

Pensar la intel·ligència artificial responsable: una guia de deliberació

introducció	2
Primera part	
Intel·ligència artificial	4
Què és la IA?	4
El que fa la IA	5
El que no fa la IA	10
Segona part	
L'ètica de la IA	11
Què és l'ètica de la IA?	11
Alguns reptes ètics i socials	12
Dels principis a l'acció	14
Casos d'ús relatius a l'ètica de la IA	15
Tercera part	
Deliberar sobre l'ètica de la IA	22
Implicar a la ciutadania	22
Per què deliberar sobre l'ètica de la IA?	23
Crèdits	26
Agraïments	27

Algorismes, dades, intel·ligència artificial (IA): molts termes que ja formen part de la nostra vida quotidiana i que, tanmateix, requereixen una certa explicació. Com comprendre aquestes tecnologies? Susciten, certament, un ampli entusiasme i comporten promeses d'un futur millor. Però, quins riscos correm si perdem el control del seu desenvolupament?

Per fer front als reptes d'un desenvolupament responsable de la IA, nombrosos organismes privats, institucions públiques i organitzacions internacionals han publicat cartes de bones pràctiques, declaracions de principis i recomanacions que han posat de manifest l'existència de convergències al voltant de principis clau (justícia, autonomia, benestar). De vegades, però, els principis són abstractes i no sempre es defineixen a tot el món de la mateixa manera.

Queda, doncs, encara molt per fer. Ara, es tracta de reflexionar sobre la implementació dels principis ètics generals, fer-los aplicables a sectors d'activitat específics (ensenyament, ciències, notícies, sanitat, etc.) i, d'una manera concreta, dur a la pràctica les propostes sorgides d'aquesta reflexió. Després, és indispensable implicar més a les ciutadanes i als ciutadans per definir les directrius d'un ús responsable de la IA i de les dades massives, així com recollir les opinions informades d'aquesta mateixa ciutadania. De fet, el desplegament de la IA afecta a tota la ciutadania, homes i dones, i planteja qüestions ètiques i polítiques que haurien de ser objecte d'una deliberació ciutadana. Finalment, és fonamental consolidar l'alfabetització digital del gran públic; és a dir, informar i donar als ciutadans les eines necessàries per poder comprendre millor els reptes del desenvolupament responsable de la IA i participar en les deliberacions públiques

sobre els principis i les normes del seu desplegament.

La participació en els tallers de deliberació sobre la IA i les tecnologies digitals es recolza en una bona comprensió, tant dels reptes ètics i socials de la IA com de les regles de la deliberació. Aquesta guia, precisament, té per vocació fer accessibles la IA i les qüestions ètiques relacionades amb ella i introduir a la deliberació sobre l'ètica de la IA. Per això, conté definicions, il·lustracions i casos d'ús, que haurien de permetre crear un llenguatge comú al voltant de l'ètica de la IA. Aquesta guia també aspira a oferir a les comunitats les eines necessàries perquè organitzin les seves pròpies deliberacions sobre la IA i el desplegament de les tecnologies digitals en el seu entorn social. S'ha dissenyat per tal de permetre als docents, a representants comunitaris, ciutadans i administradors, poder fer viure el debat i definir estratègies comunes.

Aquest plantejament, alhora deliberatiu i participatiu, es fonamenta en la confiança en les capacitats dels individus de concebre i dissenyar el seu futur i el tipus de societat en què volen viure, de formular els principis ètics i polítics que haurien d'organitzar-la i elaborar propostes pertinents de política pública.

Finalment, aquest document té certes limitacions inevitables. Senzill, per a una major claredat i eficàcia, també ha rebut certes influències culturals. És el motiu pel qual serà adaptat a les diferents realitats geogràfiques i culturals dels llocs on

s'organitzin deliberacions. Tothom està convidat a enriquir-lo. Confiem que aquesta guia podrà fomentar la deliberació entre ciutadanes, ciutadans, parts interessades i persones encarregades de la gestió dels assumptes públics, i que els tallers i fòrums de deliberació que proposa contribuiran a un desenvolupament més responsable i més democràtic de la IA.

L'equip de l'Algora Lab
Montreal, el 23 de juny de 2020



Primera part

Intel·ligència artificial

QUÈ ÉS LA IA?

La IA és el conjunt de tècniques informàtiques que permet a una màquina (per exemple, un ordinador, un telèfon) realitzar tasques que generalment requereixen intel·ligència, com ara el raonament o l'aprenentatge. També es parla d'automatització de tasques intel·ligents. Els avenços científics en el camp de la IA, com les tècniques d'aprenentatge profund,

han facilitat la posada a punt de màquines intel·ligents molt eficients gràcies a l'accés a enormes quantitats de dades i a una potència de càlcul cada vegada més elevada. Aquestes noves tècniques s'han desplegat ràpidament a gran escala en tots els àmbits de la vida social, en els transports, l'ensenyament, la cultura i la sanitat.



EL QUE FA LA IA

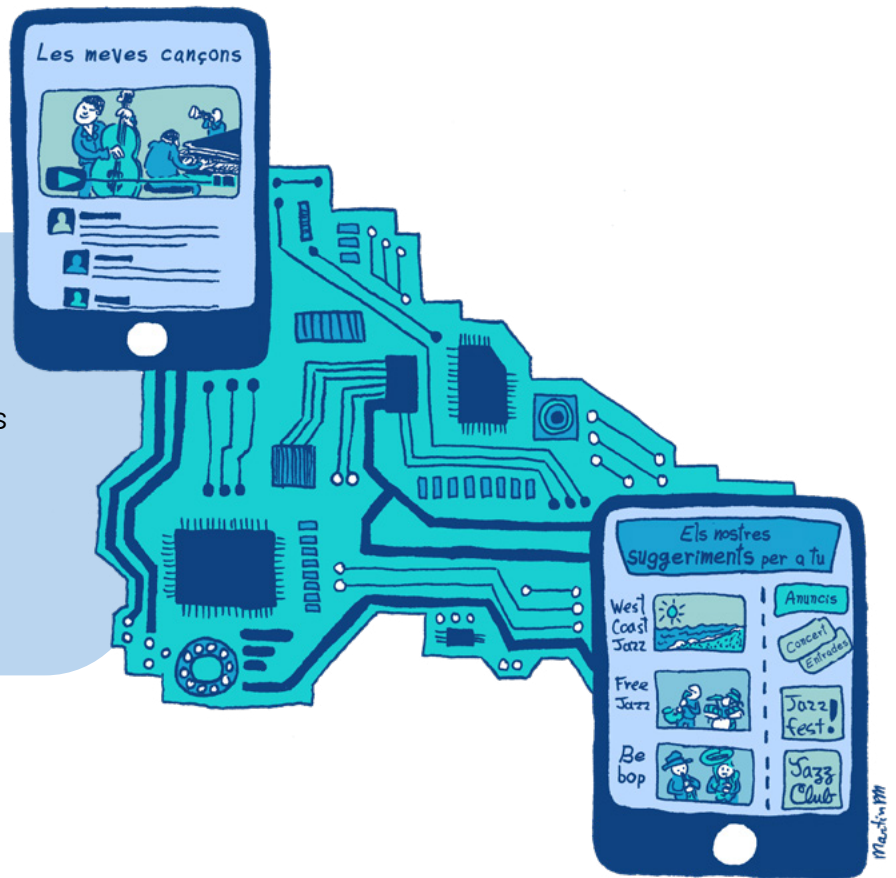
La intel·ligència artificial està basada en la utilització d'algorismes que processen dades. Un algorisme és un conjunt d'instruccions que permet resoldre problemes i realitzar tasques complexes. Aquesta sèrie d'etapes transforma una informació d'entrada (input) en un resultat útil (output). Així, una recepta de cuina és un tipus d'algorisme: per cuinar un plat

