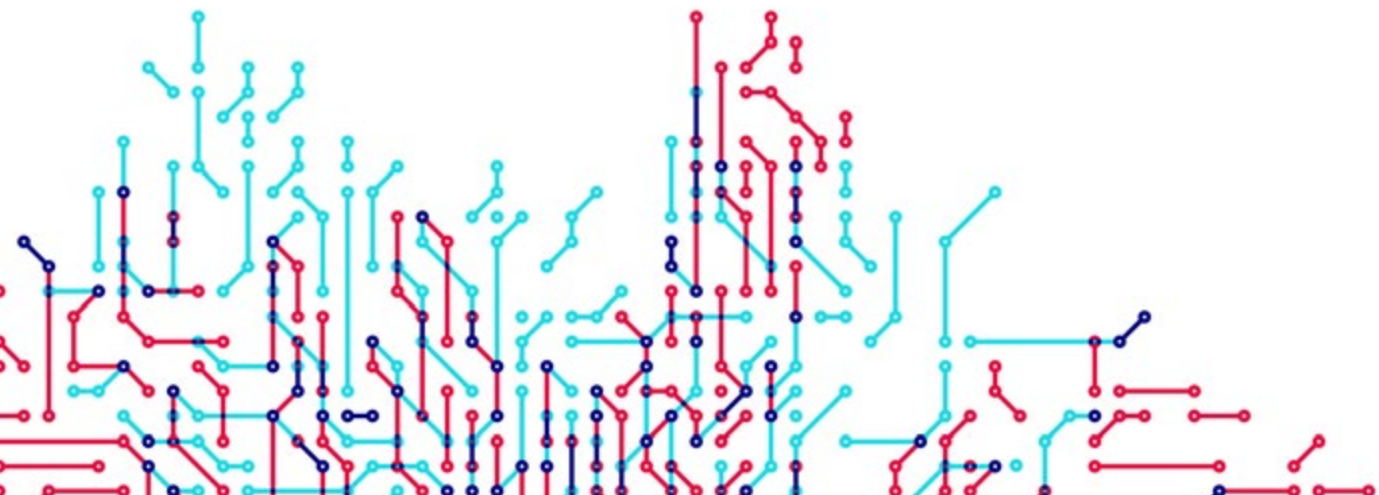
“Amb els patrons de moviment de les persones es podria optimitzar el transport en les ciutats, promoure-hi una vida més activa i menys sedentària, cosa que estalviaria diners a la sanitat pública, perquè evitaria certes malalties, o se’n podria fer prevenció més acurada. Però no ho podem fer amb dades de tota la població perquè els investigadors no tenen accés a aquesta informació. M’agradaria participar en alguna campanya per promoure la donació de dades, com les que es fan d’òrgans o de recollida de sang. S’ha de promoure la donació de dades amb transparència i explicacions, assegurant-ne un bon ús. Al Regne Unit hi ha un banc de dades accessible, de manera gratuïta, als investigadors. És una iniciativa lloable per avançar en la ciència.”

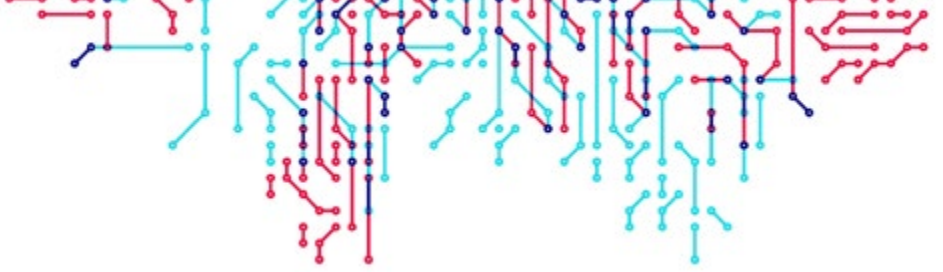
“Quan un organisme o una universitat et demana dades (per exemple, de la teva activitat física durant l’últim mes o any), la reacció inicial molt sovint és de por o de rebuig. No pensem que poden servir per avançar en malalties que salvaran la vida dels nostres fills o nets. Però, després, les cedim tranquil·lament a qualsevol empresa a Internet a canvi d’un entreteniment. Per què?”

102

Més enllà de la tecnologia

“Un exemple. Com pot ser que Espanya, estàndard de la dieta mediterrània, tingui la taxa més alta d’obesitat infantil de tot Europa? Tenim una taxa d’obesitat infantil altíssima perquè no tenim dades d’hàbits d’alimentació, i no podem avançar. Tan fàcil com seria que, a través dels mòbils, les famílies i les escoles, les recollissin i les donessin per avançar contra aquesta malaltia. La tecnologia ha donat un gran pas en les capacitats de col·lecció de dades i el seu processament. Ara falta que ho faci la societat.”





**Professor investigador distingit del Departament de Tecnologies de la Informació i les Comunicacions de la Universitat Pompeu Fabra (UPF).
Lidera el grup d'investigació de Ciència de la Web i Computació Social**

CARLOS CASTILLO:¹¹⁶ “És molt complicat que la gent sàpiga que un algorisme està esbiaixat”

El grup d'investigació de Ciència de la Web i Computació Social de la UPF, juntament amb la Universitat Tècnica de Berlín i el Centre Tecnològic Eurecat, va crear un algorisme que detecta i mitiga biaixos d'altres algorismes. El sistema l'han anomenat FA*IR¹¹⁷ i actua sobre altres sistemes automatitzats (de cerques o recomanació) que no prenen en consideració la representació de diferents grups.

“Per exemple, una persona que fa servir aquest sistema pot determinar que hi hagi com a mínim un 20% o un 40% de dones en els resultats, o de gent de menys de 25 anys, o de persones de tal origen, o de professionals de tal sector”, explica Carlos Castillo. “Els biaixos, que sempre hi són, es poden detectar i mitigar.”

Clar, que si els algorismes s'estan aplicant ja en molts sectors, caldrà molta gent detectant i mitigant els biaixos. Com es pot fer? “Necessites transparència”, respon Castillo. “I que hi hagi la capacitat que el sistema rendeixi comptes i expliqui les seves decisions. Però, com a individu, tens molt poques eines per saber si al darrere hi ha un algorisme amb un biaix potent que et discrimina. Si cerques a LinkedIn, periodistes a Barcelona, segurament no hi sortiràs mencionada i segurament hi ha un biaix de gènere, en funció de les dades del passat. Fins que algú no fa un estudi gran, com en aquest cas nosaltres, i se saben els resultats, és impossible.”

103

Justícia algorísmica

“Algun organisme públic català hauria de pensar en això. Potser l'Autoritat Catalana de Protecció de Dades –o una institució similar– pugui donar una solució perquè els ciutadans, consumidors tinguin més referències sobre els algorismes que prenen decisions. Per exemple, a França, la Comissió Nacional d'Informàtica i Llibertats¹¹⁸ (CNIL) s'involucra en aquests temes. També la normativa europea de protecció de dades (*General Data Protection Regulation*)¹¹⁹ accepta denúncies de col·lectius o associacions ciutadanes.”

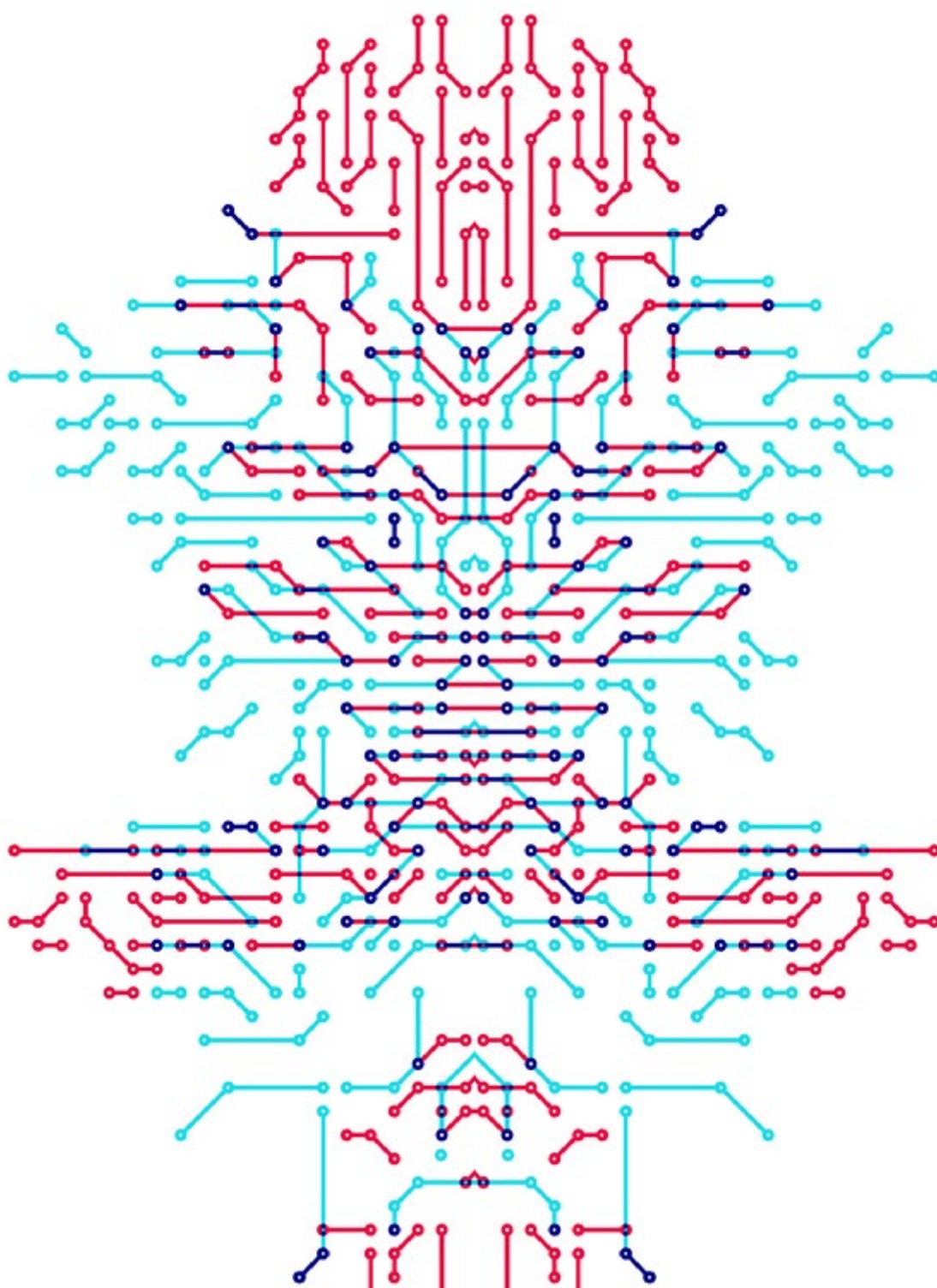
¹¹⁶ Carlos Castillo (<http://chato.cl/research/>).

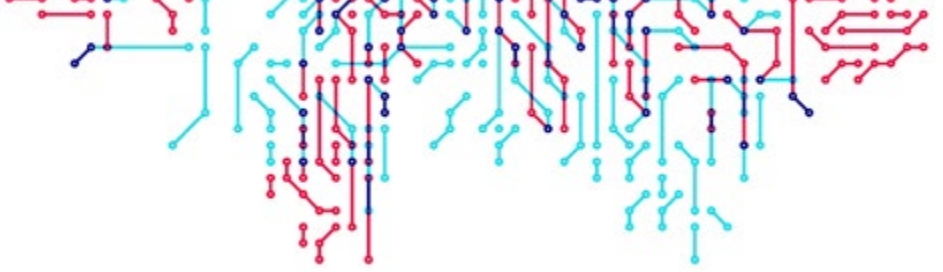
¹¹⁷ M. Zehlike, F. Bonchi, C. Castillo, S. Hajian, M. Megahed, R. Baeza-Yates. “FA*IR: A Fair Top-k Ranking Algorithm” (<https://arxiv.org/pdf/1706.06368.pdf>).

¹¹⁸ CNIL (<https://www.cnil.fr/>).

¹¹⁹ GDPR (<https://gdpr-info.eu/>).

“La Llei de protecció de dades europea (GDPR) també obliga a un processament just de les dades; per tant, s’hauria d’obligar que es compleixi aquesta justícia en els algorismes”, afegeix el director del grup d’investigació de Ciència de la Web i Computació Social de la UPF. “L’RGPD permet a associacions ciutadanes posar denúncies. Per exemple, si creus que les periodistes estan discriminades a LinkedIn, podries ajuntar dades de molts periodistes i denunciar-ho.”





Catedràtic del Departament de Ciències Polítiques i Socials de la Universitat Pompeu Fabra (UPF)

CARLES RAMIÓ:¹²⁰ “La utilització intensiva de la IA i la robòtica és l’única manera d’assegurar l’estat del benestar”

Carles Ramió insisteix que l’Administració pública ja arriba tard a la implementació de la intel·ligència artificial (IA). Aquest estiu va publicar un codi ètic amb 24 punts,¹²¹ un esborrany del que s’hauria d’estar pensant ja. Entre els punts més polèmics hi ha el d’apostar per un sistema de validació dels algorismes que utilitzi el sector públic. “S’ha de controlar el procés de disseny i entrenament del sistema automatitzat. No pot ser que estigui al càrrec només d’enginyers. Calen filòsofs, sociòlegs, gent que pensi en qüestions ètiques, que tinguin en compte les discriminacions de tota mena”, explica el catedràtic de la UPF. “I si comprem algorismes a l’empresa privada, els hem de fer públics també.”

L’autor del llibre *La intel·ligència artificial i l’administració pública*¹²² considera que “les caixes negres no poden existir. Jo no dic que tothom hagi de conèixer la fórmula de l’algorisme. Però un sistema que dicta sentència ha de poder ser revisat, almenys pels experts i els advocats, saber quin ha estat el procés d’entrenament o per què no se l’ha ensenyat més diversitat.”

“L’Administració s’hauria de dotar de sistemes que certifiquessin els algorismes perquè fossin plurals. Hauria de tenir un equip de visió ètica i tècnica, mirant els biaixos, que fes transparent l’algorisme i el sistema d’entrenament.”

Una gran oportunitat

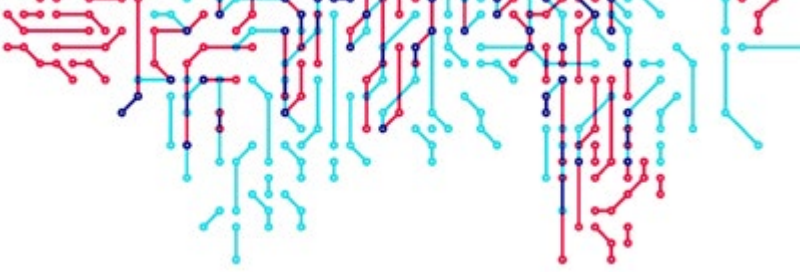
“Veig una enorme oportunitat en la IA. S’ha d’implementar com més aviat millor. Els algorismes, ben dissenyats, són imprescindibles. Si totes les dades d’una unitat de cures intensives les tens monitorades, i alimentes bé l’algorisme, el sistema no pot fallar. Detectarà qualsevol anomalia o malaltia. Després, ja vindrà l’escrutini del metge que valorarà i confirmarà, o no, el diagnòstic de l’algorisme.”

“És cert que això implica una inversió d’entrada. Però l’única forma que continuem tenint un cert estat del benestar és amb la utilització intensiva per part de l’Administració

120 Carles Ramió (<https://www.upf.edu/es/web/politiques/entry/-/-/1182/401/carles-ramio>).

121 Carles Ramió. *Estatuto ético para la implantación de la inteligencia artificial y la robótica en la Administración pública* (<https://www.administracionpublica.com/estatuto-etico-para-la-implantacion-de-la-inteligencia-artificial-y-la-robotica-en-la-administracion-publica/>).