(output), és necessari disposar dels bons ingredients (input) i seguir pas a pas una recepta que dona instruccions per utilitzar-los correctament (l'algorisme). La sèrie d'instruccions que un ordinador utilitza per predir l'edat d'una persona (output) a partir de la seva fotografia (input), també es defineix mitjançant un algorisme.

Els algorismes que s'utilitzen en el camp de la IA permeten realitzar funcions com ara:

LA PREDICCIÓ

(per exemple, pronosticar l'interès de qualsevol usuari d'Internet per un tipus de contingut cultural a partir de les seves dades de navegació)



LA DETECCIÓN

(per exemple, detectar si hi ha una cara i el lloc en què es troba partint d'una fotografia)

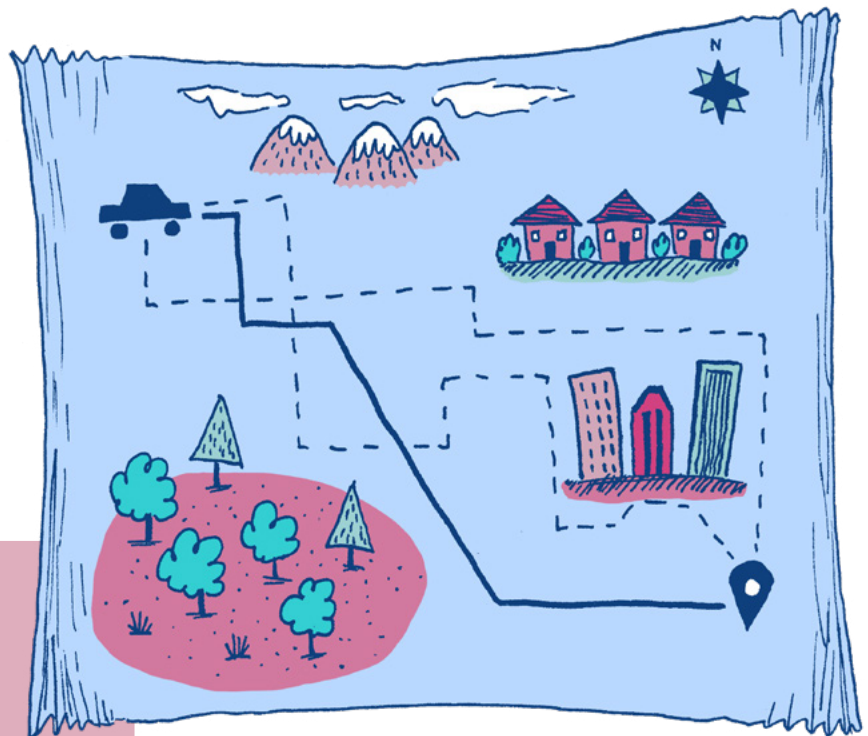
LA IDENTIFICACIÓ

(per exemple, trobar el nom d'una persona a partir de la seva foto)



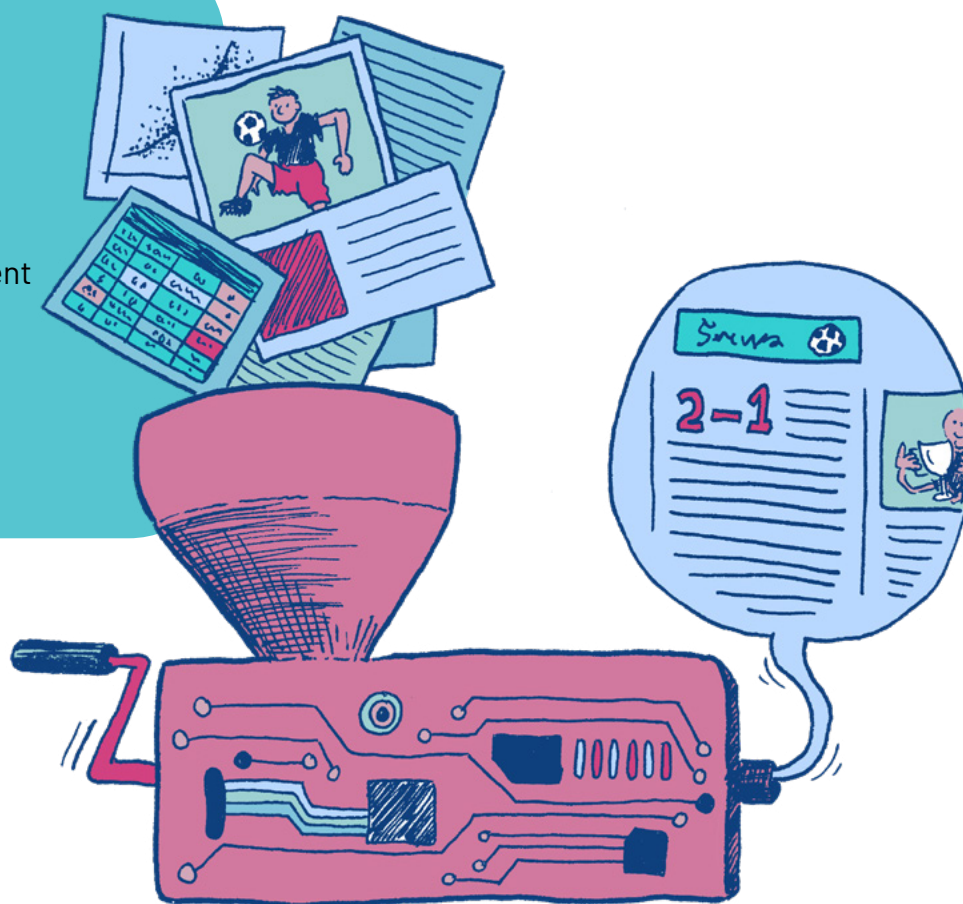
LA PLANIFICACIÓ I LA PRESA DE DECISIONS

(per exemple, escollir el trajecte més ràpid per anar a l'hospital a partir de dades del trànsit)