122 *La intel·ligència artificial i l’Administració pública*. Los Libros de la Catarata (https://www.todostuslibros.com/libros/inteligencia-artificial-y-administracion-publica_978-84-9097-590-9).



pública de la IA i la robòtica. Amb l'envelliment de la població, no podem fer front a totes despeses que impliquen en sanitat i serveis socials. La IA ben planificada i orientada es pot assolir. Però hem de canviar el discurs i ser més proactius com a administració. L'alerta és que els algorismes estan entrant en l'Administració pública de la mà d'empreses privades. Però, llavors, com podem ser transparents, si depenem de la privada?"



Catedràtic distingit de Ciència de la Computació de la Universitat Rovira i Virgili (URV) de Tarragona. Investigador ICREA Acadèmia de la URV. Director de la Càtedra UNESCO de Privadesa de Dades. Director del CYBERCAT - Centre de Recerca en Ciberseguretat de Catalunya

JOSEP DOMINGO:¹²³ “Intel·ligència vol dir ‘entendre’, i la màquina no entén”

Com la resta d'experts entrevistats, el catedràtic de la Universitat Rovira i Virgili declara que tenim un problema greu amb la intel·ligència artificial i, sobretot, amb l'aprenentatge profund. Perquè malgrat que són tremendament eficients, no hi ha manera de saber com ha après la màquina. “En el moment que una persona pugui sol·licitar un crèdit des d'una web, i després d'haver introduït unes dades el sistema automatitzat li respongui que no l'hi concedeixen, voldrà tenir una explicació. I aquí vindrà el problema.”

“Cada cop tenim menys coneixement de per què passen les coses. El *broker* borsari s'està acabant, costa molt lluitar contra un inversor robòtic que agafa variables econòmiques de tot el món i en dècimes de segon decideix què fa. L'explicabilitat¹²⁴ o transparència algorítmica és un intent de recuperar el coneixement. No et pots confiar únicament de la màquina. Has d'entendre de què va el món; si no, en perds el control.”

107

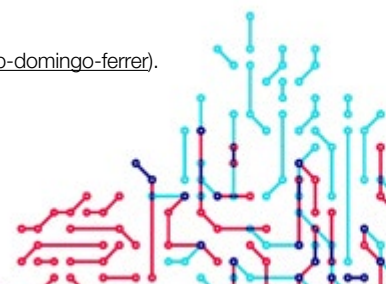
El dret a l'explicació

“La màquina no és intel·ligent. Tampoc no sap per què ha decidit el que ha decidit. Intel·ligència vol dir ‘entendre’, i la màquina no entén. Se l'entrena amb unes dades històriques (per exemple, dades de crèdits concedits en el passat que els sol·licitants van tornar o no van tornar) per ajustar els paràmetres de l'algorisme de decisió. Aquí conclou l'aprenentatge. Després, se li introdueix un cas concret sobre el qual cal prendre una decisió (per exemple, una nova sol·licitud de crèdit) i la màquina-algorisme calcula la decisió fent servir els paràmetres apresos. Continuant amb l'exemple del crèdit, si per la raó que sigui, l'algorisme troba que el nou sol·licitant s'assembla als que no van tornar el crèdit en el passat, la decisió serà refusar la sol·licitud, i a l'inrevés.

“La llei de protecció de dades europea (RGPD) consagra el dret a l'explicació, que en el fons és per protegir la democràcia. L'article 22 ha estat la pedra de toc. Amb el requisit legal ens adonem que no hi ha manera de donar explicacions satisfactòries. Cada vegada que algú impugni la decisió d'una màquina, uns tècnics hauran d'intentar

¹²³ Josep Domingo Ferrer (<http://www.urv.cat/es/universidad/conocer/personas/profesorado-destacado/2/josep-domingo-ferrer>).

¹²⁴ “Explicabilitat o transparència algorítmica” (https://en.wikipedia.org/wiki/Explainable_artificial_intelligence).



reconstruir de manera artesanal el que hagi fet. Però això no és escalable. Falten eines per explicar de manera automàtica.”

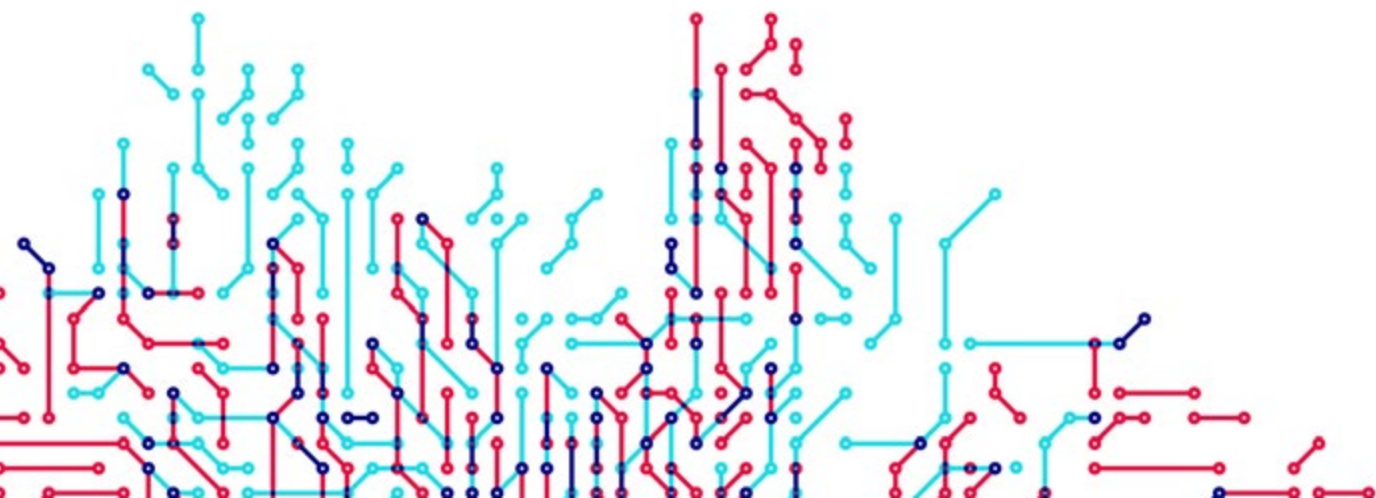
“Per sort per als qui fan servir algorismes de decisió automatitzada, la ciutadania encara no entén bé el que està passant. Va passar el mateix amb la protecció de dades. Però no he vist cap manifestació al carrer –ni tan sols quan va passar el cas de Cambridge Analytica amb Facebook–¹²⁵ per vulneració de la privadesa.”

La responsabilitat de l'algorisme

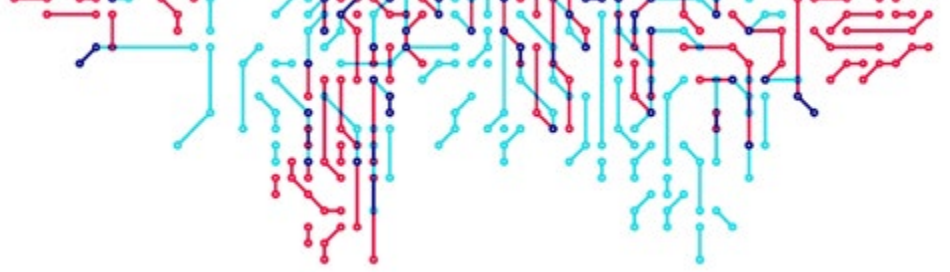
Les caixes negres preocupen i, per això, es reclama una transparència algorísmica. Però també amoïna definir de qui és la responsabilitat de la decisió de l'algorisme. Quan funcionin els cotxes autònoms (sense conductors) i hi hagi un atropellament, qui en serà el responsable? L'analista que l'ha dissenyat? El programador? El fabricant del cotxe, per haver-lo venut?

Josep Domingo Ferrer aporta una altra visió: “El fabricant de cotxes, el banquer, el metge o el polític fan l'encàrrec de l'algorisme. I tenen una idea vaga del que ha de fer el sistema automatitzat. L'analista fa un disseny i el passa al programador informàtic. Pel mig, hi ha moltes instruccions poc precises. Finalment, l'algorisme hauria de ser revisat de nou per l'analista i per qui n'ha fet l'encàrrec, però això passa poques vegades. D'altra banda, si no es trien bé les dades d'entrenament, pot ser que l'aprenentatge que fa l'algorisme només valgui per a homes de raça blanca i cristians, i que s'hi hagin negligit les particularitats de les dones, dels negres o dels musulmans, posem per cas.

108



¹²⁵ Karma Peiró. “L'escàndol Facebook-Cambridge Analytica: un cas per revisar la protecció de dades i molt més” (<https://www.naciodigital.cat/noticia/151744/escandol/facebook-cambridge/analytica/cas/revisar/proteccio/dades/molt/mes>).



Investigadora del grup d'Enginyeria del Coneixement i Aprenentatge Automàtic del Centre d'Investigació en Ciència Intel·ligent de Dades i Intel·ligència Artificial de la UPC. Vicedegana de Big Data, Ciència de Dades i Intel·ligència Artificial del Col·legi Oficial d'Enginyeria Informàtica de Catalunya

KARINA GIBERT:¹²⁶ “El cost energètic de la IA és enorme”

“La tecnologia, en principi, és neutra i el que té usos més o menys ètics és el que fem amb ella”, explica la investigadora Karina Gibert. “Els problemes ètics que cal tenir en compte avui són els biaixos, sí, però també l'explicabilitat, que és una qüestió molt més profunda i que ja s'ha comentat en aquest capítol.”

Gibert vol puntualitzar que només hi ha una part de la IA que no és explicable, la que es coneix com a *subsimbòlica*. “La part simbòlica, que imita la forma com els humans resolen els problemes, és totalment explicable. Però és terriblement cara, des del punt de vista computacional, i mai no ha fet un salt a problemes reals d'una certa complexitat perquè col·lapsa.”

“La part subsimbòlica intenta aconseguir que els ordinadors resolguin problemes que requereixen intel·ligència amb més qualitat que els humans, des del punt de vista dels resultats que aporta”, continua explicant la investigadora, “sense importar gaire si la manera com resol el problema imita o no la manera com ho fa l'humà. I aquesta és la IA no explicable, la que, entre altres coses, ha donat lloc al *deep learning*. Com a contrapartida, el *deep learning* ha aconseguit donar resultats respecte a problemes grossos que fins ara no s'havien resolt. Però està clar que, perquè el resultat d'una intel·ligència artificial tingui impacte en un context real, cal argumentar com sigui la decisió que s'ha pres. I això ho sabem els que –com jo– fa trenta anys que treballem en aplicacions reals.”

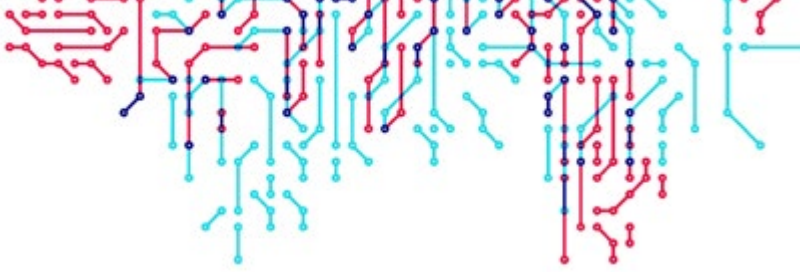
109

La petjada ecològica de la IA

Karina Gibert també aporta una reflexió nova sobre l'ètica, no mencionada fins ara. “El cost energètic que cal per executar tots aquests algorismes i per hostatjar totes les dades que participen en aquests processos és enorme. I té petjada ecològica. La pregunta és: necessitem que el teu rellotge mesuri la temperatura corporal cada cinc minuts, durant un any, quan se sap que la temperatura evoluciona d'una manera monòtona i hi ha canvis visibles cada hora?”

“Això té molt a veure amb un dilema que l'investigador Ricardo Baeza-Yates menciona sovint: ‘Dades massives o dades apropiades?’ (*big data* o *right data*). Potser necessi-

¹²⁶ Karina Gibert (<https://www.eio.upc.edu/en/homepages/karina>).



tem tenir-ne menys i que siguin informatives i representatives, fet que ens faria estalviar energia per emmagatzemar-les i processar-les. Des del punt de vista de la sostenibilitat és molt important tenir en compte quin tipus d'algorismes i dispendi de dades en fem.”

De qui és la responsabilitat si falla l'algorisme?

“Veig dos escenaris. 1. El de les intel·ligències artificials no adaptatives (no canvien els mecanismes de funcionament i mantenen el disseny), en què la responsabilitat recau sobre qui dissenya l'algorisme. Ho veig clar: qui dissenya l'algorisme ha de tenir clars quins són els valors de risc associats a les decisions o les recomanacions que el sistema pugui fer. Però també penso que potser no es visibilitzen prou els nivells d'incertesa associats a les solucions de l'algorisme. Això és com a les *guidelines* de la medicina, que diuen: ‘Quan arribi una persona amb un atac de cor, cal administrar aquesta dosi’. La dosi estàndard també està lligada a un munt d'incertesa, i a una persona en concret li pot anar malament. I quan això passa, el metge no assumeix cap responsabilitat si ha seguit la *guideline*. No veig gaire diferència de com s'hauria de tractar el que recomana una intel·ligència artificial, respecte al que decideix un doctor, que també es basa en recomanacions estàndards que poden no funcionar en un cas concret.”

110

“Escenari 2. Quan la IA és adaptativa. És complicat tenir ben acotats els tipus d'accions que pot desenvolupar a llarg termini. Si una intel·ligència artificial va guardant l'històric del que fa, i modificant el seu comportament segons el que va millor o pitjor, en funció de les accions prèvies, ens podem trobar que després d'un cert temps apareguin raonaments o recomanacions que no ens imaginem. És a dir, que la IA estigui anant més enllà, tota sola, i creant resolucions noves. Aquí és molt difícil exigir o pensar que qui dissenya l'algorisme pot tenir controlades totes les situacions i que pugui predir i limitar les que poden ser perilloses per als humans. Això planteja un debat sobre fins a quin punt volem que les intel·ligències artificials s'expressin amb tota la seva potència quan tracten amb persones o entorns on interaccionen amb l'humà i poden generar riscos.”

Karina Gibert proposa que es pugui auditar la intel·ligència artificial (igual que els col·legis auditen el programari en general), en cas de disfunció, per tal de determinar si hi ha hagut una mala descripció del problema, un mal disseny, una mala implementació o una denegació del sistema. En cada cas, la responsabilitat recau sobre una persona o sobre una altra. “Però sí que és responsabilitat del desenvolupador fer pensar en tots els límits, riscos i problemes a qui encarrega l'algorisme (o la seva funció), per tal que el disseny sigui el més complet possible, i el seu ús el més correcte possible. Això s'hi fa poc, i s'hi hauria de fer més.”

2.6. Dilemes pendents

La científica Marie Curie va dir en algun moment de la seva dilatada carrera: “No hem de témer res a la vida, només hem d'intentar comprendre-ho. És moment, ara, d'entendre més per témer menys”.

En l'apartat anterior, els principals investigadors d'intel·ligència artificial del nostre país han apuntat sobradament els dilemes que tenim avui amb la IA, com ara l'existència de biaixos en els algorismes, la falta d'explicabilitat quan actuen amb aprenentatge profund o la falta de consens sobre qui ha d'assumir la responsabilitat en cas d'un error de la màquina.

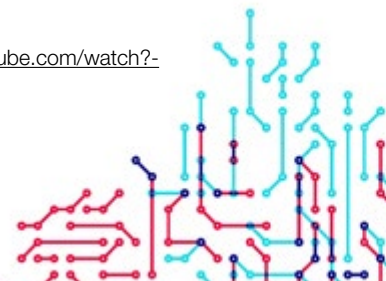
Amb el pas del temps sorgiran altres preguntes que, avui en dia, tampoc no tenen una resposta clara. “Com podem assegurar que les decisions automatitzades (o les actuacions que se'n deriven) no tenen un impacte negatiu per a les persones? Quins nivells de seguretat tenen aquests sistemes intel·ligents per garantir que no són vulnerables als ciberatacs o a un ús maliciós? Què passarà quan un algorisme ens conegui millor que nosaltres mateixos i pugui manipular el nostre comportament subliminalment?”, preguntava la doctora en intel·ligència artificial per l'Institut de Tecnologia de Massachusetts Institute of Technology (MIT), Núria Oliver, en la lliçó inaugural del curs 2019-2020 del sistema universitari català.¹²⁷

111

En paral·lel als avenços dels algorismes de decisió automatitzada (ADA), hi ha dos conceptes que no estem cuidant prou i que cada cop tenen més importància en les societats tecnològiques en què vivim: la confiança i la transparència.