LA GENERACIÓ DE CONTINGUT

(per exemple, generar un article de periòdic sobre resultats d'un esdeveniment esportiu)



La combinació d'aquestes funcions permet realitzar diferents tasques d'alt nivell, com ara:

> LA PERCEPCIÓ

Quan s'utilitza la IA per analitzar les mesures, com la imatge d'una càmera o la gravació d'un micròfon. Així, un algorisme pot reconèixer paraules pronunciades en veu alta o escrites a mà, com també determinar si hi ha una concentració de gent a partir de la detecció i identificació de cares en una fotografia.

> EL PROCESSAMENT DEL LLENGUATGE NATURAL

Quan la IA s'utilitza per processar el significat de les paraules. Una IA és capaç de comprendre quan se li dona una ordre, com trucar a un contacte o predir la propera paraula que s'escriurà, com en els teclats dels telèfons intel·ligents, per exemple. També és possible detectar si, a les notícies, un article interessarà una certa persona, o fins i tot generar la traducció d'un paràgraf en una llengua estrangera.

> LA ROBÒTICA

Quan la IA controla una màquina que pot actuar en el món físic. A partir de la informació captada pels seus sensors, la IA ha de planificar i prendre la millor decisió per assolir el seu objectiu. Un cotxe autònom, per exemple, ha de conduir de manera segura fins a un destí donat, i un robot quirúrgic ha de reproduir exactament els moviments de l'humà que el controla.

> L'OPTIMITZACIÓ I LA RESOLUCIÓ D'UN PROBLEMA

Quan la IA ha de resoldre una situació en un context definit. Per exemple, prendre decisions en un videojoc o planificar el millor trajecte possible per a un cotxe o els desplaçaments de taxis.

Una gran varietat d'**àmbits d'aplicació** es pot beneficiar de la possibilitat d'automatitzar aquestes funcions: l'ensenyament, el periodisme, la ciberseguretat, els videojocs, l'art (música, cinema, etc.), les finances, les prestacions sanitàries, els transports, el camp militar, l'ecologia, la climatologia, etc.

PER APROFUNDIR UNA MICA

Hi ha **dues branques tècniques** principals en IA. Els dissenyadors de l'algorisme poden donar totes les **instruccions** a la màquina per endavant o bé permetre a la màquina que **aprengui** les etapes per si sola. Escriure les regles per endavant requereix dissenyadors que ja sàpiguen **com resoldre** el problema que demanen que resolgui la IA. En certs casos és possible: així, la utilització de models científics basats en lleis de la física permeten preveure el temps que farà. Tanmateix, quan el problema és massa **complex**, els dissenyadors de la IA s'estimen més utilitzar l'aprenentatge màquina. En aquest cas, és necessari **entrenar** el model. És molt difícil, per exemple, escriure les normes que permeten a un ordinador diferenciar un gat d'un gos en una imatge ja que, per un costat hi ha una gran diversitat de cada una d'aquestes dues espècies i, a més, les fotografies poden haver estat preses des d'angles diversos, amb il·luminacions diferents, etc. Els algorismes d'aprenentatge màquina permeten resoldre el problema, per això es mostren diverses imatges de gossos i gats a un ordinador i se l'enseny a diferenciar-los per si mateix. Per descomptat, el resultat òptim de l'algorisme depèn de la **quantitat** d'imatges i de la seva **representativitat** —si ha estat entrenat únicament amb gats negres vistos de cara, l'algorisme no sabrà què fer amb una fotografia d'un gat blanc presa des de dalt. Aquest és el motiu pel qual els algorismes d'IA requereixen, generalment, una gran quantitat de dades i que la manera en què han estat col·lectades també és important.

EL QUE NO FA LA IA

Actualment, no existeix una IA general, o «forta», que, com la intel·ligència humana, sigui capaç d'executar tasques tan diferents com jugar a escacs, conduir un vehicle i reconèixer o detectar tumors. Per a alguns investigadors en aquest camp, és un objectiu desitjat; tanmateix, els sistemes actuals més avançats estan lluny d'aquestes possibilitats. Es diu que la IA actual és «feble», ja que, si bé pot realitzar algunes tasques de manera més eficient que un ésser humà, només ho fa en aquelles tasques i funcions per a les quals ha estat dissenyada i pensada. D'altres també consideren el fet que una IA general podria, algun dia, tenir emocions o consciència de si mateixa. Encara no estem a aquest nivell, i tot el que una IA feble és capaç de fer és identificar emocions i simular-les.

NOCIONS ESSENCIALS

INTERNET DELS OBJECTES

Infraestructura d'objectes interconnectats i capaços de comunicar entre si sense intervenció humana.

ALGORISME

Sèrie d'instruccions que permet transformar una entrada en una sortida. Per exemple, una recepta de creps permet transformar ingredients en un àpat deliciós tot seguint unes etapes precises. Les etapes per resoldre el cub de Rubik també són un algorisme.

DADES MASSIVES, MACRODADES O BIG DATA

Conjunt de dades tan voluminos que és impossible col·lectar-les, emmagatzemar-les i analitzar-les utilitzant mètodes tradicionals. Nombrosos algorismes d'IA utilitzen dades massives.

APRENTATGE MÀQUINA

Capacitat que té una màquina d'aprendre a fer una tasca sense instruccions, sinó més aviat gràcies a una experiència adquirida durant un procés d'entrenament.

APRENTATGE PROFUND

Aprenentatge màquina amb una tècnica específica i especialment potent que utilitza xarxes de neurones connectades a diverses capes. Aquestes xarxes de neurones artificials, que s'inspiren del funcionament de les neurones del cervell humà, consisteixen en un tractament matemàtic de les dades entrants.

Segona part

L'ètica de la IA

Si les tecnologies d'IA són prometedores, la seva utilització posa de manifest reptes ètics i socials sobre els quals és primordial reflexionar de manera col·lectiva, sobretot si considerem la seva creixent influència en les nostres societats. Es tracta d'un dels objectius de l'ètica aplicada a la IA que pretén identificar i impedir les males utilitzacions de la IA, alhora que n'afavoreix els beneficis que se'n desprenen.

QUÈ ÉS L'ÈTICA DE LA IA?

L'ètica designa un conjunt de normes (principis, regles) a les quals ens hem de sotmetre si volem actuar bé, com el principi segons el qual no s'ha de perjudicar els altres o la regla que prohibeix mentir. Es diu que l'ètica és prescriptiva perquè prescriu el que ha de ser, el que s'ha de fer o el que és acceptable segons els valors adoptats. Les normes ètiques articulen valors que reconeixem com a morals. L'ètica també designa la disciplina filosòfica que intenta establir aquestes normes i aquests valors.

L'ètica de la IA és el conjunt de regles i valors aplicats al desenvolupament i a la utilització de la IA. Es tracta doncs d'una àrea restringida d'aplicació de l'ètica encara que, atès que les tecnologies de la IA alteren l'organització social i poden tenir conseqüències perjudicials molt profundes, aquest camp de l'ètica es presenta com crucial i coneix un desenvolupament molt important.

L'ètica de la IA, finalment, forma part de l'ètica pública, és a dir, l'ètica aplicada a controvèrsies socials que requereixen una solució acceptable per als individus que estan en desacord. És el que passa també amb l'ètica del medi ambient o de la bioètica.

ALGUNS REPTES ÈTICS I SOCIALS

Experts, investigadors, així com ciutadans, ja han manifestat moltes preocupacions sobre el desenvolupament i la utilització de la IA. Entre altres citarem:

> EL RISC D'INGERÈNCIA EN LA VIDA PRIVADA

La vida privada és un valor social que ha sorgit fa poc en les societats que aspiraven a la democràcia. Actualment, aquest valor ja està àmpliament acceptat i moltes són les persones preocupades per la violació de la seva vida privada. El risc d'ingerència en la vida privada es veu afavorit pels resultats, sense precedents, dels mètodes aplicats per col·lectar dades, cada vegada més invasius, i per la multiplicació dels llocs on són col·lectades aquestes dades personals (per exemple, telèfons mòbils o aparells connectats a les llars). Aquest risc també augmenta com a conseqüència de les noves possibilitats d'anàlisi que ofereix la IA. Els algorismes, per exemple, són capaços d'identificar individus, creuant diferents dades que no obstant havien estat anonimitzades.

> EL RISC DE LIMITAR LA LLIURE ELECCIÓ I L'AUTONOMIA

Generalment, valorem la lliure elecció i l'autonomia, és a dir, la capacitat de ser l'autora o l'autor de les seves pròpies decisions. Les màquines, tanmateix, poden restringir considerablement la nostra llibertat en prendre decisions en nom nostre sense que en siguem conscients o sense poder-nos oposar a elles. Podem pensar, entre altres, en els efectes «filtre bombolla» dels algorismes que proposen als usuaris continguts que sempre estan d'acord amb el seu comportament digital (és a dir, les seves eleccions anteriors), limitant d'aquesta manera la diversitat del contingut que se'ls proposa o les oportunitats de descobrir noves preferències. Un bon exemple d'això és quan llocs web musicals proposen cançons i músiques sempre similars a les que els usuaris acostumen a escoltar.

> EL RISC DE DISCRIMINACIÓ

La discriminació consisteix en tractar diferentment casos semblants sense una bona raó. És el contrari de la justícia, que consisteix en tractar amb igualtat casos similars. Una màquina intel·ligent pot reproduir o generar discriminacions quan el seu algorisme, o les dades a partir de les quals aprèn, contenen errors o biaixos que fan que es tracti diferentment, sense justificació acceptable, persones o grups respecte de la resta de la població. D'aquesta manera, persones

o grups queden exclosos dels beneficis de la IA. Una màquina entrenada a partir de dades que no tracten ni inclouen a tota la població privarà una part d'aquesta població dels seus beneficis; aquest és el cas quan una màquina que detecta els càncers de la pell ha estat entrenada únicament per a pells clares: no detectarà amb la mateixa fiabilitat els càncers en pells més fosques.

> **LA MANCA D'EXPLICABILITAT**

Sovint, és difícil explicar com un algorisme arriba a tal decisió o a tal recomanació. Les màquines intel·ligents són com «caixes negres», que no se sap com funcionen. Ara bé, sembla important estar en situació de comprendre una decisió que ha pres un algorisme per poder conservar el control de les decisions que ens afecten, així com per poder qüestionar-les o fins i tot canviar-les. L'explicabilitat és un valor estretament lligat a la lliure elecció i a l'autonomia, i també a la justícia, com quan la màquina intel·ligent reproduïx o genera discriminacions.

> **L'AUTOMATITZACIÓ DEL TREBALL**

Per bé que les màquines intel·ligents permeten efectuar tasques repetitives i redueixen d'aquesta manera la penositat del treball, tendeixen a substituir a les treballadores i als treballadors humans. A més de les conseqüències socials i econòmiques d'aquesta substitució, com ara un augment de l'atur, aquesta substitució

afecta el sentit de la solidaritat i de l'autoestima. Un dels reptes ètics consisteix aleshores en mantenir relacions humanes, sobretot en sectors com el de la salut o l'ensenyament; un altre consisteix en preservar les condicions del respecte i de l'autoestima que es recolzen en part en el sentiment de la seva utilitat social.

> **ELS RISCOS PER AL MEDI AMBIENT**

La protecció del medi ambient i la lluita contra el canvi climàtic són reptes socials esdevinguts molt importants. El desplegament de la IA pot ajudar a millorar les nostres accions col·lectives i individuals per reduir el nostre impacte negatiu sobre el medi ambient, però també pot augmentar aquest impacte negatiu si ens en despreocupem. Interessar-se pel repte mediambiental de la IA implica tenir en compte l'impacte de tots els objectes que fan possible la utilització de la IA, com ara els telèfons intel·ligents amb els seus components contaminants o els centres de dades. Per exemple, els centres d'emmagatzematge de dades massives (data center), en especial els que permeten l'aprenentatge màquina, són grans consumidors d'energia.

DELS PRINCIPIS A L'ACCIÓ

Per tal de donar respostes socialment acceptables a les diferents preocupacions ètiques, socials i polítiques que suscita el desplegament de la IA, nombroses declaracions ètiques de principis i línies directrius han aparegut a tot el món. Aquests documents tendeixen a definir els principis ètics que permetran guiar la reflexió per tal de limitar les conseqüències negatives de la utilització de la IA. És el cas, per exemple, dels principis adoptats per l'Organització per a la Cooperació i el Desenvolupament Econòmic (OCDE) i pels països del G20, de les directrius per a una IA digna de confiança de la Comissió Europea o de la Declaració de Montreal per a un desenvolupament responsable de la IA, que es distingeix pel procés deliberatiu que ha alimentat el treball dels experts. Aquesta declaració promou 10 principis ètics per a un desenvolupament responsable de la IA, com ara el principi del benestar:

El desenvolupament i la utilització dels sistemes d'intel·ligència artificial (SIA) han de permetre augmentar el benestar de tots els éssers sensibles.