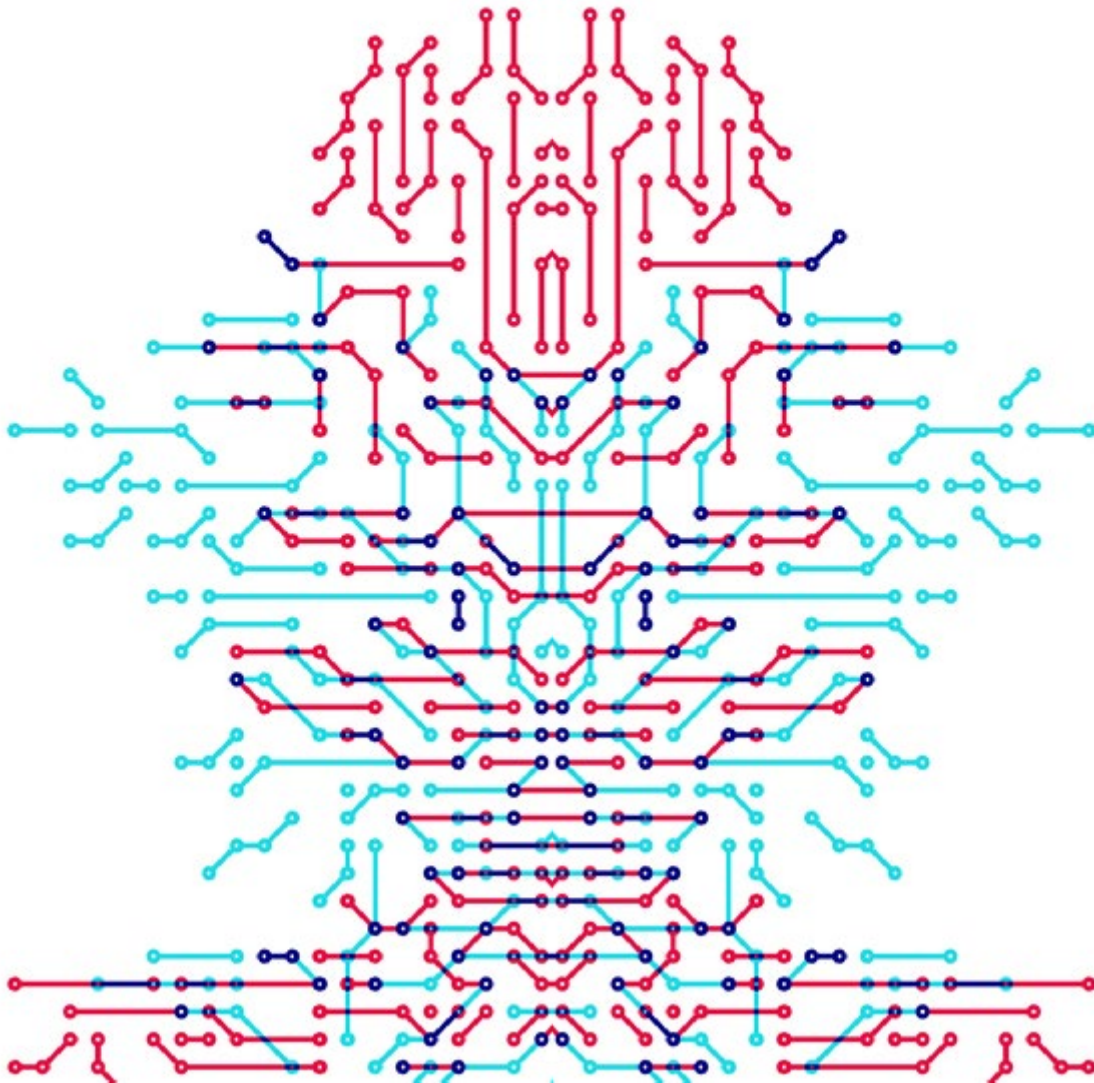
1. **La confiança** és un pilar bàsic en les relacions entre humans i institucions de qual-sevol societat. La tecnologia necessita la confiança dels usuaris, que deleguem cada cop més les nostres vides en serveis digitals. Però, en els darrers anys, hem canviat el sentit del terme. A través de la tecnologia, confiïem en desconeguts (quan lloguem una habitació de casa mitjançant AirB&B i obrim la porta a estranys; quan ens porten a un destí sense saber si condueix bé o malament qui porta el volant, sigui en BlaBlaCar o en Uber; quan transferim diners a través d'una aplicació de mòbil mitjançant una empresa de la qual no tenim referències precedents. Però la confiança se'ns ha colat entre els dits, sobretot després d'escàndols com el de Cambridge Analytica, que va manegar de manera il·legal dades de cinquanta milions de persones i va fer trontollar el sentit de la paraula *democràcia*.
2. I la **transparència**... Un terme molt emprat en els darrers anys, però molt poc consistent, perquè més aviat tenim opacitat tecnològica. “Un sistema computacional és

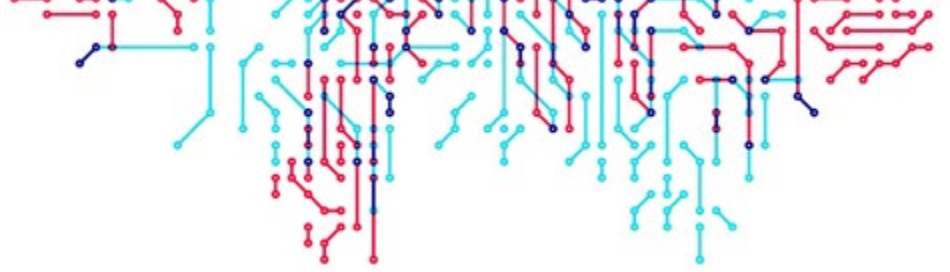
¹²⁷ Lliçó de Núria Oliver en la inauguració del curs 2019-2020 del sistema universitari català (<https://www.youtube.com/watch?v=DwCOKDwliXc>).



transparent quan una persona no experta, l'observa, l'entén i sap com funciona", explica Núria Oliver. "Però aquesta no és la realitat avui dia." Segons la investigadora, cada cop tenim més opacitat, bé perquè les empreses privades volen protegir la propietat intel·lectual dels seus algorismes, bé perquè la ciutadania no tenim un mínim de coneixement tecnològic per entendre les explicacions, bé perquè les administracions no concedeixen a la transparència la importància que es mereix (malgrat que la paraula surti en gairebé tots els programes electorals), o bé perquè l'aprenentatge profund impedeix tenir una explicació del perquè ha pres aquella decisió.

La IA s'està desenvolupant a tot el món a marxes accelerades. Qui la domini, no només tindrà poder econòmic, sinó també polític i social. Arribats a aquest punt, l'únic camí que hauríem de contemplar des de la ciutadania és aconseguir (i exigir) una intel·ligència artificial **confiable**, dissenyada i pensada **per a les persones**, amb un **sentit ètic profund**, que compleixi amb els valors de justícia, transparència (de la real) i equitat. Una intel·ligència artificial segura, que es pugui auditar i sobre la qual es puguin demanar explicacions.





2.7. Lectures recomanades per aprofundir-ne més

Aquesta secció recull normes públiques que regulen l'ús de la intel·ligència artificial a Europa, però també textos recomanats pels investigadors esmentats que han servit de context i documentació per a l'elaboració d'aquest treball de recerca.

Estratègies, declaracions i recomanacions

Estratègia d'intel·ligència artificial de Catalunya (<https://participa.gencat.cat/processes/estrategialA>).

“Barcelona Declaration for the proper development and usage of artificial intelligence in Europe” 2017 (<https://www.iiia.csic.es/barcelonadeclaration/>).

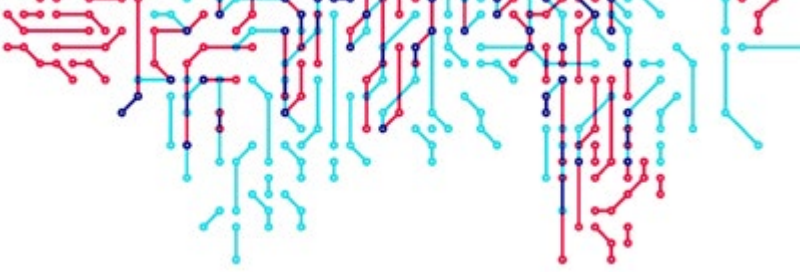
“Declaració sobre ètica i protecció de dades en intel·ligència artificial” (https://apdcat.gencat.cat/web/.content/04-actualitat/noticies/documents/ICDPPC-40th_AI-Declaration_ADOPTED.pdf).

Directrius ètiques per a una IA fiable (https://ec.europa.eu/spain/barcelona/news/press_releases/190408_ca).

Estrategia española de I+D+I en inteligencia artificial (http://www.ciencia.gob.es/stfls/MICINN/Ciencia/Ficheros/Estrategia_Inteligencia_Artificial_IDI.pdf)

Recomanacions de la OCDE sobre intel·ligència artificial (<https://www.oecd.org/centro-demexico/medios/cuarentaydospaísesadoptanlosprincipiosdelaocdesobreinteligencia-artificial.htm>).

Resolució del Parlament Europeu “Una política industrial global europea en matèria de IA i robòtica”. Febrer 2019 (https://www.europarl.europa.eu/doceo/document/A-8-2019-0019_ES.html).



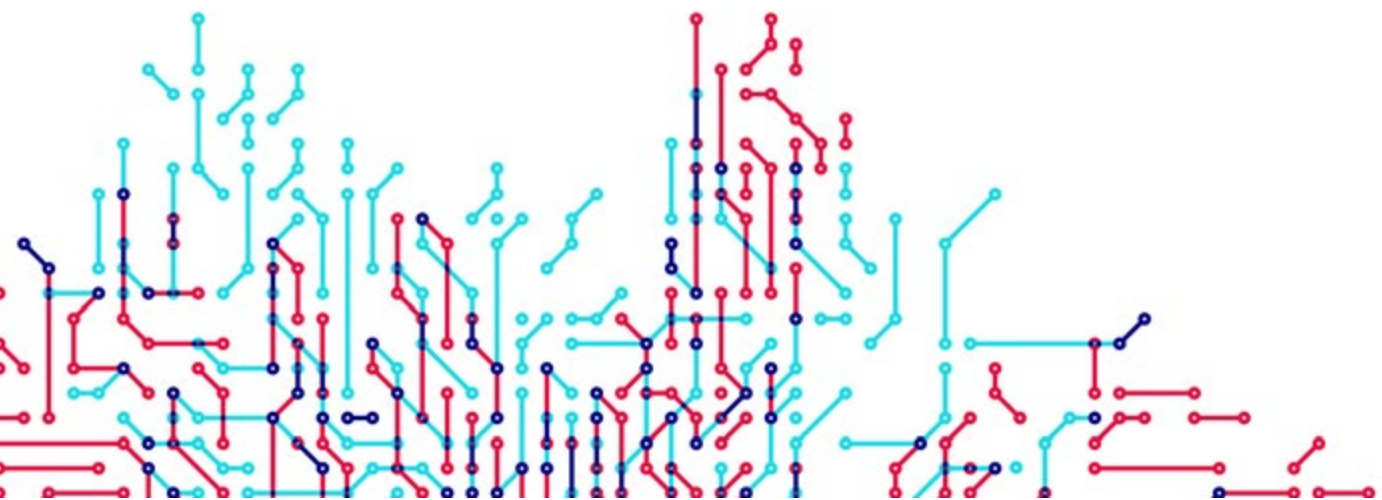
EU artificial intelligence ethics checklist ready for testing as new policy recommendations are published (<https://ec.europa.eu/digital-single-market/en/news/eu-artificial-intelligence-ethics-checklist-ready-testing-new-policy-recommendations-are>).

Regulating big data. The guidelines of the Council of Europe in the context of the European data protection framework (<http://isiarticles.com/bundles/Article/pre/pdf/106203.pdf>).

Artificial Intelligence and Data Protection: Challenges and Possible Remedies (<https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6>).

Principis establerts pel High Level Expert Group sobre Intel·ligència Artificial de la Comissió Europea (<https://ec.europa.eu/digital-single-market/en/high-level-expert-group-artificial-intelligence>).

Alessandro Mantelero. *Artificial Intelligence and Data Protection: Challenges and Possible Remedies* (<https://rm.coe.int/artificial-intelligence-and-data-protection-challenges-and-possible-re/168091f8a6>).



Informes

Algorithm Watch. *Automating society. Taking stock of Automated Decision-Making in the EU* (<https://algorithmwatch.org/en/automating-society/>).

Artificial Intelligence in Education. Challenges and Opportunities for Sustainable Development. UNESCO (<https://unesdoc.unesco.org/ark:/48223/pf0000366994>).

World Health Organisation releases first guideline on digital health interventions. WHO (<https://www.who.int/news-room/detail/17-04-2019-who-releases-first-guideline-on-digital-health-interventions>).

La intel·ligència artificial a Catalunya. Acció. Generalitat de Catalunya (<https://www.accio.gencat.cat/ca/serveis/banc-coneixement/cercador/BancConeixement/la-intel·ligencia-artificial-a-catalunya>).

La ciberseguretat a Catalunya Acció. Generalitat de Catalunya. 2018 (https://www.accio.gencat.cat/web/.content/bancconeixement/documents/informes_sectorials/ciberseguretat-informe-tecnologic.pdf).

Ethically Aligned Design. A Vision for Prioritizing Human Wellbeing with Artificial Intelligence and Autonomous Systems. IEEE. Advancing Technology for Humanity. 2016 (http://standards.ieee.org/develop/indconn/ec/ead_v1.pdf).

Partnership on AI to benefit people and society. Website (<https://www.partnershiponai.org/#>).

European Civil Law Rules in Robotics. European Parliament. 2016 ([http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU\(2016\)571379_EN.pdf](http://www.europarl.europa.eu/RegData/etudes/STUD/2016/571379/IPOL_STU(2016)571379_EN.pdf)).

La Internet de les coses a Catalunya. Acció. Generalitat de Catalunya (<https://www.accio.gencat.cat/web/.content/bancconeixement/documents/pindoles/iot-cat.pdf>).

AI Pascual. *2015 Data Breach Fraud Impact Report* (<https://www.javelinstrategy.com/coverage-area/2015-data-breach-fraud-impact-report>).

Roy Wedge, James Max Kanter, Kalyan Veeramachaneni, Santiago Moral Rubio, Sergio Iglesias Perez. *Solving the false positives problem in fraud prediction using automated feature engineering* (<http://www.ecmlpkdd2018.org/wp-content/uploads/2018/09/567.pdf>).

Dillon Reisman, Jason Schultz, Kate Crawford, Meredith Whittaker. *Algorithmic Impact Assessments: A Practical Framework For Public Agency Accountability*. AI Now Institute (<https://ainowinstitute.org/aiareport2018.pdf>).

AI in Context. Data&Society (<https://datasociety.net/output/ai-in-context/>).

The Cybersecurity Campaign Playbook. 2018 (<https://www.ndi.org/publications/cybersecurity-campaign-playbook-global-edition>).

Itziar de Lecuona. *Evaluación de los aspectos metodológicos, éticos, legales y sociales de proyectos de investigación en salud con datos masivos (big data)* (http://scielo.isciii.es/scielo.php?script=sci_abstract&pid=S0213-91112018000600576).