Els principis tenen per objecte guiar l'acció. D'aquesta manera, poden servir de base per a diferents mesures concretes de la gestió i del desenvolupament ètic de la IA. Poden ser, per exemple:

- > lleis;
- > polítiques públiques;
- > auditories o certificacions;
- > formacions;
- > actors institucionals;
- > codis de conducta; o
- > solucions tècniques.

NOCIONS ESSENCIALS

DILEMA ÈTIC

Una situació presenta un dilema ètic quan exigeix una elecció entre dues opcions incompatibles i ambdues moralment insatisfactòries. Això es produeix quan hi ha un conflicte de principis o de valors i que s'ha de sacrificar un valor moral important.

GOVERNANÇA ALGORÍSMICA

El terme pot al·ludir tant a la manera de governar els algorismes, emmarcar el seu desenvolupament i regular-ne la utilització, com a la manera com els algorismes ens governen o regulen la nostra vida, les nostres relacions socials, les nostres institucions públiques.

CAIXA NEGRA

L'expressió caixa negra designa un sistema d'IA les decisions o recomanacions del qual són difícils o impossibles d'explicar. Aquest terme destaca la manca de transparència del funcionament de les màquines intel·ligents, en especial aquelles que utilitzen algorismes d'aprenentatge profund (*deep learning*).

ÈTICA EN EL DISSENY (ETHICS BY DESIGN):

Una manera d'emmarcar l'ús ètic de les eines digitals i dels sistemes d'IA consisteix en considerar els principis ètics tan bon punt són dissenyats per investigadors i enginyers (homes i dones), impedint d'aquesta manera, en les etapes inicials, la seva utilització no ètica o socialment no desitjable.

BIAIX

S'entén per biaix el mecanisme del pensament que altera i falseja un judici. Un algorisme pot presentar biaixos si afavoreix certs resultats sense una bona justificació moral.

CASOS D'ÚS RELATIUS A L'ÈTICA DE LA IA

EINES PRÀCTIQUES

Entre els nombrosos reptes ètics, socials i polítics que presenten les aplicacions de la IA, sembla urgent prioritzar els sis reptes següents:

1. **l'equitat, la inclusió i la diversitat en el procés de reclutament;**
2. **les dades i la vida privada en l'àmbit sanitari;**
3. **l'automatització en l'educació;**
4. **la intel·ligència artificial per gestionar les crisis sanitàries;**
5. **la lluita contra els canvis climàtics;**
6. **la desinformació i les xarxes socials.**

Per reflexionar col·lectivament sobre aquests reptes, cal partir de situacions concretes en les quals la IA ha estat desplegada (casos d'ús) i ens afecta tant individualment com col·lectiva. En aquestes situacions, hem de fer una elecció social en consonància amb les exigències ètiques i polítiques que haurem reconegut de manera col·lectiva.

Per donar llum a aquests casos d'ús, hem posat a disposició una il·lustració i un guió prospectiu. Un guió prospectiu és la descripció d'una situació futura. No es tracta ni d'una història vertadera, ni de ciència ficció, sinó d'una situació que es pot produir. En vigílies de transformacions

molt importants amb la IA en les nostres societats, el guió prospectiu ens permet descobrir, crear, pensar i prevenir de manera diferent.

ELS OBJECTIUS

Els sis casos d'ús que presentem a continuació tenen diversos objectius.

- > Són **desencadenadors**: constitueixen punts de partida per a la deliberació.
- > Els casos d'ús són **prospectius**: descriuen situacions tipus susceptibles de presentar-se en un futur proper.
- > Descriuen **dilemes ètics**, és a dir, situacions en què els valors i els principis poden entrar en conflicte o, com a mínim, generar tensions.
- > Els casos d'ús són **oberts**: no indiquen quina és la solució moral que cal adoptar, no prescriuen cap norma o obligació moral en particular.

COM?

Podeu utilitzar el guió o la il·lustració, o els dos junts, per iniciar converses o deliberacions.

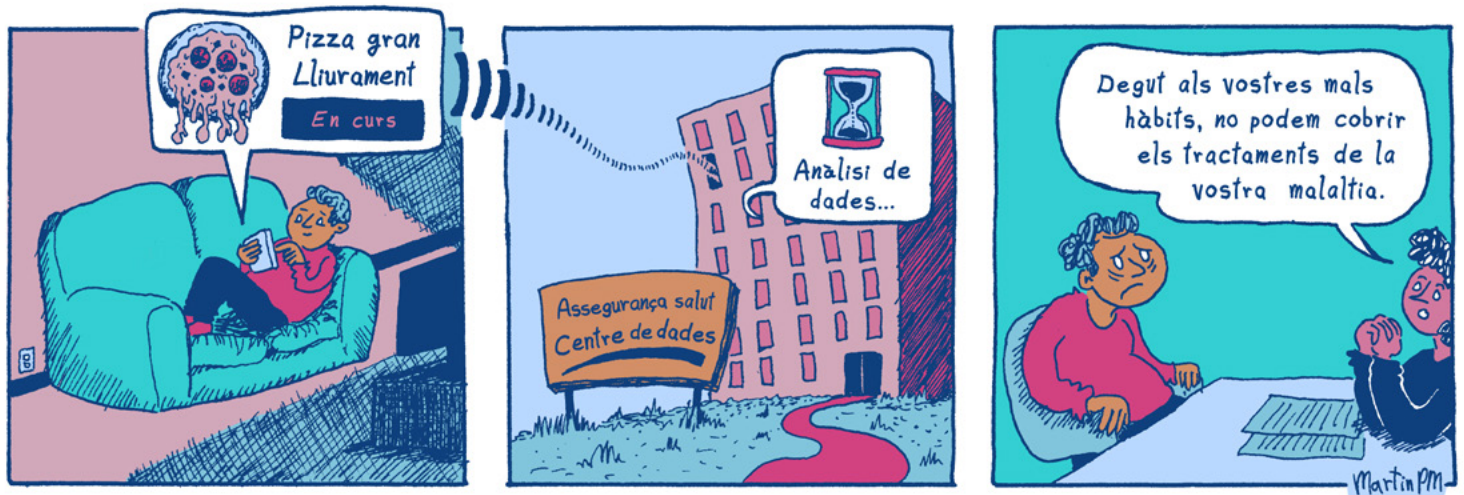
1. L'EQUITAT, LA DIVERSITAT I LA INCLUSIÓ EN EL PROCÉS DE RECLUTAMENT



EManage és una empresa recentment creada que acompanya les petites empreses a fer el pas cap a les noves tecnologies. L'empresa s'ha desenvolupat i està cercant nous talents. L'equip directiu es preocupa molt per tenir una diversitat de candidates i candidats, per a la qual cosa promou la inclusió i l'equitat en la seva oferta de llocs de treball. Per tal d'entrevistar-se amb els candidats per a l'empresa, *EManage* ha utilitzat un procés algorímic de reclutament per, d'aquesta manera, limitar els biaixos humans. Dit d'una altra manera, es tracta d'utilitzar un algorisme per separar i preseleccionar les candidatures amb el propòsit de limitar les influències dels prejudicis sobre la selecció de les dones i