El vehicle connectat a Catalunya. Acció. Generalitat de Catalunya. 2019 (<https://www.accio.gencat.cat/ca/serveis/banc-coneixement/cercador/BancConeixement/vehicle-connectat-a-catalunya>).

Liliana Arroyo. *Trustful and trustworthy: Manufacturing trust in the digital era* (<https://www.esade.edu/researchyearbook/node/231>).

Eli Pariser. *The Filter Bubble: What The Internet Is Hiding From You* (https://hci.stanford.edu/courses/cs047n/readings/The_Filter_Bubble.pdf).

116

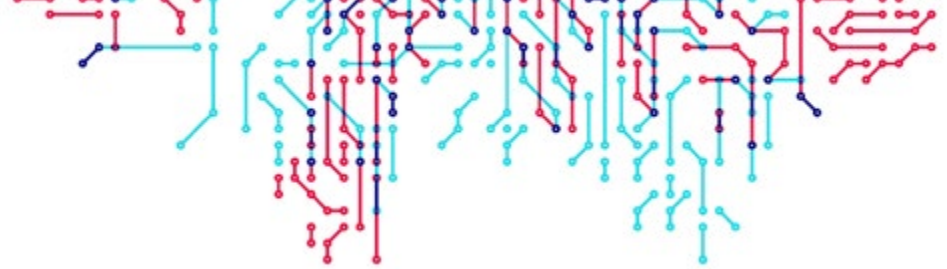
Articles

Itziar de Lecuona. “La oportunidad de la investigación con datos masivos en salud” (<http://www.gacetasanitaria.org/es-evaluacion-los-aspectos-metodologicos-eticos-articulo-S0213911118300864>).

Ramón López de Mántaras. “La ética en la inteligencia artificial” (<https://www.investigacionyciencia.es/revistas/investigacion-y-ciencia/el-multiverso-cuntico-711/tica-en-la-inteligencia-artificial-15492>).

Ricardo Baeza-Yates, Karma Peiró. “És possible acabar amb els biaixos dels algorismes? (1a i 2a part)”. 2019 (<https://www.karmapeiro.com/2019/06/17/es-possible-acabar-amb-els-biaixos-dels-algoritmes-1a-part/>).

Tomer Hochma. “The Ultimate List of Cognitive Biases: Why Humans Make Irrational Decisions” (<https://humanhow.com/en/list-of-cognitive-biases-with-examples/>).



“Machine Bias”. *ProPublica* (<https://www.propublica.org/article/machine-bias-risk-assessments-in-criminal-sentencing>).

Sandra Wachter. “How to make algorithms fair when you don’t know what they’re doing”. *Wired*. (<https://www.wired.co.uk/article/ai-bias-black-box-sandra-wachter>).

Jayshree Pandya. “Hacking Our Identity: The Emerging Threats From Biometric Technology” (<https://www.forbes.com/sites/cognitiveworld/2019/03/09/hacking-our-identity-the-emerging-threats-from-biometric-technology/#353ed3505682>).

Loren Grush. “Google apologizes after Photos app tags two black people as a gorillas” (<https://www.theverge.com/2015/7/1/8880363/google-apologizes-photos-app-tags-two-black-people-gorillas>).

Kyle Wiggers. “MIT researchers: Amazon’s Rekognition shows gender and ethnic bias (updated)” (<https://venturebeat.com/2019/01/24/amazon-rekognition-bias-mit/>).

Karma Peiró, Ricardo Baeza-Yates. “Algoritmo, yo también existo” (<https://www.karma-peiro.com/2019/09/24/algoritmo-yo-tambien-existo/>).

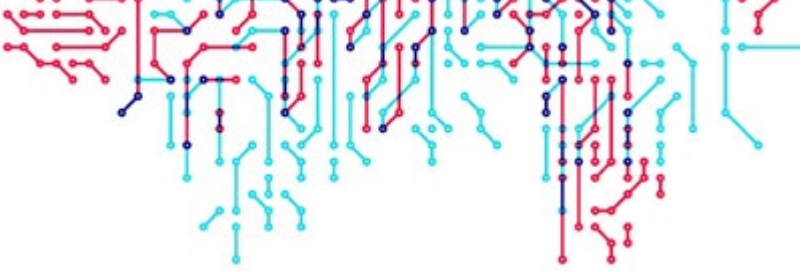
Meike Zehlike, Francesco Bonchi, Carlos Castillo, Sara Hajian, Mohamed Megahed, Ricardo Baeza-Yates. “FA*IR: A Fair Top-k Ranking Algorithm” (<https://arxiv.org/pdf/1706.06368.pdf>).

Carles Ramió. “Estatuto ético para la implantación de la inteligencia artificial y la robótica en la administración pública” (<https://www.administracionpublica.com/estatuto-etico-para-la-implantacion-de-la-inteligencia-artificial-y-la-robotica-en-la-administracion-publica/>).

La intel·ligència artificial i l’administració pública. Los Libros de la Catarata (https://www.todostuslibros.com/libros/inteligencia-artificial-y-administracion-publica_978-84-9097-590-9).

Karma Peiró. “L’escàndol Facebook-Cambridge Analytica: un cas per revisar la protecció de dades i molt més” (<https://www.naciodigital.cat/noticia/151744/escandol/facebook-cambridge/analytica/cas/revisar/proteccio/dades/molt/mes>).

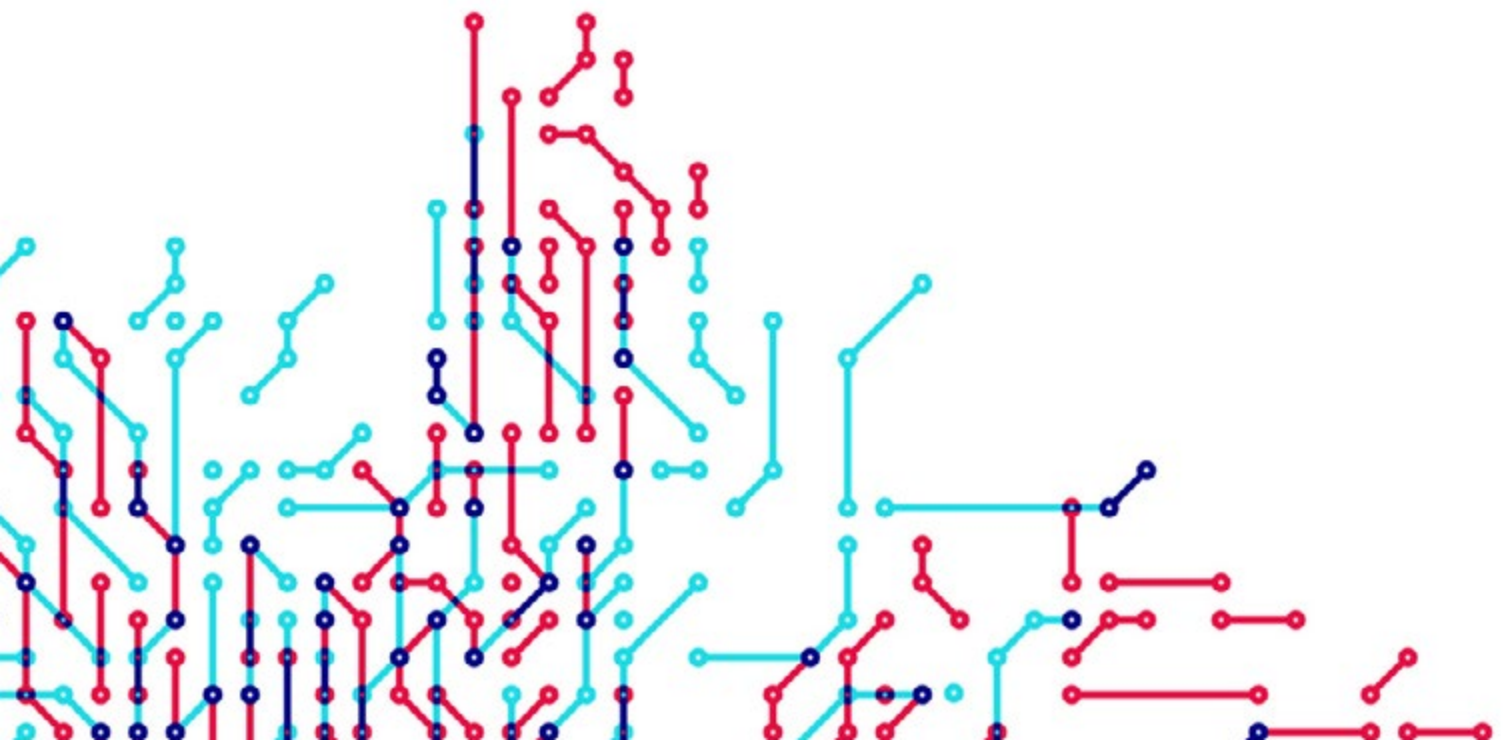
Agustí Cerrillo. “Com obrir les caixes negres de les administracions públiques? Transparència i rendició de comptes en l’ús dels algoritmes” (<http://revistes.eapc.gencat.cat/index.php/rcdp/article/view/10.2436-rcdp.i58.2019.3277>).



Vídeos

Lliçó inaugural de Núria Oliver del curs 2019-2020 del sistema universitari català (<https://www.youtube.com/watch?v=DwC0KDwliXc>).

Por la razón y la ciencia. Documental de la televisión chilena. Intervención de Ricardo Baeza-Yates (https://www.youtube.com/watch?time_continue=1253&v=7PCC7tRyM2I).



2.8. Agraïments

Per acabar, correspon afegir un agraïment sincer als experts en intel·ligència artificial i ciència de dades que han estat consultats i que ens han permès entendre el moment actual de l'aplicació dels algorismes de decisió automatitzada a Catalunya, pel que fa tant als beneficis com als riscos. L'agraïment també als responsables de departaments de l'Administració pública i als responsables d'empreses privades que han fet fet possible explicar més de cinquanta exemples.

La intenció era aprofundir en aquesta tecnologia, defugint l'alarmisme habitual de titulars sense context. Les entrevistes personalitzades i el temps concedit per totes les persones que mencionem a continuació és impagable.

Ricardo Baeza-Yates, cap de Tecnologia de NCENT. Catedràtic d'Informàtica de la Universitat Pompeu Fabra i de la Northeastern University.

Anton Bardera, professor del Departament d'Informàtica, Matemàtica Aplicada i Estadística de la Universitat de Girona (UdG).

Meritxell Bassolas, directora de Coneixement i Transferència de Tecnologia al Centre de Visió per Computació de la Universitat Autònoma de Barcelona (UAB).

Marga Bonmatí, directora del Consorci AOC.

Marco Bressan, ex-científic de dades en cap del BBVA. Director de Satellogic.

Victòria Camps, catedràtica d'Ètica i Filosofia del Dret Moral i Polític de la Universitat Autònoma de Barcelona (UAB).

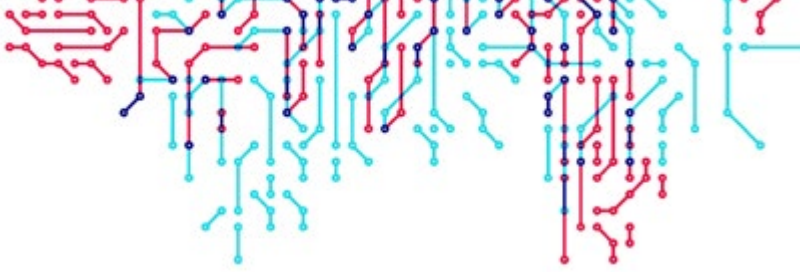
Carlos Castillo, professor investigador distingit del Departament de Tecnologies de la Informació i les Comunicacions de la Universitat Pompeu Fabra (UPF). Lidera el grup d'investigació de Ciència de la Web i Computació Social.

Ulises Cortés, director científic del grup d'Intel·ligència Artificial d'Alt Rendiment en el Centre de Supercomputació de Barcelona. Catedràtic d'Intel·ligència Artificial de la Universitat Politècnica de Catalunya.

Fernando Cucchiatti, director d'Analítica de Dades i Visualització del Barcelona Supercomputing Center.

Josep Domingo, catedràtic distingit de Ciència de la Computació de la Universitat Rovira i Virgili (URV) de Tarragona. Investigador ICREA Acadèmia de la URV. Director de la Càtedra UNESCO de Privadesa de Dades. Director del CYBERCAT - Centre de Recerca en Ciberseguretat de Catalunya.

Ricard Gavalrà, professor i coordinador del laboratori de recerca de la Universitat Politècnica de Catalunya (UPC). CEO i director científic d'Amalfi Analytics.



Karina Gibert, investigadora del grup d'Enginyeria del Coneixement i Aprenentatge Automàtic del Centre d'Investigació en Ciència Intel·ligent de Dades i Intel·ligència Artificial de la Universitat Politècnica de Catalunya (UPC). Vicedegana de Big Data, Ciència de Dades i Intel·ligència Artificial del Col·legi Oficial d'Enginyeria Informàtica de Catalunya.

Elisabeth Golobardes, doctora en enginyeria informàtica i professora catedràtica d'Intel·ligència Artificial i Ciència de les Dades de La Salle - Universitat Ramon Llull. Membre del grup de recerca en Data Science for the Digital Society (DS4DS).

Àlex Hinojo, ex-director general de l'associació Amical Wikimedia. Cofundador del projecte Drets Digitals.

Alberto Labarga, científic de dades d'IOMED.

Itziar de Lecuona, professora del Departament de Medicina de la Universitat de Barcelona (UB). Subdirectora de l'Observatori de Bioètica i Dret de la UB.

David Llorente, CEO de Narrativa.

Jorge López, responsable de l'Àrea d'Investigació i Desenvolupament d'UBinding.

Ramon López de Mántaras, professor investigador del Consell Superior d'Investigacions Científiques (CSIC). Exdirector de l'Institut d'Investigació d'Intel·ligència Artificial (IIIA).

Gabriel Maeztu, doctor i científic de dades. CEO d'IOMED.

Alessandro Mantelero, relator d'Intel·ligència Artificial i Privacitat del Consell d'Europa. Professor de dret de la Universitat de Torí.

Jordi Mas, membre de Softcatalà i pioner de la Internet catalana.

Manel Medina, catedràtic de la Universitat Politècnica de Catalunya (UPC). Fundador i director de l'esCERT-UPC.

Jordi Navarro, CEO & co-founder de Cleverdata.io.

Adina Nedelea, responsable de l'Equip de Matemàtics d'UBinding.

Ivan Ostrowicz, enginyer i expert en intel·ligència artificial aplicada al sector educatiu.

Petia Radeva, professora del Departament de Matemàtiques i Informàtica de la Universitat de Barcelona (UB). Directora del grup consolidat de recerca Machine Learning and Computer Vision de la UB.

Carles Ramió, catedràtic del Departament de Ciències Polítiques i Socials de la Universitat Pompeu Fabra (UPF).

Teresa Roig Sitjar, assessora tecnològica i científica.

Pier Paolo Rossi, director de l'Advanced Customer Marketing & Analytics del Banc de Sabadell.

Giusseppe Scionti, CEO i fundador de NovaMeat.

Carles Sierra, director de l'Institut d'Investigació en Intel·ligència Artificial (IIIA).

Antonio Andrés Pueyo, investigador principal del Grup d'Estudis Avançats en Violència (GEAV) de la Universitat de Bracelona (UB). Catedràtic de Psicologia de la UB.