homes candidats. Com funciona això? L'algorisme és entrenat amb les dades de reclutaments anteriors i amb els criteris del perfil de la candidata o del candidat ideal. És a dir que s'utilitzen les dades existents de candidates i candidats que ja han obtingut un treball a l'empresa. Tanmateix, i malgrat la intenció de l'empresa d'obtenir un caràcter mixt i paritari en la seva selecció, semblaria ser que l'algorisme ha reproduït una certa imatge de la candidata ideal o del candidat ideal seleccionant únicament homes per a una entrevista de treball.

2. LES DADES I LA VIDA PRIVADA EN L'ÀMBIT SANITARI



Fa poc, l'Elias ha canviat d'assegurança mèdica: la seva nova companyia d'assegurances regala un rellotge connectat i una aplicació Salut que anima els seus clients a adoptar un estil de vida que redueixi els riscos mèdics. Aquesta aplicació té accés als seus hàbits d'alimentació, als seus desplaçaments, al seu ritme cardíac i a d'altres informacions que permeten definir un perfil de risc dels clients, homes i dones. Un dia, diagnostiquen a l'Elias una malaltia que requereix atencions i tractaments llargs i costosos. Avisa la seva assegurança per iniciar el seu tractament. La seva companyia li anuncia que no cobrirà les seves despeses mèdiques. Quan intenta comprendre el que passa, el servei en línia de l'assegurador li diu que l'algorisme és qui determina la

seva admissibilitat segons paràmetres importants, però ignoren com s'ha arribat a aquest resultat. Davant de la insistència de l'Elias, el servei informàtic de l'assegurador l'informa finalment que el sistema l'ha classificat com a persona d'alt risc: alimentació no sana, sedentarisme. És cert, l'Elias té per costum fer-se lliurar els menjars a domicili gràcies a l'aplicació Deliverfood, i li agraden les pizzes. I les seves activitats físiques? L'aplicació de GymGym revela que ha anat tres vegades al gimnàs en dos anys. La companyia d'assegurances opina que no se la pot considerar financerament responsable dels mals hàbits de vida de l'Elias i que les conseqüències, pel que fa a la seva cobertura sanitària, estaven previstes en la seva pòlissa d'assegurances.

3. L'AUTOMATITZACIÓ EN L'EDUCACIÓ



El col·legi Albert Einstein sempre ha estat a l'avantguarda de les tecnologies pedagògiques. Així, quan la companyia AthenIA li proposa provar el seu nou producte, un programa d'ensenyament de llengües personalitzat, AthenIA va ser adoptat amb entusiasme per a les classes d'anglès. Com de costum, els alumnes segueixen les seves classes a l'aula... però sense professora o professor. Es lliura a cada alumna i a cada alumne una diadema electrònica que permet gravar la veu i llegir el que escriuen. Després es transmet tot a un programa informàtic que elabora un perfil detallat de cada alumne. Amb aquestes informacions, el

programa adapta les activitats en grup i dona deures personalitzats, tenint en compte els progressos a nivell individual. Aquestes dades també s'utilitzaran per a l'avaluació final i per a la promoció a una classe superior.

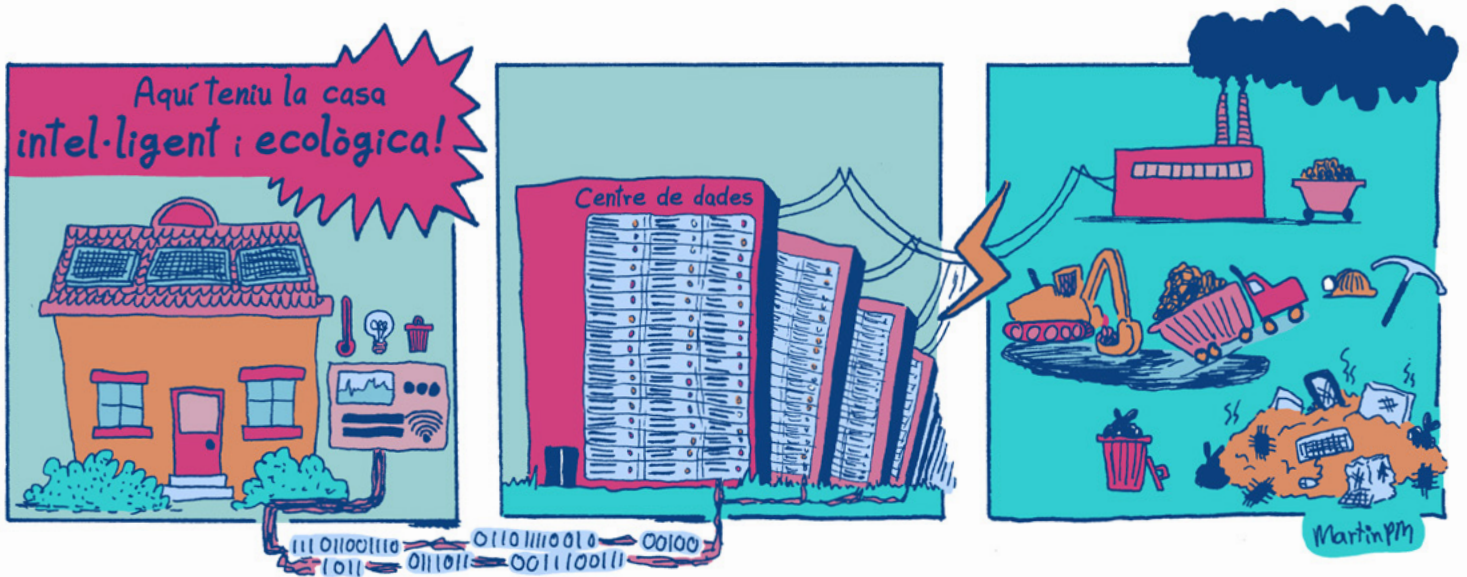
4. LA INTEL·LIGÈNCIA ARTIFICIAL EN LA GESTIÓ DE LES CRISIS SANITÀRIES



Una pandèmia s'ha estès als cinc continents; no hi ha cap país fora de perill. Per limitar la propagació del virus, les autoritats sanitàries han adoptat estrictes mesures de confinament. En constatar la disminució del nombre de contaminacions, els governs decideixen procedir a un desconfinament progressiu. Per fer-ho de manera segura, un dels mitjans emprats és la utilització d'una aplicació mòbil, ANTIV, que col·lecta dades sobre la salut dels individus i registra els seus desplaçaments i els seus contactes. A continuació ANTIV avalua automàticament el nivell de risc d'infecció de cada usuari. Permet a les autoritats investigar després sobre casos de contaminació i avisar si algun ciutadà ha estat en contacte

amb una persona contaminada. Hakim, repartidor per a una cadena de supermercats, utilitza ANTIV. Amb la pandèmia, el volum d'entregues a domicili és molt important i Hakim fa més hores que un rellotge. Durant un dels seus repartiments, es fa una ferida a un turmell. Preocupat, va immediatament a l'hospital. Quan arriba, a l'entrada, el servei de triatge verifica el seu risc d'infecció gràcies a l'aplicació. Segons aquesta, presentaria un nivell de risc elevat, donats els seus nombrosos contactes durant els seus repartiments. Encara que no tingui símptomes, podria ser portador del virus i contaminar els altres pacients de l'hospital. Com la ferida no sembla ser massa crítica, l'hospital no l'admet a urgències i el fa tornar a casa seva.

5. LA LLUITA CONTRA ELS CANVIS CLIMÀTICS



Mei i Pablo es volen fer construir una casa. Un arquitecte els proposa l'últim model de casa ecològicament responsable: intel·ligent i ecològica. Com la parella és sensible a la causa mediambiental, la idea els agrada i demanen més detalls.

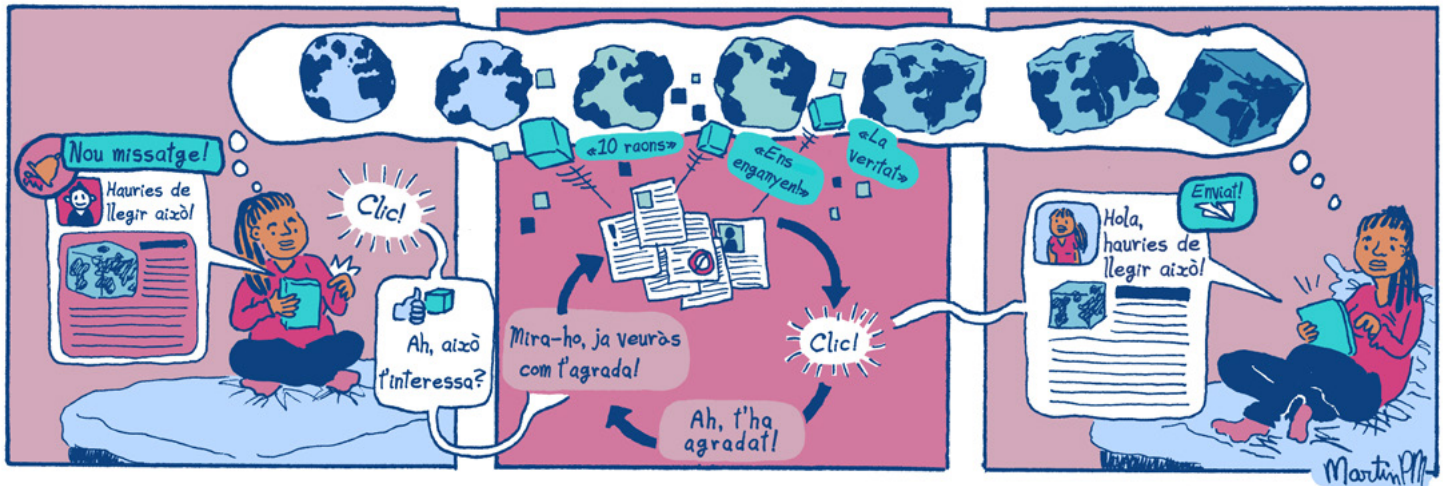
L'arquitecte els explica que la casa utilitzarà sistemes d'intel·ligència artificial per tal d'optimitzar el consum d'energia, la despesa d'aigua, el tractament dels residus domèstics i fins i tot la il·luminació natural. No només estalviaran en reduir el consum d'energia, sinó que el seu impacte ecològic serà insignificant.

Entusiasmats, Mei i Pablo parlen del seu projecte amb una amiga que treballa en una organització de protecció del medi ambient.

La resposta d'aquesta els sorprèn: «Heu pensat en l'impacte mediambiental dels aparells intel·ligents de la vostra casa ecològicament responsable? Comença per l'explotació de les terres rares que s'utilitzen en la fabricació d'aquests aparells electrònics que, a més, tenen un cicle de vida curt. També és necessari reemplaçar-los regularment, cosa que genera molts residus contaminants... I això sense comptar amb el consum astronòmic d'energia dels gegantins centres de processament de dades als què la vostra casa estarà connectada.»

Sens dubte Mei i Pablo s'ho hauran de pensar millor.

6. LA DESINFORMACIÓ I LES XARXES SOCIALS



La Sofia freqüenta les xarxes socials. Els seus amics i parents comparteixen regularment articles sobre temes diversos. Un dia, rep un missatge del seu amic José: «Realment hauries de llegir això». Amb el missatge va adjunt un article que la intriga: «La Terra és cúbica. Una veritat incòmoda.» Escèptica, però curiosa, clica damunt l'enllaç. L'article està ben redactat, el lloc web que el publica (Lesvraisnews.com) està ben fet, però no està convençuda del tot. «Si la Terra no fos rodona, se sabia», pensa. L'endemà, quan es connecta a la seva xarxa social favorita, apareixen en la seva secció de notícies nombrosos articles sobre el mateix tema: «Ho sabia? La Terra és un cub!», «Ens menteixen: les 10 raons per creure que la Terra és cúbica», i més encara «Per què la Terra no és

rodona?: la veritat en un vídeo». En efecte, l'algorisme de la xarxa social ha registrat el seu interès pel primer dels articles i, per tal de presentar-li un contingut personalitzat d'acord amb les seves preferències, li proposa articles semblants. La Sofia comença a plantejar-se algunes preguntes. Durant els següents dies, llegeix tots els articles que apareixen en la seva secció de notícies. Una setmana després escriu a la seva amiga Lenú el missatge següent: «Llegeix això! A la vida cal fer-se algunes preguntes!», i li envia un article titulat «10 proves de que la Terra és cúbica».