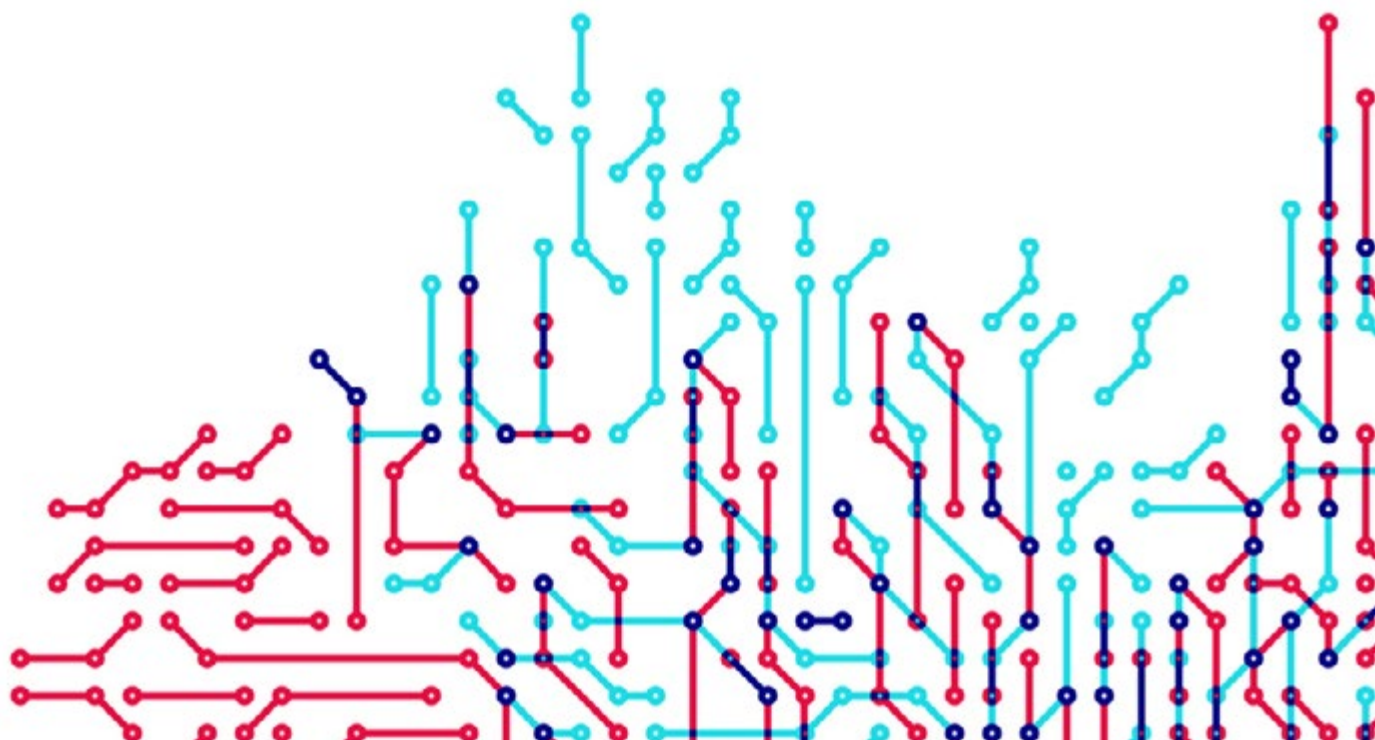
Carne Torras, doctora en informàtica i professora d'investigació a l'Institut de Robòtica i Informàtica Industrial (CSIC-UPC).

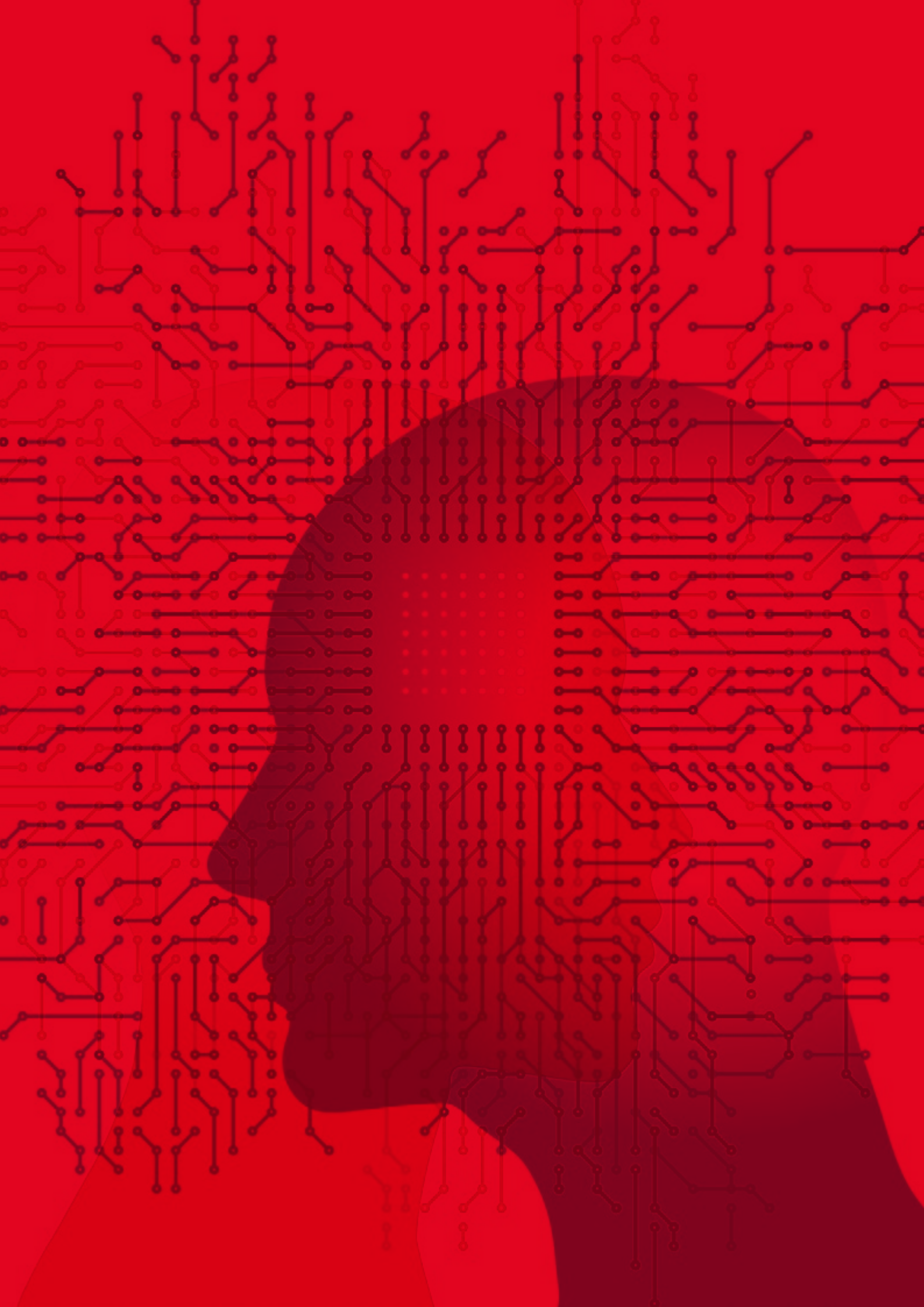
Jordi Torras, fundador i CEO d'Inbenta.

Lluís Torrens, director d'Innovació Social de l'Àrea de Drets Socials, Justícia Global, Feminismes i LGTBI de l'Ajuntament de Barcelona.

Jordi Vitrià, catedràtic de Llenguatges i Sistemes Informàtics de la Universitat de Barcelona (UB). Membre del Departament de Matemàtiques i Informàtica de la UB.

Responsables de la **Secretaria Tècnica del Servei d'Ocupació Català**.





3.

PROTECCIÓ DE DADES I LA IA

En la introducció, quan parlàvem dels riscos associats als algorismes de decisió automatitzada (ADA), hem destacat dos tipus de riscos: els relacionats amb l'aplicació concreta (per exemple, el risc que ens deneguin un crèdit) i els relacionats únicament amb el tractament de dades personals (per exemple, el risc que es filtrin les nostres dades). El treball de camp sobre aplicacions dels ADA a Catalunya s'ha centrat en els riscos associats a cadascuna de les aplicacions presentades. En aquesta secció, de l'informe, ens centrarem en els riscos associats a la protecció de dades.

El respecte a la protecció de dades és un punt imprescindible per fer un ús ètic dels algorismes de decisió automàtica. En el marc de la Unió Europea, això implica complir amb l'RGPD.

Quan parlem de riscos associats a la protecció de dades, ens referim al fet que no es compleixi alguna de les disposicions de l'RGPD. L'ús intensiu de dades personals que comporten moltes aplicacions dels ADA fa que hi hagi friccions entre els ADA i l'RGPD. En aquest escenari, l'objectiu de l'APDCAT i de les altres autoritats supervisoras és que es compleixin les disposicions de l'RGPD.

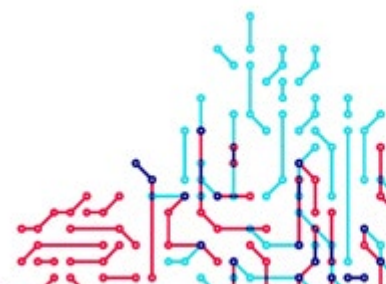
123

Recollida i tractament de dades de forma il·legítima

L'RGPD demana que tot tractament de dades tingui una base legal que el legitimi. N'hi ha diverses i totes són igualment vàlides: consentiment, obligació legal, interès legítim, etc. Independentment de la base legal usada, cal que l'interessat estigui informat del tractament de les seves dades.

Actualment, les dades personals tenen molt de valor. S'ha arribat a dir que són el nou petroli. En aquest context, seria ingenu pensar que totes les organitzacions que recullen dades personals ho fan de manera apropiada. Els alts nivells de personalització que permeten els ADA són un incentiu per recollir la quantitat més gran possible de dades personals, ja que són la base per al seu funcionament.

D'exemples de recollida i tractament de dades de manera il·legítima, n'hi ha molts. Posant el focus sobre els dispositius mòbils (que, pel fet tenir-lo al costat gran part del dia tenen un gran potencial per generar informació personal), trobem dos treballs molt



interessants que fan referència a l'ús il·legítim de la nostres dades. El primer¹²⁸ fa una anàlisi de dues tècniques que utilitzen algunes aplicacions mòbils que trobem al mercat d'aplicacions per evitar els controls que estableix Android per accedir a certes dades. El segon¹²⁹ explica com peces de software adjuntades durant el procés de manufactura del mòbil aprofiten els seus privilegis per llegir i distribuir informació sense el coneixement dels usuaris.

Dret a no ser objecte de decisions automatitzades

Els ADA prenen decisions basades en els perfils de les persones (les característiques d'una persona que els dissenyadors de l'algorisme han considerat rellevants). El fet de limitar la presa de decisions a un conjunt de característiques prèviament definides és una limitació important respecte a la capacitat que té un jutge humà d'avaluar altres aspectes que, en principi, poden semblar poc rellevants. Per exemple, parlant amb un interlocutor humà una persona pot explicar el perquè de la importància o la irrellevància d'alguna de les característiques del seu perfil o, fins i tot, remarcar coses que el perfil no reflecteix.

124 L'RGPD reconeix aquesta problemàtica i estableix el dret a no ser objecte de decisions automatitzades si aquestes decisions poden tenir un efecte significatiu sobre la seva persona. En particular, l'article 22 diu:

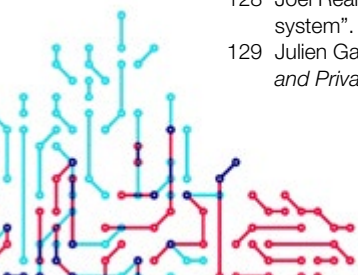
“Tota persona té el dret a no ser objecte d'una decisió basada únicament en el tractament automatitzat, inclosa l'elaboració de perfils, que tingui efectes jurídics sobre l'interessat o li afecti de significativament.”

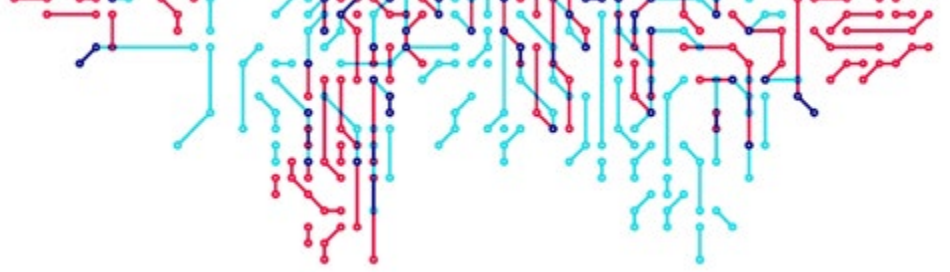
L'article 22 no exclou, en cap cas, l'ús de mitjans automàtics en prendre una decisió, però exigeix que, si la decisió té efectes significatius sobre les persones, hi hagi intervenció humana.

Tot i que a la redacció de l'article 22 parla del dret a no ser objecte d'una decisió automatitzada, les directrius publicades per a la seva interpretació afirmen que no és un dret que hagin d'exigir els interessats, sinó una prohibició. La diferència és substancial perquè si fos una mena de dret d'oposició, les organitzacions podrien fer ús de decisions automatitzades de forma generalitzada i ajustar el seu comportament únicament en els casos específics en què un interessat ho demani.

128 Joel Reardon, Álvaro Feal, et al. “50 ways to leak your data: an exploration of apps' circumvention of the Android permission system”. *28th USENIX Security Symposium*, 2019.

129 Julien Gamba, Mohammed Rashed, et al. “An analysis of pre-installed Android software”. *41st IEEE Symposium on Security and Privacy*, 2020.





El principi de transparència

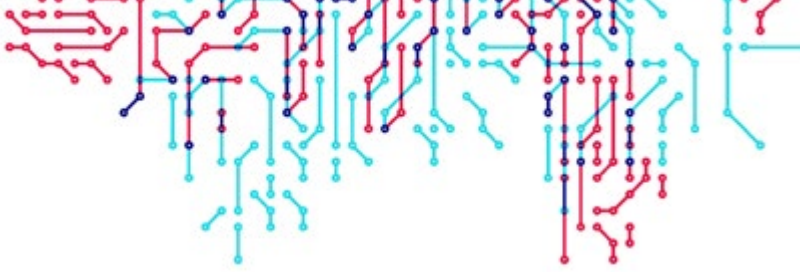
Perquè les persones puguin exercir els drets que tenen sobre les seves dades, cal que el tractament sigui transparent. Segons l'RGPD, la transparència exigeix que tota la informació que es dona a les persones (les dades que es tracten, l'objectiu del tractament, els drets de les persones, etc.) sigui concisa, fàcilment localitzable i faci ús d'un llenguatge entenedor.

Pel que fa a la transparència de les decisions preses pels algorismes, les disposicions de l'RGPD són limitades. És cert que en els considerants (la part no dispositiva de la norma que serveix per interpretar-la) es parla del dret a obtenir una explicació de la decisió, però la part dispositiva és limitada a demanar que els interessats rebin informació rellevant sobre la lògica que s'hi aplica. Mentre que donar una explicació d'una decisió requereix tractar amb el cas concret, donar informació rellevant sobre la lògica és pot fer d'una manera més general, amb relació a l'algorisme.

Possiblement, la decisió de no demanar una explicació per les decisions individuals es fonamenta en la incapacitat tècnica que hi ha per fer-ho en alguns casos rellevants. Cada algorisme d'IA té un nivell específic d'explicabilitat. En general, l'increment de la complexitat dels algorismes ha anat acompanyat d'una reducció en la capacitat d'explicar-ne els resultats. Per exemple, mentre que explicar una decisió resulta senzill en un sistema basat en regles, pot ser molt complicat en un sistema d'aprenentatge profund (on una estructura complexa de neurones artificials s'autoregula en la fase d'entrenament del model sense cap intervenció externa).

Donada la rellevància de l'aprenentatge automàtic i, en particular, de l'aprenentatge profund, podem dir que l'explicabilitat és un dels punts febles de la intel·ligència artificial. Però per què és important que el resultat d'un algorisme sigui explicable? A vegades, no ens interessa saber el perquè d'una decisió; n'hi ha prou amb saber que l'algorisme té un nivell de precisió prou bo. Per exemple, en un sistema de recomanació de pel·lícules, no sembla especialment important explicar el perquè d'una recomanació concreta. Ara bé, quan la decisió pot tenir efectes importants sobre les persones, tenir una explicació concreta és més necessari. Te'n refiaries d'un metge que no dona cap explicació del diagnòstic o d'un jutge que dicta sentències sense cap justificació? Llavors, si exigim una explicació als experts humans, per què no l'hauríem de demanar als ADA que fan funcions equivalents?

L'RGPD evita el problema anterior exigint intervenció humana en aquelles decisions que puguin afectar les persones d'una manera important; és a dir, en les decisions en què l'exigència d'una explicació és més necessària.



D'altra banda, la necessitat d'explicar algunes decisions està fomentant la investigació en intel·ligència artificial explicable (XAI). Hi ha dues aproximacions. La primera busca generar la explicació d'un comportament observat d'un ADA. En certa manera, aquesta aproximació segueix el procediment de la ciència experimental: donat un comportament observat, construïm un model que l'explica. El problema és la validesa d'aquest model. Com va dir l'estadístic britànic George Box el 1978, tots els models per explicar la realitat són incorrectes, però alguns són útils. Tenir un model útil per explicar el funcionament d'un ADA no vol dir que l'explicació que obtinguem per a un cas concret sigui la correcta. La segona aproximació evita el problema de la validesa de l'explicació exigint que l'algorisme de decisió sigui fàcilment interpretable. Això s'aconsegueix, principalment, reduint-ne la complexitat. El preu a pagar en aquest cas és la pèrdua de precisió de l'ADA.

De la mateixa manera que no totes les decisions són senzilles, tampoc no és necessari tenir una explicació per a totes les decisions. El primer pas és determinar si cal que les decisions d'un ADA siguin explicables i valorar quina és la millor manera d'obtenir les explicacions. Per exemple, en el cas d'un cotxe autònom, les explicacions de les decisions poden ser interessants a l'hora de determinar la responsabilitat en cas d'accident.

126

El principi de lleialtat

El principi de lleialtat exigeix fer un ús de les dades que sigui previsible per part de les persones (amb relació a la finalitat buscada) i que no hi hagi conseqüències adverses injustificades per a les persones .

El principal punt de fricció entre els ADA i el principi de lleialtat és la potencial discriminació que poden patir les persones com a conseqüència d'un algorisme que pren decisions esbiaixades. Cal que els ADA no facin un tracte desigual cap a un grup de persones en funció del gènere, del color de la pell, de les creences religioses, de l'orientació sexual, etc.

Podem pensar que els algorismes, a diferència de les persones, no tenen prejudicis i que, per tant, les decisions que prenen no són discriminatòries. Aquesta és una visió que no s'ajusta a la realitat: els ADA tenen biaixos. L'origen d'aquests biaixos és divers.

Tot i que no podem descartar que el biaix d'un ADA sigui intencionat, en general, no hi ha males intencions per part del programador, del dissenyador o de l'organització responsable. És més aviat una qüestió de desconeixement, de manca de diligència. La manca d'una visió ètica de la computació de les persones que hi intervenen n'és la causa principal.

Els biaixos en un algorisme poden tenir l'origen en l'ús de dades errònies o esbiaixades. Com ja s'ha comentat, el problema és que l'entrenament dels algorismes d'aprenentatge automàtic necessita dades i, lamentablement, moltes dades històriques presenten biaixos. Cal ser conscient d'aquest fet i cal aplicar tècniques per minimitzar-ne l'impacte.

Els biaixos també poden venir d'un algorisme mal dissenyat. És un fet comú dissenyar un algorisme pensant en el grup majoritari de persones. Encara que aquest algorisme sigui acurat amb les persones d'aquest grup, pot no ser-ho per a altres grups minoritaris. En aquest cas, estaríem tractant de forma diferent els grups minoritaris; és a dir, hi ha un risc de discriminació. Si el pes d'aquests grups en el conjunt de les dades és petit, pot tenir poc impacte global i passar desapercebut.

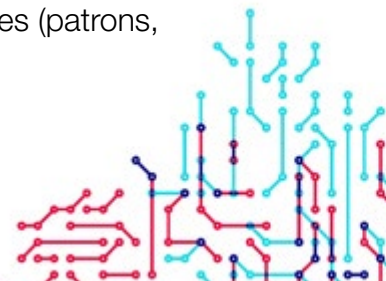
Una tècnica bàsica per dissenyar algorismes no esbiaixats és evitar l'ús de variables potencialment discriminatòries com ara el gènere, la raça, etc. Això no sempre és possible; hi ha algorismes que en poden fer un ús legítim. Per exemple, si s'ha demostrat que les persones d'una raça concreta són més propenses a patir un cert tipus de malaltia, seria irracional no aprofitar aquest coneixement. D'altra banda, eliminar aquestes variables no garanteix que no hi hagi discriminació. Per exemple, hi pot haver una correlació entre àrees concretes d'una ciutat i la nacionalitat de les persones. En aquest cas, el lloc de residència és una font potencial de discriminació.

Hi ha tècniques més sofisticades per detectar i mitigar el risc de discriminació. La recerca en el camp de l'antidiscriminació s'ha incrementat en els darrers anys. Ara bé, l'ús efectiu d'aquestes tècniques pot presentar alguns problemes. Per exemple, és necessari determinar els grups de persones en risc de patir discriminació; si no, no és pot avaluar si hi ha discriminació ni mitigar-la. El fet que, moltes vegades, aquests grups estiguin definits a partir de dades de categories especials (que són objecte de limitacions addicionals al tractament) dificulta l'aplicació d'aquestes tècniques.

Principi de limitació de la finalitat

El principi de limitació de la finalitat diu que les dades s'han de recollir per a una finalitat específica i explícita, i que no s'han d'utilitzar les dades de manera incompatible. Aquest principi és essencial perquè les persones puguin fer un control efectiu de l'ús que se'n fa de les seves dades.

Alguns dels algorismes d'intel·ligència artificial s'entrenen amb una finalitat específica; per exemple, decidir si s'ha de concedir un crèdit, detectar una malaltia, etc. En aquests casos, l'objectiu de l'aprenentatge és clar. Ara bé, altres algorismes entren dins la categoria de l'"aprenentatge no supervisat", i extreuen informació de les dades (patrons,



correlacions, etc.). Com que aquesta informació no es coneix per avançat, és difícil satisfer el principi de limitació de finalitat.

Tot i la facilitat amb que es recullen dades actualment, la gran avidesa de dades de la IA fa que els seus promotors vegin amb bons ulls la possibilitat de recórrer a conjunts de dades ja existents. Aquest recurs pot reduir els costos i, fins i tot, permetre l'accés a certs tipus de dades que poden ser difícils de recollir. La reutilització de les dades recollides per part d'un altre actor amb una finalitat diferent pot xocar amb la limitació de finalitat que estableix l'RGPD.

Els sectors que propugnen una flexibilització en la possibilitat de reutilitzar dades argumenten que una reducció en les dades disponibles pot donar lloc a una intel·ligència artificial menys precisa i amb més biaixos.

El principi de limitació de la finalitat té algunes excepcions. En particular, l'article 5 de l'RGPD diu que el tractament amb finalitat d'investigació científica o històrica, o amb finalitat estadística és compatible amb el propòsit inicial de recollida de dades. Per tant, el recurs a la investigació científica és una possible via per a la reutilització de dades en IA. La pregunta és: què és investigació científica? L'antecedent 159 de l'RGPD diu que el tractament amb finalitat d'investigació científica s'ha d'interpretar d'una manera àmplia, i això inclou, per exemple, el desenvolupament i la demostració de noves tecnologies, la investigació fonamental, la investigació aplicada i la investigació amb finançament privat.

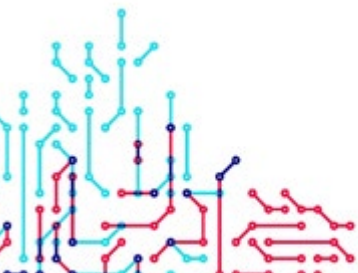
128

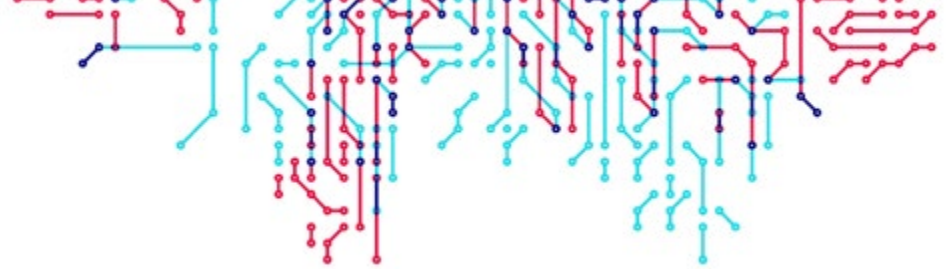
Principi de minimització

El principi de minimització diu que les dades utilitzades en un tractament han de ser adequades, pertinents i limitades a l'estrictament necessari per assolir la finalitat del tractament.

Com en el cas del principi de limitació de la finalitat, les dificultats que introdueix aquest principi en els ADA rau en el fet que aquests algorismes necessiten una gran quantitat de dades per aprendre i prendre decisions intel·ligents. Si volem que un nen reconegui un cotxe, li expliquem característiques típiques i li ensenyem alguna foto. Per fer que un algorisme d'aprenentatge automàtic reconegui els cotxes amb precisió, l'hem d'entrenar amb molts exemples.

Seguint el principi de minimització, s'han de tractar únicament les dades que són estrictament necessàries. Ara bé, cal tenir en compte l'efecte que una reducció de dades pot tenir sobre la precisió del sistema i sobre l'aparició de biaixos.





Hi ha diverses tècniques que permeten reduir l'ús de dades personals:

- Selecció dels atributs. La precisió d'algunes de les tècniques d'aprenentatge automàtic té una gran dependència dels atributs que es consideren. En aquests casos, afegir-hi atributs que no són pertinents pot ser contraproduent. Altres tècniques, com l'aprenentatge profund, no són tan sensibles als atributs considerats. En qualsevol cas, la selecció d'atributs rellevants dona lloc a models més senzills i més fàcils d'entrenar.
- Aprenentatge federat. Suposem que volem entrenar un model d'intel·ligència artificial sobre les dades d'un conjunt de persones. En comptes de lliurar les dades a una entitat que fa l'entrenament, l'aprenentatge federat proposa que l'entrenament es faci de forma distribuïda: cadascú fa l'entrenament amb les seves dades. Això s'aconsegueix donant el model actual a cada persona, que l'actualitzarà amb les seves dades i retornarà l'actualització feta. Un cop recollides les actualitzacions de tots els usuaris, aquestes s'agregaran per generar el model entrenat.
- Anonimització i ús de dades sintètiques. L'objectiu és evitar l'ús de dades personals en l'entrenament, tant si són dades anonimitzades (dades en què s'ha trencat l'enllaç amb la persona que les ha originat) com si són dades sintètiques (dades inventades que reproduïxen les característiques de les dades originals).

129

Tot i que tenir dades abundants beneficia els ADA, és igualment important que les dades siguin acurades: pot ser millor tenir menys dades, si són més acurades. Això va en la línia del principi d'exactitud.

Com hem comentat anteriorment, no sempre és possible definir amb claredat la finalitat. El principi de minimització està formulat en termes de la finalitat que es busca. Per tant, si no podem definir la finalitat amb claredat, tampoc no podem avaluar si se satisfà el principi de minimització.

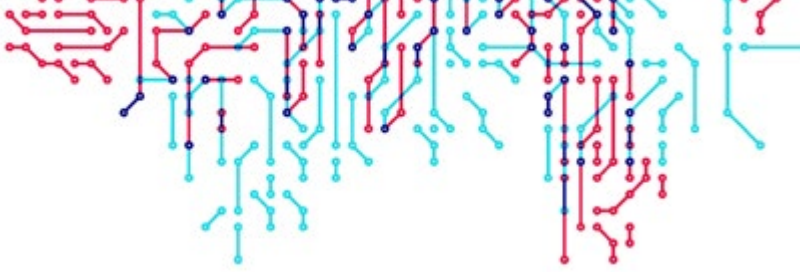
Anàlisi d'aplicacions dels ADA

Per concloure aquesta secció, revisem, des del punt de vista de la protecció de dades, algunes aplicacions d'ADA presentades en l'estudi de camp. Per a aquesta anàlisi s'han escollit aplicacions que tenen un impacte important sobre les persones.

Avaluació del risc de violència

En el capítol d'aplicacions de la intel·ligència artificial en el sistema judicial hi ha diverses eines que busquen estimar el risc de reincidència.

- RisCanvi: estima el risc que una persona torni a delinquir un cop ha sortit de la presó.



- SAVRY: estima el risc de reincidència juvenil.
- VioGen: estima el risc de reincidència en el cas de violència contra les dones.

L'avaluació del risc de reincidència és una tasca necessària per determinar les mesures preventives o de protecció que cal implantar. Abans de la introducció d'aquests sistemes, el risc de reincidència l'avaluaven de manera no estructurada els professionals amb coneixements específics com ara psicòlegs, psiquiatres o treballadors socials. El problema d'aquestes avaluacions és que depenen en molt bona part de la persona que les fa: dels factors que considera rellevants, dels biaixos que té, del seu estat d'ànim, etc. Els estudis mostren que aquestes estimacions de risc tenen una precisió molt baixa.

Les eines RisCanvi, SAVRY i VioGen busquen reduir la dependència de la persona que fa l'avaluació. Basant-se en una anàlisi de les dades de reincidència, aquestes eines determinen les característiques que incrementen el risc i la interrelació entre elles. Un cop determinades, s'utilitzen en totes les estimacions de risc. El fet que aquestes característiques siguin públiques dona un bon nivell de transparència al sistema. La transparència afavoreix que no s'utilitzin característiques que poden ser una font clara de discriminació de minories (per exemple, el color de la pell o el país d'origen).

130

L'estimació que donen aquests sistemes no és definitiva. El personal la pot modificar si consideren que hi ha raons de pes (situacions específiques rellevants d'un individu que no es tenen en compte en els paràmetres del sistema). Ara bé, qualsevol canvi en la valoració feta pel sistema, cal justificar-la.

En aquests casos, la decisió presa pel sistema té un impacte important sobre les persones i, per tant, segons l'RGPD, tenen el dret a obtenir intervenció humana, a expressar el seu punt de vista i a impugnar la decisió.

Sistemes de scoring

Abans de concedir un crèdit, els bancs estudien el risc que té. El mateix passa quan volem fer una assegurança, per exemple, per al cotxe. Actualment, no és el personal del banc o de l'asseguradora qui determina el risc, sinó un algorisme.

Els bancs tenen un perfil molt detallat dels seus clients: saben on treballen, el seu sou, en què gasta els diners, etc. Els sistemes de *scoring* tenen la capacitat d'utilitzar tota aquesta informació a l'hora de determinar el risc associat a un crèdit. Per exemple, les transaccions fetes amb targeta tenen un codi que identifica el tipus d'activitat: farmàcies, apostes, serveis de cites, etc. Algunes d'aquestes categories poden indicar un increment en el perfil de risc de la persona. Hi ha algorismes que utilitzen dades demogràfiques, dades de xarxes socials, etc.

En acceptar les clàusules de protecció de dades, els clients estan consentint una gran quantitat d'usos, entre els quals l'ús de les nostres dades amb finalitats comercials (oferir productes i serveis). Lamentablement, hi ha molt poca gent que llegeixi aquestes clàusules.

Encara que tinguin una base legal, el fet que l'algorisme sigui una caixa negra és un problema a l'hora de determinar si són sistemes justos. Alguns estudis suggereixen que hi ha canvis aparentment irrelevants (com ara modificar la resolució del mòbil) que afecten el resultat. Altres estudis suggereixen que es podrien estar perpetuant discriminacions històriques pel que fa a l'accés al crèdit.

Donada la rellevància de la decisió, segons l'RGPD, les persones tenen dret a obtenir intervenció humana, a expressar el seu punt de vista i a impugnar la decisió.