Tercera part

Deliberar sobre l'ètica de la IA

IMPLICAR A LA CIUTADANIA

Existen varias maneras para incluir la ciudadanía en los debates sobre la ética de la IA. Según su grado de implicación, se distinguen generalmente tres procesos tipo:

> CONSULTAR

La consulta és un procés que consisteix en recollir les opinions ja formades de les persones consultades sobre un tema prèviament definit. La consulta permet a les persones consultades formular preguntes, manifestar les seves preocupacions, les seves expectatives, fer comentaris i donar les seves opinions a fi i efecte de millorar el procés de presa de decisions.

> DELIBERAR

La deliberació és una discussió raonada mitjançant un intercanvi d'arguments amb l'objecte d'assolir una decisió col·lectiva. La deliberació ha de permetre augmentar el coneixement de cadascuna i cadascun, comprendre els seus propis interessos així com els interessos del grup. Pot dur-nos a modificar les nostres preferències inicials. No pretén necessàriament arribar a un consens, sinó més aviat identificar orientacions comunes basades en les convergències i divergències d'opinions i en els motius que les sustenten.

> COCONSTRUIR

La coconstrucció és un mètode que consisteix en integrar les ciutadanes i els ciutadans al conjunt del procés d'ideació i creació. Es tracta d'un procés col·laboratiu i interactiu en què els ciutadans i els intervinents intercanvien i creen junts.

PER QUÈ DELIBERAR SOBRE L'ÈTICA DE LA IA?

El desplegament de la IA té repercussions en totes les esferes de la vida personal i social; afecta a tothom, i ningú pot mesurar totes les conseqüències d'un fenomen tecnològic i social complex. És essencial multiplicar les experteses: experteses científiques, per descomptat, però també experteses ciutadanes, experteses dels usuaris de la IA, com també de les persones que en pateixen els seus efectes. És el motiu pel qual és essencial implementar **la intel·ligència col·lectiva** i implicar el major nombre possible de persones en el procés de reflexió sobre les qüestions socials i ètiques de la IA, més enllà dels cercles d'experts i de prenedors públics de decisions.

La deliberació no només permet aprofundir els nostres coneixements de la IA com objecte tecnològic que transforma les nostres relacions socials i polítiques, sinó que també permet prendre millors decisions, ja que els confereix una legitimitat que sovint no tenen les decisions dels experts. Això requereix la participació d'un important nombre de persones i d'una gran diversitat de participants, homes i dones. La riquesa cultural i social del món és l'únic límit.

Finalment, participar a les deliberacions permet a cadascuna i cadascun de vosaltres i de nosaltres **fer sentir la seva veu** i donar-se l'oportunitat de posar la IA al servei del bé comú i dels nostres interessos fonamentals.

NOCIONS ESSENCIALS

ARGUMENT

Un argument és un raonament que permet demostrar o justificar una afirmació (una opinió). Per defensar una opinió i convèncer amb bones raons les persones a les quals ens adreçem hem de recórrer a arguments coherents.

BÉ COMÚ

El bé comú remet a una realitat compartida per tothom, sigui quina sigui l'organització social. El «comú» implica la idea d'un vincle entre els membres del grup. Una innovació per al bé comú suposa que aquesta es desenvolupi en interès de tots.

INTEL·LIGÈNCIA COL·LECTIVA

L'expressió «intel·ligència col·lectiva» designa la capacitat, per a un grup, d'assolir solucions més adaptades, prendre millors decisions i aprofundir els seus coneixements a través de la discussió, l'intercanvi d'arguments i la posada en comú dels coneixements dels individus. L'exercici de la intel·ligència col·lectiva exigeix als membres del grup que comparteixin objectius i interessos comuns, així com un espai col·laboratiu (físic o virtual).

PER DELIBERAR BÉ, RECORDEM QUE...

... les i els participants són iguals en la discussió, i la participació en un taller deliberatiu exigeix un esperit de respecte mutu.

... cada opinió, compta: no es poden excloure, sense discussió, opinions defensades de bona fe per les i els participants.

... les opinions que es manifesten han de ser recolzades per arguments i l'intercanvi d'arguments ha de ser públic.

Una **deliberació** agrupa una diversitat de persones que tenen opinions diferents sobre un mateix tema.



Implica una reflexió col·lectiva que comporta intercanvi i argumentació.



Pensem de manera col·lectiva vers un consens.

Un consens no significa que tothom estigui d'acord. Un consens és una posició o una opinió que sembla la més raonable per a la col·lectivitat.



MartinPM

AGRAÏMENTS I CRÈDITS

PER A LA REDACCIÓ

DILHAC, Marc-Antoine

Université de Montréal, OBVIA, Mila-Institut Québécois d'intelligence artificielle, Chaire CIFAR, Algora Lab.

MAI, Vincent

Université de Montréal, Mila-Institut Québécois d'intelligence artificielle, Algora Lab.

MÖRCH, Carl-Maria

Université de Montréal, OBVIA, Algora Lab.

NOISEAU, Pauline

Université de Montréal, OBVIA, Algora Lab.

VOARINO, Nathalie

Université de Montréal, OBVIA, Algora Lab.

PER A LES IL·LUSTRACIONS

PATENAUDE-MONETTE, Martin

martinpm.info

PER AL DISSENY

HAUSCHILD, Stéphanie

stephaniehauschild.com

AMB LA PARTICIPACIÓ DE

ALANOCA, Sacha

The Future Society

ILIADIS, Niki

The Future Society

LAMOURI, Jihane

IVADO – Institut de valorisation des données

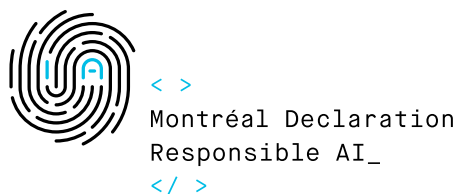
RUBEL, Sasha

UNESCO

TAY RU JIN, Melissa

UNESCO

GRÀCIES A LES INSTITUCIONS COL·LABORADORES



Volem agrair el suport financer dels governs del Quebec i del Canadà, com també de la Comissió canadense per a la UNESCO, del Consell de recerca de ciències humanes del Canadà, del Fons d'investigació del Quebec i del Consell nacional de recerques del Canadà.