Detecció de frau en les targetes de crèdit

La detecció de frau és, indubtablement, una aplicació que busca el benefici tant del banc com dels seus clients. Aquests sistemes analitzen les característiques de les compres per determinar si són fraudulentament. El problema d'aquests sistemes és que encara tenen un percentatge de falsos positius molt elevat.

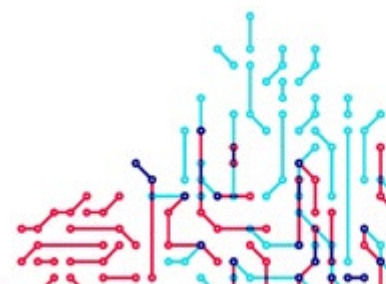
131

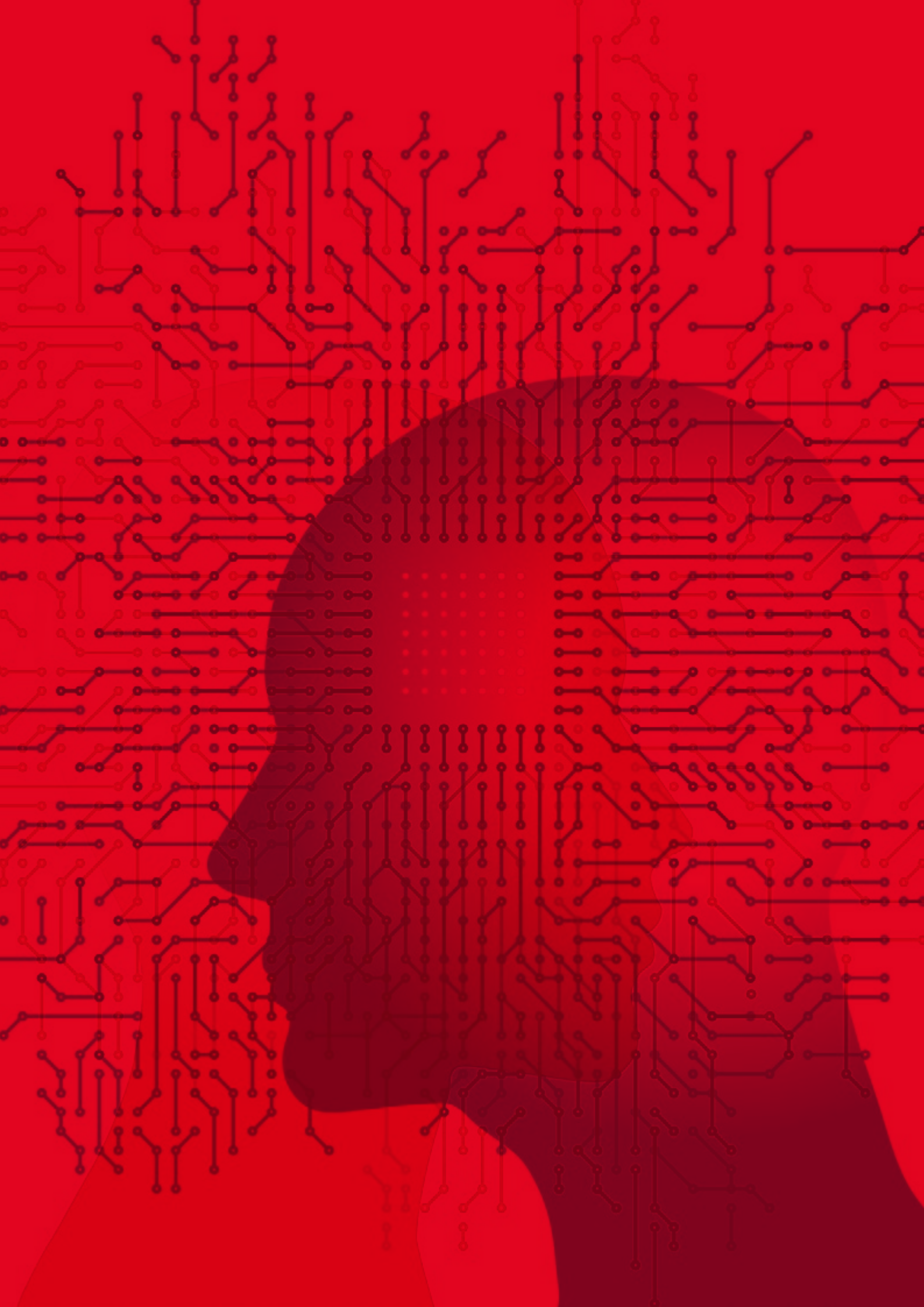
Tot i que les conseqüències d'un fals positiu, en general, no haurien de ser greus. Convé que s'estableixin mecanismes per impugnar la decisió.

Detecció de malalties

La detecció de malalties és una de les grans aplicacions de la intel·ligència artificial en la medicina. En el capítol d'aplicacions en salut se n'han presentat tres casos: la detecció del càncer de colon, la detecció del dolor exagerat i la detecció de la cirrosi.

En funció de l'aplicació concreta, l'ús de la intel·ligència artificial pot tenir diversos avantatges: millora de la precisió en la detecció, detecció primerenca, ús de tècniques menys invasives, etc. Ara bé, la importància que poden tenir aquestes aplicacions amb relació a la vida de les persones fa que sigui imprescindible la intervenció d'un metge, que té l'última paraula. Afortunadament, en el camp de la medicina, aquests sistemes es plantegen com a sistemes d'ajuda a la decisió.





4.

RECOMANACIONS FINALS

La capacitat de prendre decisions de forma autònoma s'ha mirat amb cautela des de l'inici de la IA. De fet, el primer codi ètic el va proposar l'autor de ciència ficció Isaac Asimov el 1942 a *Runaround*, un relat curt. Això va ser catorze anys abans que John McCarthy encunyés el terme *intel·ligència artificial* el 1956.

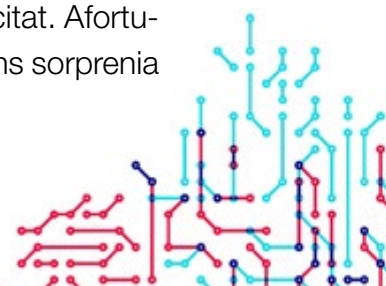
La gran penetració que han assolit els ADA i la IA en els darrers anys, junt amb l'impacte que pot tenir sobre les persones, ha donat lloc al desenvolupament de múltiples codis ètics, tant per part d'empreses com d'altres organitzacions. Per exemple, l'OCDE ha proposat un codi ètic basat en els principis següents:

- La IA ha de beneficiar les persones i el planeta a través de creixement inclusiu, desenvolupament sostenible i benestar.
- Els sistemes IA han de dissenyar-se per respectar la llei, els drets humans, els valors democràtics i la diversitat, i han de incloure el mecanismes necessaris per garantir una societat justa.
- Hi ha d'haver transparència per assegurar-nos que la gent entén els resultats de la IA i que s'hi poden oposar.
- Els sistemes IA han de funcionar de manera robusta i segura durant tot el cicle de vida, i els riscos potencials s'han d'avaluar i gestionar de manera continua.
- Les organitzacions i les persones que desenvolupen, despleguen o operen els sistemes d'IA són responsables del seu bon funcionament, d'acord amb els principis anteriors.

El nostre objectiu en aquesta secció és desenvolupar la part de protecció de dades personals que s'ha inclòs en molts dels codis ètics mencionats. Això ho fem donant una sèrie de recomanacions que han d'aplicar els diferents actors que intervenen en un sistema ADA/IA (persones, organitzacions i autoritats supervidores).

Recomanacions per a les persones

L'RGPD és una legislació de protecció de dades avançada que busca tornar a les persones el control sobre l'ús de les seves dades. Ara bé, per tal que l'RGPD sigui efectiu cal un canvi de mentalitat en les persones; cal crear una cultura de la privacitat. Afortunadament, el desconeixement que hi havia fa deu anys ja no hi és: abans ens sorprenia



obtenir serveis gratuïts a Internet, ara sabem que no ho eren (volien les nostres dades). Desafortunadament, encara som massa permissius amb l'ús que es fa de les nostres dades: la majoria de persones accepta les condicions d'ús sense llegir-les.

En aquest escenari, les recomanacions que donem a les persones respecte a l'ús de les seves dades en aplicacions d'intel·ligència artificial són força generals:

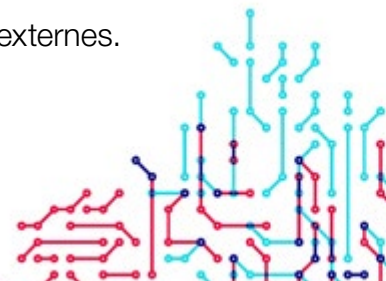
- Prendre consciència que les dades personals que una organització tracta no són propietat de l'organització, sinó una informació que hem cedit les persones.
- Cal conèixer els drets que tenim respecte a les nostres dades.
- En cas que no es respectin els drets que tenim sobre les nostres dades, cal conèixer els mecanismes per fer-los efectius.
- Cal una mentalitat crítica a l'hora de consentir el tractament les nostres dades: entendre la finalitat d'un tractament, conèixer quines dades són necessàries, entendre les conseqüències potencials del tractament i decidir sobre el consentiment amb coherència. Per exemple, un client d'un gimnàs pot considerar que haver de donar l'empremta dactilar per accedir a les instal·lacions és desproporcionat.
- Prendre consciència que els algorismes ADA/IA poden ser especialment intrusius. Un cop es detecta que les nostres dades poden ser objecte de tractament per part d'un algorisme ADA/IA, convé analitzar quines són les conseqüències potencials que se'n poden derivar.

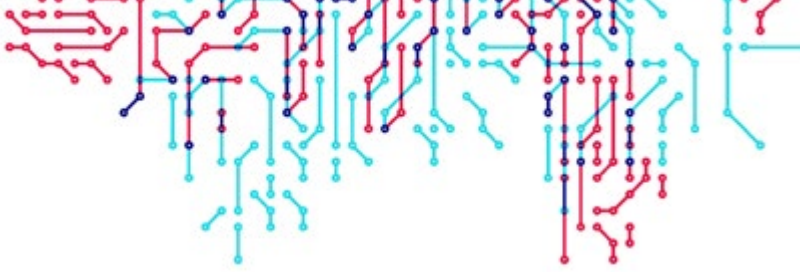
Recomanacions per a les organitzacions que fan ús de la IA

La intel·ligència artificial i els ADA suposen un avantatge competitiu que les organitzacions no poden deixar passar. Ara bé, fer-ne un mal ús pot tenir conseqüències importants. Per exemple, Cambridge Analytica era una empresa de consultoria política que va fer un ús il·lícit de dades personals per fer publicitat electoral personalitzada, ressaltant a cada persona aquella part del programa que encaixava més amb les seves preferències. De resultes d'aquest escàndol, Cambridge Analytica ha deixat d'existir i Facebook (la font de les dades) ha hagut de pagar una multa de cinc mil milions de dòlars.

L'ús dels algorismes de decisió automatitzada s'ha de fer des d'una **perspectiva ètica** i, en particular, **respectant l'RGPD** en els seus principis i en els drets que reconeix a les persones. Tot i que, el disseny i el desenvolupament d'algorismes de decisió automatitzada és primordialment una tasca tècnica, cal exigir uns mínims coneixements d'ètica i de protecció de dades a les persones que se n'encarreguen. És necessari el seguiment dels codis ètics existents o el desenvolupament d'altres de propis.

- Cal avaluar l'impacte sobre les persones. Els algorismes de decisió automatitzada són una tecnologia innovadora. L'ús d'aquestes tecnologies és un factor que s'ha de tenir en compte a l'hora de determinar la necessitat de fer una **avaluació d'impacte sobre la protecció de dades** (AIPD). Per tant, es més probable que aquest tipus de tractaments requereixin una AIPD. A banda d'això, la realització de l'AIPD per part del responsable és una manera de demostrar responsabilitat en el tractament de dades personals en el cas que aquesta sigui demandada per l'autoritat de protecció de dades.
- **Dret a no ser objecte de decisions automatitzades.** D'acord amb l'article 22 de l'RGPD, en els casos que la decisió té conseqüències importants per a les persones, no s'ha de prendre cap ADA sobre la base exclusiva de mitjans automatitzats.
- En els casos en què no se n'exclou la possibilitat de presa de decisions automatitzades, convé que l'organització consideri altres maneres d'assolir la finalitat buscada que siguin menys invasives.
- El **principi de lleialtat** exigeix que l'ús de les dades estigui dins del que les persones poden esperar raonablement, i que les conseqüències per a les persones no siguin injustificades. En el cas dels algorismes de decisió automatitzada pren una rellevància especial el biaix dels algorismes, que pot donar lloc a decisions discriminatòries. És molt difícil poder garantir que un algorisme està lliure de biaixos, però hi ha algunes tècniques que poden ajudar a reduir-los:
 - Utilitzar dades no esbiaixades. L'ús de dades amb biaixos en l'entrenament d'un algorisme de decisió automatitzada és una de les principals formes d'introduir biaixos en les decisions dels algorismes. Cal analitzar les dades d'entrenament per detectar i, posteriorment, mitigar els biaixos.
 - No utilitzar característiques que poden donar lloc a decisions discriminatòries: sexe, edat, color de la pell, etc. Cal tenir en compte que això no sempre és possible i que el fet de no usar aquestes característiques no garanteix que no hi hagi discriminació indirecta, a partir d'altres característiques que siguin conseqüència de les anteriors.
 - Analitzar els resultats de l'algorisme per detectar potencials efectes discriminatòris de l'algorisme.
- El **principi de transparència** exigeix que les persones rebin una informació clara respecte al tractament de dades. En el cas d'algorismes de decisió automatitzada que tenen un impacte significatiu sobre les persones, hi ha l'exigència de donar informació sobre la lògica de l'algorisme. Per afavorir la transparència de l'algorisme:
 - Convé utilitzar algorismes que estiguin disponibles per a revisions externes.





- Quan sigui factible, convé utilitzar algorismes explicables en comptes d'algorismes que funcionin com una caixa negra.
- En qualsevol cas, la informació proporcionada sobre la lògica de l'algorisme ha de ser clara i ha d'incloure els aspectes bàsics del seu funcionament.
- Del **principi de minimització de dades**, en resulta una limitació de les dades disponibles en un camp que en fa un ús extensiu. Per minimitzar l'impacte d'aquests principis, convé fer ús de tècniques que redueixen la quantitat de dades necessàries. Entre aquestes dades hi ha les que es comenten en la secció anterior: anonimització, generació de dades sintètiques, selecció d'atributs i aprenentatge federat.
- Del principi de limitació de la finalitat també en resulta una limitació de les dades disponibles. Malgrat aquesta limitació, cal garantir que l'ús de les dades no és incompatible amb la finalitat que va motivar-ne la recollida.
- En el cas que el responsable del tractament identifiqui una finalitat compatible, cal informar els interessats perquè puguin prendre decisions informades respecte a l'ús de les seves dades.

136

Recomanacions per a les autoritats supervisores

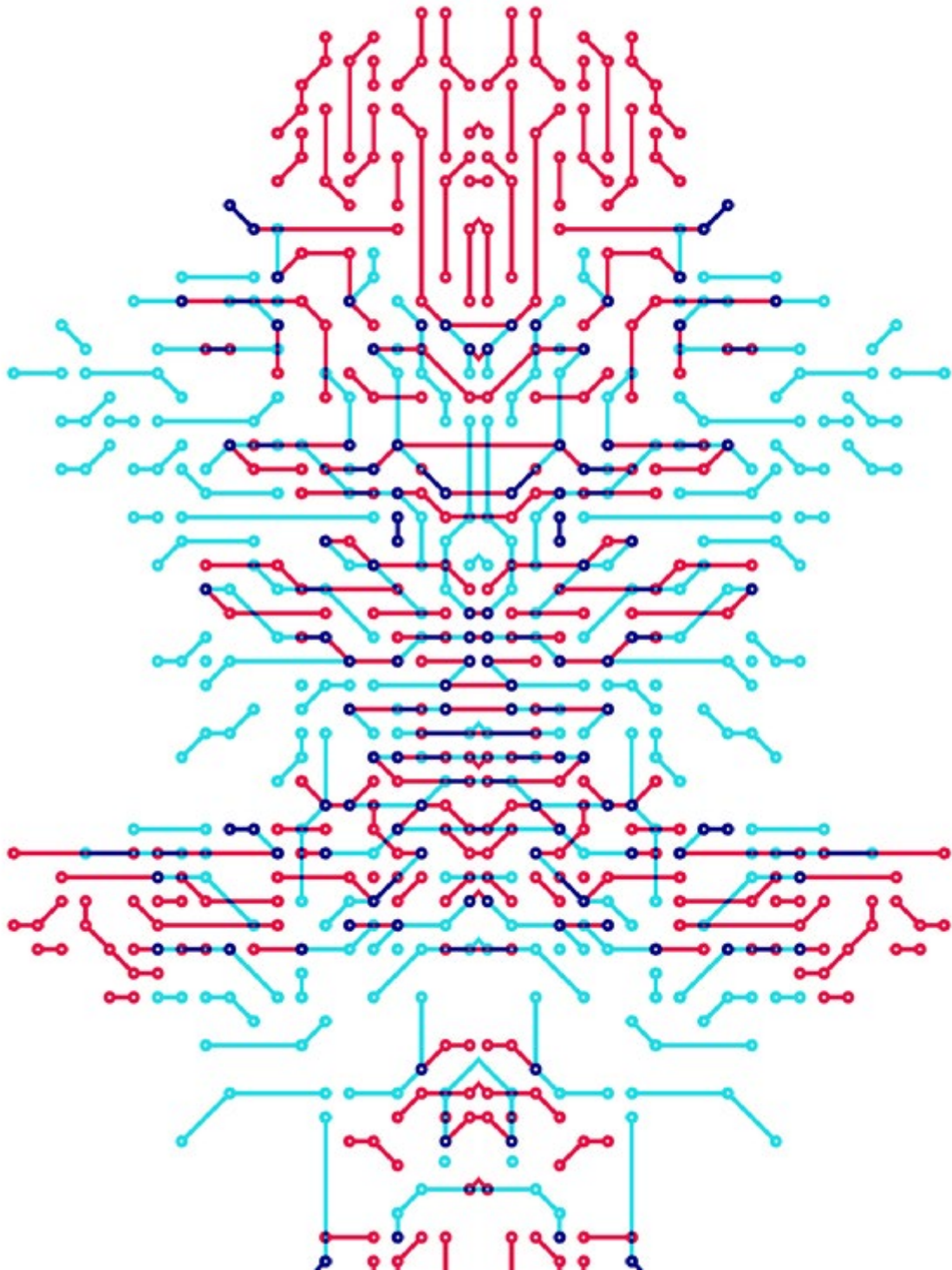
Les friccions dels algorismes d'intel·ligència artificial amb els principis de l'RGPD, com també l'increment del seu ús, fan previsible un increment de les reclamacions davant de les autoritats de protecció de dades, que tenen el mandat de fer complir el reglament.

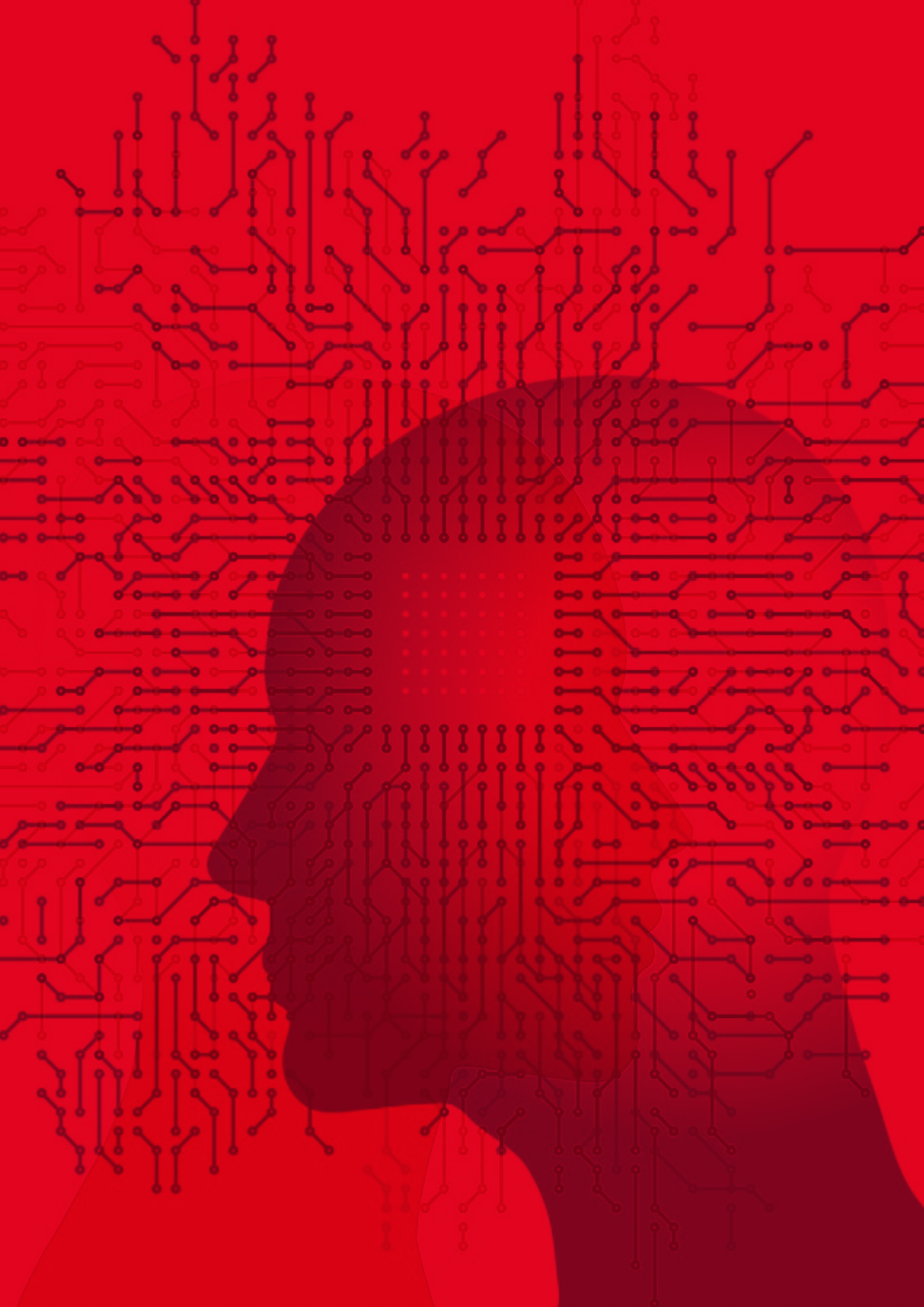
En la seva competència investigadora, les autoritats de protecció de dades poden tenir algunes dificultats relacionades amb la complexitat dels sistemes de decisió automatitzada i d'intel·ligència artificial. Per exemple: a l'hora d'avaluar si el sistema discrimina (principi de lleialtat); a l'hora d'avaluar la necessitat de tractar un tipus de dades (principi de minimització). Per portar a terme aquestes tasques, pot ser necessari comptar amb l'ajuda d'experts.

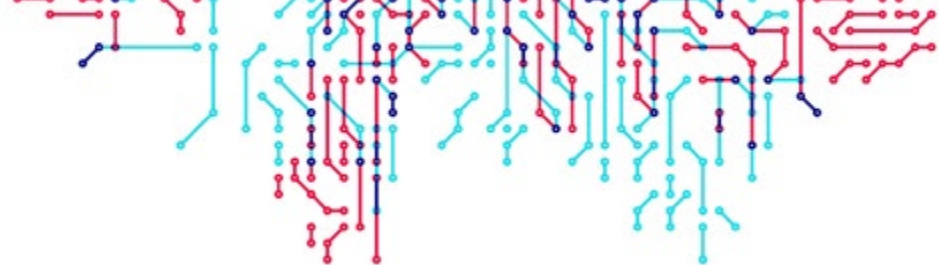
Donada la importància de les autoritats supervisores per garantir el compliment de l'RGPD i per demanar canvis en el tractament quan no es compleix, cal:

- Tenir el coneixement necessari per portar a terme de forma efectiva les tasques de supervisió que tenen encomanades. Els ADA, la IA i, més en general, la tecnologia és un camp en ràpida evolució. Cal que les autoritats de protecció de dades tinguin prou coneixements o la capacitat d'adquirir-los quan sigui necessari.

- Cal fer un esforç divulgatiu per conscienciar la ciutadania sobre els beneficis associats als ADA i a la IA, però també sobre els riscos que comporten .
- Cal fer un esforç divulgatiu adreçat a les organitzacions que fan ús dels ADA i la IA, amb l'objectiu de conscienciar-les dels problemes amb relació a la protecció de dades i fomentar la incorporació d'aquesta temàtica en els codis de conducta.







5. GLOSSARI

A

Algorisme. Seqüència precisa d'instruccions que cal seguir per fer una tasca determinada. Els algorismes són a tot arreu. Per exemple, els passos necessaris per matricular-se en una escola o per demanar una ajuda social no deixen de ser algorismes. En computació, els algorismes són essencials. De fet, l'únic que fa un ordinador és seguir les instruccions dels algorismes amb què ha estat programat.

Algorisme de decisió automatitzada (ADA). Algorisme que pren decisions de manera automatitzada sense que sigui necessària cap intervenció humana. En els darrers anys, aquests algorismes han aconseguit una gran extensió. Per exemple, es calcula que un 70% de les transaccions financeres les realitzen algorismes. Els efectes potencials d'aquests algorismes sobre les persones han posat de manifest qüestions ètiques importants, com ara la responsabilitat de les decisions (per exemple, en un accident d'un cotxe autònom), etc.

Antidiscriminació. Conjunt de tècniques que s'utilitzen per avaluar i mitigar el tractament discriminatori respecte a certs grups de persones.

Aprenentatge automàtic (*machine learning*). Subconjunt d'algorismes d'IA que s'utilitzen per ensenyar als ordinadors a fer tasques sense haver-hi de donar instruccions precises. L'aprenentatge automàtic és particularment útil per ensenyar els ordinadors a fer tasques complexes per a les quals no es coneix un algorisme determinat. Per exemple, no coneixem cap algorisme que determini si un correu electrònic és un correu brossa. El que fa l'aprenentatge automàtic és que a partir d'un conjunt de correus electrònics assenyalats com a són brossa, analitza les dades per determinar les característiques que són indicatives de correu brossa.

Aprenentatge federat (*federated learning*). Tècnica usada en aprenentatge automàtic per entrenar el sistema de forma descentralitzada, de manera que no és necessari que una entitat central tingui accés a totes les dades d'entrenament. El model actual es distribueixen el conjunt d'entitats que participen en l'entrenament. Cadascuna d'aquestes entitats l'actualitza sobre la base de les seves dades i retorna l'actualització. L'aprenentatge federat té dos avantatges importants: millora la privacitat de les persones (atès que no cal una entitat que tingui accés a totes les dades) i distribueix el cost computacional de l'entrenament entre diverses entitats.

Aprentatge profund (*deep learning*). Fa referència a l'ús de xarxes neuronals complexes que compten amb múltiples capes de neurones. L'elevat cost computacional d'aquesta tècnica fa que el seu ús pràctic sigui força recent.

Aprentatge no supervisat. Tècnica usada en aprenentatge automàtic per entrenar el sistema. Donat un conjunt de dades d'entrenament, l'aprenentatge no supervisat busca patrons en aquestes dades (relacions de causa i efecte, correlacions, grups tipus, etc.).

Aprentatge supervisat. Tècnica usada en aprenentatge automàtic per entrenar el sistema. Considerem un algorisme que fa ús d'un model matemàtic que, donades unes dades d'entrada, produeix un resultat. A partir d'un conjunt de dades d'entrenament que tenen les entrades i la sortida que s'espera de l'algorisme, l'aprenentatge supervisat ajusta al model per minimitzar l'error sobre les dades d'entrenament.

C

Caixa negra (*black box*). La complexitat inherent a molts dels models usats en aprenentatge automàtic fa que sigui difícil explicar-ne el comportament. En aquests casos, es diu que el model es comporta com una *caixa negra*: rep unes entrades i produeix una sortida, però desconeixem com es genera la sortida a partir de les entrades. Cada model té un nivell específic de transparència. Per exemple, un sistema que aprèn regles a partir de les dades d'entrenament és molt transparent; mentre que un sistema basat en aprenentatge profund és una caixa negra (la complexitat del model fa que sigui difícil determinar quines característiques de les dades d'entrada tenen rellevància a l'hora de determinar el resultat).

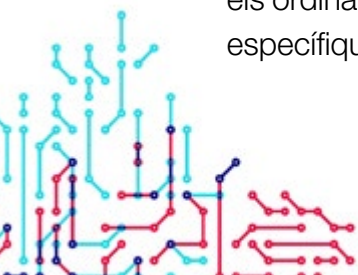
Consentiment. Expressió clara, informada i lliure que fa l'interessat respecte a l'acceptació del tractament de les seves dades.

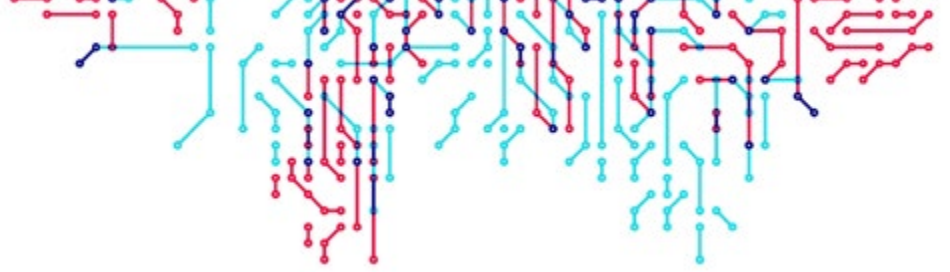
D

Dades personals. Dades associades, o que és possible associar, a una persona.

I

Intel·ligència artificial (IA). És un terme genèric que engloba aquells algorismes que busquen dotar els ordinadors d'un comportament intel·ligent. La definició de *comportament intel·ligent* és evasiva. Hi ha dues visions principals: IA general (que busca dotar els ordinadors de capacitats intel·lectuals humanes) i IA dèbil (que se centra en tasques específiques com ara el reconeixement de veu, d'imatges, etc.). Actualment, la intel·li-





gència artificial ha superat les capacitats humanes en moltes tasques específiques, però la IA general encara és llunyana.

Interessat. Persona a qui fan referència les dades.

L

Lleialtat (*fairness*). Principi bàsic de l'RGPD. Un tractament de dades és lleial quan fa un ús de les dades previsible per part dels interessats i d'aquest tractament no se'n deriven conseqüències que siguin injustificades. Per exemple, un tractament lleial no ha de discriminar per gènere, color de la pell, etc.

M

Model. Representació de la realitat subjacent que fan servir els algorismes de decisió automatitzada a l'hora de fer prediccions i prendre decisions. Aquest model s'ajusta a cada cas particular en un procés d'aprenentatge basat en les dades disponibles.

R

Reglament general de protecció de dades (RGPD). Regulació europea sobre l'ús de dades personals. Va entrar en vigor el maig del 2018. Substitueix la directiva de protecció de dades 95/46/EC i reforça les garanties que donava.

Responsable del tractament. Persona o organització que determina la finalitat i la manera en què s'han de processar les dades.

T

Transparència. Principi bàsic de l'RGPD. Perquè un tractament sigui transparent cal que els interessats rebin informació que sigui clara i que faci ús d'un llenguatge entenedor.

X

Xarxa neuronal. Model d'aprenentatge automàtic inspirat en el funcionament de les xarxes de neurones presents en els organismes vius.

apdcat

Autoritat Catalana de Protecció de Dades



**Generalitat
de Catalunya**